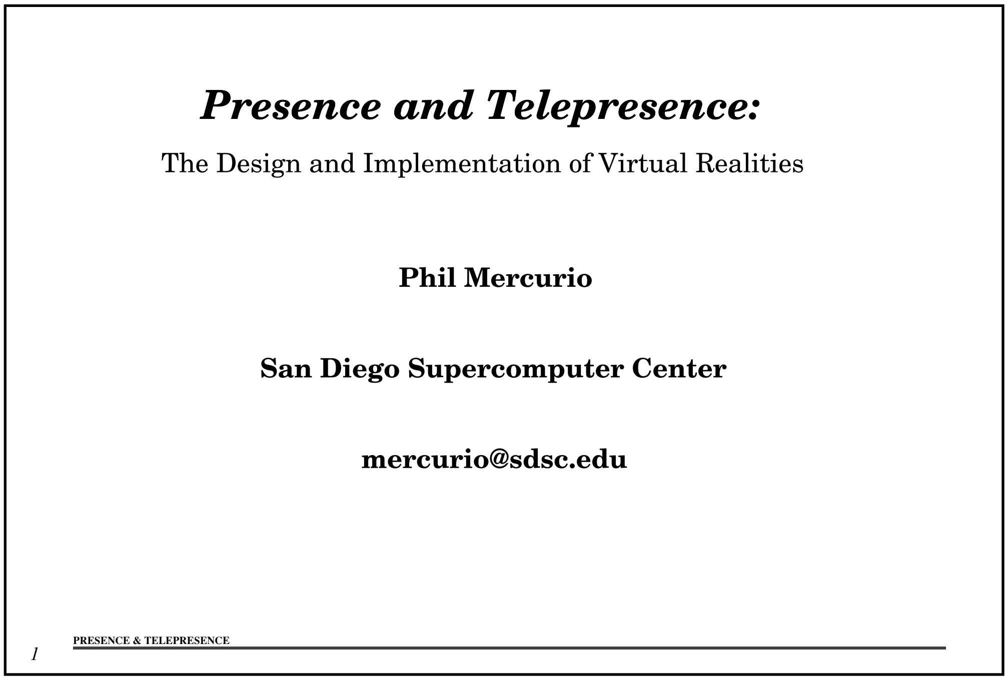


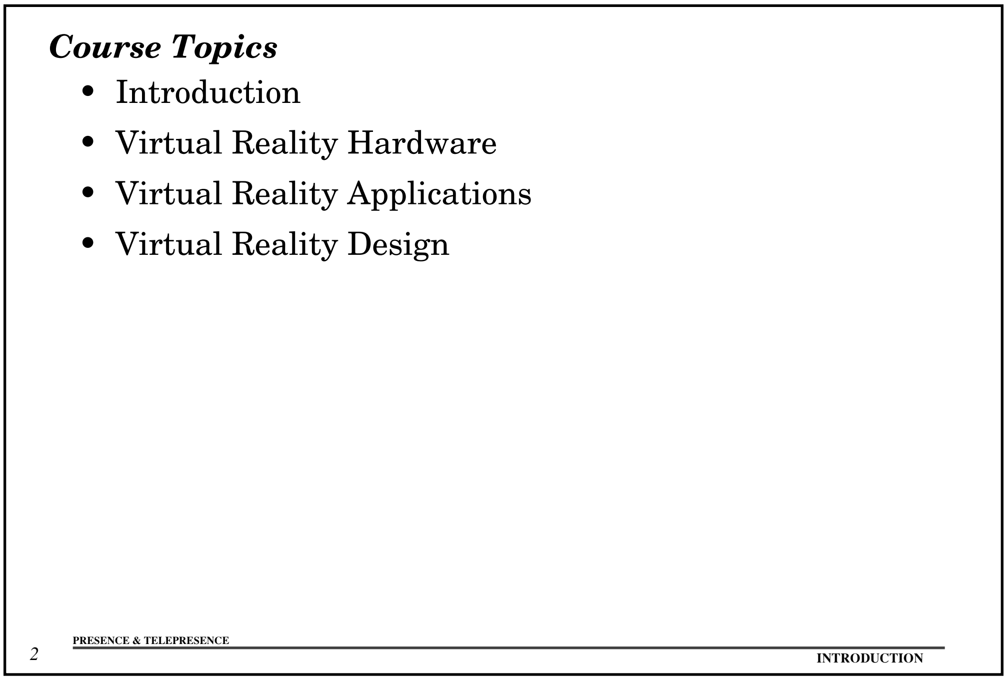
Presence and Telepresence:

The Design and Implementation of Virtual Realities

Phil Mercurio

San Diego Supercomputer Center

mercurio@sdsc.edu



Course Topics

- Introduction
- Virtual Reality Hardware
- Virtual Reality Applications
- Virtual Reality Design

Introduction

- What is Virtual Reality?
- Cyberspace
- Presence and Telepresence
- Additional Terms
- The History of Virtual Reality

What is Virtual Reality?

- The public notion of VR, largely propagated by Jaron Lanier, is of a hardware setup involving a head-mounted binocular video display and a glove with sensors.
- Many have seen this as an exclusive definition, and some shy away from the term, in favor of
  - virtual environment,
  - virtual world, etc.
- A better definition: a multi-sensory, computer-mediated simulation, usually with a first-person point-of-view
- VR as Metaphor

VR as Metaphor

- Let's break it down:
  - virtual: Existing or resulting in essence or effect though not in actual fact
  - reality: That which exists objectively and in fact
- Another definition of VR: virtual reality is a metaphor for system design in which elements of the system are made to appear and behave like their real-world counterparts
  - Note that computers are not mentioned in this definition
  - VR is a metaphor that can be applied to any medium
- The user interface community has a similar notion: user-centered design

Cyberspace

- Cyberspace was originally coined by William Gibson in the 1984 novel Neuromancer:
  - "Cyberspace. A consensual hallucination experienced daily by billions of legitimate operators, in every nation, by children being taught mathematical concepts...A graphic representation of data abstracted from the banks of every computer in the human system."
- Some less poetic definitions might be:
  - an electronically-mediated reality
  - a place where data lives
  - a space defined, not by what matter it contains, but by what data it contains or can support
- The telephone network is a cyberspace, so is the network of electronic mail
- Cyberspace (continued)

Cyberspace (continued)

- We already live in cyberspace
- Sometimes used interchangeably with virtual reality, but there's no reason why a cyberspace has to be designed to appear virtually real

Presence and Telepresence

- Presence is the subjective perception of being in a cyberspace
- The degree to which a virtual reality is effective can be measured (again, subjectively) as the degree of presence
- Achieving presence is what VR is all about
- Telepresence is the subjective perception of being in another physical space
- Also known as "tele-existence" or "telerobotics"

Additional Terms

- Immersion
  - In this context, a virtual reality which uses a head-mounted display and sensor glove
  - Some mandate that only an immersive virtual reality should be called a virtual reality; this is a narrow-minded view
- Affordance
  - A term from the user interface community referring to a (usually visible) element of an interface that implements a task
  - A doorknob is an affordance for opening the door
  - The close box in the corner of a Macintosh window affords closing the file or exiting the application

The History of Virtual Reality

- Prehistory to Morton Heilig
- Ivan Sutherland
- Tom Furness
- Myron Krueger
- Scott Fisher & Michael McGreevy
- Jaron Lanier
- Randy Pausch

Prehistory to Morton Heilig

- The quest for reality in media can be traced back to cave paintings, and has been part of every medium invented
- For example, stereoscopy (3-D visuals) is older than photography (Wheatstone's stereoscope, 1833)
- 1962: Morton Heilig patented the Sensorama, a device which played a multisensory cinema with stereophonic sound, stereoscopic visuals, tactile feedback, and olfactory stimulation
  - Not interactive, but in many ways has yet to be duplicated

Ivan Sutherland

- Generally regarded as the father of VR (and of computer graphics in general)
- Built the first interactive graphics system (Sketchpad) in 1965
- Built the first head-mounted display in 1968
- First HMD was binocular, but not stereoscopic
- Wrote a seminal article outlining much of the field of VR, "The Ultimate Display", in 1965
  - "The screen is a window through which one sees a virtual world. The challenge is to make that world look real, act real, sound real, feel real."

Tom Furness

- Extensive work on the human factors of VR as part of the SuperCockpit project, begun in 1966
- Used head-mounted display as interface to flight simulator
- All instruments, sensors, and flight path displayed in real-time
- "Look and shoot" feature, capitalizing on innate human abilities
- Used glove with piezoelectric vibrotactile actuators
- 25 years of research into human factors of VR
- Later founded the Human Interface Technology Lab at U Washington

Myron Krueger

- Author of "Artificial Reality" in 1983, updated in 1991 as "Artificial Reality II"
- Developed many art pieces involving reactive environments, rooms with large displays, sensors; whatever is necessary to give the participant an interactive experience
- Strong proponent of the notion that virtual realities don't need to be encumbering, that one shouldn't need to wear special clothing or hardware to enter a virtual world

Scott Fisher & Michael McGreevy

- Researchers at NASA Ames who pioneered VR within NASA, primarily for telerobotic space exploration: VIVED (Virtual Environment Display)
- Developed their own head-mounted display, using LCDs from portable televisions, in '86
- Contracted with VPL Research to develop the DataGlove
- NASA Ames was also the birthplace of the Convolvotron 3-D audio system
- VIVED evolved into VIEW (Virtual Interface Environment Workstation) which used and developed the BOOM

Jaron Lanier

- The most visible figure in VR in the past 5 years, claims to have coined the term "virtual reality"
- Developed the DataGlove as an interface to a visual programming language, hence the name VPL Research
- Developed the EyePhone, an HMD using LCDs, based on the work at NASA Ames
- VPL turned into a very litigious company, attempting to patent everything
  - One patent claims the rights to any manipulation of a graphic object in a virtual space--they have never defended it successfully
- VPL is now out of business, their patents the property of a French company, Thomson Digital

Randy Pausch

- Researcher at the University of Virginia Computer Science department
- Developed a VR system for $5000, using PCs and Private Eyes, essentially starting the homebrew VR field
- Wrote "Virtual Reality on $5 a Day"
- Active in human factors and engineering analysis of current VR systems, and in making VR research part of the computer-human interaction field

Virtual Reality Hardware

- Computers
- Displays
- Input Devices
- Haptic Devices
- Sound and Speech
- Other Devices

Computers

- Graphics Capabilities
- Workstation vendors
- Personal Computers
- Game Machines

Graphics Capabilities

- Most personal computers and workstations are equipped for raster graphics
- Color capabilities
  - 1-bit (black & white)
  - 8-bit (256 colors)
  - 24-bit (16M colors--true color)
- Resolutions
  - 640 x 480 (video resolution)
  - 1280 x 1024 (workstation resolution)
- Speeds
  - Vectors per second
  - Polygons per second
  - Pixels per second

Workstation vendors

- Silicon Graphics
- SGI Graphics Systems
- Sun Microsystems
- Other vendors

Silicon Graphics

- Iris Crimson: $35K to $130K (& up)
  - MIPS R3000
  - Elan, VGX or RealityEngine graphics
- Iris Indigo: $10K & up
  - MIPS R4000 or R4400 processor
  - Indigo^2 Extreme: $35K
    - 415K shaded, lighted polys/sec
- Onyx: $110K to $615K
  - Multiple R4400s (up to 24)
  - VTX and RealityEngine^2 graphics
- Power Challenge: $100K & up
  - TFP 64-bit MIPS super-scalar processor(s) (up to 18)
  - up to 5.4 GFLOPS

SGI Graphics Systems

- RealityEngine
  - highest quality graphics option for SGIs
  - very realistic imagery
  - full-frame stereo
  - two output video signals (for stereo)
  - 1.2M polygons/second
- RealityEngine^2 (4Q '93)
  - Up to 2M polygons/second
  - 900K textured, anti-aliased triangles/second
  - Up to 3 graphics systems per Onyx
- VTX (4Q '93)
  - 1/2 of RealityEngine^2 performance

Sun Microsystems

- SparcStation
  - desktop, $4K to $15K
  - 8-bit graphics, some acceleration
- Evans & Sutherland Freedom
  - 500K to 3M polygons/second
  - $50K to $75K
- Fujistu graphics system
  - 5 Mpolys/sec

Other vendors

- Hewlett-Packard
  - systems up to 880K polys/sec, $105K
- IBM RS6000
  - Use SGI hardware for graphics
  - fast processing speed
- DEC Alpha
  - also fast processors
  - Kubota-Pacific graphics accelerators, slightly better price/performance than SGI

Personal Computers

- Many homebrew or low-cost VR systems are built around PCs
- Amiga
- IBM
- Macintosh

Game Machines

- Nintendo Powerglove
- Sega Genesis Home VR
  - $200 add-on, 4Q '93
  - Full-color head-mounted display
  - Proprietary head-tracking (left-right & up-down only)
  - $80 body tracking???
- 3DO
  - $700, 32-bit, double-speed CD, 16M colors
  - Backed by Electronic Arts, Time-Warner, Matsushita, others
  - Christmas '93?

Displays

- Monitor-based Stereoscopy
- Head-Mounted Displays
- HMD Advantages/Disadvantages
- New Consumer HMDs
- Private Eye
- BOOM
- CAVE
- Laser Scanning
- Comparison Chart*

Monitor-based Stereoscopy

- What is Stereoscopy?
- Side-by-side
- Side-by-side (example)
- Anaglyphic
- Pulfrich Effect
- LCD Shutter Glasses
- LCD Shutter Glasses (continued)
- Polarizing Plate
- Parallax Illumination

What is Stereoscopy?

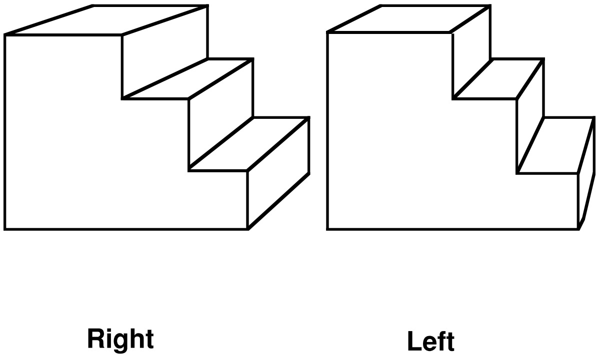
- Humans are binocular--they perceive a 3-D world via two 2-D views, separated by ~6.5 cm
- This sense is called stereopsis, from the Greek, meaning solid viewing
- Stereoscopy delivers the left view to only the left eye of the viewer, and the right view to only the right eye
- Although there are several other cues that lead to 3-D perception, stereopsis is a very powerful technique for achieving presence
- 5% to 20% are stereo-blind

Side-by-side

- Images are placed next to each other on display
- Free viewing
  - L R Viewer goes wall-eyed (difficult)
  - R L Viewer crosses eyes (easier)
- Assisted
  - Mirror box
  - Prism glasses
  - Even a piece of cardboard placed vertically between views will aid the viewer

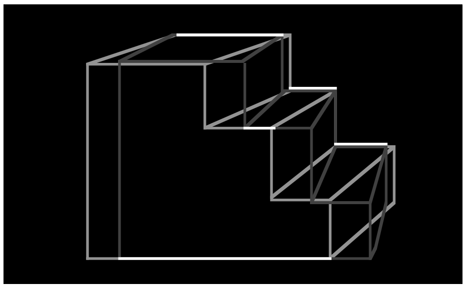
---

Side-by-side (example)



Anaglyphic

- Left/right images displayed on top of each other in different colors (red/green or red/cyan are common)
- Viewer wears glasses with colored filters
- Very cheap, but essentially monochrome



Pulfrich Effect

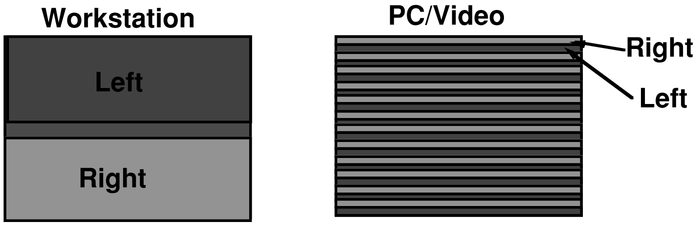
- When viewing a moving image, the user can see two different images if a filter is placed over one eye
  - Usually a neutral-density filter
  - Also works with sunglasses with one lens removed
- The perception in the covered eye is slowed, thus that eye sees an older image
- If the images are created correctly (for example, rotating the scene about a vertical axis), the current image and the delayed image will form a stereo pair
- Cheap and capable of being broadcast, but requires constant motion

LCD Shutter Glasses

- The computer display alternates between the left and right views (time-muliplexing)
- The viewer wears glasses with liquid crystal shutters, synchronized to the display
- Workstations
  - Often implemented by drawing the left image in the top half of the display and the right image in the bottom half
  - Workstation runs at twice the normal frame rate (120 Hz) and expands each half-image to full screen
  - SGI's Reality Engine first to provide two full-screen stereo buffers

LCD Shutter Glasses (continued)

- Personal computers, video
  - Glasses synchronized to video field rate
  - Left image drawn in even scanlines, right image drawn in odd scanlines
- Both techniques result in half the vertical resolution

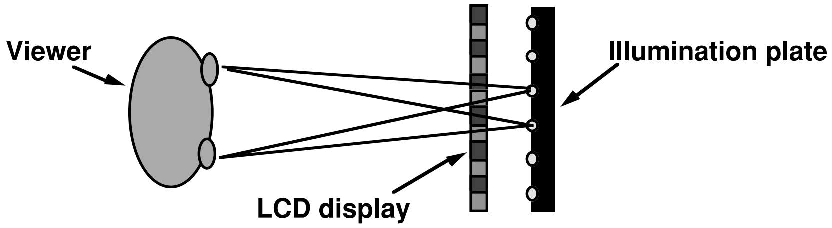


Polarizing Plate

- Computer time-multiplexes left/right views
- Monitor is fitted with a polarizing plate which alternates between clockwise and counterclockwise polarization
- Viewer wears passive glasses with matching polarizing filters
- Not as popular as LCD-shutter-based systems, because plate interferes with non-stereo use

Parallax Illumination

- Autostereoscopic--requires no glasses or other equipment to view
- A 640 x 480 pixel LCD is back-lit by an illumination plate with 320 fine, bright, vertical lines
- Because of parallax, a viewer at the correct distance sees the illumination through the odd columns in the left eye, even columns in the right eye



Head-Mounted Displays

- Two displays mounted one in front of each eye
- Older HMDs used CRTs, sometimes with mirrors or fiber optics
- Most modern HMDs use LCDs, similar to those used in hand-held televisions
- Typical resolutions
  - CRTs: 640x480 to 1280x1024 (mono)
  - LCDs: 320x240 to 700x500 (color)
- Higher resolutions are much more expensive
- Most existing HMDs (VPL EyePhone, Virtual Research Flight Helmet) use 320x240 LCDs

HMD Advantages/Disadvantages

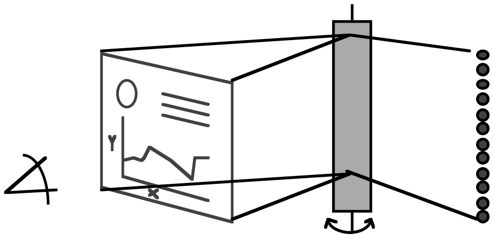
- Advantages
  - One display per eye (two NTSC video inputs) makes stereopsis easy
  - Immersion in virtual world, outside world is occluded
  - Wide field of view
  - Head position can be tracked to enable natural navigation of virtual world
- Disadvantages
  - Currently, extremely poor resolution
  - Occluding outside world may not be desired (look-through displays are under development)
  - Weight of display and cables lead to fatigue and possible hazards in use

New Consumer HMDs

- Sega HMD (described above, slide 27)
- Virtual Vision
  - Single display look-thru video
  - Targeted for walk-around television viewing
  - $900, 2Q '93?
- Sony Visortron
  - 2 0.7", 103K pixel displays, bud earphones
  - 50 degree field of view
  - 9 ounces!
  - Targeted toward mass-market, for television viewing, but 2 displays are separate and could display stereo images

Private Eye

- Lightweight, monochrome, monoscopic display worn on the head
- 720x280 pixel red-on-black display, see-through, $500
- Column of 280 red LEDs, swept horizontally by a vibrating mirror
- Two Private Eyes can be paired to make a cheap HMD



BOOM

- Binocular display supported by a mechanical linkage
- Advantages
  - Linkage supports weight of display, allowing use of CRTs, thus higher resolution
  - Linkage is also fitted with sensors to provide head tracking; more accurate than spatial trackers
  - Linkage is counterbalanced to remain stationary, allowing another person to examine same view
- Disadvantages
  - Current BOOM (Fake Space Labs) uses monochrome CRTs
  - Color CRTs too heavy, too delicate, and emit too much radiation to be used
  - Mechanical linkage limits some movement

CAVE

- Room consisting of rear-projected video screens--4 walls, floor, and ceiling
- Each screen displays a stereoscopic image, delivered to the user via LCD shutter glasses
- User's head position is tracked to provide highly accurate perspective
- Advantages
  - High-resolution
  - Immersive, but not intrusive
- Disadvantages
  - Expensive
  - Difficult to build/maintain

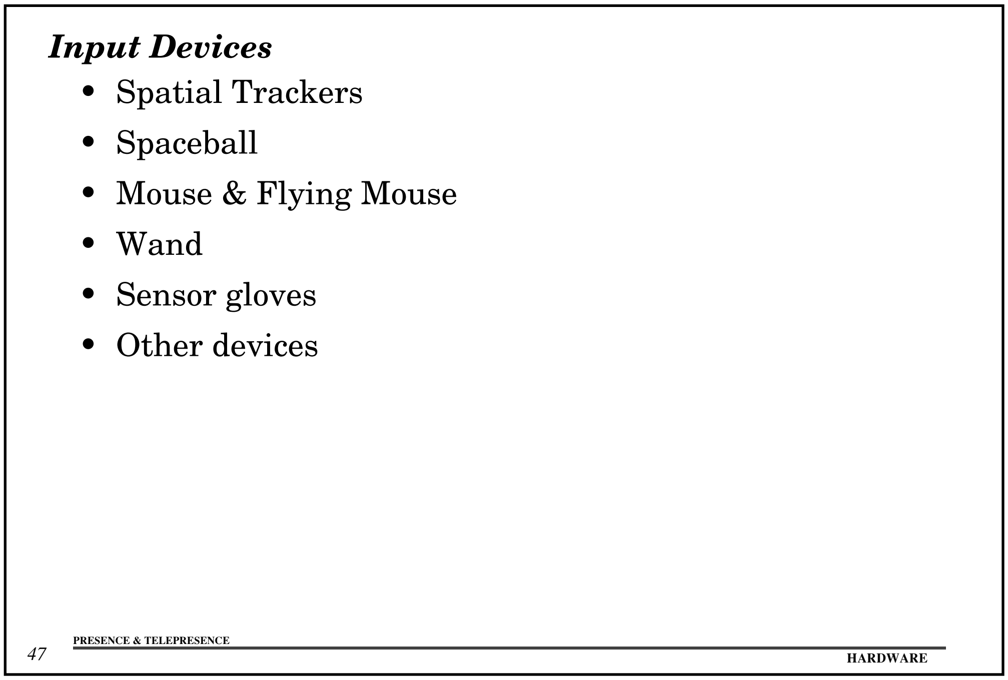
Laser Scanning

- The Human Interface Technology Lab at the U of Washington has, under development, a device to scan an image directly onto the retina via a low-power laser
- Current device uses a red laser and is built on an optical bench, viewer must position self so that laser enters eye
- Image is ~400 x 400, red on black
- Eye tracking will be required
- Could lead to future high resolution, low cost HMD

*Comparison Chart**

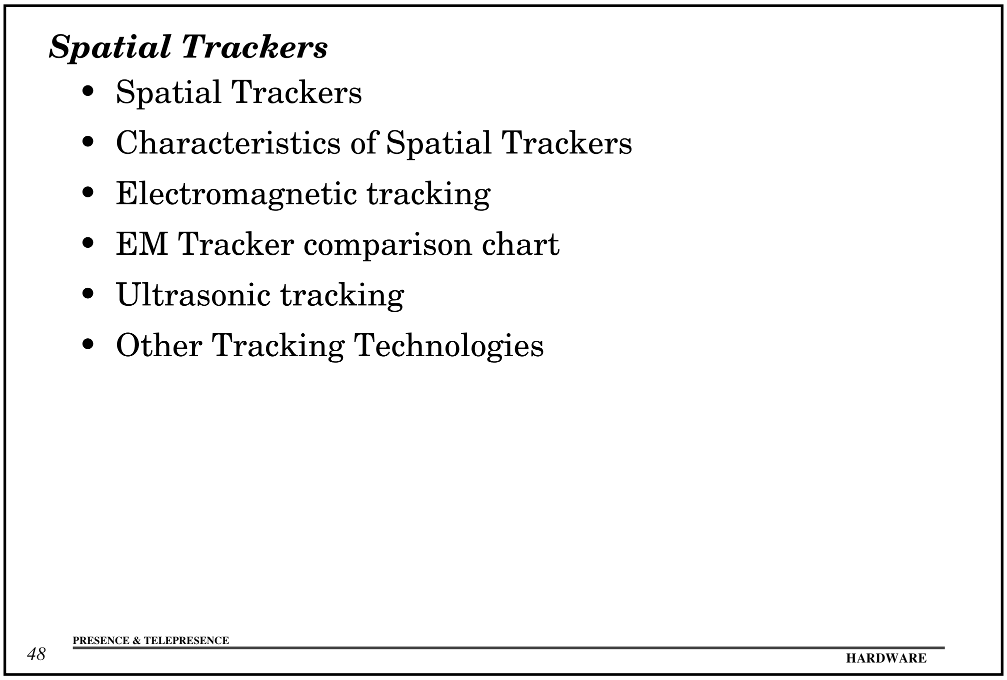
Display	Field of View (degrees)	Visual Acuity	Immersion	Intrusion
CRT	45	20/45	None	None
BOOM	90 to 120	20/85	Nearly Full	Partial
HMD	100 to 140	20/425	Full	Full
CAVE	Full	20/110	Full	None

*** Derived from Cruz-Neira, et. al., "The Cave", Comm. of the ACM, June 1992**



Input Devices

- Spatial Trackers
- Spaceball
- Mouse & Flying Mouse
- Wand
- Sensor gloves
- Other devices



Spatial Trackers

- Spatial Trackers
- Characteristics of Spatial Trackers
- Electromagnetic tracking
- EM Tracker comparison chart
- Ultrasonic tracking
- Other Tracking Technologies

Spatial Trackers

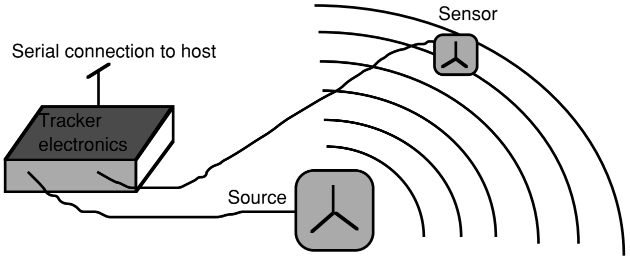
- Used to detect the position of the head, hand, or other parts of the body
- Vital for immersing the user's head or hand in the virtual world
- Most report 6 degrees of freedom
  - X, Y, Z position of tracker
  - orientation of tracker
    - roll, pitch, yaw or
    - quaternion (axis/angle)

Characteristics of Spatial Trackers

- Range--volume of space in which position and orientation can be recorded
- Accuracy--fractional error in making a measurement
- Resolution--smallest detectable quantity
- Sampling rate--how often readings are taken
- Latency--time between movement and reading

Electromagnetic tracking

- Both source and sensor contain 3 orthogonal electromagnetic coils
- Source coils emit radio frequency signal, in sequence
- 3 sensor coils measure signals simultaneously
- 9 measurements required to get position and orientation (9 equations with 6 unknowns)



EM Tracker comparison chart

Tracker	Polhemus Isotrak	Polhemus Fastrak	Ascension Bird
Range	3'	10'	3' (8' optional)
Position Accuracy	0.25" @ 30"	0.03"	0.1"
Position Resolution	0.006"	0.0002"	0.03"
Orientation Accuracy	0.85°	0.15°	0.5°
Orientation Resolution	0.35°	0.025°	0.1° @ 12"
# of trackers	1	1..4	1..10
Sampling rate	24 Hz	120 Hz/trackers	100 Hz/trackers
Latency	25 msec	4 msec	N/A
Price	$3200	$5000..$7500	$2000

Ultrasonic tracking

- Source unit contains 3 sound emitters
- Sensor contains 3 microphones
- Not susceptible to electromagnetic interference like EM trackers
- Limited range of orientations, must be approx. facing source
- 50 Hz, 25 msec latency, position accurate to 2% of distance to source
- ~$1000
- Good choice for head tracking for seated user in front of a workstation

Other Tracking Technologies

- Mechanical
  - Linkage, like that built into BOOM
  - Shooting Star Technologies makes a separate mechanical tracker that can be used with an HMD or in front of a workstation
  - 300 Hz, 2 msec latency, 0.2", 0.3° accuracy, $1500
- Gyroscopic
  - Orientation only, range limited by cable only
  - Prototype from Gyration, Inc.
  - $1500
- Video-based
  - Experimental only

Spaceball

- An isometric joystick in the shape of a ball, about the size of a tennis ball
- Senses pressure and twist in all directions
- 6 degree of freedom device--reports translations and rotations applied to ball
- Difficult for some users, who expect it to move, but can be learned easily
- Good control choice for workstation environment
- $1600

Mouse & Flying Mouse

- The mouse is a common 2-D pointing device, with 1, 2, or 3 buttons, usually standard on any computer with graphics capabilities
- The Flying Mouse (Simgraphics Inc.) is a 3-button mouse with an EM tracker inside
  - Compatible with SGI, Sun, other workstations
  - Behaves like 2-D mouse while on mousepad
  - Behaves like 6-D tracker when picked up
- Ascension Technologies has a similar device (6-D only)

Wand

- Many research sites have placed a 6-D tracker inside some sort of hand-held unit with buttons; such a device has come to be called a "wand"
- UNC used a pool ball
- Several other sites have used parts of joysticks or flashlights
- SDSC's Banana uses part of a toy guitar
- Virtual Research has a commercial version under development

Sensor gloves

- A glove fitted with sensors that can detect the bends of the finger joints
- Glove is usually fitted with a 6-D tracker to report hand position
- VR normally displays an image of the hand, which affords a strong sense of presence in the virtual world
- VPL Dataglove
- Nintendo Powerglove
- Virtex Cyberglove
- Exos Dextrous Hand-master

VPL Dataglove

- One of the first sensor gloves, invented by Thomas Zimmerman
- Sensors are loops of optical fiber, with a nick over the joint. The amount of light transmitted by the fiber is attenuated by the amount of bend
- Specifications
  - Standard glove contains 10 sensors, 2 per finger or thumb
  - 6-D tracker mounted on back of hand
  - $8800 (includes Polhemus Isotrak)
- Limitations
  - Glove needs to be calibrated for each user and may need to be recalibrated during a session
  - Sensors are not very linear

Nintendo Powerglove

- Technology licensed by VPL to Nintendo
- 3 bend sensors (thumb and first two fingers), very low accuracy
- Bend sensors use electrically conductive ink on flexible plastic strips
- Ultrasonic position sensor
- No longer in production, originally sold for $100, now $50
- Third parties have made an RS-232 interface for it, for use in homebrew VR systems
  - Petrochem Software, 206 Roycroft Rd, DeWitt, NY 13214

Virtex Cyberglove

- Developed by Jim Kramer at Stanford for use in sign language research
- Sensors are strain gauges, very linear and reliable
- Specifications
  - 18 sensors: 10 joints, 5 abduction (finger spread), 2 palm, 1 wrist
  - Sensor mounted behind wrist (more stable than back of hand)
  - $6500 (no tracker included)
- Needs calibration once per user; calibration reliable even between sessions

Exos Dextrous Hand-master

- Exoskeleton with sensors to measure all joint angles and abduction
- Used primarily in industrial design and telerobotics projects requiring very precise hand position measurement
- $15,000

Other devices

- Video
  - Mandala (Amiga)--video input of user, outline of image used to control VR
  - Several labs are working on real-time image understanding to sense hand position, etc.
- Reactive Environments
  - Much of Myron Krueger's work employs rooms equipped with sensor pads under the floor, video recognition systems, and other means to create a non-intrusive VR

Haptic Devices

- Objects in the real world have surfaces and weights which cannot be felt when using a sensor glove or other input device
- Force feedback attempts to allow the sensing of the boundaries and weights of virtual objects by pushing back and resisting movement
- Tactile feedback attempts to display the textures of surfaces by stimulating the skin
- Tactile feedback is often used to notify the user that they've made contact with a virtual object, when force feedback is unavailable
- Force feedback devices
- Tactile feedback devices

Force feedback devices

- Force feedback joysticks have been experimented with for some time, usually employing motors and brakes applied to each axis of movement
- Margaret Minsky at MIT developed a 3-axis force feedback joystick
- Several labs and manufacturers are working on piston-based force feedback devices for use with a sensor glove

Tactile feedback devices

- Three common means of stimulating the skin
  - Piezoelectric vibrators
  - Memory metals--metals which change shape under electrical current
  - Air bladders
- A Flying Mouse can be ordered with an array of pins in one of the buttons, vibrated using memory metals
- Global Devices has a Spaceball-like device with the ability to vibrate the ball under computer control
- Generally cheaper and safer than force feedback, but less effective

Sound and Speech

- Sound is an important sensory modality lacking from most computer applications
- Sound input is a powerful, hands-free means of issuing commands
- Sound output is an important modality for establishing a sense of presence, especially for "displaying" the status of things that are not currently visible or attended to
- Voice Input
- Speech Recognition
- Speech Output
- 3-D audio

Voice Input

- Voice input (without speech recognition) can be used to issue commands, select from menus, and augment pointing gestures
- The computer need not understand what is being spoken, it only needs to apply pattern recognition to select which, from a set of templates, matches each utterance
- Commercial packages exist for voice input to trigger macros and other events in workstation environments, could also be used for VR
- Systems recognize up to 1000 utterances
- System must be trained for each user

Speech Recognition

- True speech recognition, where the computer turns speech into text, is not yet a reality
  - Some systems exist which can recognize words spoken slowly by a particular speaker
  - Few systems can recognize speech at a normal speed, or from arbitrary speakers
- Recent advances in research labs approaching vocabularies of 5K to 20K words, 90% recognition
  - 20K vocabulary requires recognition time 10-30x length of input
- Commercialization of these technologies is just beginning
- Once it exists, full speech recognition will rapidly become an expected feature of VR systems

---

Speech Output

- In contrast to speech recognition, computer speech output has been commonplace for years
  - For example, text-to-speech has been standard on all Amigas, since 1985
- Speech synthesizers are controlled by codes for the ~40 phonemes (in English)
- Most also accept plain text and attempt to compute the phonemes, with varying success
- While few speech synthesizers would fool a listener into thinking a human was speaking, most are readily understandable

3-D audio

- Stereophonic sound is nearly as important to the perception of a 3-D space as stereoscopic vision
- Much of our ability to localize a sound in space is due to the effects of the outer ear (pinnae). The effect of the pinnae can be represented as a Head-Related Transfer Function (HRTF)
- HRTFs which work for a large portion of the population have been found, enabling the generation of 3-D audio with very high fidelity
- Two systems
  - Convolvotron, Crystal River Eng.--custom VLSI processor, 4 sound sources, $15,000
  - Focal Point--DSP board for Mac or PC, 1 source, $1800

Other Devices

- Biomuse
  - Direct sensing of brain or muscle electrical activity
  - Headband for sensing brain electrical signals and eye movements
  - Arm/leg bands for sensing muscle signals
  - RS-232 output
  - ~ $10K
- Temperature Display device
  - Senses and displays temperature for use especially in telerobotics environments

Virtual Reality Applications

- Presence
- Telepresence

Presence

- Using commonplace hardware
- Using VR hardware

Using commonplace hardware

- Habitat
- Dungeon Master
- Interactive Multimedia
- Xerox' 3-D Rooms

Habitat

- C-64 multiplayer game from Lucasfilm in '85
- Multiplayer cyberspace managed by QuantumLink, 300 baud dial-in service
- Uses sideways-scrolling graphics common in arcade games; joystick for input
- Graphics stored locally on C-64
- Achieves presence via
  - Persistence of objects--virtual world is maintained by host, can be entered and exited at will
  - Realistic interaction with other active entities, which are controlled by other humans
- Habitat quote

Habitat quote

- "The essential lesson that we have abstracted from our experiences with Habitat is that a cyberspace is defined more by the interactions among the actors within it than by the technology with which it is implemented."
  - Chip Morningstar & F. Randall Farmer
  - "The Lessons of Lucasfilm's Habitat" in
  - Cyberspace: First Steps

Dungeon Master

- Developed by FTL Games in '86 to '88
- First first-person point-of-view role playing game
- Mouse icon was hand, to cue user into direct manipulation capabilities of dungeon
- Although motion constrained to right-angle turns, movement controls become second-nature
- Amiga version had 3-D audio to cue off-screen events

Interactive Multimedia

- Joe Henderson, Dartmouth Medical School
- Used first-person point-of-view video to engage student in simulation of combat medical situation
- Used constrained input choices to allow focus on the story, rather than the interface
- Effective use of audio, especially while looking at interface screen, to maintain sense of presence and thus the urgency of the situation

Xerox' 3-D Rooms

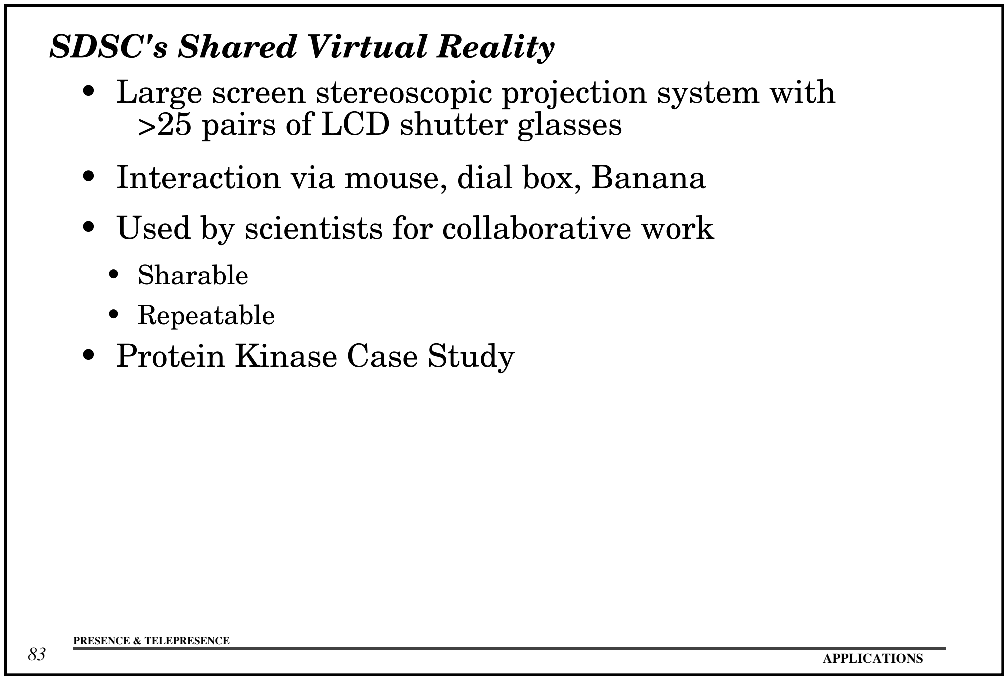
- Used SGI to create full-motion 3-D space for information visualization
- Mapped information display needs onto 3-D display capabilities
  - Perspective wall--uses foreground-background continuum inherent in perspective view to implement content/context distinction in browsing of large dataset
  - Cam tree--uses innate ability to perceive and manipulate a 3-D structure to afford browsing of tree-structured data
  - Point-of-interest navigation--capitalizes on strengths of 2-D pointer to implement constrained 3-D navigation

Using VR hardware

- Krueger's Videoplace
- SDSC's Shared Virtual Reality
- UNC's GROPE Molecular Visualization System
- NASA's Virtual Windtunnel
- VPL's Virtual Kitchen & Virtual Subway
- Virtuality

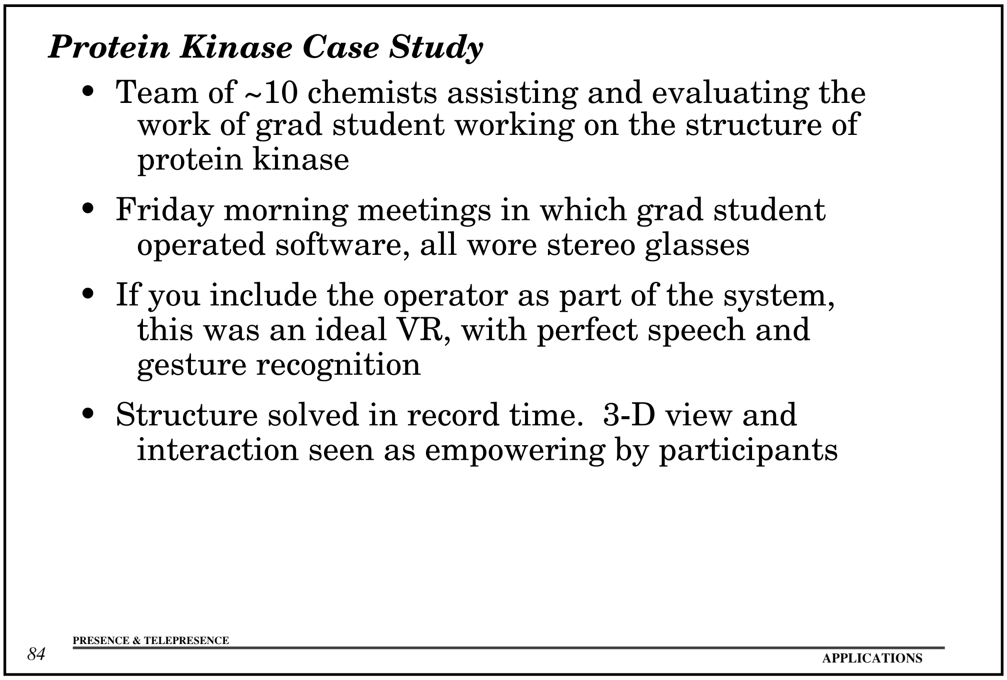
Krueger's Videoplace

- Uses real-time video recognition of outline of participant (similar to Mandala)
- Two setups, networked
  - Videoplace--projection TV, full-body image
  - Videodesk--desk-sized, image of hands only
- Sense of presence achieved through interaction with other human
- "Critter"--autonomous computer-generated agent which interacts with user



SDSC's Shared Virtual Reality

- Large screen stereoscopic projection system with >25 pairs of LCD shutter glasses
- Interaction via mouse, dial box, Banana
- Used by scientists for collaborative work
  - Sharable
  - Repeatable
- Protein Kinase Case Study



Protein Kinase Case Study

- Team of ~10 chemists assisting and evaluating the work of grad student working on the structure of protein kinase
- Friday morning meetings in which grad student operated software, all wore stereo glasses
- If you include the operator as part of the system, this was an ideal VR, with perfect speech and gesture recognition
- Structure solved in record time. 3-D view and interaction seen as empowering by participants

UNC's GROPE Molecular Visualization System

- Projection stereo system with force-feedback manipulator arm
  - Originally remote end of manipulator system for handling radioactive materials
- Used to dock molecules into minimum energy configuration
- Docking task is a multidimensional minimization problem, very difficult for a computer to perform
- Results show that subjects could dock molecule twice as fast with force feedback alone (no visual display) than with visual alone
- Excellent example of tight fit between problem domain and interface design

NASA's Virtual Windtunnel

- NASA Ames uses a Fake Space BOOM as interface to a computational fluid dynamics simulation of air flow over surfaces (space shuttle)
- A DataGlove is used to allow the user to position virtual smoke emitters in the stream
- The BOOM allows rapid and natural viewpoint positioning, while the glove allows selective query of the simulation

VPL's Virtual Kitchen & Virtual Subway

- Two examples of applications developed with "standard" gloves & goggles hardware
- Virtual Kitchen
  - Japanese shoppers can visit a store and experience their future kitchen in VR--30,000 different appliances in database
  - Forerunner of use of VR in point-of-sale marketing applications
- Virtual Subway
  - When an abandoned subway under the Berlin wall was about to be reopened, its design was so politically sensitive that they prototyped it in VR first, to speed achieving consensus on the design
  - One of many applications of VR to architecture

Virtuality

- W Industries markets a two-player video game using two platforms, each with an HMD and a wand called Virtuality
- The current game is called "Dactyl Nightmare" and consists of a multilevel virtual playing field
- The wand has two buttons, one walks forward in the direction of view, the other fires. The tracker in the wand aims the gun.
- Intelligent use of the platforms to confine the user and integrate the tracking hardware into a stand-alone unit; difficult to deliver VR to the mass market otherwise

Telepresence

- SDSC's Distributed Laboratory
- NASA's Remote Space Exploration
- SRI's Telesurgery

SDSC's Distributed Laboratory

- We recently designed a software environment for remote control of an electron microscope from a workstation
- We decided early on that direct control of the stage motors would not work due to network latency
- Instead, the user views a precollected image calibrated with respect to the microscope
- All image requests are given relative to the calibrated image
- To provide an additional sense of telepresence with the lab, audio is transmitted in both directions over the network

NASA's Remote Space Exploration

- NASA has developed prototypes of telerobotics applications for space exploration, extra-vehicular repair work, etc.
- Began with head mounted displays, now tend to use BOOMs instead
- One of many applications of tele-operated robots in hazardous environments
  - Other applications: fire-fighting, bomb defusing

SRI's Telesurgery

- Many surgical procedures are now being performed using fiber optic cameras and instruments inserted through small openings in the body
  - endoscopy--uses existing orifices
  - laproscopy--small slits in belly, inflated abdominal cavity
- While a significant benefit for the patient, the surgeon has lost 3-D view and haptic feedback--"like operating with sharp sticks in a bag"
- Philip Green at SRI and others are working on the "Nintendo surgeon": 3-D laproscopic cameras, force-feedback remote-operated forceps
- Currently only open table prototypes, yet to solve many of the problems of telesurgery

Virtual Reality Design

- System Architecture
- VR Toolkits
- Presence and Telepresence in Design

System Architecture

- Modeling
- Computation and Simulation
- Viewing Specification
- Interaction in 3-D
- Practical Constraints

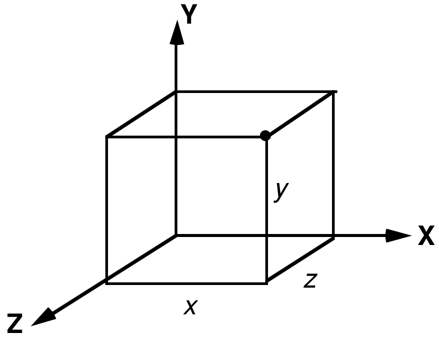
Modeling

- Coordinates
- Transformations
- Coordinate Systems
- Representing Objects
- Complicating Objects

Coordinates

- Three-dimensional Coordinates
- Homogeneous Coordinates
- Advantages of Homogeneous Coordinates

Three-dimensional Coordinates



Any point in 3-space can be represented by a 3-element vector of real values

(X Y Z)

Homogeneous Coordinates

A 3-space point can also be expressed as a 4-element vector

(X Y Z W)

where W can be thought of as the *scale* of the other three values

Homogeneous	3-space
(X Y Z W)	$\left(\frac{X}{W}\ \frac{Y}{W}\ \frac{Z}{W}\right)$

Advantages of Homogeneous Coordinates

- Can be expressed as integers
- All transformations can be expressed as 4x4 matrices

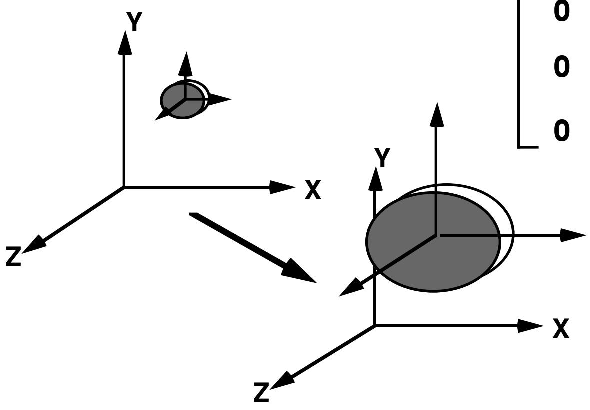
$$(\mathtt{X\ Y\ Z\ W})\ \mathtt{x} \begin{bmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \\ m_{41} & m_{42} & m_{43} & m_{44} \end{bmatrix} = (\mathtt{X'\ Y'\ Z'\ W'})$$

Transformations

- Scaling
- Translation
- Rotation

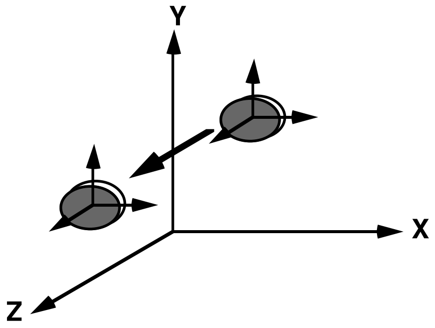
Scaling

$$\begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



Translation

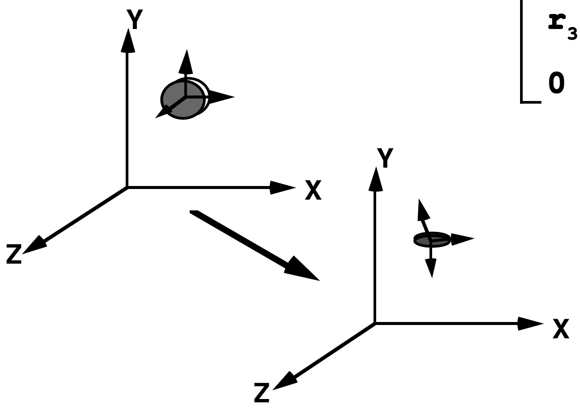
$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ t_x & t_y & t_z & 1 \end{bmatrix}$$



Rotation

- Euler Angles
- Axis-Angle

$$\begin{bmatrix} r_{11} & r_{12} & r_{13} & 0 \\ r_{21} & r_{22} & r_{23} & 0 \\ r_{31} & r_{32} & r_{33} & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$



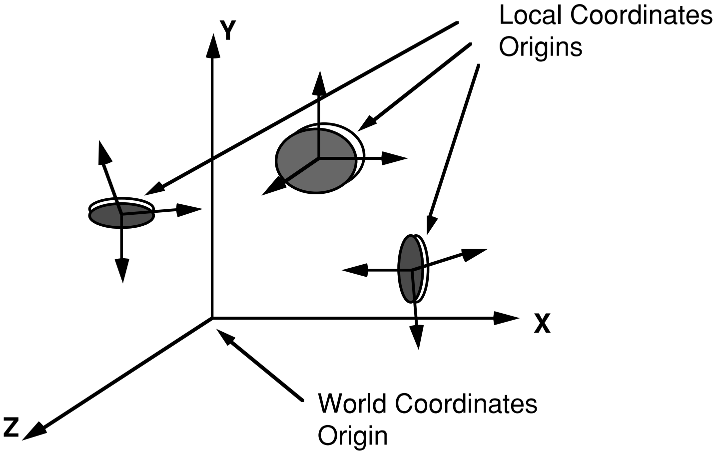
Euler Angles

- Rotation can be specified with three values:
  - angle about X axis
  - angle about Y axis
  - angle about Z axis
- Problem: dependent on order of rotations

Axis-Angle

- Rotation can also be specified with four values:
  - arbitrary axis of rotation (X Y Z)
  - angle about axis
- Avoids all of the problems of Euler angles
- Can be expressed mathematically as a quaternion, which can be easily multiplied, divided, interpolated, etc.
- Quaternions used by most position tracking devices

Coordinate Systems



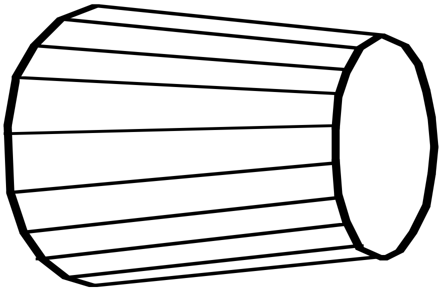
Revised 930414

Representing Objects

- Polygons
- Surfaces of Revolution
- Constructive Solid Geometry
- Parametric Patches
- Procedural Models

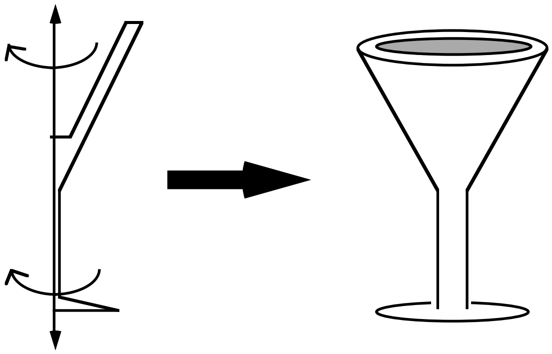
Polygons

- Surface decomposed into polygons
- Each polygon is closed & convex



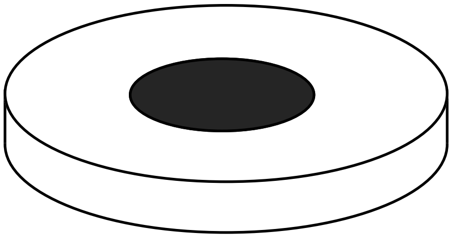
Surfaces of Revolution

- 2-D Polyline or curve revolved about an axis



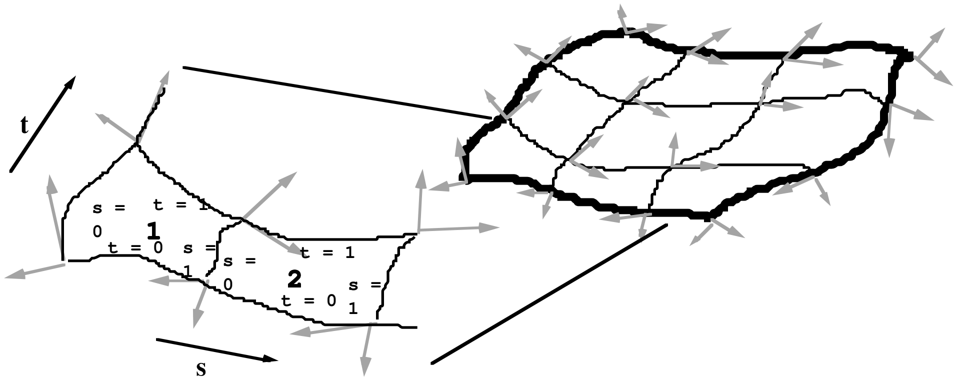
Constructive Solid Geometry

- Union and intersection of solid primitives
  - spheres, cylinders, cubes, etc.
  - Example: a wheel constructed from a cylinder subtracted from another cylinder



Parametric Patches

- Curved surfaces broken into patches, each specified by control points or equations



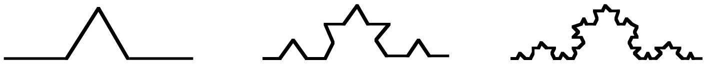
Procedural Models

- Some objects can be modeled as the result of an equation or algorithm
- Examples
  - Fractals
  - Lindenmayer Grammars

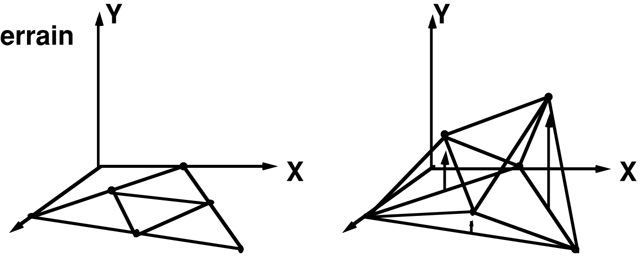
Fractals

- Simple primitives operated on by simple rules

Koch Curve



Fractal Terrain



Lindenmayer Grammars

- alphabet + grammar + axioms
- model plants very well

Alphabet

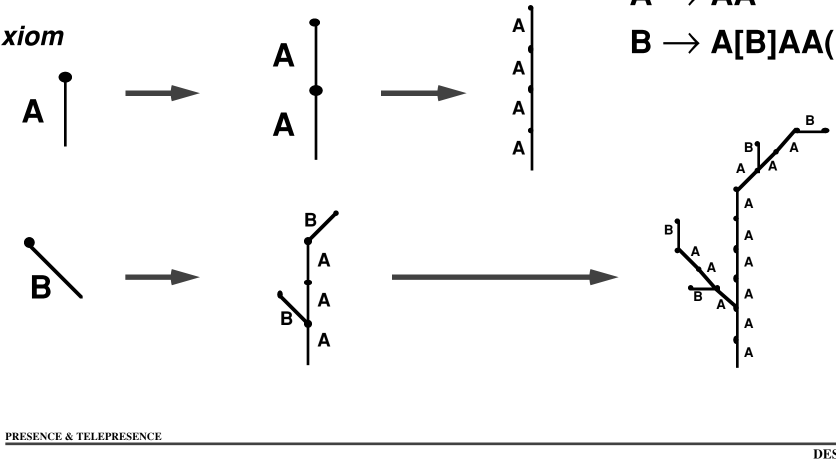
A B [ ] ( )

Grammar

A → AA

B → A[B]AA(B)

Axiom



Complicating Objects

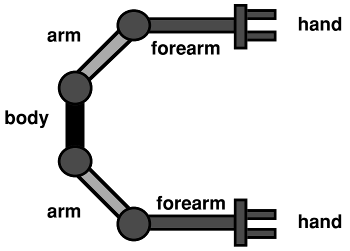
- Surface Mapping
- Hierarchical Models
- Masters and Instances
- Levels of Detail

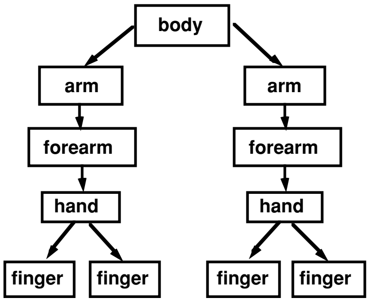
Surface Mapping

- Texture mapping
  - Maps an image onto a surface
- Bump mapping
  - Uses an image to perturb the normals to a surface (affects shading only)
- Displacement mapping
  - Uses an image to perturb the actual geometry of a surface
- All are techniques used to increase the visual complexity of a scene

Hierarchical Models

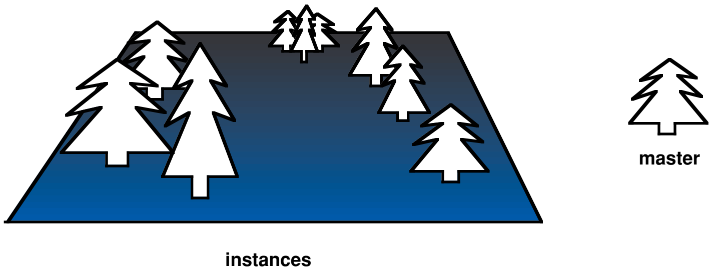
- Rather than specify the position of each object explicitly, hierarchies can be constructed where each object's position is specified relative to its parent
- Example: robot arms





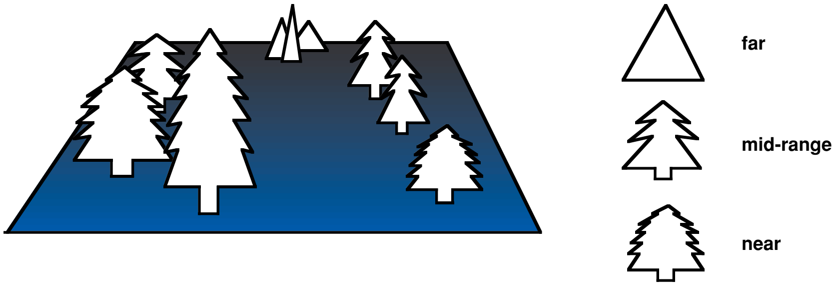
Masters and Instances

- World complexity can be reduced by describing a master object once, then instancing it multiple times



Levels of Detail

- An object may be represented by several models with varying detail
- Example: near, mid-range, and far



Computation and Simulation

- Rendering
- Simulation
- Decoupling

Rendering

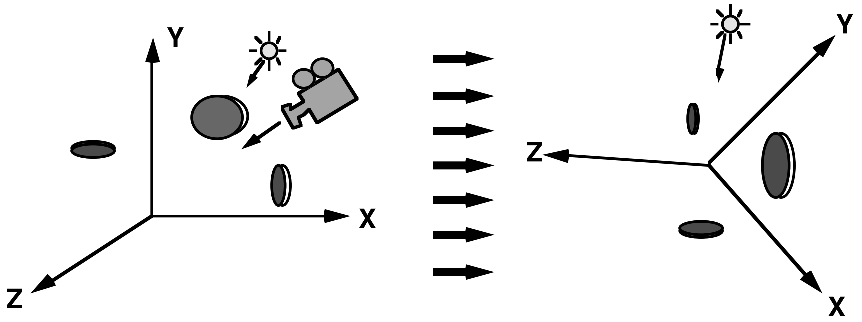
- Graphics Pipeline
- Image Quality
- Polygon count
- Progressive Refinement

Graphics Pipeline

- Rendering involves a sequence of operations
- Often implemented in hardware as a pipeline
- Each polygon can be operated on independently
  - Polygonization and Transformation
  - Clipping, Shading, Perspective
  - Rasterization
  - Z Buffering

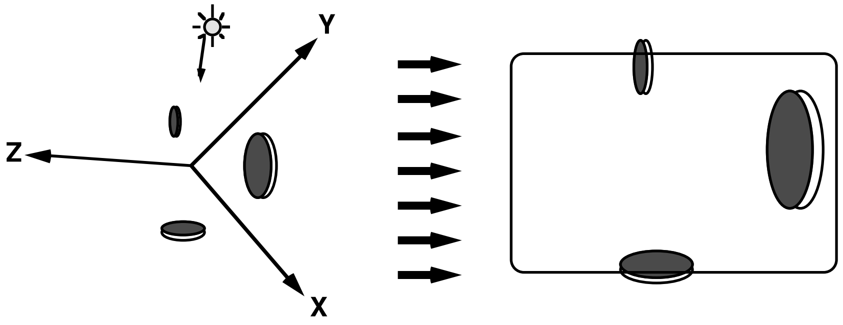
Polygonization and Transformation

- All primitives are turned into polygons and transformed so that the viewpoint is at the origin



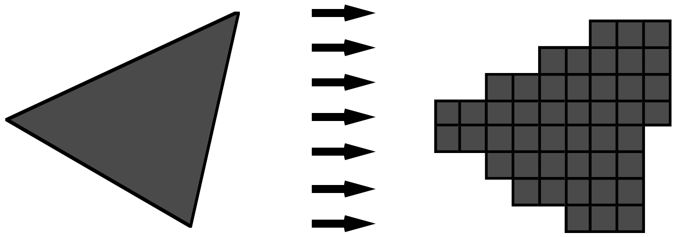
Clipping, Shading, Perspective

- Each polygon is clipped by the viewing window, shaded by the light sources, and distorted by perspective



Rasterization

- Each transformed polygon is "sliced" and "diced" into pixel-sized fragments



Z Buffering

- Each pixel fragment is sent to a Z buffer, where they are sorted by depth
- The resulting pixel is the color of the nearest fragment

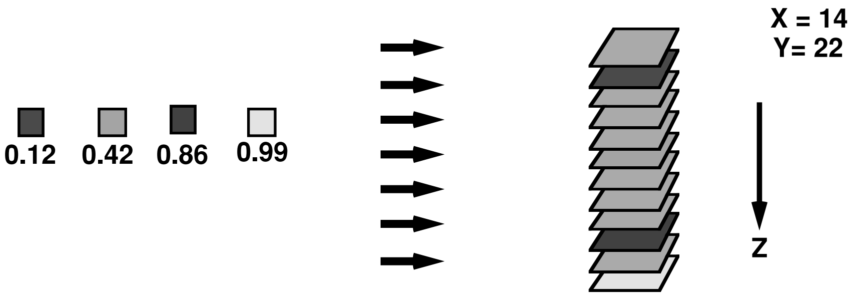


Image Quality

- Rendering time is inversely proportional to image quality
- A number of ways of rendering polygons, in order of increasing rendering time
  - wire-frame
  - flat-shaded
  - smooth-shaded
- Additional features, such as texture mapping, smoothing of lines (anti-aliasing), etc. all add to the rendering time
- Balancing rendering time with quality requires a good knowledge of the capabilities of the hardware
  - Example: on an SGI, a tree consisting of one polygon with a texture map might display faster than a tree-like object

Polygon count

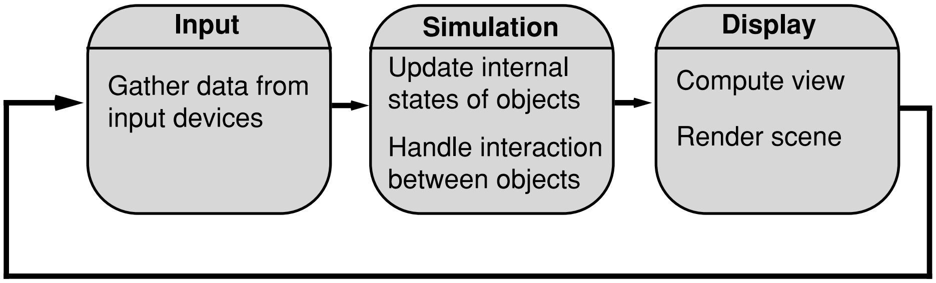
- Given a particular image quality, the major factor affecting display speed is the number of polygons in the scene
- Graphics hardware is rated by the number of polygons/second it can display, usually measured with polygons of a certain size, and with a certain image quality
- Alvy Ray Smith (Pixar) once said that reality is 80 million polygons at 30 Hz, or 2400 Mpolys/sec; nearest available system is 3 Mpolys/sec (Sun Freedom 3000)
- In designing a virtual world, give yourself a total polygon allotment that will give the desired frame rate, then partition it among the objects fairly

Progressive Refinement

- In order to provide a pleasing interactive display, a minimum of 10 frames/sec is needed. 30 frames/sec, video rate, is desirable
- One technique is to display a simpler scene while the user is interacting, filling in the detail when motion has stopped--this is called progressive refinement
- One example is to use a wireframe rendering during motion, then draw shaded polygons when still
- Unfortunately, this doesn't work well in many VR systems where interaction is continuous, the user may rarely be still enough to fill in the detail

Simulation

- Every VR application needs to have an underlying simulation--something which manages the behavior of objects, the interaction with the user, etc.
- The top level architecture of a VR application can be viewed as a loop



Input

Gather data from input devices

Simulation

Update internal states of objects

Handle interaction between objects

Display

Compute view

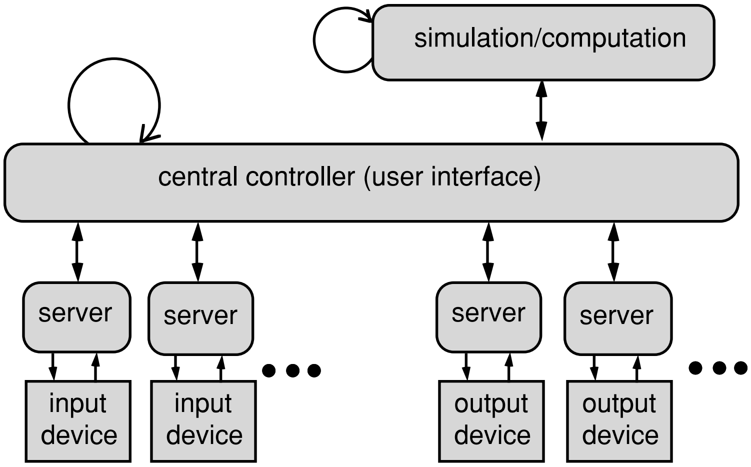
Render scene

Decoupling

- The problem with this model is that each frame doesn't get displayed until all processing is complete
  - If the simulation is compute-intensive, the whole system slows, causing significant display lag
- In a multitasking environment, such as Unix, it's best to decouple the input, display, and computational portions
  - Each becomes an autonomous process running its own loop, accepting input from the other processes when it is made available, operating on the previous input if no new data is ready
  - This also works well in a multiprocessing environment, where each task can run on a separate process
- Decoupled Simulation Model

Decoupled Simulation Model

- The VR is coordinated by a central controller which implements the user interface



simulation/computation

central controller (user interface)

server

server

server

server

input device

input device

• • •

output device

output device

• • •

Viewing Specification

- Specifying the view consists of two tasks:
  - Viewing transformation
  - Stereo Perspective

Viewing transformation

- The task of the viewing transformation is to transform the entire scene so that the eyepoint is at the origin, and we're looking in the right direction
- The viewing transformation consists of two parts, given the coordinates provided by the head tracker:
  - Translation: if the eye point is (X, Y, Z), translate by the inverse (-X, -Y, -Z) to bring that point to the origin
  - Rotation: turn the quaternion representing the orientation of the head into a rotation matrix and apply that matrix

Stereo Perspective

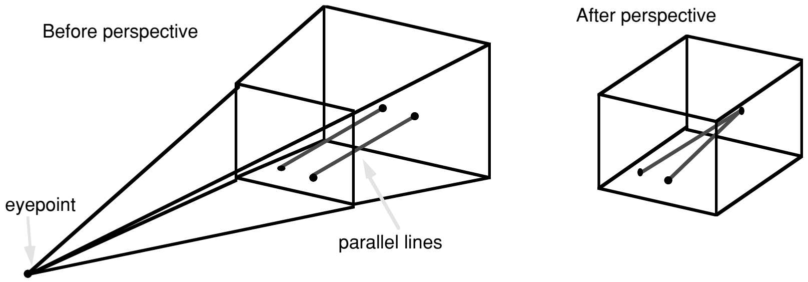
- Orthographic projection
- Perspective transformation
- Parallax and retinal disparity
- Vertical parallax problem
- Accommodation/Convergence
- Parallax control
- Stereoscopic Composition
- Specifying perspective
- Parallel projections
- Off-axis projections

Orthographic projection

- Once a viewpoint has been chosen and the scene has been transformed so that the origin is the viewpoint, the final transformation is to take the 3D world and produce a 2D view; essentially, we need to eliminate the Z information
- The simplest approach is to throw away the Z; this is called an orthographic projection
  - The resulting scene will have no perspective foreshortening; this is usually only suitable for viewing a single object
  - A scene of a room or many objects needs perspective in order to look real
- However, stereo is simple--just rotate the object about a vertical axis +/- N degrees, where N is usually 3 to 5

Perspective transformation

- A normal perspective transformation selects a frustum-shaped portion of the world and squashes it to form a rectangular solid, the front face of which is the view
  - A frustum is a truncated pyramid
  - This is roughly equivalent to dividing the X & Y coordinates of each point by its Z coordinate

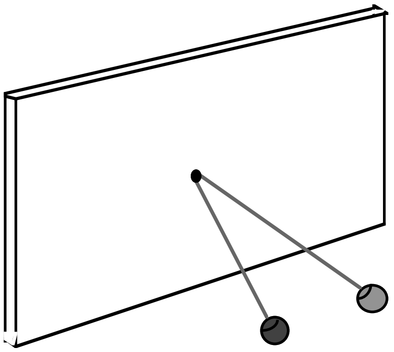


Parallax and retinal disparity

- Retinal disparity
  - We see stereo because a point in a 3-D world is projected onto different locations on the left and right retinas
- Parallax
  - is the disparity of two images, measured at the screen
  - Often expressed as an angle
- Parallax: Zero parallax setting
- Parallax: Positive
- Parallax: Divergent
- Parallax: Negative

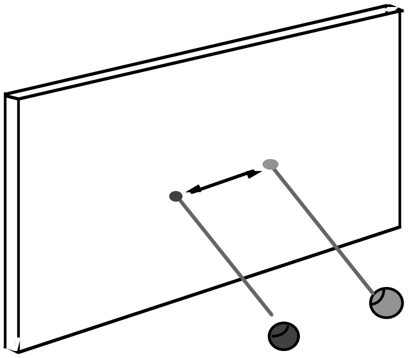
Parallax: Zero parallax setting

- Object appears in plane of display
  - Also known as the convergence plane



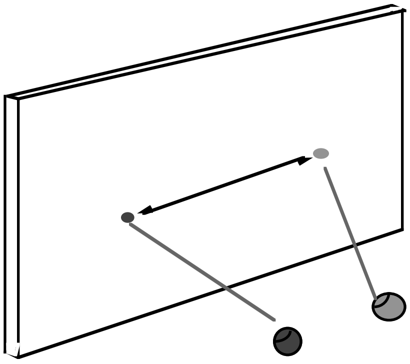
Parallax: Positive

- Object appears distant, in display space (behind screen)



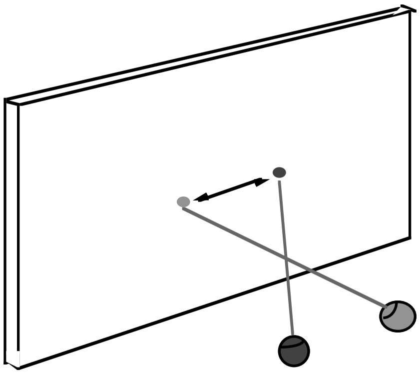
Parallax: Divergent

- No physical analog; not used



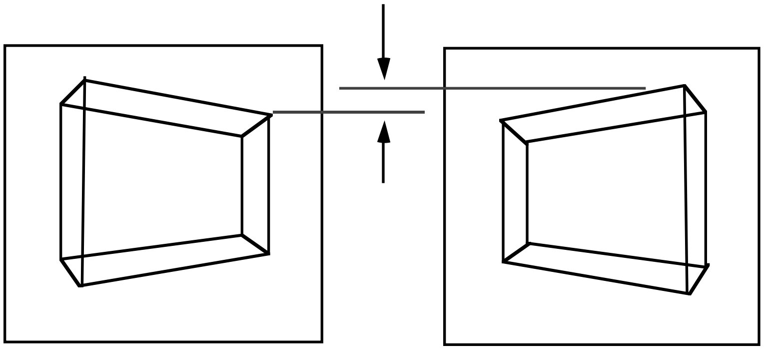
Parallax: Negative

- Object appears near, in viewer space (in front of screen)



Vertical parallax problem

- When using perspective, we can no longer simply rotate the scene to produce stereo views, vertical parallax results



Accommodation/Convergence

- Convergence: rotation of eyes to intersect object
- Accommodation: focusing on object
- Learned relationship: we both focus and converge on objects
- Breaks down in stereoscopic displays

Parallax control

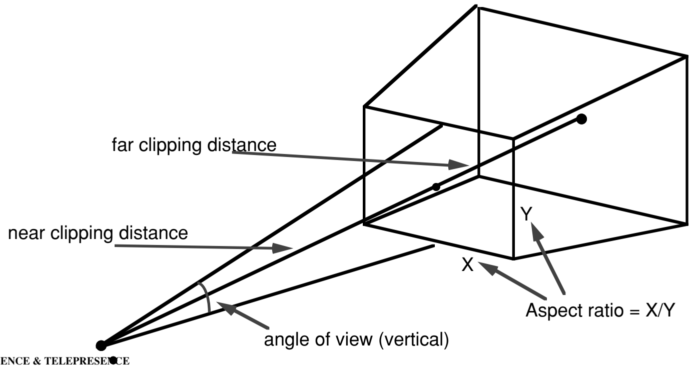
- Low parallax lessens accommodation/convergence problem
- Both positive and negative parallax OK
- Rule of thumb: don't exceed 1.5 degrees
  - Typical workstation: 0.5" from 1.5 ft.
- Distance of viewer important
- Rule may be broken (especially for negative parallax images)

Stereoscopic Composition

- Keep parallax low (<1.5 degrees)
- Stress perspective cue via wide-angle "lens"
- Use depth cuing and other monoscopic cues
- Set the ZPS plane in center of object
- If viewing single object, keep ZPS in center
- Use parallel axes camera model

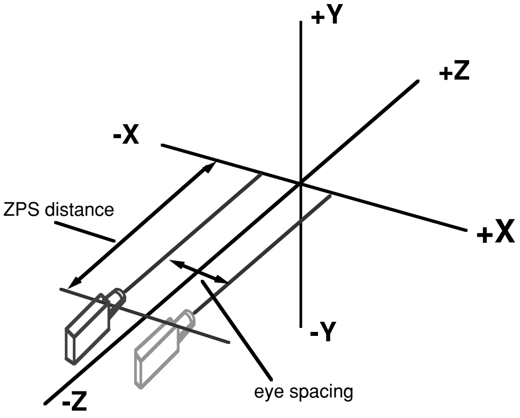
Specifying perspective

- Perspective is usually specified in a graphics language via the distances of the near and far clipping planes, the angle of view, and the aspect ratio of the window
  - The example here is from SGI's GL, which measures the angle of view in the vertical direction



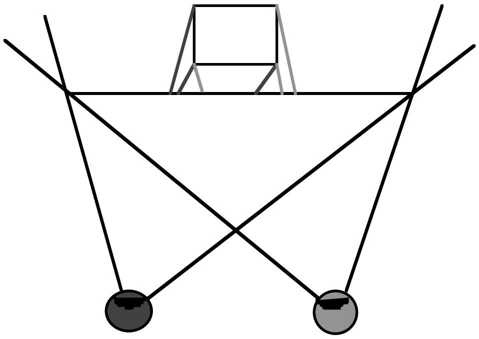
Parallel projections

- The correct way to compute stereo perspective views is with two parallel cameras



Off-axis projections

- Computing parallel projections requires two off-axis projections
  - Off-axis projection: the line of sight is not halfway between the left and right edges
  - Computing off-axis projections



Computing off-axis projections

- Most graphics languages, such as SGI's GL, allow one to specify the frustum of view by giving the left, right, top and bottom edges
  - window(left, right, bottom, top, near, far)
- See the appendix for a source code example of computing the window parameters from the normal perspective parameters (field of view in Y, aspect ratio) plus two additional parameters
  - conv: the distance of the zero-parallax (convergence) plane from the eye point
  - eye: 1/2 the separation between the two eyes

Interaction in 3-D

- Workspace Mapping
- Device Calibration
- Filtering
- Hand Gesture Recognition
- Selection
- 2-D Techniques
- Navigation
- Behavior

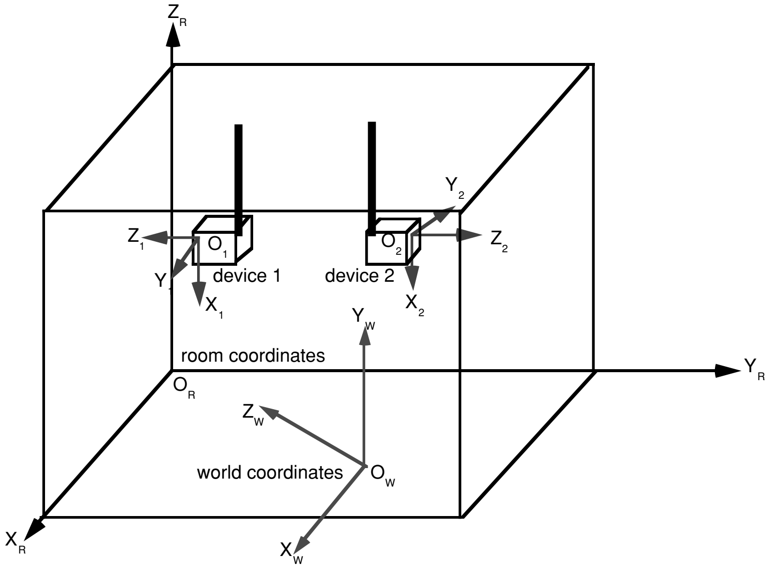
Workspace Mapping

- In a VR application employing one or more spatial trackers, each tracker has its own coordinate system defined by the location of the source
  - Multi-sensor (one source) trackers like the Polhemus Fastrak or the Ascension Tech. Flock of Birds ease this problem somewhat
- The ideal solution is to map all of the coordinate systems in the workspace to one primary coordinate system
- Workspace coordinate systems

Workspace coordinate systems

- Three coordinate systems in the workspace
  - Device--the coordinates reported by a device, different for each device
  - Room--a common coordinate system for all of the devices in the room
  - World--the world coordinate system used in the application
- We need to be able to perform the following mappings
  - device coordinates -> room coordinates
  - room coordinates -> world coordinates
- Workspace coordinates diagram
- Mappings

Workspace coordinates diagram



Mappings

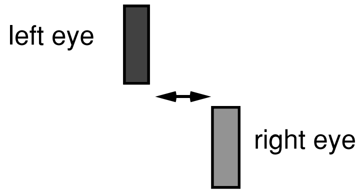
- The mapping from device coordinates to room coordinates is static and does not vary from one application to the next
  - Locations of the device coordinates can be read from a configuration file shared by all applications--only this file need be changed if the devices are moved
- Room coordinate to world coordinate mapping has two parts
  - Translation: maps the room origin to world coordinates
  - Rotation and scaling: maps the orientation and size of the room coordinates to world

Device Calibration

- Calibration is performed once, at the beginning of each session
- Stereo calibration
- Glove calibration

Stereo calibration

- In any stereo viewing situation, the offset between the two views needs to be determined
  - In a monitor-based environment the offset is fixed and does not need recalibration
  - With a head-mounted display, the offset needs to be adjusted for each user, each session
  - Display a simple image, such as a vertical bar, in each eye, and use an input device to allow the user to adjust the offset until the images coincide



Glove calibration

- A sensor glove needs to be calibrated for each user, because of different hand sizes and glove peculiarities
  - DataGloves need to be calibrated each session, and may fall out of calibration during a session
  - CyberGloves usually only need to be calibrated once per user
- One approach is to deal, not with joint angles directly, but with normalized ranges
  - maximum flexion = 1.0
  - minimum flexion = 0.0
- The user can be prompted to make several representative gestures (flat hand, fist, etc.)
- Linear interpolation between these readings is good enough for most applications

Filtering

- Two major problems with spatial trackers and head tracking are noise and lag
  - Primary display lag perceived by the user is due to lag in the orientation data
    - Additional lags are due to data communications and drawing time
  - Noise in the position data leads to image jitter
- The solution is to filter the tracker data, and to apply different filters to position and orientation
  - Position Filtering
  - Orientation Filtering

Position Filtering

- Most movement is along the line of sight, thus
  - Lag is most noticeable along the line of sight
  - Image jitter is mostly due to noise perpendicular to the line of sight
- We want to remove the high-frequency noise by using a low-pass filter
- To trade off speed and accuracy, we use an anisotropic filter
  - more accurate high-order filter perpendicular to the line-of-sight
  - faster low-order filter along the line of sight
- The amount of filtering should increase the further the sensor is from the source, since the noise increases

Orientation Filtering

- For orientation, we want a predictive filter, based on the assumption that most motion is continuous
- One recommendation from the literature is to apply a Kalman filter to each of the four elements of the orientation quaternion, then renormalize the elements to produce a valid quaternion
- Major problem is with reversals of motion
- Predicts orientation effectively between 100 and 120 milliseconds

Hand Gesture Recognition

- Gesture recognition actually consists of two parts
  - Postures--static hand positions where beginning and ending a posture is analogous to pressing and releasing a button
  - Gestures--moving finger or hand data, such as waving, "come-hither", etc.
- Posture recognition
- Gesture recognition

Posture recognition

- Hand posture recognition is a function of the finger joint angles only, and is independent of the hand tracker data
- Template matching
- Posture classification
- Minimum distance

Template matching

- One approach is to encode a template for each posture
    - Each joint either has an active range (minimum and maximum flexion) or is marked "don't care"
    - If all of the current joint flexions fall within the range of a template, that template is matched
- Postures can be defined by example by having the user give the posture several times and deriving the range
- Instead of min/max, could use mean and match if flexion is within 1 standard deviation of user-provided samples
- Too many templates leads to ambiguity, maximum of about 10

Posture classification

- Another approach is to classify each joint angle
- If we start with flexion values normalized to [0..1]
  - [0.0 .. 0.2] = Extended
  - [0.2 .. 0.8] = Neutral
  - [0.8 .. 1.0] = Flexed
- We can then define a standard set of gestures based on which joints are out of the neutral range
  - "Fist" might be defined as "all joints flexed"
  - "Pointing" might be defined as "index finger extended, all others flexed"

Minimum distance

- A third approach is to compute the minimum distance to a set of canonical examples
- Each posture is represented as a N-dimensional vector, where N is the number of joint angles sensed by the glove
- The current set of joint angles is also a N-vector
  - Compute the square of the distance from the input vector to each of the postures as
  - $(i_1-p_1)^2 + (i_2-p_2)^2 + (i_3-p_3)^2 \ldots$
  - Minimum is the closest posture
- Quick and dirty, either of the first two approaches is probably preferable

Gesture recognition

- The simplest gesture to detect is that made with one or more fingers, independent of the hand tracker data, such as a "come-hither" motion
  - Similar case: using one joint angle to control a continuous variable, such as speed
- True gesture recognition should
  - Recognize hand motion in 3-space
  - Require no explicit start or stop signal (continuous recognition)
  - Require as little training as possible
- Gesture feature extraction

Gesture feature extraction

- Based on work of Dean Rubine at CMU on 2-D (pen) gestures, extended to 3-D by David Sturman
- Operates on list of input device samples form the last N seconds
- Features are extracted, such as diagonal of the bounding box, path curvature, direction of motion, average speed of motion, etc.
- Classifier then matches the parameters from the input motion against prototype gestures
  - Each prototype gesture is a set of weights, one for each feature
  - For each gesture (g), a score is computed as the sum of each weight (W) multiplied by the input feature (F); largest score wins
  - $\text{Score}(g) = \text{Sum}(W_{(g,i)} * F_i)$

Selection

- The user needs the means to select which objects in the virtual world are to be manipulated
- This is simple in 2-D environments, the user simply points with the mouse and clicks
- In 3-D, the object may be visible, but out of reach, or it may be invisible, obscured by another object or out of the field of view
- Grabbing
- Pointing
- Calling

Grabbing

- If the user is provided with a sensor glove or wand, reaching out and grasping an object is a very natural means of selection
  - When a fist is made while the virtual representation of the user's hand is intersecting an object, the object is grabbed, and any subsequent hand movements cause the object to move also
- Grabbing should be accompanied by some sort of feedback, either visual, such as changing the color of the object or drawing a box around it, or auditory
  - If available, tactile feedback upon contact with an object will make grabbing more natural
  - An audio or visual cue can also be used to indicate contact, if tactile feedback is unavailable

Pointing

- If the desired object is visible but out of reach, pointing at it is another way to select it
- The user can be presented with a ray which emanates from the virtual representation of the glove or wand
- When the ray intersects an object, that object is selected
- Selection feedback is necessary, just as with grabbing
- Since objects far away are small and difficult to select accurately, it may be necessary to require that the user confirm a selection before allowing the object to be manipulated

Calling

- Moving objects or objects not currently visible can be selected by calling
- A hand gesture is assigned to each type of object; when the gesture is given, all objects of that type approach the user
- Once near the user, the objects can be selected among via grabbing or pointing
- The system could also afford calling only those objects that are visible, or causing the called objects to go away rather than come

2-D Techniques

- Many of the interaction techniques common to 2-D interfaces, such as buttons, menus, sliders, etc., can be extended to 3-D
- If working in with a workstation environment, these techniques may be best implemented by overlaying the view 2-D interaction devices
  - Allows use of the controls provided with the windowing system
  - Makes it possible to remove or recall the overlay easily
- 2-D techniques can be implemented in 3-D, by projecting 2-D drawing onto a polygon that lives in 3-D
  - Use a ray emanating from the glove or wand to operate on 3-D controls

Navigation

- Although the user's position can be tracked using a spatial tracker, this does not afford a large enough range for a reasonably-sized virtual world
  - The range of a Polhemus Isotrak is a 2 meter hemisphere centered at the source, and gets pretty noisy beyond 1 meter
- Means must be implemented to allow the user to navigate within the world
  - Motion must be easily controllable, afford rapid access to any portion of the space, and be accurate
- Line movement
- Pointing to fly
- Point of interest movement
- Vehicles

Line movement

- An approach implemented in many workstation environments is to constrain motion to be along one axis (X, Y, or Z) at a time, or to map three controllers, such as three dials, onto the three axes
- This approach can also be taken in immersive environments
- Can also map controllers onto movement along line of sight, pitch, and yaw
- In general, approaches which decompose movement into separately controlled dimensions are good for slow, fine movement, but poor for larger and faster navigation

Pointing to fly

- The first navigation technique implemented by most VR developers is "point to fly"
- When a pointing hand posture is recognized, the eyepoint begins flying in the direction of pointing
  - Some systems fly along the line of sight instead. This is simpler, but less powerful
  - One can use the flexion of the thumb, for example, to control velocity
- Both "fly where I'm pointing" and "fly where I'm looking" can be difficult to control
- It is very easy to give the posture accidentally and take off unintentionally

Point of interest movement

- User selects a point of interest (POI) in the scene and a flight path to that point is computed, with a logarithmic approach
  - We saw this in the Xerox Rooms videotape
  - In a workstation environment, a 2-D pointer (mouse) can be used to select the POI by casting a ray from the cursor location into the scene, and marking what it hits
- Logarithmic approach starts out fast and slows as POI nears
- For each time step
  - $eye_x = eye_x - k * (eye_x - POI_x)$
  - $eye_y = eye_y - k * (eye_y - POI_y)$
  - $eye_z = eye_z - k * (eye_z - POI_z)$
  - $k = 0.15$

Vehicles

- The range of navigation afforded by the tracker can be represented as the boundaries of a vehicle, which can be rendered as part of the scene
  - The user can be given separate control over the navigation of the vehicle, perhaps via a control panel, or via special hand gestures
  - By making the boundaries of the vehicle explicit, the user is aided in distinguishing the two modes of movement, walking and flying
- Navigating the vehicle can be made easier by reducing the degrees of freedom
  - For example, the vehicle may be modeled as a car, constrained to motion along a surface
  - Another example is a helicopter, which can fly but still usually remains parallel to the ground

Behavior

- To make a virtual world interesting, the objects within it must behave
- Objects can respond to several events
  - Time--the progression of time can cause the object to change its internal state and actions
  - User interaction--user actions can trigger corresponding reactions
  - Other objects--objects can interact with each other and their environment, such as obeying "physical" laws
- Rule-based behavior

Rule-based behavior

- Object behavior can be determined by a set of rules; each rule consists of
  - a condition--the circumstances or internal state which triggers the rule
  - an action--the result of the rule's firing, can change internal state or start a behavior
- At each time step, all rules are evaluated to find matching conditions
- Multiple rules may be triggered at once, some heuristic must be applied to select which one to fire

Practical Constraints

- Environment Design
- Occupational Hazards

Environment Design

- In designing a VR, care must be taken to deal with the constraints of the hardware
- An immersive VR is almost impossible to operate alone
  - Calibrating the HMD usually requires aid
  - Can't see through the HMD, constant danger of running into things
  - Require multiple cables, which easily tangle
- The space in which the system is going to be used should be considered as part of the design

Occupational Hazards

- Simulator sickness
  - Due to too much or too little presence
  - For example, mismatch between visual motion cues and (lack of) inner ear motion cues can lead to nausea, disorientation
  - Eased somewhat by careful design, for example, by displaying a floor
- Some reports of epileptic seizures in Nintendo users
- Initial test indicate that prolonged use of HMDs can lead to headaches, eyestrain
  - Largely due to accommodation/convergence disparity, all images at fixed focus distance

VR Toolkits

- There are a number of toolkits available for implementing virtual worlds
- Commercial products include the WorldToolKit from Sense8 and the tools provided by Vream
- In the Unix world, there are two freely-available toolkits
  - Both toolkits are written in C
  - Both use primarily SGIs for rendering
  - Good starting points for coding VR
- Minimal Reality Toolkit
- Virtual Environment Operating Shell

Minimal Reality Toolkit

- Developed by a team at the University of Alberta
- Available for noncommercial purposes only
- Send email to mr-help@cs.ualberta.ca and request a license agreement
- MR Functionality
- MR Structure
- Programmer's view of MR
- MR Source code example

MR Functionality

- Distributed computing--implements a decoupled simulation model on multiple workstations
- Data communication--handles the interprocess communication at a high level
- Workspace mapping--see above
- Timing and performance monitoring
- Supports multiple devices
  - VPL DataGlove and EyePhone, sound generators, CyberGlove driver under development by SDSC
- Various interactive techniques
  - Gesture recognition, 2-D user interface in 3-D, Navigation

MR Structure

- MR consists of three layers
  - Client/server pairs for interfacing with input and output devices
  - A number of packages, each of which encapsulates some service or functionality
  - A top layer which provides a consistent programmer interface to all of the packages
- The package layer consists of two parts
  - Required packages that provide standard services
  - Optional packages which support particular devices or interaction techniques

Programmer's view of MR

- An MR program is a collection of processes, each of which has one of three roles
  - Master--only one, controls everything
  - Slave--responsible for some aspect of interaction with the user, either input or output
  - Computation--performs computation that does not require interaction
- The master process starts up all the other processes and manages communications
  - Not the best implementation of distributed computing, but it works
- All processes share the same source code, the role is selected at run-time

MR Source code example

```
#include <MR/mr.h>

Program slave;
Hand hand;

main(int argc, char** argv)
{
    double r;

    MR_init(argc, argv); /* Configuration */
    MR_set_role(MR_MASTER);

    slave = MR_start_slave("beast","firstS");

    MR_clear_device_set();
    MR_add_device_set(EyePhone);
    MR_add_device_set(RightDataGlove);
    MR_split_glove(TRUE);

    EyePhone_slave(slave);

    MR_configure();
        /* end of configuration section */
    set_room_reference(0.0, 0.0, 0.0);
    map_reference_to(1.0, 1.0, 2.2);

    MR_calibrate();

    while (getbutton(LEFTMOUSE) == 0) {
        update_hand(RIGHT);
        navigate();

        MR_start_display();

        hand = get_hand(RIGHT);

        drawScene();
        draw_hand(hand, RIGHT);

        r = MR_end_display();
    }

    MR_terminate();
}
```

Virtual Environment Operating Shell

- Developed by the Human Interface Technology Lab at the University of Washington, Seattle
- Available via anonymous ftp from ftp.u.washington.edu (140.142.56.2) as pub/user-supported/veos/veos.tar.Z
- VEOS itself is an environment for implementing distributed applications under Unix
    - The VR system implemented under VEOS is called FERN
    - VEOS may be used for other than VR work
- Implemented in C, uses XLisp as the programmer's interface to interprocess communication
- More symmetric implementation of distributed processing than MR, but less direct VR support

Presence and Telepresence in Design

- Virtual Reality as Metaphor
- Ideal Virtual Reality
- Computers as Theatre
- Interfaces as Boundaries
- The First Virtual Reality

Virtual Reality as Metaphor

- At the start of the course, we defined VR as a metaphor for designing systems
- People perform best in environments similar to those in which they developed their skills
- VR provides a natural mapping between intention, action, and feedback
  - behavioral, not symbolic, interaction
- VR design requires attention to peculiarities of human perception
- While print tells and film shows, VR embodies reality

Ideal Virtual Reality

- Meredith Bricken, HITLab, in an article titled "Virtual Worlds, No Interface to Design", wrote
  - "The technology itself is invisible and carefully adapted to human activity so that we can behave naturally."
- In fact, there is an interface to design, because no virtual reality is ideal
  - No virtual reality is truly indistinguishable from reality
- We can attempt to judge, at least qualitatively, how much presence is afforded by a VR
  - Judging presence is like judging a beauty contest, guidelines can be used but it is ultimately subjective

Computers as Theatre

- Plato stated that all we ever see is an illusion based on reality
- Aristotle spoke of mimesis in drama, the psychological resonance enabling drama to move people emotionally
- Brenda Laurel wrote about computers as theatrical, first-person experiences
  - Creation of artificial realities in which the potential for action is cognitively, emotionally, and aesthetically enhanced
- The key point is that the user, if motivated, will willingly suspend disbelief and join the reality we've created, as long as we don't violate the rules of that reality and break the spell

Interfaces as Boundaries

- An interface is a boundary between the user and the system
- Presence occurs when the boundary is no longer perceived
- John Walker, from Autodesk, described the evolution of human-computer interfaces as the successive removal of boundaries
- Walker's Taxonomy of Computers

Walker's Taxonomy of Computers

Phase	Decade	Implementation	Barrier
1st	40's	plugboards	front panel
2nd	50's	punch cards, batch processing	countertop
3rd	60's	screens & keyboards, time-sharing	terminal, commands
4th	70's	menus	menu heirarchy
5th	80's	direct manipulation, point & click	screen
6th	90's	virtual reality, cyberspace	???

The First Virtual Reality

- 30,000 years ago civilization was beginning, we were making the transition from hunting/gathering to agriculture and technology
- Life had been pretty constant for the previous 200,000 years, now there were many things to be learned, a tribal encyclopedia
  - seed saving, herbal medicine, stargazing, animal husbandry, etc.
- The cave paintings in Lascaux, France were the first virtual reality, created to invoke a state of consciousness in which profound learning could take place
- The First VR (continued)

The First VR (continued)

- Novices were led down to the caves, where ochre paintings and tricks of light were used to produce a 3-D light and sound show
- Pictures of bison, birds, humans, symbols, but no natural lighting
- Many paintings are anamorphic, producing 3-D when viewed from the correct location
- Confined spaces, obstacles, hidden images, mind-altering drugs, all led to an imprinting of information on the novice's mind, a rite of initiation
- Tim Leary said: "Capture their eyes, and you'll capture their mind"