Virtual Environments: A Survey of the Technology

TR93-033

Richard Holloway
Anselmo Lastra

Sitterson Hall   CB-3175
University of North Carolina at Chapel Hill
Chapel Hill, NC   27599-3175

email: holloway@cs.unc.edu
lastra@cs.unc.edu

September 1993[†]

[†] These notes were actually prepared in April and May of 1993 for distribution in September at Eurographics.

1.   Introduction

Other names for virtual environments (VE):

- • Virtual reality (VR)
- • Virtual presence
- • Artificial reality
- • Virtual worlds
- • Cyberspace

Some early highlights:

- • Flight simulators: GE (in late 50s!), Singer-Link, Evans & Sutherland, McDonnell Douglas

- • 1965: Ivan Sutherland postulates "The Ultimate Display:"

    "The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal."   [Sutherland 65].

- • Sutherland builds HMD system with:
    - • see-through stereo head-mounted display
    - • head tracking
    - • hand tracking

- • Tom Furness and researchers at Wright-Patterson Air Force Base develop VCASS

- • NASA VIEW project

- • MIT- Callahan's HMD

- • UNC systems: GROPE, STHMDs, Walkthrough

Developments in mid-1980s:

- • Advances in small LCD television screens

- • Performance and cost improvements in image-generation systems

- • Cost improvements in magnetic tracking systems

- • VPL and others popularize "Virtual Reality;" lots of media attention

Sutherland's Problems = Today's Problems

Even after all of this progress, the subset of the "ultimate" capabilities that Sutherland implemented in 1968—stereo images, head and hand tracking—remains prevalent even today. Typical modern systems are essentially just improved versions of Sutherland's system, and are still plagued by the same problems:

1. Head-gear display devices have to produce high-resolution images and wide-angle views for two eyes.

2. Image generation to each eye has to be achieved at real-time rates.

3. Tracking of the head and hand has to be determined in real time and with considerable accuracy.

This tutorial will begin with an overall model for virtual environments and will therefore touch on all of the associated technologies, but will then focus on the technology associated with the three problems listed above.

2.   How VE Works: Fooling the Senses

2.1   The Display/Detection Model

• The real environment is perceived by our *senses*; our *actions* change our viewpoint in the environment and cause changes in it.

• A virtual environment (VE) is created by presenting alternative stimuli to one or more of our senses. The system must also monitor the user's movements and other actions in order to keep the viewpoint up to date and to initiate virtual actions such as grabbing or flying.

There are therefore two essential tasks: *Display* (presenting the alternative stimuli) and *detection* (monitoring user actions). Another way of looking at these two processes is as a user-centered input/output model: Displays provide sensory input to the user, and the detectors sense any user actions, which can be thought of as outputs.

The following table gives a description of displays for each sense.

Sense	Description	Display
visual	sensing of visible light	screens, optics, image-generation system
auditory	sensing of audible sound waves	computer-controlled sounds played through headphones or speakers
olfactory	sensing chemical concentration in air	odor transport system[1]
gustatory	sensing chemical content in solution	(unimplemented)
haptic[2]:	the general "touch" sense, with the following subcategories:	
• tactile	touch[3], temperature, texture, pressure sensed by skin	tactile display- controls small-scale forces or temperature variations
• kinesthetic	force sensed by muscles, joints, tendons	medium- to large-scale force-feedback display
• proprioceptic	sense of limb/torso positions and angles	(sense is internal; may not be possible to display to this sense directly)
vestibular	sense of balance; inner ear's sensing of linear and angular accelerations of the head	motion platform

Table 2.1.   Displays broken down by each sense

---

[1] Mort Heilig's Sensorama is the only system of which we are aware that has olfactory display.

[2] Haptics- "Pertaining to sensations such as touch, temperature, pressure, etc. mediated by skin, muscle, tendon, or joint." -*Webster's New International Dictionary*

[3] Touch- "The special sense by which pressure or traction exerted on the skin or mucous membrane is perceived." - *Webster's 7th Dictionary*

Accordingly, there is also a set of detection devices in current use. The following table is organized by user actions, and lists the types of detectors that are used. We will discuss trackers in depth in section 5.

User Action	Detector
head movement	head tracker
body or limb movement	tracker, force-feedback device, Spaceball
finger movement	glove, push-button device, joystick, keyboard
eye rotation	eye tracker
speech	speech-recognition system
forces exerted	force-feedback device with force detection

Table 2.2.   Actions and detectors

In addition to the actions listed in Table 2.2, one can imagine a host of other attributes or actions that future systems may monitor: facial expression, heart rate, blood pressure, etc.

The key idea is that there are two types of tasks in a VE system: display and detection, and a VE system must have components for each.

2.2   Perceptual Issues in Matching Displays to Senses

Visual

- Visual sensory organs are localized at the eyes, so displays can be localized.

- Field of view: Human eye has approximately a 208° instantaneous horizontal field of view [Levine 85]. (19" CRT viewed at 18" subtends about 56°)

- Illumination range: Visual system can operate over a range from roughly 1 to 10^{10}, or about 10 orders of magnitude [Jain 89].

- Acuity: Eye can resolve a separation of about 0.5 minutes of arc under optimal lighting conditions [Bruce & Green 90].

- Brightness (perceived luminance): Perceived luminance is not strictly a function of the light emitted from a surface, but also depends on its surround. Leads to *Mach banding*.

- Color system of eye uses three primaries ⇒ Displays can use three primaries

- Critical fusion frequency: A light flashing above this frequency is perceived as constant/steady (this frequency depends on many factors)

- Accommodation vs. convergence

- Depth perception ability decreases with distance

- Acuity falloff in fovea ⇒ high-res insets in displays

Auditory

- Limited frequency response means auditory display doesn't have to reproduce all frequencies

- Sensory organs (ears) are localized, so easy to center displays on them

- Human can process many sounds in parallel $\Rightarrow$ must be able to display multiple sounds simultaneously

- Location is one of the most important characteristics in differentiating sounds $\Rightarrow$ display should be able to localize sound

- Ideal display should account for individual differences: Pinnae, head, etc.

- Elevation accuracy is lower than azimuthal accuracy $\Rightarrow$ less sampling in elevation necessary

Haptic

- Sensory organs *not* localized; no way to place one or two displays to fool haptic sense

- Human is sensitive to many different types of input: texture, temperature, pressure, forces, etc.

- Grounded vs. ungrounded forces

- Hard-surface forces and textures require high frequency response and non-linear damping [Deyo 89]

Vestibular

- Vestibular system is located in inner ear: no localized display

- Fooling this sense implies real body motion (although fooling other senses can give illusion of self-motion)

- Motion sickness may result if vestibular input does not match input from other senses; interactions between other senses are critical

- Frequency response of vestibular system vs. that of motion platforms

- Individual differences in susceptibility to motion sickness are large

Smell

- Sense is localized in nose

- Smells once generated must be cleared away

Taste

- Sense is localized at tongue

- Basis functions may be available: Sweet, sour, salty, bitter

- Problem: Tasting is usually followed by swallowing

3.   Displays

3.1   General

Function: Provide stimulus to sensory organ

General characteristics:

- Display is usually located close to sensory organ (e.g., displays in front of eyes, earphones next to ears, etc.).

- Display's output should ideally match stimulus range of organ (e.g., visual displays need only display visible light, not infrared, ultra-violet, etc.).

- There is often the option of *replacing* or *augmenting* the real environment (e.g., visual displays may be see-through or opaque;   earphones may block outside noises or let them through).

3.2   Visual Displays

3.2.1   Overview

Function: Provide images of virtual objects to user's eyes that either replace or augment the real environment.

Types:
- monitor-based (non-immersive)
- head-mounted (immersive)
- arm-mounted displays (semi-immersive).

Monitor-based displays are typically fixed within the environment and give a "window" onto a virtual world, rather than immersing the user in it. A conventional CRT is the simplest example of this type of display. Other systems use some stereo mechanism to deliver different images to the user's two eyes in order to display objects in depth. Realism can further be enhanced by monitoring head movement so that the "through the window" image can be updated according to head movements.

Head-mounted displays (HMDs) are typically headsets incorporating two small LCD or CRT screens with optics to bring their images to the user's eyes. Head motion is usually monitored with a *tracker* (discussed in section 5) so that the image can be updated to reflect the current head position and orientation, which gives an illusion of presence in the virtual environment.

Arm-mounted displays consist of a monocular or binocular display mounted on a mechanical arm. The arm serves both as a counterbalance for the display to allow it to be moved easily, and as a tracking device which measures the display's position and orientation. Thus, they are like a cross between a HMD and a monitor: They allow the user to look around the virtual environment like a HMD does, but can still be used on a desktop like a monitor. Current systems (discussed below with HMDs) offer higher resolution than most LCD-based HMDs, and allow the user to get in and out of a virtual environment quickly. This type of display does not, however, allow the user to walk around the virtual environment with the same freedom and sense of immersion afforded by a HMD.

In keeping with this course's focus on immersive VE systems, we will concentrate on head-mounted and arm-mounted displays.

3.2.2 How Immersive Displays Work

Typical opaque configuration:
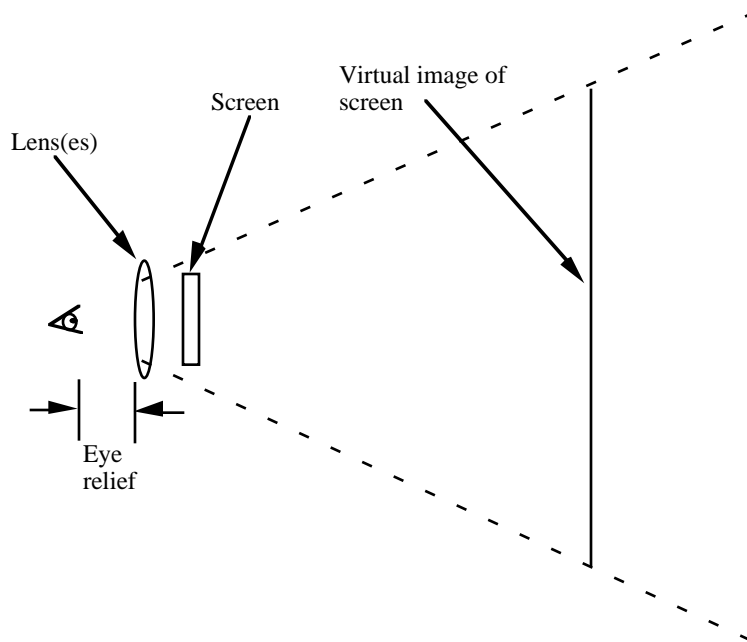
Figure 3.1   Opaque HMD optics for a single eye

Stereo viewing:
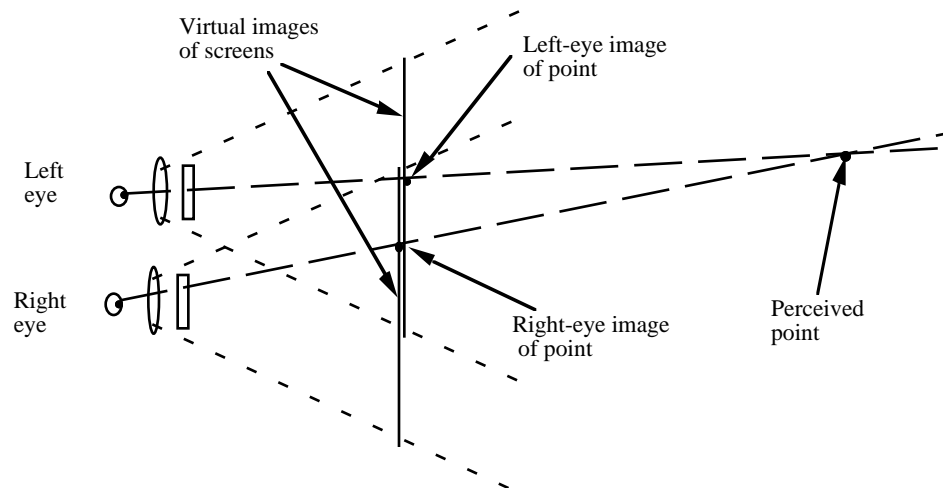
Figure 3.2   Stereo images on a HMD

Field of view:
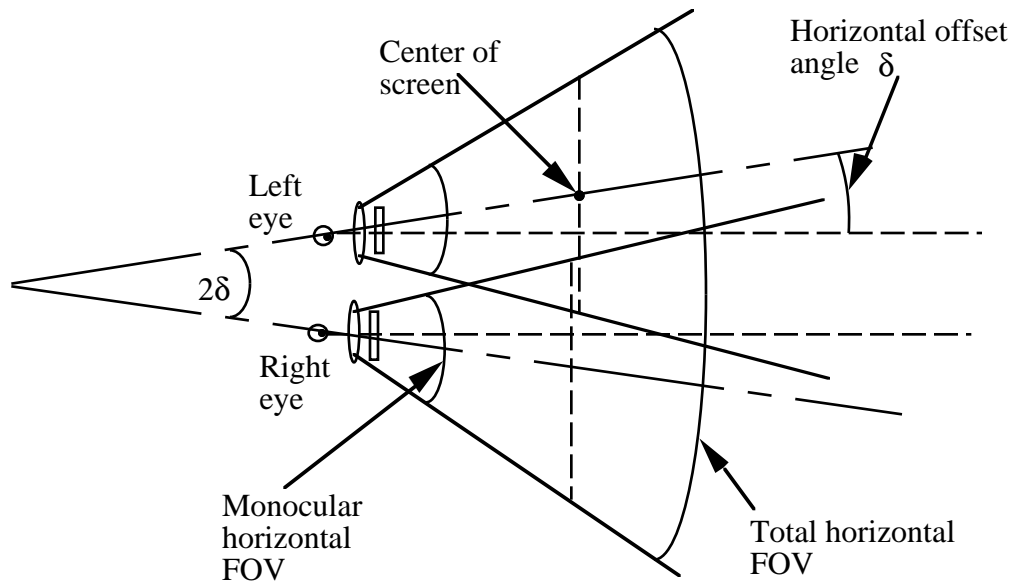
Figure 3.3 Horizontal field of view

FOV formulas:

If we define:

FOV$_m$ = the horizontal field of view for one eye
FOV$_T$ = the total horizontal field of view = binocular field of view

δ = the horizontal offset angle

then the following relations hold:

FOV$_T$ = FOV$_m$ + 2δ

binocular overlap angle = FOV$_m$ - 2δ = 2 * FOV$_m$ - FOV$_T$
binocular overlap percentage = (FOV$_m$ - 2δ) / FOV$_m$

Note: If center of screen is straight ahead, δ = 0, binocular overlap = 100%, and FOV$_T$ = FOV$_m$.

See-through HMD optics:
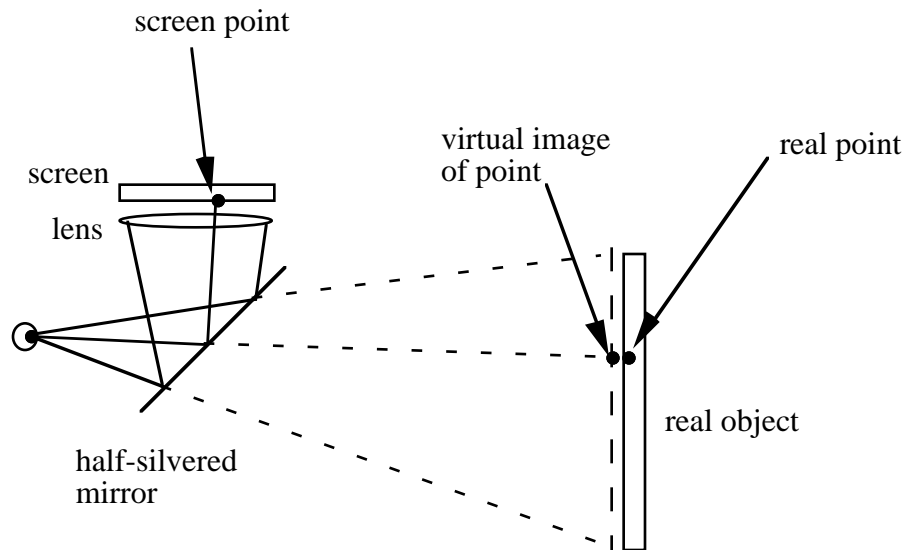
Figure 3.4 See-through HMD optics

3.2.3   Immersive Display Characteristics

General
- Stereo/Mono- Can the device display stereo images, mono images, or either?
- See-through/Opaque: Does the display obscure the real environment or superimpose the virtual objects into it?
- Price

Image source
- Resolution
- Type: CRT, LCD, ELD, etc.
- Luminance: overall brightness; critical for see-through HMDs (may be mediated by optics; e.g., pancake window)
- Contrast
- Color/Monochrome
- Refresh rate and lag
- Screen or image-source size

Optics
- Field of view (horizontal and vertical)
- Binocular overlap
- Image distance
- Eye relief
- Exit pupil
- Magnification: Screen image size
- Distortion and other aberrations
- Overall size/weight of optics

Wearability
- Weight
- Moments of inertia
- Balance
- Comfort
- Range (length and flexibility of cable, etc.)
- Safety

3.2.4   Immersive Display Issues and Problems

Major problems with currently available displays:

- Resolution: Most popular units have resolution of less than 200x200: Leaves user legally blind in virtual world [Piantanida et al. 93].

- Contrast/Brightness: LCD systems are difficult to use in see-through applications; some other systems use optics that dim display considerably. Dynamic range of current displays is nowhere near that of the eye, nor of the real world.

- Weight/Comfort: Current systems tend to be heavy and claustrophobic.

- Color: Small, color CRTs are just becoming available, but only at high cost.

- Field of view: Current systems are not even close to matching human visual system's FOV. Tends to diminish the sense of presence in VE.

- Distortion: Current wide-angle optics distort images considerably. Software correction is difficult and adds to latency of system. Distortion tends to increase with wider FOV.

- Expense: Some military-oriented units cost > $100,000.

- Range: Current HMD systems require the user to be tethered by cables of varying lengths, which limits range of use and encumbers the user.

Many design goals are in opposition with each other:

- Increasing FOV usually increases distortion and weight.

- Higher-resolution images are harder to transmit without cables (vs. NTSC).

- Color displays typically have less resolution than monochrome, since color pixels may take up three times as much area as monochrome. Field-sequential color systems don't have this problem, but are harder to update quickly and harder to make bright.

- CRTs have better images but are less safe than LCDs due to high voltages and strong magnetic fields.

- Stereo images take up to twice as long to generate as mono images.

- Larger fields of view mean more work for the image-generation system.

3.2.5 Desired Display Specifications

Ideally, a visual display should meet or exceed all of the limitations of the human visual system, so that the virtual image would be indistinguishable from a real one. Practically speaking, this will not happen any time soon. The following is our wish list for a visual display that we might see within the next 10-15 years.

- Stereo
- 180˚x100˚ FOV
- 4,000 x 4,000 resolution, with optics that concentrate resolution in center of FOV
- Can electronically blend between completely opaque and completely see-through on a per-pixel basis
- < $10,000
- Full color
- High contrast and variable brightness
- 100 Hz update rate
- Negligible optical distortion (or easily correctable in software)
- Lightweight, low moments of inertia: Like a pair of glasses
- No encumbering cables

3.2.6 Commercial Systems

Commercial HMDs can be divided into two major categories: LCD-based, general-purpose opaque HMDs and CRT-based, military, see-through HMDs.

This section gives a listing of many of the current vendors of immersive displays for comparison and contrast. This listing is intended to be representative of the current state of the art, but is not necessarily a complete listing of all vendors. A good listing of vendor contact information is Appendix B of [Pimental & Teixeira 92].

General notes:

- See-through HMDs can typically be converted to opaque by adding blinders, so a classification of "see-through" can be interpreted as "either see-through or opaque".

- Stereo/Mono: Many stereo HMDs cannot be used for mono images; i.e., feeding the same video signal to both eyes will not work because the screen images are not centered in front of the eyes; generation of two different viewpoints or simple lateral shifting is required.

- The *range* category is intended to give an idea of whether the display itself (not the tracker) limits the range of movement of the user. Systems whose inputs are NTSC are listed as unlimited, since these cables can be quite long and the NTSC signal can be broadcast as RF without cables. Other systems limited by RGB cable or fiber optic cable length are so indicated.

- Some vendors also list "field of regard" or "total field of view", which is essentially a tracker specification.

- CRTs typically have much higher contrast, brightness, and resolution than LCDs.

Name	Type	Stereo/ Mono	Color	Image Source	Reso- lution HozxVert	Binocular FOV (HxV)	Overlap	Range	Price ($US)	Notes
Virtual Research Flight Helmet	opaque	stereo	color	LCD	360x240†	93x61°	79%	unltd	6,000	a
LEEP Cyberface 2	opaque	stereo	color	LCD	479x234†*	140x94°*		unltd	8,100	b
VPL Eyephone LX	opaque	stereo	color	LCD	442x238†	108x76°	74%	unltd	9,150	c
VPL Eyephone HRX*	opaque	stereo	color	LCD	720x480†	106x75°	77%	unltd	49,000*	c
VRontier Worlds tier 1 HMD	opaque	both	color	LCD	479x234†	112° diag	100%	unltd	6,500	
Liquid Image Corp Mirage HMD	opaque	mono	color	LCD	720x240†	~110°	N/A	unltd	9,500	
W Industries Visette	opaque	both	color	LCD	376x276†	90-120°	50-100%	unltd	N/A*	d
Polhemus Labs	either	both	color	see notes	~500x500 TV line prs	50x40°	0-100%	8'	35-50,000	e
Private Eye	opaque	mono*	mono	red LED	720x280	22x14° (one eye)	N/A	unltd	795	f
Virtual Reality, Inc HMD 121	see-thru	both	gray-scale (green)	CRT	1280x1024	50-77° diag	35-100%	unltd	60,000	
n/Vision High-Resolution HMD*	see-thru	both	color*	CRT	1280x1024	50-83°	35-100%	10'	75,000	g
CAE FOHMD	see-thru	both	color	Light-valve* projector	1024x1024*	127x66°*	46%	6'	~1 million	h
Kaiser SIM EYE	see-thru	both	color*	CRT	640x480 to 1280x1024	60x40°	50-100%	6'	200,000	i
Honeywell IHADSS	see-thru	mono	gray-scale	CRT	1280x1024	40x30° (one eye)	N/A	6'	see notes	j
Honeywell WFOV	see-thru	both	gray-scale	CRT	1280x1024	80-110° x 60°	17-67%	6'	~ 200,000	k
Honeywell MONARC	see-thru	both	gray-scale	CRT	1280x1024	35-52° x 35°	51-100%	6'	~ 150,000	
Fake Space BOOM2	opaque	both	gray-scale	CRT	~480x480	140x90°*	43%	ltd*	35,000	l
Fake Space BOOM2C	opaque	both	pseudo-color	CRT	1280x1024	140x90°*	43%	ltd*	74,000	l
LEEP Cyberface 3	opaque	mono	color	LCD†	720x240	80° diag	N/A	ltd*	9,740	

* See product-specific notes.

† LCD display resolutions are typically quoted in "primary colored pixels," meaning that each red, green, and blue pixel is counted individually, which multiplies the real resolution by a factor of three. The real resolution depends on the layout used for the color triads in the display. Dividing the horizontal resolution by 3 or both resolutions by √3 should give some idea of the real resolution. Keep this in mind when comparing CRT-based HMDs with LCD-based HMDs- CRTs are currently MUCH sharper.

Product-Specific Notes:

    a- FOV numbers for this HMD were actually measured by Jannick Rolland of UNC. Center of FOV is taken to be first nodal point of eye, estimated at 25mm.

    b- FOV spec given is for "corneal field:" The distance from the surface of the lens to the corneal surface (20mm) is used as the center of the FOV. The "direct" field uses the eye's center of rotation as the center of the FOV; its FOV values are 109.5˚x 84.5˚. Not clear which point other vendors use for their calculations, so comparisons are difficult. Optics radially compress image of screen to enhance resolution at center of image.

    c- VPL was not selling the HRX at the time these notes were prepared (4/93) due to problems with their supply of hi-res LCDs.

    d- Screens are actually mounted on side of head; optical path is folded. HMD only sold as part of entire arcade system.

    e- Prototype fiber-optic HMD: can use projector, workstation, VCR as image source. Not related to Polhemus Inc.

    f- Display is sold as a single-eye unit, but two units can be combined to make a stereo display.

    g- Formerly Virtual Reality Group. Also makes a monochrome unit for $60,000.

    h- CAE also offers a lower-cost version that uses CRTs as image sources. Both versions feature a high-resolution, 24˚x18˚ inset in the center of the FOV of each eye with 500,000-1,000,000 pixels.

    i- The color SimEye is supposed to ship in June. The monochrome version has similar specs, but is cheaper. Kaiser also has several HMD products for military applications. Kaiser is the parent company of Polhemus Inc., makers of trackers discussed in section 5.

    j- A complete 2-station system with head tracking costs ~$150,000.

    k- Price is approximate; includes drive electronics but no head tracking.

    l- Color: Pseudo-color is red-green combinations; blue signal is not used. In pseudo-color mode, vertical resolution is halved. Resolution: the 140˚ figure is for "peering"; if the user doesn't move his head *relative to the Boom*, the hoz. FOV is about 100˚. Range: Tracker is built in; range is 5 ft dia. by 2.5 ft high cylinder with excluded center core.

3.2.7 Research Systems

Research in head-mounted displays can be broken into three categories:

1. *Image source:* Improving the resolution, color, size, etc. of the displays in the HMD

2. *Optics:* Creating wide-field-of-view optical systems with large exit pupils, adequate eye relief, minimal distortion and other aberrations, while minimizing size and weight.

3. *Mounting:* Developing the part of the HMD worn on the head to make it lightweight, comfortable, and easily adjustable.

A few of the many projects going on worldwide are listed below; this is by no means a complete listing.

NASA-Ames

Stephen Ellis and Urs Bucher have developed an "electronic haploscope" for performing perceptual experiments for better understanding depth perception in VEs. They use a HDTV monitor configured for NTSC to achieve a horizontal resolution of better than 4.5 arcmin/pixel with a monocular FOV of about 25˚. The CRT images are relayed to the user by a partially silvered mirror. The haploscope has adjustments for overlap, vergence and accommodation, and is mounted on a table similar to an optical bench. The display is used in experiments for making depth comparisons between real and virtual objects.

HITLab Virtual Retinal Display

The HITLab VRD uses modulated light from a red laser to scan out a computer-generated image directly onto the retina of the user. An optical bench prototype has been demonstrated with a 500x1000 monochrome image subtending a 45˚ monocular field of view (approximate).

The system's goal is to achieve a display with the following characteristics:
- high-resolution (4,000x3,000)
- low-profile
- portable
- full color
- wide field of view (> 100˚)

Air Force Institute of Technology

The group at AFIT (Wright-Patterson AFB) has done research on HMDs since at least 1988, when Captain Robert Rebo built an LCD-based bicycle-helmet-mounted display, described in [Rebo 88]. The third-generation HMD was designed by Jim Mills and Phil Amburn and addresses the current lack of color CRTs for HMDs. Their approach is similar to that used for projection TVs: Separate red, green and blue CRTs are optically combined to achieve a high-resolution, high-contrast color image. A block diagram of the optics is given in Figure 3.5.
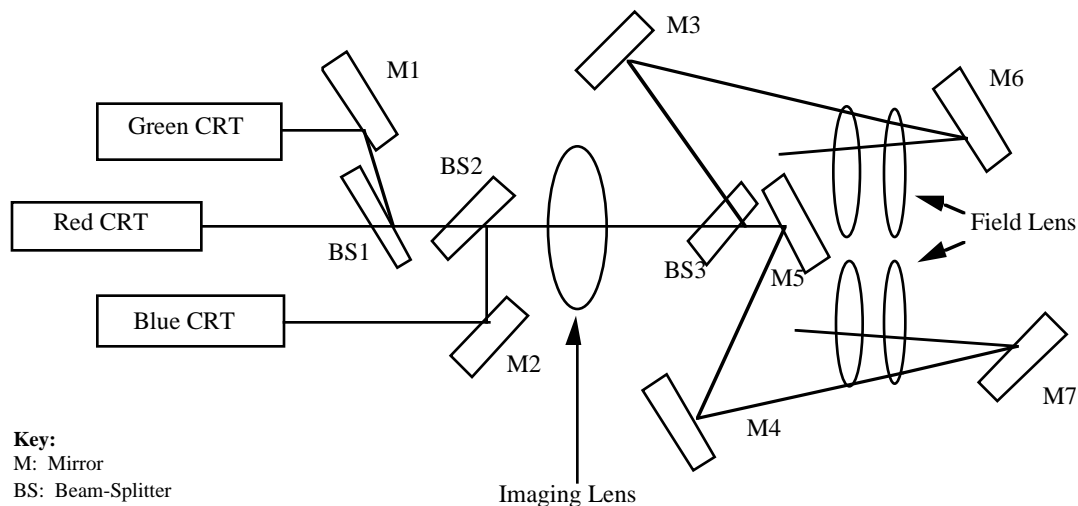

Figure courtesy of Phil Amburn, AFIT

Figure 3.5   Block diagram of AFIT HMD-III optics

This system uses 1" CRTs, beam splitters, dichroic filters and mirrors to form a 640x480 image through the LEEP optics.

University of Central Florida

The HMD being developed at UCF will use holographic optical elements (HOEs) to provide a lightweight, wide-field-of-view, see-through design. The system will warp the image prior to display to maximize the FOV and concentrate detail near the center. The optics then perform the inverse warping to create a correct, wide-FOV image. The HOEs are based on duPont's new holographic recording film materials. The warping is done either in hardware or software by the image generator. The HOE is both lightweight and inexpensive to fabricate, and could lead toward the goal of HMDs as comfortable as eyeglasses. [Clarke 93]

UNC See-Through HMDs

Current 30˚ FOV model:
- Finished in Spring 1992
- Built from off-the-shelf parts
- Components:
  - Sony color LCD screens
  - Off-the-shelf optics
  - Custom-built head-mount w/ rigid, repeatable adjustments

- 30-degree FOV
- Resolution: 360x240 primary-colored pixels
- 100% Overlap
- Design goals: Rigid frame, adjustments for IPD, focus, fore-aft and up-down for images, calibrated settings

A 60˚ FOV model using custom optics and color CRTs has been designed and should be operational by late Fall '93.

FDE Associates: Working on a high-resolution, small color CRT. One of the problems in building a tiny color CRT system is that a traditional three-gun system is too large to fit into the neck of the small tubes used for such CRTs. Their idea is to use a single gun and a moving shadow mask in order to selectively excite phosphor stripes of each color. FDE's goal is a 1000x1000, 1" color display.

Tektronix, Inc: Another approach to the color-CRT problem. The EX100HD uses a pair of one-inch monochrome CRTs and color shutters to create frame-sequential color images. The NuCOLOR shutter is an electrically switchable color filter made of fast optical LCD switches and color filters that allow its color to change rapidly between red, green and blue. A full color image is created by drawing the R, G and B fields of an image sequentially in sync with the filters. The field rate for each individual color is 180 Hz, and the overall frame rate is 60 Hz.

Texas Instruments: TI is developing a display based on "deformable-mirror spatial light modulators" [Hornbeck 89]. This deformable-mirror display (DMD) uses micromechanical mirrors supported on two diagonal corners and free to rotate about the axis between the supports. The other two (unsupported) corners act as electrodes and can be pulled to one side or the other, thus rotating the mirror. Angular deflections of about 10˚ allow arrays of

such mirrors to form high-resolution displays when coupled with colored light sources. A 700x500 array has already been demonstrated.

Final Note: There are probably a dozen research efforts underway trying to produce a high-resolution, small, color display for use in VE. At the time of preparation of these notes, some of the first such products are entering the marketplace; more will doubtless make their debut by September '93.

3.3 Auditory Displays

Function: Provide auditory feedback to user that either replaces or augments auditory input from real environment. System should ideally be able to present any acoustic waveform to either ear in real time.

Types: Speakers mounted on head (earphones) or mounted in environment. The latter has limited imaging capabilities.

Issues:
- User-specific sound modulation/head-related transfer functions (HRTFs)
- Echoic vs. anechoic sound
- Isotropic vs. non-isotropic sources
- Multiple vs. single sources: Processing power
- Sampled vs. computer-generated sounds
- Synchronization vs. playback speed
- Modeling reflective acoustic surfaces, Doppler shift and other physical characteristics
- Scheduling: Queued or interruptable

Commercial Systems

3D Sound:

- Crystal River Engineering *Convolvotron-* Stored spatial filters are convolved with sound sources to produce up to four localized, isotropic sounds at a time. Simple reflective environment can be modeled with additional hardware. $15,000.

- Crystal River Engineering *Beachtron-* Low-cost version of the Convolvotron. Two channels; $1,900.

- Focal Point- Similar to Convolvotron/Beachtron. Two channels. $1,800.

- Visual Synthesis *Audio Image Sound Cube-* C-language library for MIDI-based sounds and 3D positioning. $8,000.

Text-to-Speech Systems:

- AICOM *Accent SA-* PC-based text-to-speech synthesis board; $500 and up.

- Voice Connexion *Micro IntroVoice-* Text-to-speech and voice recognition of up to 1,000 words; $1,200.

3.4   Haptic Displays

Function: Display and measurement of forces on and from the user.

Types:
- Force-feedback joysticks
- Force-feedback arms
- Force-feedback exoskeletal devices for hand, arm, other
- Tactile displays
    - shape-changing devices:
        - shape memory actuators
        - pneumatic actuators
        - micro-mechanical actuators
    - vibrotactile
    - electrotactile

Issues:
- Sensors scattered throughout human body- no localized display possible as with vision, etc.
- Physical characteristics of objects must be modeled: Hardness, texture, temperature, weight, etc.
- Collision detection must be done in real time
- Hard surfaces require high frequency response, non-linear damping
- User input must be measured and reflected in environment
- Grounded vs. ungrounded displays
- Degrees of freedom (DOFs) of device vs. human
- Safety

Commercial Systems

Tactile

- ARRC/Airmuscle *Teletact II*- Pneumatic shape-changing tactile array. 30-cell device integrated with glove device includes large palm pocket that can be inflated up to 30psi. $4,900; control system is an additional $13,400.

- Xtensory *Tactools XTT1*- Small transducers create tiny vibrations or impulses; basic system has one tactor and costs $1,500; system can support up to 10.

- CM Research *DTSS X/10*- Temperature display/detector device using thermodes. A solid-state heat pump can generate hot or cold sensations for 8 channels on the user's fingertips. $10,000.

- Telesensory Systems *Opticon*- A vibrotactile display for blind reading; 20x5 pins on 1.5"x.75" area; $3495.

- Begej Corporation *Tactile Stimulator*- Fingertip and tool-mounted tactile arrays. Vibrotactile or shape-changing display with 37-cell finger display or 128-cell tool display.

- Exos *TouchMaster*- Tactile feedback device which uses miniature voice-coil vibrators on fingertips. Variable amplitude and frequency. Approx. $2,000 per stimulator.

- TiNi Corp *Tactors*- Shape-memory alloy tactile stimulators, points and arrays. Monitor and 3x3 tactile display cost $7000.

- Digital Image Design *Cricket*- See entry under 3D mice in Section 5.3.

Kinesthetic/Proprioceptic

- TeleTechnologies *TT-2000 Force Reflecting Hand Controller*- Earth-grounded, 6-DOF generalized teleoperator master, joystick uses electric motors and cable transmission to give up to 34N at handgrip. $30-50,000.

- Shilling *Omega*- Earth-grounded, 6-DOF (+ grip) teleoperator arm using DC torque motors acting through harmonic reducers.

- Sarcos Research *Dextrous Arm Master*- Exoskeletal 10-DOF force-feedback arm using 3,000 PSI hydraulic lines to apply forces to the user's arm and hand. There are seven degrees of freedom in the arm and three degrees of freedom in the end effector. Maximum force when arm is full extended horizontally is 10 lb. Can be used to control the *Sarcos Dextrous Arm* for telerobotics applications.

- Cybernet Systems *PER-Force Handcontroller*- A compact, 6-DOF, earth-grounded force reflection device. Forces are generated by six DC servo motors.

Research Systems

- Margaret Minsky's Virtual Sandpaper- Minsky implemented a force-feedback joystick for feeling virtual sandpaper, stirring virtual ice cubes in virtual molasses, etc. Details in [Minsky et al. 90].

- UNC Argonne Remote Manipulator- This force-feedback arm can output three forces and three torques at the handgrip where the user holds it and has a working volume of about one cubic meter. It has been used in a number of different studies; most recently for finding a minimum energy docking configuration for drugs in the active site of a protein molecule [Ouh-Young, Beard and Brooks 90] and for feeling the atomic-scale peaks and valleys on a surface imaged by a scanning-tunneling microscope [Taylor et al. 93].

- Tsukuba University Master- [Iwata 90] describes a 9-DOF force-feedback system with force display for the thumb, two fingers and the palm. The feedback is provided by electric motors based on a solid model of the virtual space.

- Rutgers Portable Dextrous Master with Force Feedback- This system uses micropneumatic actuators place in the palm of a VPL DataGlove to create a light, simple, and relatively inexpensive manual force-feedback device. The air-piston actuators are grounded at the palm and give feedback to the thumb, index, and middle fingers [Burdea et al. 92].

- Hull Electrorheological Tactile Display- [Monkman 92] describes a system using fluids whose viscosity changes as a function of the applied electric field to create a tactile feedback device. An electrode array provides an addressable electric field underneath a layer of electrorheological fluid. The top layer is a compliant insulator layer with a compliant conducting layer underneath at ground potential. By selectively energizing the electrodes in the bottom layer, the fluid at that tactor can be made rigid. By energizing a pattern of electrodes, a tactile "image" can be created.

Prediction: Haptic displays will remain more application-specific than other VE displays for the foreseeable future.

4. Image Generation

4.1 Introduction

Most of the commercial 3D graphics systems are sold as desktop or deskside workstations. Unfortunately, the demands on image generators of VE and immersive displays are quite different than those of CRTs for typical desktop graphics uses.

- VE demands high frame rates.

- Due to the high frame rates, graphical models tend to be small. There is no need to support arbitrarily large datasets.

- Low-resolution video rather the standard 1280 x 1024 is the norm for driving HMDs. Most require NTSC or PAL.

- Usually two double-buffered channels are necessary for stereo HMDs.

- Windowing systems are not necessary.

These differences make it difficult to use many of the excellent graphics accelerators on the market. Many of the low and mid-level desktop workstations do not provide NTSC, either standard or as an option. Most cannot produce two channels. Thus they would require two graphics heads and two scan converters to be useful. This often pushes the price to the levels of the high-end image generators.

This section first reviews graphics performance, illumination, and shading. We then survey systems specifically for VE, and the subset of commercial image generators that can provide good service for HMDs, followed by a section on experimental graphics systems. We conclude with the special demands of immersive VE on image generators.

4.2 Graphics Performance

Polygon performance and price are probably the first considerations when choosing an image generator. Not only is VE work very demanding of performance, but the image generator is likely to the most expensive part of a system for VE. Other factors beside raw polygon performance must then be considered, such as capabilities for texturing. As we shall see, for some applications, texturing can make for a much richer visual environment. The improvement in realism afforded by texturing is often greater than that provided by more complex geometry. Most flight simulators render few polygons, but can create very rich scenery with textures.

We have found that it takes at least 15 frames per second to provide user comfort in a head-mounted display. A higher frame rate, say 20 to 25 fps, is noticeably better. When using stereo, the work of rendering is essentially doubled. This means that we'd like to render at least 30 frames each second and would prefer 50. We would also like to maintain low latency, but do not have much control over this factor in most image generators. Typically a higher frame rate results in lower latency.

Consider these possible models and sizes:

- A 1000 polygon model can represent a room with some detail, or perhaps a house with very little detail.

- A moderately detailed room can be modeled with 5,000 polygons.

- A house with some detail, such as that modeled by the UNC Walkthrough project consists of approximately 50,000 polygons.

Using our two frame rates, we can get an idea of the rendering performance that we'd like to achieve.

# of Polygons	Frame Rate	
	30 fps	50 fps
1,000	30,000	50,000
5,000	150,000	250,000
50,000	1,500,000	3,000,000

We should note several things:

- Many models for VE, such as architectural models, use quads as the normal polygonal primitive. Since quads are more expensive to render than triangles, we have to scale the quoted performance of image generators appropriately.

- Maximum image generator performance can often be obtained only by using triangle or quad meshes. This puts more of a burden on the modeler.

- It is often convenient, or necessary, to apply several levels of transformations: For example, to initially place objects and to allow for their movement. This often takes considerable computation, especially in parallel systems (see below).

- On the positive side, one can subdivide models such as a house into sets of polygons potentially visible from the current viewing position [Teller & Sequin 91] and only render the necessary polygons. Another technique is to group primitives into objects, such as rooms or furniture and cull entire objects based on their bounding boxes.

- Advanced shading techniques, such as texture mapping, can produce a realistic image with fewer polygons. However, the textured polygons are more expensive to render. Image generators with texturing capabilities are becoming more common.

Since published benchmarks, with one notable exception, do not take into account the complex nature of models to be rendered, be sure to carefully take your particular needs into account when comparing image generators.

4.2.1   The cost of graphics performance

It is naive to think that one can render a rich scene at 30 frames per second on a workstation without graphics acceleration hardware. To illustrate this fact, let us examine the number of operations that it takes to render a triangle (this analysis is adapted for our needs from [Molnar & Fuchs 90]).

Figure 4.1 illustrates the classical graphics pipeline (for more information, refer to the references listed below). This particular case is for Gouraud shading and z-buffering, the standard for current graphics hardware.

```
┌─────────────────────────────────┐
│     Modelling Transformation     │
└─────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐
│          Trivial Reject/          │
│        Backface Culling           │
└─────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐
│             Lighting             │
└─────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐
│       Viewing Transformation      │
└─────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐
│             Clipping             │
└─────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐
│     Division by w and mapping    │
│            to viewport           │
└─────────────────────────────────┘
                  │
                  ▼
┌─────────────────────────────────┐
│           Rasterization          │
└─────────────────────────────────┘
```
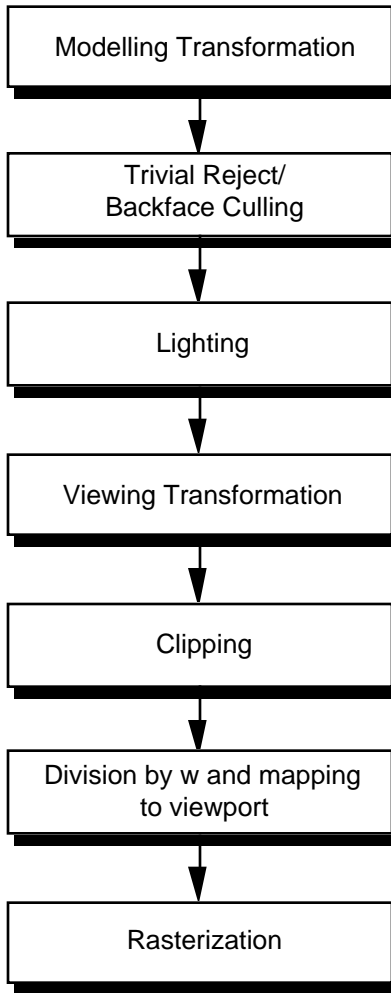
Figure 4.1

The computational costs of the individual steps are:

- **Transformation** - Triangle and normal vectors are transformed from object to world coordinates. This takes 25 floating point multiplications and 18 floating point additions per vertex.

- **Trivial Reject -** Examine the primitive to see if it is completely outside of the viewing frustum. This takes 18 floating point multiplications and 14 additions per vertex.

- **Lighting -** For Gouraud shading, a color must be computed for each vertex. For a single light source, this costs 12 floating point multiplications and 5 additions per vertex.

- **Viewing Transformation -** This costs 8 multiplications and 6 additions per vertex.

- **Clipping -** The cost of clipping is difficult to compute. The number of primitives that need clipping varies with the dataset and clipping parameters. For our

example, let us assume that scissoring is being performed. This transfers the burden of computation to the rasterization stage.

- **Division by w** - The cost is three divisions per vertex. Mapping to a viewport costs 2 multiplications and 2 additions per vertex.

- **Rasterization -** It is especially difficult to characterize the computational requirements at this stage. For each pixel, we will have to calculate a z value, and access the stored depth value from the z-buffer. If the pixel is visible, we must store the new z value. Gouraud shading will require interpolation of each of red, green, and blue across the polygon and storage of the color in the frame buffer.

Total Cost

Let us assume that we are rendering, in stereo, the middle example shown above, the 5000 triangle dataset. For simplicity, let us assume that these are rendered as individual triangles, though the cost would be reduced by using triangle strips. Assume that half of the triangles will be trivially rejected. The floating point costs will be:

	Mult. Div.	Add	Times (vertices, stereo)	Total Mult. Divisions	Total Additions
Transformation	25	18	$2 \times 3 \times 5000$	750,000	540,000
Culling	18	14	$2 \times 3 \times 5000$	540,000	420,000
Lighting	12	5	$2 \times 3 \times 2500$	180,000	75,000
Transformation	8	6	$2 \times 3 \times 2500$	120,000	90,000
Clipping	-	-	0	0	0
Division	2	2	$2 \times 3 \times 2500$	30,000	30,000
Total per frame				1,620,000	1,155,000
Total per second				24,300,000	17,325,000

This shows a total of approximately 42 megaflops for a relatively simple dataset. While some of the newer microprocessors promise this level of performance, you'd be hard pressed to achieve it in reality. Furthermore, note that we have assigned no cost to database traversal, or to the application program. As a practical matter, it takes multiple processors to achieve this level of front-end performance.

The back-end rasterization costs are also very high. To clear the frame buffers, each z (one 32 bit word) and color value (perhaps a 32 bit word) must be written, for a total of $2 \times 640 \times 480 \times 2 = 1,228,800$ 32-bit writes to the frame buffer. Let us assume that the average size of our triangles is 100 pixels, and that one half of them is visible. Since some pixels are initially visible, and are then obscured by another pixel, let us assume that 3/4 of the pixels must perform the complete z-buffer access, while the remaining 1/4 will just read the z value.

Using forward differencing, calculation of a z value takes one addition. Therefore, our invisible pixels cost $2 \times 625 \times 100 = 125,000$ integer additions and the same number of frame buffer cycles. The remaining initially visible pixels cost four additions per pixels and three frame buffer cycles. The total costs for each stereo frame are 1,625,000 additions and 2,845,800 frame buffer cycles. When multiplied by 15 frames per second and two channels, the total costs are 48 million additions, and 84 million word-wide frame buffer cycles per second.

Using the latest processors, the biggest problem is limited memory bandwidth. The memory subsystems of high-end workstations would have a difficult time sustaining this performance. Coupled with the front-end demands, it becomes necessary to provide specialized hardware. Low-end systems may use a processor, such as the Intel i860 which is enhanced with rasterization specific instructions, to accelerate the rendering task.

All high performance graphics accelerators use parallelism in one form or another. The first stage of the pipeline that needs parallelism is rasterization and frame buffer access. These functions are usually performed with specialized hardware. At some point the transformation, lighting, and clipping become the bottleneck and must also be parallelised.

4.2.2 Antialiasing

Even with the poor quality of current head-mounted displays, aliasing is clearly visible. With better displays in the future, antialiasing will be necessary. We see the trend toward antialiasing in high-end graphics systems of today such as the Reality Engine from Silicon Graphics. Given the normal progression of features we expect to see lower cost systems with strong antialiasing capabilities soon. Note that antialiasing is very expensive computationally.

4.3 Illumination Models

Most graphics hardware and software support illumination models consist of the following components:

- *Ambient* - models a source of light that is non-directional and arrives from all directions. The ambient light term is a simple approximation to the light reflected by objects in the environment.

- *Diffuse* - models the dull reflections of light from matte surfaces due to point or directional light sources. The brightness of a surface depends on the angle between the surface normal and the light, and is independent of viewing angle.

Some may also include:

- *Specular* - models the shininess of materials such as plastic and metal. The brightness of these materials is dependent on the viewing angle. The *Phong* illumination model attempts to capture the specular reflections of shiny surfaces. It is more expensive to compute than a simple diffuse model.

The illumination models may be calculated for each vertex, or perhaps only once per polygon, depending on the shading model to be executed during rasterization. Note that these models are fairly simple, and do not include the effects of reflections of lights within an environment. *Global illumination* lighting models take into account the contribution of light from other objects in the environment.

Radiosity Illumination

Radiosity methods calculate global illumination based on physical models of reflection and can provide very realistic images. They are too computationally expensive to be calculated in real-time on most graphics systems, as is necessary for HMDs. However, the illumination is view independent, so may be pre-computed off-line if the graphical database

is static. This is a good illumination model for applications such as architectural walkthroughs.

4.4   Shading

By *shading*, we refer to algorithms used to calculate color for individual pixels. The color is normally based on illumination models that determine the effects of light on geometry.

- *Flat* shading - This is the simplest shading model. It was used in early systems, and is still used in low-end systems because the computational requirements are much lower. Only one normal vector must be computed per polygon to calculate the lighting, and no color interpolation is performed. Unfortunately, this causes very objectionable effects because the color of polygons changes at polygon edges.

- *Gouraud* shading - This is the common shading model provided by graphics systems today. Normal vectors at each vertex are used to compute lighting, but only the color is interpolated during rasterization. The interpolation avoids the color discontinuities at polygon edges.

- *Phong* shading - A lighting model is calculated at each pixel using a normal interpolated across the polygon from normals at the vertices. This shading method produces much more accurate specular highlights than those possible with Gouraud shading. Commercial systems typically do not perform Phong shading at the frame rates required for VE.

Texture Mapping

Texture mapping refers to the mapping of an image to geometry. Usually the image is stored in dedicated memory on the graphics accelerator. Extra processing must be done to properly map the images onto the polygons. Hardware texture mapping is a feature found mostly on high-end graphics accelerators. In many simulations, the sense of realism achieved by texture mapping can replace many polygons.

The most realistic textures are obtained from hardware capable of tri-linearly interpolating pre-filtered images. The usual technique used is known as Mip mapping and is due to Williams [Williams 83]. Some hardware point-samples the image-based texture, with a resulting loss in image quality. Note that for highest quality, textures should also be corrected for perspective distortion.

Shadows

Shadows can provide visual cues that enhance the effects of an image. Unfortunately, computing shadows can be quite expensive. Using pre-computed radiosity illumination, one can obtain shadows for static datasets. To provide shadows for dynamic datasets, it is possible to use a two-pass z-buffering algorithm [Williams 78].

4.5   Graphics Libraries

There are two main subdivisions of graphics libraries, or Application Programmers Interfaces (APIs) as they are also known. One type, *immediate mode*, requires that the application program generate the graphical primitives for every frame. The second type of API, *retained mode*, maintains a display list (initially generated by the application, of

course) which the image generator traverses every frame. The application interacts with the database of primitives by executing editing commands.

4.5.1 Retained mode

This type of API is well suited for applications that do not change the graphics database very much. A disadvantage of retained mode that is often cited, the fact that image complexity is limited by memory size, is not a problem for work in VE since the size of the dataset is limited by frame rate. Retained-mode systems can load the display list onto the graphics accelerator, thus largely divorcing rendering from the main CPU. The best known retained mode API is PHIGS+ which supports a hierarchical display list.
The main disadvantage with retained mode is that, for some applications, you must replicate the database. For example, the 3DM application [Butterworth et al. 1992] at UNC (a 3D MacDraw like tool for use under the HMD) maintains an application specific database and issues PHIGS editing commands. This results in two independent copies of the dataset.

4.5.2 Immediate mode

Immediate mode is well suited for data that change a great deal every frame. For example, a mesh of triangles describing an airplane wing during a simulation. It certainly involves the host CPU, therefore that processor must be fast enough both to run the application and to generate the polygons to be rendered. The best known immediate-mode API is IRIS GL from Silicon Graphics. An evolution of the GL into an open standard called OpenGL is underway. The standard is now under the control of a committee with a broad base of industry representation. Many manufacturers have agreed to support OpenGL .

4.6 Image Generation Systems

We begin by examining performance specifications for graphics systems. We then divide the systems into those specifically targeted to VE (normally sold as turnkey systems) and general graphics workstations that may be used for VE.

4.6.1 Performance Specifications

It is very difficult to compare the published performance of graphics systems. As a rule, specifications provided by vendors may not be directly compared because they are obtained using different testing methodologies. For example, one vendor will measure rendering rate using 50 pixel triangles, while another will use 100 pixel triangles. Furthermore, the specifications cite absolute peak performance.

Of course, the best way to compare computers in general is to benchmark them on the problems that one is interested in solving. Since this is rarely possible, the next best approach is to compare the results of standard benchmarks executing code similar to that necessary to solve the target problems. This has been done for many years with CPUs.

Vendors of graphics-oriented hardware have cooperated to form the Graphics Performance Characterization Committee (GPC) which is administered by the National Computer Graphics Association (see address below). The GPC publishes a benchmark called the Picture Level Benchmark (PLB). The PLB consists of a specification of geometry and actions to be performed, a methodology for performing the tests, and a standard suite of tests. A sample implementation is available from:

National Computer Graphics Association
2722 Merrilee Drive
Suite 200
Fairfax, VA 22031 USA
(703) 560-2752

or over the internet by ftp from swedishchef.lerc.nasa.gov (139.88.54.33).

Benchmark results are reported in units called GPCmarks, and can be literal (instructions followed exactly), or optimized (some optimizations were performed to increase performance). GPCmarks are calculated as

Normalization Factor / Elapsed Time

The normalization factor is meant to capture the difficulty of the benchmark. Higher GPCmarks indicate lower elapsed time and higher performance. Of the standard benchmark suite, three are applicable to VE. They are:

head - a scanned human head of 60,000 triangles in triangle strips lit by four directional light sources. The head rotates around the vertical axis for 240 frames. The difficulty factor is 4800.

shuttle - A space shuttle rendezvous with a satellite. The models consist of a mix of triangle strips, quad meshes, and polygons. There are also 2283 markers used to represent stars. The difficulty factor is 4000.

studio - An architectural walkthrough of a radiosity-lit artist's studio modeled with 7518 quads. A total of 300 frames are rendered during the walkthrough. The difficulty factor is 2500.

Note that the standard benchmarks are computed using a 900 by 720 window, somewhat larger than the normal NTSC resolution. Unfortunately, not all vendors supply PLB specifications.

4.6.2   Commercial VE Systems

The systems in this section are targeted specifically for work in VE with HMDs. Unlike systems for the desktop graphics market, these feature the NTSC video necessary to drive the current HMDs as standard. These firm's intents are to sell complete systems, including the image generator, HMD, tracker, software, and sometimes a sound generator.

Division

Division provides hardware and software for VE. They sell complete standard configurations, as well as custom systems. The Division software is also available for use on SGI image generators.

Systems include:

- Image generator providing two channels for stereo

- Virtual Research Flight Helmet HMD (see section 3.2.6)

- Polhemus Fastrak (see section 5.2.2.1)

- Division 6D mouse

- Beachtron (see section 3.3)

- dVS software

- A separate i860 for collision detection

System	Triangles per sec.	Textured Triangles	Base Price US$
100VRX	35K	N/A	$64,000
100VRXT	35K	35K	$70,000
Supervision	280K	280K	$180,000

Test conditions:

- Triangles are 100 pixel, 24-bit color.

- Textured triangles are 100 pixel, 24-bit color, point sampled.

Notes:

- Performance is per eye.

- Host processor is an Intel 80486 running UNIX System V.

- Fill rate of 100VRT is 8M pixels per eye.

- *Supervision* rendering system consists of a communications ring (for each eye), supporting renderers based on i860s. A frame grabber for generating composite images of live and synthetic video is an option.

- The rendering board set (3 PC cards) used in the 100VRX is sold separately as the *dView* for approximately $13,000. This set generates two channels of video.

Sense8

The main focus of Sense8 is their software for VE, the *WorldToolKit*. However, they resell the SPEA Fire PC-peripheral Graphics board that uses an i860 for rendering. The *WorldToolKit* also runs on SGI hardware.

Graphics	Polygons per sec.	Textured Polygons	Price US$
SPEA Fire	10K	5K	$2,795

Test conditions:

- Polygons are 100 pixel, 24-bit color.

- Textured polygons are 100 pixel, not perspective corrected, 24-bit color.

Notes:

- 8 Mb SPEA Fire memory upgrade available for $800.

VPL

VPL is developing the Microcosm system hosted by a Macintosh Quadra 950 with graphics accelerator boards (manufactured by Division, but sold exclusively by VPL). The graphics boards are essentially the dView boards (see above) with a Macintosh rather than a PC interface. The current pre-production system includes an Eyephone LX HMD (see section 3.2.6), and a Polhemus Fastrak (see section 5.2.2.1).

4.6.3   Commercial Image Generators

By far the largest market segment in 3D graphics is for desktop workstations. The typical uses of these workstations are in fields such as computer-aided design, where the performance needs and demands are much different from those in VE. Since these workstations are sold in moderately high volume and there is much competition for the market, price/performance tends to be good.

There are some problems with using desktop graphics systems for VE, however. One is that the standard video provided on workstations is not NTSC (which is used for most HMDs). To obtain NTSC, optional video cards or expensive scan converters must be purchased. Since a good scan converter costs upwards of $20,000US, it is not an attractive solution. Furthermore, generating two channels of video to provide stereo for the two displays in a typical HMD is another problem. The low-end workstation systems are not equipped with multiple channels of output video. The combination leaves many good systems, especially in the middle of the performance range, unusable.

There have been custom solutions to the problem of generating two channels of video for HMDs: For example, Folsom Research, a manufacturer of high quality scan converters (see section 4.7.6) has built custom scan-converter / video-splitter products for customers. Unfortunately, these are not cheap.

General notes:

- Some of the following systems are configured with two accelerators to provide two channels of video.

- Specifications are those given by the manufacturer for a single graphics subsystem. *Care must be taken when making comparisons between systems with one graphics accelerator, and those containing two.*

- Graphics workstations will normally include one or two CRT monitors.

- Real-time operation of UNIX systems can be difficult.

Silicon Graphics

Silicon Graphics (SGI) is the best known vendor in the graphics workstation arena. They have a range of processor products from the Indigo with a MIPS R3000 CPU, to the Onyx with a maximum of 24 MIPS R4400 CPUs. Matching graphics processors range from the XS24 to the RealityEngine2.

The high-end systems are all very good for work with HMDs.

Processor	Graphics	Triangle Mesh	Textured Polygons	Base Price US$
Crimson	VGXT	1.1M	35K	$88,000
Crimson	Reality Engine	1.1M	600K	$118,000
Onyx (2 CPU)	VTX	1.1M	450K	$138,000
Onyx (2 CPU)	Reality Engine 2	1.6M	900K	$178,000

Test conditions:

- Triangle Mesh: 50 pixel, unlighted, flat-shaded, z-buffered

- Textured Triangles: 50 pixel triangles in a mesh, antialiased, Mip-mapped, tri-linearly interpolated.

Notes:

- All provide 24 bits for color. Extra frame buffer memory for advanced shading.

- 32 bits for z, except for the VGXT which provides 24.

- At least four channels of NTSC is provided via a "video splitter" option at a cost of $19,000US. Price for a splitter is included in the table.

- The Reality Engine is primarily an image generator for generating textured map polygons.

- SGI software is specialized for real-time operation.

- Crimson CPU is a MIPS R4000 benchmarking at 60.5 SPECfp92, and 58.3 SPECint92.

- Onyx is a multiprocessor with R4400 CPUs. Can be configured with a minimum of 2 and a maximum of 24 processors. Some Onyx configurations can support two graphics pipelines.

- SGI supports IRIS GL, and will introduce OpenGL in the near future.

- Higher-level programming tool kits are available from SGI.

The new Indigo^2 Extreme is a very interesting desktop graphics system with performance of 450,000 triangles per second (meshed as above). It is priced at $35,000. An option called Galileo and priced at $6,500 is available that can provide NTSC for the Extreme. However, unlike the splitter, it only provides one channel. Two Extremes can not be placed in one Indigo. It's possible to use two networked Indigo ^{2} Extremes, one for each eye. However, this may not be an attractive option because of the tight synchronization that must be performed to switch both frame buffers simultaneously.

Evans & Sutherland

The E & S Freedom series was introduced this year. They function as graphics accelerators for Sun Microsystems SPARCstations.

At the time this report was prepared, the PLB results for the model 3200 were available. They are:

	Literal	Optimized
head	239	
shuttle	78.2	102.5
studio	198.3	

Manufacturer's specifications:

Graphics	Triangle Mesh	Polygons	Textured Polygons	Base Price US$
1050-02	500K	100K	40K	$65,595
1100-10	1M	200K	80K	$82,995
3100-10	1M	200K	80K	$117,245
3150-10	1.5M	300K	120K	$139,245
3200-40	2M	400K	160K	$177,245
3300-40	3M	600K	240K	$229,245

Test conditions:

- Triangle strips contain 25 pixel triangles, flat shaded.

- Polygons of size 100 pixel, Gouraud shaded.

- Textured polygons are in a quad mesh, sized 50 pixels Mip-mapped, tri-linearly interpolated.

Notes:

- NTSC or PAL video output is standard, but only one channel.

- Two accelerators must be used to generate stereo for a HMD. Provision for this is made in the hardware and software.

- The prices include a SPARCstation 10 and two graphics accelerators.

- The 1000 series models are shown with a SPARCstation 10/30 configured with 32 Mb or main memory, and 424 Mb of disk, priced at $13,995.

- The 3000 series accelerators are shown with a SPARCstation model 10/41 with 64 Mb of memory, and 1 Gb of disk, priced at $23,245. The 10/41 is rated at 53.2 SPECint92, and 63.4 SPECfp92.

- The specifications were obtained with only one graphics accelerator. It may take a faster SPARCstation host achieve full performance using two accelerators. Multiprocessor SPARCstation 10s are available.

- Suffixes to the model numbers indicate the amount of texture memory. It ranges from 256 Kb to a maximum of 4 Mb and is expandable.

- Available APIs include SunPHIGS and Sun's XGL. E & S plans to support OpenGL in the future.

Kubota

Recently Kubota announced the Denali line of graphics accelerators, designed to interface to Digital Equipment Corporation workstations based on the Alpha microprocessor architecture. Combined workstation/graphics systems are named Kenai.

Picture Level Benchmark Results:

Processor	Graphics	head	shuttle
3400	E25	83.54	26.03
	P510	196.00	46.92
	V620	233.55	45.65
3500	E25	84.12	28.69
	P510	189.65	51.71
	V620	237.51	57.28

Manufacturer's specifications:

Processor	Graphics	Triangle Mesh	Textured Triangles	Base Price US$
3400	E15	200K	100K	$55,145
3400	E25			$63,145
3500	E25			$80,695
3400	P510	900K	400K	$98,845
3500	P510	1M	400K	$116,395
3400	V620	1M	600K	$109,845
3500	V620	1.2M	600K	$127,395

Test conditions:

- Triangle strips contain 50-pixel Gouraud-shaded triangles.

- Textured triangles, of size 50 pixels, are perspective corrected, and point sampled.

Notes:

- NTSC and PAL video is made via an optional card at $3000. One channel.

- Two accelerators must be used to generate stereo for HMDs. Provision for this is made in the hardware and software.

- Prices are for one Kenai workstation, two Denali accelerators, and two NTSC video options.

- Processor for the 3400 is a 133 MHz Alpha benchmarking at 75 SPECint92, and 112 SPECfp92.

- Processor for the 3500 is a 150 MHz Alpha benchmarking at 84 SPECint92, and 128 SPECfp92.

- The specifications were generated with only one graphics accelerator.

- Dual 24-bit frame buffers.

- 24 bits for z.

- The graphics accelerator model number designates the series, E, P, and V, followed by the number of transformation and frame buffer modules: For example, P510 is a P series accelerator with 5 transformation modules (TEM) and 10 frame buffer modules (FBM). Transformation and frame buffer modules may be added for higher performance, up to a limit of 5 FBMs and 3 TEMs for the E series, 10 FBMs and 5 TEMs for the P series, 20 FBMs and 6 TEMs for the V series.

- Both immediate (GL) and retained (PHIGS) mode APIs are available. The GL is provided by a third party vendor, Nth Graphics. Support for OpenGL is planned.

4.6.4 Experimental

Pixel-Planes 5

The first full-size prototype system built by the graphics group of the University of North Carolina at Chapel Hill was Pixel-Planes 4 [Poulton et al. 1987]. That system had a frame buffer with a 1-bit processor for every pixel on the display. It served as the main graphics engine at UNC for many years, and was used to drive several HMDs.

However, it became apparent that a processor for every pixel was not a good way to utilize computing power when primitives typically covered only a small portion of the screen. This led to the design and construction of Pixel-Planes 5 [Fuchs et al. 89], which is a much more modular machine. A block diagram is shown in Figure 4.2.

Pixel-Planes 5 Graphics Multicomputer

Graphics Processor

Powerful, math-oriented processor based on Intel i860's. The GPs traverse the graphics database performing geometric and lighting calculations. A system may have as few as 2 or as many as 50 or more.

Ring Network

5Gigabit/sec multi-token ring network carries control, object data, and pixel traffic. It is implemented as an "active backplane" in 100K ECL logic running at 160 MHz, globally synchronized by a novel "salphasic" clocking scheme.

Graphics Processor

Ring Network

Host Interface

Workstation

Renderer

128x128-processor SIMD "computing surface" based on custom, logic-enhanced memory chips. Renderer also contains two other full custom chips, a controller for the SIMD array and a data "corner turner". Each Renderer can rasterize over 120K Phong- shaded triangles per second, and over 100K spheres per second. A system may contain from one to 25 or more renderers.

Renderer

Frame Buffer

Frame Buffers

System can accept a variety of Frame Buffers and other external interfaces. Currently, these include HiRes (1280x1024, 74Hz, 24bit) and NTSC (640x512, 30Hz, 24bit). Applications are supported by a Host Interface to a Sun4 and a HiPPI interface. A video frame grabber is under construction.
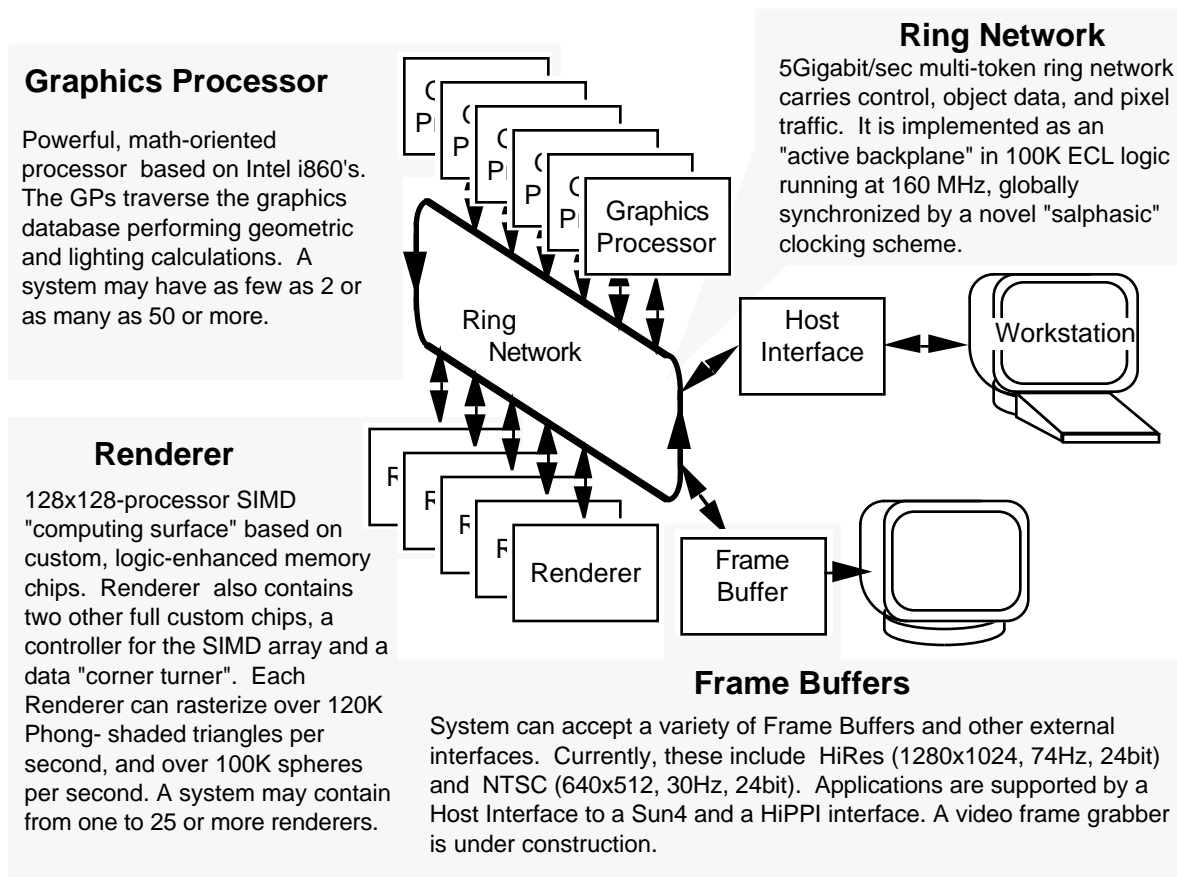
Figure 4.2

Since system components are simply devices on the ring, there is much more flexibility to design special purpose frame buffers. The original frame buffer was of 1280x1024 or 640x480 resolution at 72 Hz and was meant to drive HiRes CRTs. We subsequently built NTSC frame buffers, packaged two to one card for driving HMDs. Currently we have a design for a 180 Hz field-sequential color frame buffer to drive HMDs being developed at UNC.

As with other graphics machines, the performance of Pixel-Planes 5 is difficult to quantify. We have demonstrated rendering rates of over 2 million Phong-shaded individual triangles per second. However, a better basis for comparison is the PLB. Pixel-Planes 5 achieves a verified score of 559 GPCmarks on the "head" benchmark.

PixelFlow

Most parallel graphics architectures, such as that of Pixel-Planes 5, require communications that scale with the number of primitives. This is not a major problem until the required performance of a system outpaces the practical bandwidth of a communications network. The property makes systems such as Pixel-Planes 5 inherently non-scalable. Beyond a certain point, adding more processing power does not increase performance.

To provide an architecture that is scalable with respect to the number of primitives, one can provide multiple communications paths for primitives. This is possible, but since primitives can generate pixels on any part of a display, there must still be a way to allow a primitive to influence any pixel. This is done by sorting pixels instead of primitives via a composition network.
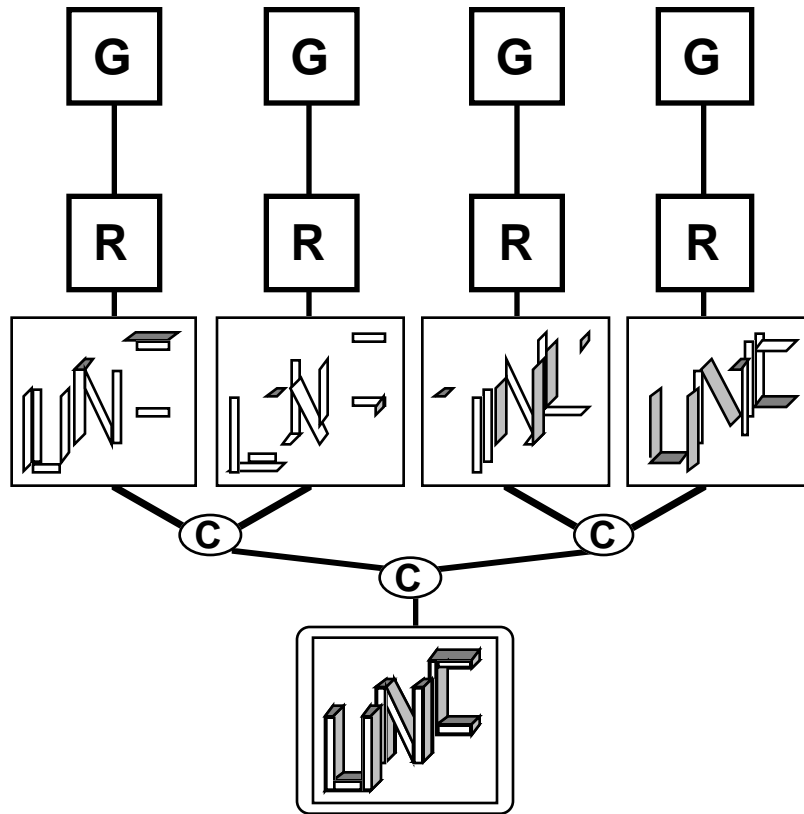
Figure 4.3

The system shown in Figure 4.3 consists of four separate graphics accelerators with geometry (**G**) and rasterization pipelines (**R**), each rendering a complete z-buffered image, but only of a subset of the primitives. Each of these pipelines looks very much like an older graphics system without much parallelism. The images produced by the individual pipelines are composited by a combining network to generate the final image. The data flowing down the tree contains depth, as well as color. The composition network performs z-buffering to generate the final image. To add more performance, one need only add more graphics processing nodes. The bandwidth of the composition network remains fixed and is determined by the product of the frame buffer size, number of subsamples for antialiasing, and the required frame rate. For ease of physical implementation, PixelFlow will have a linear composition network instead of a tree.

Another bottleneck encountered when trying to increase rendering rate is the host and host to image generator link. We expect to drive PixelFlow with a parallel host and parallel data streams to the rendering modules. The software model will generate primitives in parallel. We call this *distributed immediate mode* and have prototyped it on Pixel-Planes 5 by using some GPs as host processor nodes.

4.7   Special Requirements for VE

4.7.1   Input/Output

This is not much of a problem. Most commercial trackers interface via RS232, and one or more RS232 interfaces are standard on most computers. Trackers that require a higher data rate use more complex I/O interfaces, such as IEEE488. Usually these interfaces must be accommodated by adding interface cards.

Output to non-graphical devices, such as those to generate audio, also tends to be at low data rates. RS232 interfaces accommodate most of these needs.

4.7.2   Shared Virtual Worlds

Stereo is often used with head-mounted displays. That alone requires two channels of video. We may also want to support two or more users wearing HMDs. The performance demands on one image generator can grow drastically, though multiple channels are often used in flight simulators. We support two persons wearing HMDs on Pixel-Planes 5, but would find it difficult to sustain the pixel traffic necessary to add more HMDs on one machine.

Another way to support *shared virtual worlds* is to use separate image generators attached by a network. Changes in the dataset are transmitted over the network. This approach splits the intensive rendering and pixel traffic, but limits the dynamics of the dataset to what can be exchanged between machines. It seems better to use a tightly coupled parallel machine with multiple image generation pipelines (a SGI Onyx, for example).

4.7.3   Potential z-buffering problems

Since the hither plane for work using HMDs is often very close to the user and objects can be far away, a deep z-buffer may be necessary. Keep this in mind when choosing an image generator. Twenty four bits should be adequate for most work.

4.7.4   Low latency

Latency is a very big problem for HMD systems. We can divide latency into that caused by the tracking device, and that caused by the image generator. Unfortunately, latency and throughput in an image generator are often at odds. One can use pipelining of successive frames to increase throughput. We do that on the standard PHIGS system on Pixel-Planes 5. This, of course, increases latency. Reducing the latency by eliminating pipelining also reduces the utilization of resources, and consequently the throughput.

Note that the minimum latency is normally determined by the scanout of the display. Usually latency is measured back from the last scan line of the display. This fixes the minimum latency of an NTSC display at one field time, 1/60 of a second.

4.7.5   Correction for optical distortion

The wide field of view optics used in HMDs tends to distort the images by a large amount [Robinett & Rolland 92]. There are several ways to correct this. One is optical, but this solution can increase the weight of a HMD to an unacceptable amount. Another way is to correct by using analog circuitry in the scanning drivers for the display (if it is a CRT). This may be difficult, especially if the distortion is a complex function of screen position.

One can also use the computing power in the image generator to pre-distort the image so it appears correct when viewed through the optics.

4.7.6 Video

Scan converter

A scan converter acts as a format converter for video. It accepts as input a video stream in one format, and produces as output a video stream in a second format. For our purposes, the input is HiRes video, such as used in graphics workstations, and the output is NTSC. Scan converters are expensive, so are not a very practical way to generate video for HMDs.

Composite video

Some NTSC frame buffers generate video with three channels of color, one for each of red, green, and blue. To use this video for typical HMDs, it must be converted to composite video. Devices, called encoders, are available to perform this function. Some are relatively inexpensive with prices in the low hundreds of US dollars. Others, of high quality, can cost as much as $8000. We have found encoders priced under $1000 to be adequate for the current HMDs.

4.7.7 Combining camera video and synthetic imagery

Chroma keyer

A chroma keyer combines two video streams in a controlled fashion. One video stream, called the "key" is gated to the output by default. However, the key color is constantly examined. If a specific color, called the "chroma key color" is detected, the second video stream is inserted into the output, replacing the first.

Thus to selectively mix video, one can use the image generator's output as the key and make the default background color the chroma key color. Camera video can be used as the second video input to the keyer. The result is that any objects generated synthetically appear superimposed on the image of the "real world."

This, of course, does not result in a correct "see-through" HMD, but it approaches one. As with any see-through HMD, tracking and latency problems will be especially apparent and noticeable[Bajura et al. 92].

z-buffering

If one could calculate the distance of objects imaged by the video camera, one could combine the z information with that generated for the synthetic objects to compute visibility for the complete scene, real and virtual. Unfortunately, there are difficult research problems to be solved to compute the distance of objects, and the computational demands to compute the visibility are high.

4.8 Further Reading

For information on the general subject of graphics, the following text is the standard reference:

Foley, J. D., A. vanDam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice*, Addison-Wesley, 1990.

Another good text which provides good coverage for the field of rendering is:

Watt, A. and M. Watt, *Advanced Animation and Rendering Techniques*, ACM Press, New York, 1992.

5.   Detectors

5.1   General

As discussed in Section 2.1, a VE system may be interested in monitoring any number of attributes associated with a user: head position, posture, finger movements, blood pressure, etc. We will call any device used to monitor a user characteristic such as this a *detector*. Although detectors for all of the user actions listed in Table 2.2 have been implemented in some VE system, the most common action detected in current systems is head motion.

5.2   Trackers

5.2.1   Principles of Tracking

Trackers are fundamentally sensors whose task is to detect position and/or orientation of an object and make the information available to the rest of the VE system. The most common function is to report the position and orientation of the user's head. Hand tracking is also common.

For head or hand tracking, there are six types of motion that may be tracked:
- Translation in x, y, z
- Rotation about the x/y/z axes: Roll, pitch and yaw.

Because these motions are mutually orthogonal, there are thus six independent variables or *degrees of freedom* (DOFs) associated with any asymmetrical 3D object. These six numbers are the minimum required to completely specify the position and orientation of a rigid object[4]. A particular tracker may monitor all six or some subset, depending on the implementation. In addition, some trackers monitor only over a limited range of a particular variable. For example, a tracker might detect roll only in the range ±90˚, or x, y, and z only within a sphere of one meter in radius.

Types:

- *Magnetic-* a magnetic field emitted from a transmitter induces current in a receiver according to distance and orientation.

- *Optical-* light from a fixed source or from an object is imaged by a camera device; some number of sources and detectors is used to calculate the transformation to the object.

- *Mechanical-* measures position/orientation via a physical connection to object by jointed linkages. The angles of the joints are used to derive the transformation to the object.

---

[4] Since the hand is not a rigid object, some systems add additional degrees of freedom to account for the joint angles for each finger, as discussed later.

- *Acoustic-* ultrasonic frequency sound is bounced off object and either time-of-flight or phase information is used to derive distance to a number of points on object, which is then used to derive the transformation.

- *Inertial-* accelerometers and gyroscopes are used to detect changes in linear and angular velocity, respectively. These devices are useful for predictive tracking when coupled with other technologies, but are not currently used for full, 6-DOF tracking systems.

Issues:

- *Update rate:* How many measurements are made each second. This can limit how often we can update our display of the VE. Low update rates lead to jerky, unconvincing virtual worlds.

- *Delay/Lag/Latency:* How much time elapses from the moment a user moves until the tracker data reflecting that movement is received by the host?

- *Accuracy:* This is the amount of error in the measurement. Usually given as a bound on the magnitude of the error or as an average error amount. A tracker that has an accuracy of 0.1" will report positions that are (in theory) ±0.1" from the actual position.

- *Resolution:* This is the smallest amount of the quantity being measured that the instrument will detect. A movement smaller than the tracker's resolution will not be reflected in its output.

- *Interference/Distortion:* All trackers except for inertial systems are subject to either interference (such as blocking of the line of sight) or distortions (such as field distortions in magnetic trackers) which can reduce the accuracy or produce gross errors.

- *Absolute vs. Relative:* Trackers may report absolute position/orientation information or just send information on changes.

- *Range:* Working volume and angular coverage. Absolute trackers all have limits on working volume; many systems have limited angular range as well.

- *Size/Weight:* How easy is the tracker to wear on the head or hand?

- *Robustness:* Is the tracker subject to gross errors when its operating environment is degraded?

- *DOFs Measured:* Some trackers measure only a subset of the six DOFs cited above.

- *Safety:* Does use of the tracker pose a long-term (or short-term) health risk?

5.2.2 Commercial and Research Systems

5.2.2.1 Magnetic Trackers

Magnetic trackers have been made by Polhemus, Ascension, Honeywell, Rediffusion, Zeiss, Ferranti, and the Israeli government. The two dominant players in the VE market at present are Polhemus and Ascension.

Magnetic trackers typically consist of a control unit, some number of transmitters (or sources) and some number of receivers (or sensors). The transmitter radiates a magnetic field with is sensed by the receiver, whose measurements are used by the control unit to derive the six DOFs for that receiver.

Advantages:
- No line-of-sight constraints (particularly well-suited for hand tracking)
- Impervious to acoustic interference
- Receivers are generally small and unobtrusive

Disadvantages:
- Distortion/Interference due to metallic objects
- Current systems are very accurate only in small volume
- Cable connection required

Polhemus

Principle of Operation: Alternating-current (AC) magnetic field. Transmitter contains three orthogonal coils that emit a magnetic field when current is passed through them. The receivers also contain three orthogonal coils in which current is induced by the changing magnetic field of the transmitter. Current is supplied to one transmitter coil at a time, and three readings are given by the receiver coils, leading to nine measurements for each measurement cycle. These measurements are then processed in the control unit to compute the 6-DOF solution. Because the AC field induces eddy currents in conductive metals, tracker should be used in an environment free of metallic objects.

Ascension

Principle of Operation: Direct-current magnetic field. Transmitter emits a series of short DC pulses. After eddy currents in surrounding metallic objects have decayed, measurements are taken at receivers. Background magnetism (such as the earth's magnetic field) is subtracted off of the measurements and transformation is calculated. Advantage to this approach is that it is relatively insensitive to conductive metals and less sensitive to ferromagnetic metals, which allows the tracker to be used closer to metal than AC trackers.

Specifications for Polhemus and Ascension

Characteristics in common:

- All trackers' angular range is 360˚ about all axes

- All have RS-232 connections; the Polhemus Fastrak and the Ascension Flock both support high-speed connections as well (IEEE-488 and IEEE-485, respectively)

- Range given is the maximum transmitter-receiver distance; i.e., how far you can get from the transmitter. The working volume diameter is therefore roughly twice this distance.

The table below is a condensation of the complete specs for each product and is thus only a rough guide to each product's characteristics. Consult the manufacturer's information for full specifications.

manu-facturer	name	max # sensors	range (feet)	max update rate	static accuracy (xlate, rot)	resolution (xlate, rot)	price ($US)	price w/ two sensors	notes
Polhemus	Fastrak	4	10'	120 Hz†	0.03" rms 0.15° rms	0.0002" per inch* 0.025°	$5,750	$6,250	a
Polhemus	Tracker (discontinued)	4	5'	60 Hz†	0.1" rms 0.5° rms	0.023" avg 0.1°	N/A	N/A	b
Polhemus	Isotrak (discontinued)	1	3'	58 Hz	0.25" rms 0.85° rms	0.006"/inch 0.35°	N/A	N/A	c
Polhemus	Isotrak II	2	5'	60 Hz†	0.1" rms 0.75° rms	0.0015 per inch* 0.01°	$2,875	$3,375	
Ascension	Flock of Birds	30	3'	144 Hz	0.1" rms 0.5° rms avg	0.03" 0.1° @12"	$3,120	$5,740	
Ascension	Flock w/ ext. range transmitter	30	8'	144 Hz	0.1" rms 0.5° rms avg	0.03" 0.1° @12"	$6,520	$8,965	d

Notes from Specifications:
   a- Resolution for Polhemus trackers is a function of distance. Resolution = (number given above) * (number of inches from transmitter). Example: Fastrak at 30" has a resolution of 0.0002*30 = 0.006".
   b- No longer in production. Replaced by the Fastrak.
   c- Replaced by Isotrak II.
   d- Each receiver/transmitter requires its own control unit, with the exception of the ERT control unit, which can control two transmitters.

†*Note on Number of Sensors:* Polhemus trackers all share the property that tracking more than one sensor reduces the update rate by one over the number of sensors tracked. Tracking two sensors gives a 60 Hz update rate; three gives 40 Hz, etc. An exception to this is that up to four Fastraks can be operated within the same volume at different frequencies without multiplexing, but this requires four separate trackers. The Flock of Birds does not time-slice in this fashion, so adding more sensors does not affect the update rate unless the additional data for the other sensors bogs down the communications channel.

Delay Measurements

Mark Mine at UNC did a comprehensive study of the delay in the trackers in use at UNC. The following figures and text are excerpted (with permission) from [Mine 93]. A detailed explanation of the measurement procedures is given in the technical report.

Note: The delay numbers shown here reflect the trackers at UNC in one of many possible configurations; the delay in systems at other sites may be higher or lower, depending on the configuration used. In particular, discrepancies between manufacturer's latency numbers and the numbers given here may be due to the way the UNC system is set up.

Tracking System Delays
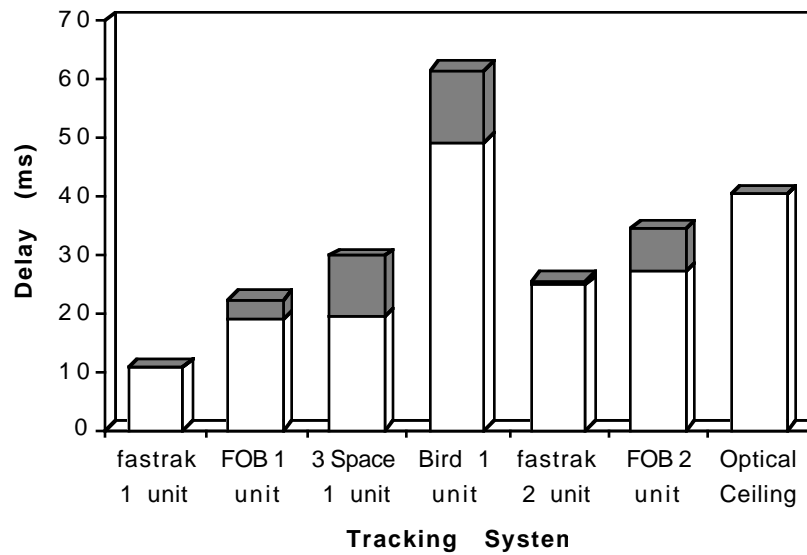
Figure 5.1: Tracker delay

Notes:
1)   Shaded area represents communication delays.
2)   Flock of Birds not in optimum communication configuration- the IEEE-485 interface should reduce the communication delay depicted above significantly.
3)   Fastrak timing with position and orientation filter on.

In tabular form the results are:

| Tracking System	Delay (ms)		
	Tracker	*Communication*	*Total*
Polhemus Fastrak - 1 Unit	10.65	0.3	10.95
Ascension Flock of Birds - 1 Unit	18.96	3.65	22.61
Polhemus 3SpaceTracker - 1 Unit	19.23	10.4	29.63
Ascension Bird - 1 Unit	49.15	12.5	61.65
Polhemus Fastrak - 2 Units	24.9	0.6	25.50
Ascension Flock of Birds - 2 Units	27.09	7.3	34.39
UNC Optical Ceiling	40.5	-	40.5

Recall that the total delay is the time between a change in the position/orientation of the HMD and the receipt of the corresponding tracking data at the host computer.

5.2.2.2   Optical Trackers

Although more optical trackers have been developed than any other technology type, they are not in common use for general-purpose VE systems. Most of the commercial systems are employed for military applications.

Types:

- Outside-In Trackers- Fixed sensors in the environment image a pattern of beacons with known geometry mounted on the tracker; object's position and orientation in the environment is derived from these images. See Figure 5.2.

- Inside-Out Trackers- Sensors are mounted on the object to be tracked, where they image some number of fixed beacons in the environment. The position and orientation is derived from a similar set of measurements as in the outside-in system. See Figure 5.3.

- Natural environment trackers- No beacons are used. Sensors are mounted on the object to be tracked and the system derives the 6D information by imaging the environment and looking at image shifts from one frame to the next. Multiple sensors can be used to cover the six degrees of freedom. No complete implementation of this concept has yet been done to our knowledge. Partial implementations are described in [Bishop 84] and [Tanner 84].
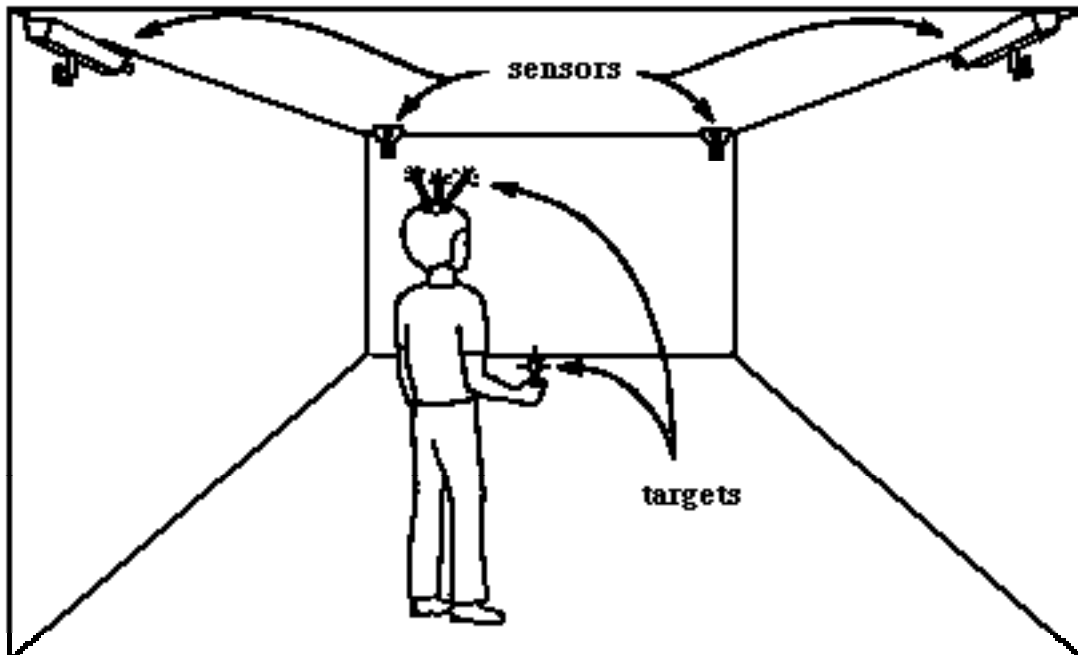
Reproduced from [Bishop 84]

5.2 Outside-In Tracking
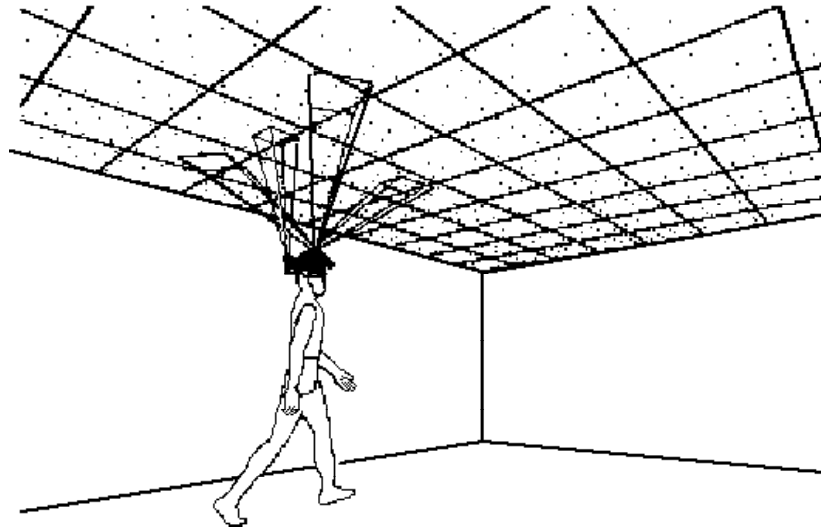
Drawing: Mark Ward

Figure 5.3.   Inside-out tracking

Advantages:
- Can be very fast
- Accurate for small volume
- Immune to magnetic, acoustic interference/distortion

Disadvantages:
- Line-of-sight restriction
- Angular range is often restricted
- Interference due to other light sources
- Current inside-out implementations are cumbersome
- Difficult to track multiple objects in same volume

Commercial Optical Systems

In general, commercial systems tend to be small-working-volume, high-cost systems suitable for military applications but not particularly well suited for general-purpose immersive VE systems. In particular, many of these systems only track 2-3 degrees of freedom (typically angular: yaw, pitch and possibly roll).

manufacturer	product name	DOFs	working volume	accuracy	update rate	cost ($US)
Northern Digital	Optotrack 3000	3	1 m cube	0.1 mm	128 Hz	$59,000
GEC Ferranti	GRD-1010	6		0.1"/0.6°	240 Hz	$50,000
Selspot	Selspot II	3	1 m cube	5 mm	variable	$44,00
Qualisys AB	MacReflex	2	1 m cube	5 mm	50 Hz	$18,000
Spatial Positioning Systems	RtPM	3 or 6		0.01" at 100'	100 Hz	$50-70,000

Note: Some of the trackers listed can be used in larger working volumes with a proportionate decrease in accuracy.

UNC Ceiling Tracker

The ceiling tracker developed at UNC is an inside-out implementation. Four lateral-effect photodiode sensors mounted on a HMD look out at a specially constructed ceiling composed of tiles with infrared LEDs mounted in them (see Figure 5.3). The LEDs are strobed in sequence, and their images on the sensor arrays are used to derive the position and orientation of the HMD. The ceiling is presently 10 feet by 12 feet, which gives a large tracker volume for walking around architectural and other large models.

The system has a linear resolution of 2 mm and an angular resolution of 0.2°. The update rate varies from 20 to 100 Hz depending on the number of LEDs visible. The delay also varies, but is typically around two frames, or 30 msec.

A complete description is given in [Ward 92].

5.2.2.3   Mechanical Trackers

Mechanical trackers are in fairly wide use in many different application areas, from stereotactic surgery to atomic research. The basic idea for a 6 DOF mechanical tracker is that there must be at least six orthogonal joints, one for each degree of freedom, and the joint angles must be measured accurately. These measurements, coupled with the knowledge of the linkage geometry, allow for very accurate tracking.

Advantages:
- Accurate
- Can serve as counter-balance to hold display
- Can be used for force feedback
- No magnetic, line-of-sight, acoustic interference constraints

Disadvantages:
- Linkages can be intrusive, heavy, and have high inertia
- Difficult to track multiple objects in same volume
- Difficult to implement for large volume
- Central part of working volume is inaccessible for some trackers
- Hard to keep DOFs orthogonal- gymbal-lock problem

Shooting Star Technologies

The ADL-1 is a low-cost, low-latency mechanical tracker. It claims an update rate of 300 Hz, an accuracy of 0.2 inches in position and a resolution of 0.025." The working range is a cylinder 1.5 feet high with a radius of 1.5 feet.

Fake Space

Fake Space's tracker is typically sold with its display, as described in Section 3.2.6. Its range is a cylinder of radius 2.5 feet and height of 2.5 feet, with a 1-foot excluded inner core. Update rate is 60 Hz; latency is determined by RS-232 communication time. Translational and rotational accuracy figures are not given; they quote the shaft encoders for each joint at 4000 counts per 360 degrees.

5.2.2.4   Acoustic Trackers

Types:
   • Time of flight (TOF): Send sound from set of transmitters to set of receivers and measure the elapsed times. These times plus the known geometry of the transmitters/receivers is used to derive the position and orientation.

   • Phase coherent (PC): Comparison of phases of emitted waves with phase of a reference wave. Only gives relative change, so errors accumulate.

Advantages:
   • No electro-magnetic fields
   • Can be implemented fairly cheaply

Disadvantages:
   • Limited range
   • Subject to acoustic interference, changes in air density
   • Line-of-sight restriction

Logitech 6-DOF Ultrasonic Head Tracker
   • Cost: $1,000
   • Accuracy: 2% of distance from source/ 0.1 deg orientation
   • Working volume: 5', 100° cone
   • Lag: 30 msec
   • Update rate: 50 Hz
   • Restricted to line-of-sight

5.3   Other Detectors

Hand/finger action detection devices
- Exos *Dextrous Hand Master*-   20-DOF hand-motion sensing device measures joint angles of hand precisely using Hall-effect sensors. Senses 3 bending DOFs per finger, plus one side-to-side angle. Accuracy is better than 1˚ in most cases.
- Virtual Technologies (Virtex) *CyberGlove*- Uses foil strain-gauge technology to measure finger joint angles
- VPL *DataGlove*- Uses fiber-optic sensors to measure joint angles; also manufactured by Greenleaf Medical Systems of Palo Alto, CA.
- Mattel *Power Glove*- [Discontinued] Uses resistive sensors to measure joint angles and ultrasonic sensors for hand tracking.

3D Mice
- Polhemus *3Ball*- Polhemus tracker sensor (6 DOF) mounted in a billiard ball equipped with a single push-button for initiating actions in VE. This is a commercial version of the UNC 3D mouse cited in [Brooks 86].
- Ascension Bird 3-Button Mouse- Ascension Bird sensor (6 DOF) mounted in a 3-button mouse housing.
- Logitech *3-D Mouse*- 2/6 DOF ultrasonic tracker with 5 push buttons
- SimGraphics *Flying Mouse*- 2/6 DOF mouse using magnetic tracking with 3 push buttons
- Gyration *GyroPoint*- 3 DOF mouse with 5 buttons. Uses gyroscopes to detect changes in orientation.
- Digital Image Design *Cricket*- Prototype hand grip device can have tracker sensor mounted inside; has tactile display vibration, pressure return at trigger and grip, and directional thumb button that returns pressure and direction.

Speech Recognition

Issues:
- Speaker-dependent/independent
- Vocabulary size
- Continuous speech vs. discrete words
- Grammar (restrictions on inputs to system)
- Effects of background noise
- Speaker enunciation
- Speech-to-text ↔ Text-to-speech

company	product name	speaker-dependent / independent	discrete / continuous	vocabulary size (words)	text to speech capability	price ($US)
Articulate Systems	Voice Navigator	dependent	discrete	|200; 1000 per file	no	1,295
Cherry Electrical	VoiceScribe 1000 Plus	dependent	discrete	1,000	option	3,600
Covox	Voice Master Key	dependent	discrete	64	option	150
Dragon Systems	Dragon Dictate 30K	independent (adaptive)	discrete	25,000	no	9,000
Dragon Systems	Dragon Writer 1000	dependent	discrete	1,000	no	3,600
Kurzweil Applied Intelligence	VoiceReport	dependent	discrete	5,000	option	26,500
Scott Instruments	SIR Model 20	independent	discrete	160	no	1,495
Speech systems	DS200	independent	continuous speech	40,000	option	33,900
Texas Instruments	TI Voice Card	dependent	continuous speech	50	yes	995
Voice Connexion	Introvoice-5 PTVC-756	dependent "	discrete "	250 250	option yes	495 2,995
Verbex Voice Systems	Verbex 5000 Verbex 6000 Verbex 7000	dependent " "	continuous " "	80 300 1980	yes option yes	5,600 5,600 9,600
Voice Control Systems	TeleRec VR-4	independent "	discrete "	16 50	no no	2,495 3,600
Voice Processing Corp	VPC-1000	independent	continuous	13	no	5,500
Voice Recognition Technologies	Voicebox	dependent	discrete	500	no	395
Votan	Voice Card VPC-2100	dependent	continuous	125	no	1,750

All systems claim accuracy of over 90%.

Eye Tracking
- Skalar Medical *IRIS*- Uses differential reflection to cover 30x10˚ range with an accuracy of 0.1˚.
- ISCAN- Uses video image-processing to measure eye movement; range is 25x20 and accuracy is 1˚.
- Applied Science Labs *Model 4000 SU-HMO*- Video image processing eye tracker; range is 50x40; accuracy is 1˚.

Other Detectors
- ARRC/Airmuscle *Datacq II*- Input device to the ARRC/Airmuscle Teletact II. Forces generated by gripping an object are measured and recorded with this device for later display with the Teletact II glove.

- BioControl Systems *Bio-signal Processing Unit*- Uses dermal electrodes to track muscle activity or brain waves. Has 8 independent channels.

- CM Research *DTSS X/10*-   See entry in Section 3.4 for temperature sensing.

6.   Acknowledgments

7.   References

Bajura, Michael, Henry Fuchs, and Ryutarou Ohbuchi. 1992. Merging Virtual Objects with the Real World. *Computer Graphics: Proceedings of SIGGRAPH 92,* 203-210.

BBN. 1992. Bolt, Beranek and Newman Report No. 7661: Virtual Environment Technology for Training. Prepared by The Virtual Environment and Teleoperator Research Consortium affiliated with MIT. March 1992.

Bhatnagar, Devesh. 1993. Position trackers for head mounted display systems: A survey. UNC Technical Report TR93-010.

Bishop, Gary, and Henry Fuchs. 1984. The Self-Tracker: A smart optical sensor on silicon. *1984 conference on advanced research in VLSI.*   MIT. 1/23/84.

Brooks, Frederick P. Jr. 1986. Walkthrough- A dynamic graphics system for simulating virtual buildings. *Proc. 1986 Workshop on Interactive 3D Graphics.* Chapel Hill.

Bruce, Vicki, and P. Green. 1990. *Visual perception: Physiology, psychology and ecology.*   Lawrence Erlbaum. E. Sussex, U.K.

Burdea, Grigore, J. Zhuang, E. Roskos, D. Silver, and N. Langrana. 1992. A portable dexterous master with force feedback. *Presence.* 1:1.

Butterworth, Jeff, Andrew Davidson, Stephen Hench, and Marc T. Olano. 1992. 3DM: A Three Dimensional Modeler Using a Head-Mounted Display, *Proc 1992 Workshop on Interactive 3D Graphics,* 135-138.

Clarke, Tom. 1993. Handout describing the UCF/IST HMD.

Deyo, Roderic, and D. Ingebretsen. 1989. Notes on real-time vehicle simulation. *ACM SIGGRAPH '89 Course Notes: Implementing and interacting with real-time microworlds.*

Ferrin, Frank. 1991. Survey of helmet tracking technologies: Large-screen-projection, avionics, and helmet-mounted displays. *Proceedings of SPIE 1456*

Fuchs, Henry, John Poulton, John Eyles, Trey Greer, Jack Goldfeather, David Ellsworth, Steve Molnar, Greg Turk, Brice Tebbs, and Laura Israel. 1989. Pixel-Planes 5: A Heterogeneous Multiprocessor Graphics System Using Processor-Enhanced Memories. *Computer Graphics: Proceedings of SIGGRAPH 89*, Vol. 23, No. 3, 79-88.

Holloway, Richard, Henry Fuchs, and Warren Robinett. 1992. Virtual-World Research at the University of North Carolina at Chapel Hill as of February 1992. *Proceedings of Computer Graphics International.* Japan. June.

Hornbeck, Larry. 1989. Deformable-mirror spatial light modulators. Proc. *SPIE Volume 1150: Spatial Light Modulators and Applications III.* August, San Diego.

Jain, Anil K. 1989. *Fundamentals of digital image processing.* Prentice-Hall, Englewood Cliffs, NJ.

Iwata, Hiroo. 1990. Artificial reality with force-feedback: Development of desktop virtual space with compact master manipulator. Proc. *ACM SIGGRAPH* 1990.

Levine, Martin D. 1985. *Vision in man and machine.* McGraw-Hill. New York.

Meyer, Kenneth, H. Applewhite, and F. Biocca. 1992. A survey of position trackers. *Presence*. 1:2.

Mine, Mark. 1993. Characterization of end-to-end delays in head-mounted display systems. UNC Technical Report TR93-001. Chapel Hill.

Minsky, Margaret, M. Ouh-Young, O. Steele, F. Brooks, and M. Behensky. Feeling and seeing: Issues in force display. *ACM Computer graphics.* 24:2.

Molnar, S., and H. Fuchs. 1991. Advanced Raster Graphics Architecture, in Foley, J. D., A. van Dam, S. K. Feiner, and J. F. Hughes, *Computer Graphics: Principles and Practice,* Addison-Wesley. 1990.

Molnar, S., J. Eyles, and J. Poulton . 1992. PixelFlow: High-Speed Rendering Using Image Composition, *Computer Graphics: Proceedings of SIGGRAPH 92*, 231-240.

Monkman, G.J. 1992. An electrorheological tactile display. *Presence.* 1:2.

Ouh-Young, Ming, D. V. Beard and F. P. Brooks, Jr. 1989. Force display performs better than visual display in a simple 6-D docking task. *Proceedings of IEEE 1989 Robotics and Automation Conference.* Scottsdale, Ariz.

Piantanida, Tom, D. Boman, and J. Gille. 1993. Human perceptual issues and virtual reality. *Virtual Reality Systems.* 1:1.

Pimental, Ken, and K. Teixeira. 1992. *Virtual Reality: Through the new looking glass.* Intel/Windcrest/McGraw-Hill. New York.

Poulton, John, Henry Fuchs, John Austin, John Eyles, and Trey Greer. 1987. Building a 512 x 512 Pixel-Planes System, *Proceedings of the 1987 Stanford Conference on Advanced Research in VLSI.*   MIT Press, 57-71.

Rebo, Robert. 1988. A helmet-mounted virtual environment display system. Master's thesis, AFIT. December.

Rheingold, Howard. 1991. *Virtual reality.* Summit books. New York.

Robinett W., and J.P. Rolland 1992. "A Computational Model for the Stereoscopic Optics of a Head-Mounted Display." *Presence*, 1:1.

Robinett, Warren. 1992. Synthetic experience: A proposed taxonomy. *Presence.* 1:2.

Sutherland, Ivan. 1965. The ultimate display. *Information Processing 1965: Proceedings of IFIP Congress 65.* 506–508.

Tanner, J.E., and C. Mead. 1984. A correlating optical motion detector. In *1984 conference on VLSI.*

Taylor, Russell II, W. Robinett, V. Chi, F. Brooks, W. Wright, R.Williams, and E. Snyder. 1993. The Nanomanipulator: A Virtual-Reality Interface for a Scanning Tunneling Microscope. To appear in *Computer Graphics: Proceedings of SIGGRAPH 93.*

Teller, Seth, and C. H. Sequin. 1991. Visibility Preprocessing for Interactive Walkthroughs. Proceedings *of SIGGRAPH 91*, 61-69.

Ward, Mark, R. Azuma, R. Bennett, S. Gottschalk, and H. Fuchs. 1992. A Demonstrated Optical Tracker with Scalable Work Area for Head-Mounted Display Systems. Proc. *1992 Symposium on Interactive 3D Graphics.* Cambridge, MA.

Wang, Jih-Fang. 1990. A real-time optical 6D tracker for head-mounted display systems. PhD dissertation. University of North Carolina, Chapel Hill.

Williams, L. 1978. Casting Curved Shadows on Curved Surfaces.*Computer Graphics: Proceedings of SIGGRAPH 78*, 270-274.

Williams, L. 1983. Pyramidal Parametrics. *Computer Graphics: Proceedings of SIGGRAPH 83, 1-11.*