Unit 2

I/O INTERFACE AND TECHNIQUES IN VR 9 Hrs.

Multiple Modals of Input and Output Interface in Virtual Reality: Input -- Tracker, Sensor, Digital Glove, Movement Capture, Video-based Input, 3D Menus & 3DScanner etc. Output -- Visual / Auditory / Haptic Devices. Interactive Techniques in Virtual Reality: Body Track, Hand Gesture, 3D Manus, Object Grasp.

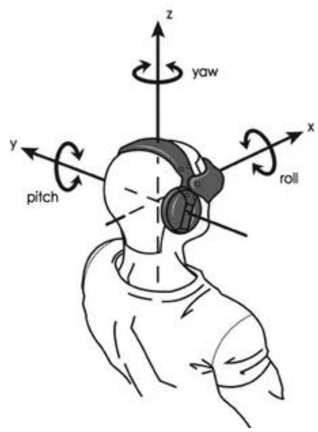
AUGMENTED AND VIRTUAL REALITY

Multiple Modals of Input and Output Interface in Virtual Reality

TRACKING

Tracking devices are intrinsic components in any VR system. These devices communicate with the system's processing unit, telling it the orientation of a user's point of view. In systems that allow a user to move around within a physical space, trackers detect where the user is, the direction he is moving and his speed.

There are several different kinds of tracking systems used in VR systems, but all of them have a few things in common. They can detect **six degrees of freedom (6-DOF)** -- these are the object's position within the x, y and z coordinates of a space and the object's orientation. Orientation includes an object's **yaw, pitch and roll**.



From a user's perspective, this means that when you wear an HMD, the view shifts as you look up, down, left and right. It also changes if you tilt your head at an angle or move your head forward or backward without changing the angle of your gaze. The trackers on the HMD tell the CPU where you are looking, and the CPU sends the right images to your HMD's screens.

Every tracking system has a device that generates a signal, a sensor that detects the signal and a control unit that processes the signal and sends information to the CPU. Some systems require you to attach the sensor component to the user (or the user's equipment). In that kind of system, you place the signal emitters at fixed points in the environment. Some systems are the other way around, with the user wearing the emitters while surrounded by sensors attached to the environment.The signals sent from emitters to sensors can take many forms, including electromagnetic signals, acoustic signals, optical signals and mechanical signals. Each technology has its own set of advantages and disadvantages.

Electromagnetic tracking systems measure magnetic fields generated by running an electric current sequentially through three coiled wires arranged in a perpendicular orientation to one another. Each small coil becomes an electromagnet, and the system's sensors measure how its magnetic field affects the other coils. This measurement tells the system the direction and orientation of the emitter. A good electromagnetic tracking system is very responsive, with low levels of latency. One disadvantage of this system is that anything that can generate a magnetic field can interfere in the signals sent to the sensors.

Acoustic tracking systems emit and sense ultrasonic sound waves to determine the position and orientation of a target. Most measure the time it takes for the ultrasonic sound to reach a sensor. Usually the sensors are stationary in the environment -- the user wears the ultrasonic emitters. The system calculates the position and orientation of the target based on the time it took for the sound to reach the sensors. Acoustic tracking systems have many disadvantages. Sound travels relatively slowly, so the rate of updates on a target's position is similarly slow. The environment can also adversely affect the system's efficiency because the speed of sound through air can change depending on the temperature, humidity or barometric pressure in the environment.

Optical tracking devices use light to measure a target's position and orientation. The signal emitter in an optical device typically consists of a set of infrared LEDs. The sensors are cameras that can sense the emitted infrared light. The LEDs light up in sequential pulses. The cameras record the pulsed signals and send information to the system's processing unit. The unit can then extrapolate the data to determine the position and orientation of the target. Optical systems have a fast upload rate, meaning latency issues are minimized. The system's disadvantages are that the line of sight between a camera and an LED can be obscured, interfering with the tracking process. Ambient light or infrared radiation can also make a system less effective.

Mechanical tracking systems rely on a physical connection between the target and a fixed reference point. A common example of a mechanical tracking system in the VR field is the BOOM display. A BOOM display is an HMD mounted on the end of a mechanical arm that has two points of articulation. The system detects the position and orientation through the arm. The update rate is very high with mechanical tracking systems, but the disadvantage is that they limit a user's range of motion.

Videometric

An alternate method of optical tracking is referred to as videometric tracking. Videometric tracking is somewhat the inverse of the cases just described in that the camera is attatched to the object being tracked and watches the surroundings, rather than being mounted in a fixed location watching the tracked object. The VR system analyzes the incoming images of the surrounding space to locate landmarks and derive the camera's relative position to them. For example, the camera could be mounted on a head-based display to provide input to the VR system, which would be able to determine the locations of the corners of the surrounding room and calcu- late the user's position from this information.

Inertial Tracking

Inertial tracking uses electromechanical instruments to detect the relative motion of sensors by measuring change in gyroscopic forces, acceleration, and inclination [Foxlin 1996]. These instruments include accelerometers which are devices that measure acceleration. Thus accelerometers can be used to determine the new loca- tion of an object that has moved, if you know where it started. Another instrument is the inclinometer, which measures inclination, or how tipped something is with respect to its "level" position (the tilt of a person's head, for instance). It is very much like a carpenter's level except that the electrical signal it provides as its output can be interpreted by a computer. Inexpensive transducers that provide angular rates using gyroscopes, combined with angular and linear accelerometers and inclinometers, can be used separately or together to provide small self-contained tracking systems. This tech- nology has been used in highly accurate, large-scale systems as a means of maritime and flight navigation via inertial navigation systems (INS). Inertial tracking operates with the same technique by which the inner ear aids in determining the head's orientation. A fluid tends to remain motionless while the surrounding structure rotates. Sensors within the structure relay information about the location of the structure relative to the fluid. The brain uses this informa- tion to calculate orientation and changes in orientation.

BODY TRACKING

Body tracking is the VR system's ability to sense position and actions of the participants. The particular components of movement that are tracked depend on the body part and how the system is implemented. For example, tracking head movement might consist of just 3-DOF location, just 3-DOF orientation, or full 6-DOF position information. Another example is tracking finger movement with a glove device, which might measure multiple joints of finger flexion or might measure just contacts between fingertips. Any component of the body can be tracked in one or more degrees of freedom, assuming a suitable tracking mechanism is available in an appropriate size and weight and is attached to the system.

The body parts and techniques of body tracking commonly used in VR appli- cations include

1. Tracking the head
2. . Tracking the hand and fingers
3. Tracking the eyes
4. Tracking the torso
5. Tracking the feet
6. Tracking other body parts
7. Indirect tracking

The Head

The head is tracked in almost every VR system, although not always the full 6-DOF. Most VR systems need to know something about the user's head orientation and/or location to properly render and display the world. Whether location or orientation information is required depends on the type of display being used. Head-based displays require head orientation to be tracked. As users rotate their heads, the scenery must adapt and be appropriately rendered in accordance with the direction of view, or the users will not be physically immersed. Location tracking, while not essential, enhances the immersion quality of these VR experi- ences. Location tracking helps provide the sense of motion parallax (the sense that an object has changed position based on being viewed from a different point). This cue is very important for objects that are near the viewer. Some VR experiences avoid the need for location tracking by encouraging or requiring the user to continuously move (virtually) through the environment.

The Hand and Fingers

Tracking the hand, with or without tracking the fingers, is generally done to give the user a method of interacting with the world . In multiparticipant spaces, hand gestures can also provide communication between participants. A hand can be tracked by attaching a tracker unit near the wrist or through the use of a tracked handheld device. If detailed information about the shape and movement of the hand is needed, a glove input device is used to track the positions of the user's fingers and other flexions of the hand. In this case, the hand position tracker is generally mounted directly on the glove. While glove input devices provide a great amount of information about a key interactive part of the user's body, they have a few disadvantages. First, they are hard to put on and take off, which is especially a problem in systems that otherwise encourage the sharing of interactive control of the world. Usually there is only a glove or two, and certainly only one or two are connected to the system at any given time. Consequently, to change control of the system requires that the first person remove the glove, the new control person don the glove, and in most cases go through a cal ibration routine with the glove.

The Eyes

Technology for tracking the direction in which the user's eyes are looking relative to their head has only recently become practical for use with virtual reality and, conse- quently, has not been tried in many applications. There are two basic areas in which eye tracking can be useful. One is monitoring the direction of the gaze to allocate computer resources. The scene displays a higher degree of detail in the directon of the tracked eye's gaze. Eye tracking could be used as part of the interface with the world itself. Objects might be selected or moved based on the movement of the eyes.

The Torso

Very few VR applications actually track the torso of the participant, although when an avatar of the user is displayed, it often includes a torso with certain assumptions made about its position, based on head and hand positions. However, the torso is actually a better indicator of the direction the body is facing than are either the head or hands. The torso's bearing might be a better element to base navigational direction on than head or hand positions (see the Placeholder application in Appendix D). The benefit of using the torso's bearing for travel direction correlates with the user's experience level in moving through an immersive virtual world. Novice users may adjust better to moving in the direction of their gaze (i.e., nose direction). However, limiting movement to the direction of gaze limits the user's ability to look around.

The Feet

Some work has been done to provide a means for tracking the feet of the user. Tracking the feet provides an obvious means of determining the speed and direction a user wishes to travel. The obvious method of determining feet movement is to track the position of each foot. The most common way to do this is to use electromagnetic trackers. The tracker is attached to the foot with a wire connecting that device to the computer or to a body pack (containing a radio transmitter). Optical tracking is another method and uses cameras to "watch" the feet. This doesn't require any sensors on the foot or attached wires. Tracking the feet this way, though, is tricky. You would most likely want to put a very high-contrast spot or sticker (a fiduciai marker) on the foot to give the camera a very specific reference for which to "look." Other less encumbering methods have been used as part of what are called platform input devices. These platforms basically are places to stand or sit while using the VR appli- cation.

Other Body Tracking

Other body parts can be tracked and used to control various aspects of a virtual world. These items include body functions, such as temperature, perspiration, heart rate, respiration rate, emotional state, and brain waves (FIGURE 3-8). These functions might be measured simply to monitor the participant's condition as they experience a world, or they might be used to control the world--to determine what experiences are most relaxing, for instance, and use that feedback to guide the user down a calmer path [Addison et al. 1995

Indirect Tracking

Indirect tracking refers to the use of physical objects other than body parts to estimate the position of the participant. These physical objects are usually props and platforms. For example, the movement of a hand- held device, such as a wand or a steering wheel mounted on a platform, are good indicators of the position of one of the participant's hands.

Sensors:

Magnetometers, Accelerometers and Gyroscopes

Before we get to the sensors themselves though, we first have to talk about MEMS or micro-electromechanical systems. The sensors used in electronic devices today are actually microscopic mechanical devices embedded in solid-state silicon microchips. Digital Light Processing or DLP projectors, for example, use millions of microscopic mirrors that tilt individually, thousands of times a second, to produce high definition images. Each mirror can be precisely tilted for fine gradation in light intensity. MEMS technology can provide the mechanical and electrical components needed to build devices such as gyroscopes at a tiny, tiny scale. The next step beyond MEMS is NENS, or nano-electromechanical systems. This takes these components out of the microscopic realm into the domain of nanotechnology. Without MEMS manufacturing methods any smartphone that needed a compass, accelerometer and gyroscope would be very bulky indeed.

A magnetometer is, as you probably can tell, a device that measures magnetic fields. Therefore it can act as a compass, by detecting magnetic North it can always tell which direction it is facing on the surface of the earth. Clever developers have repurposed the magnetometer for use with the Google Cardboard, where a magnetic ring is slid up and down another magnet, the fluctuation in the field is then registered as a button click. Metal detectors also use magnetometers, which is why they can only detect ferrous metals. That is, metals that can be magnetized. Magnetometers can work in a number of different ways. Some use permanent magnets and others use electromagnets. In either case when a magnetic field perturbs the material inside the magnetometer this is detected and the magnitude and direction of that magnetism can be measured.

An accelerometer is a mechanism that lets your device, such as a smartphone, know which way up it is. This is the sensor that tells your phone or tablet whether the screen should be in portrait or landscape mode. One accelerometer can tell whether it is in line with the pull of gravity or not, but if you combines three of them (one for each axis) you can tell which way up something is, since each axis is fixed in relation to the device it is in. MEMS accelerometers are a few hundred microns across, truly tiny devices that add an intuitive input method for computer systems. Of course, accelerometers can measure more than orientation. As the name suggests it can also measure the magnitude of acceleration along an axis. For example, they are used to trigger airbags during a crash where the g-force along the horizontal axis exceeds a certain threshold. In the case of motion sensing devices such as the Nintendo Wiimote the can tell how strong a swing in a particular direction is. The most successful MEMS accelerometers (the capacitive transduction type) consist of microscopic silicon parts that have an almost comb-like structure. When gravity (or g-forces from a fall or swing) perturb the combs they generate an electrical current that can be translated into acceleration data.

Gyroscopes

No matter how the frame (and whatever it's mounted to) changes orientation, the spinning disc in the middle stays true to its original plane. This is invaluable in, for example, aircraft where you may not know if the craft is level or not due to issues like visibility.

Of course, there are MEMS gyroscopes that achieves the same outcome, just using a different principles. The types of components used to measure rotation in electronic devices can be quite varied. Generally they detect vibration that’s translated to rotational measurement with microscopic tuning forks, vibrating wheels or resonant solid matter. These work like the organs of insects (known as halteres) that help to orient them them. These MEMS gyroscopes are also known as vibrating structure gyroscopes. Since objects that vibrate tend to continue vibrating in the same plane even when rotated, which means the vibrating mass generates a coriolis force that can be detected electronically.
In electronic devices such as smartphones, HMDs or motion controllers these sensors are often represented in multiple iterations. The device may have several gyroscopes, accelerometers and magnetometers in order to provide rich sensor input that can be interpreted by the software as highly accurate and complex movements. The goal is usually to achieve true six-degrees-of-freedom (6DoF), which covers all the degrees of motion for a rigid body in space. In virtual reality 6DoF is the gold standard for a motion sensor.

Motion capture

(sometimes referred as **mo-cap** or **mocap**, for short) is the process of recording the movement of objects or people. It is used in military, entertainment, sports, medical applications, and for validation of computer vision and robotics. In filmmaking and video game development, it refers to recording actions of human actors, and using that information to animate digital character models in 2D or 3D computer animation. When it includes face and fingers or captures subtle expressions, it is often referred to as **performance capture**. In many fields, motion capture is sometimes called **motion tracking**, but in filmmaking and games, motion tracking usually refers more to **match moving**.

In motion capture sessions, movements of one or more actors are sampled many times per second. Whereas early techniques used images from multiple cameras to calculate 3D positions, often the purpose of motion capture is to record only the movements of the actor, not their visual appearance. This *animation data* is mapped to a 3D model so that the model performs the same actions as the actor. This process may be contrasted with the older technique of rotoscoping.

Camera movements can also be motion captured so that a virtual camera in the scene will pan, tilt or dolly around the stage driven by a camera operator while the actor is performing. At the same time, the motion capture system can capture the camera and props as well as the actor's performance. This allows the computer-generated characters, images and sets to have the same perspective as the video images from the camera. A computer processes the data and displays the movements of the actor, providing the desired camera positions in terms of objects in the set. Retroactively obtaining camera movement data from the captured footage is known as *match moving* or *camera tracking*.

Optical systems

Optical systems utilize data captured from image sensors to triangulate the 3D position of a subject between two or more cameras calibrated to provide overlapping projections. Data acquisition is traditionally implemented using special markers attached to an actor; however, more recent systems are able to generate accurate data by tracking surface features identified dynamically for each particular subject. Tracking a large number of performers or expanding the capture area is accomplished by the addition of more cameras. These systems produce data with three degrees of freedom for each marker, and rotational information must be inferred from the relative orientation of three or more markers; for instance shoulder, elbow and wrist markers providing the angle of the elbow. Newer hybrid systems are combining inertial sensors with optical sensors to reduce occlusion, increase the number of users and improve the ability to track without having to manually clean up data

Reflective markers attached to skin to identify body landmarks and the 3D motion of body segments

Passive markers

Passive optical systems use markers coated with a retroreflective material to reflect light that is generated near the cameras lens. The camera's threshold can be adjusted so only the bright reflective markers will be sampled, ignoring skin and fabric.

The centroid of the marker is estimated as a position within the two-dimensional image that is captured. The grayscale value of each pixel can be used to provide sub-pixel accuracy by finding the centroid of the Gaussian.

A dancer wearing a suit used in an optical motion capture system

Markers are placed at specific points on an actor's face during facial optical motion capture.

An object with markers attached at known positions is used to calibrate the cameras and obtain their positions and the lens distortion of each camera is measured. If two calibrated cameras see a marker, a three-dimensional fix can be obtained. Typically a system will consist of around 2 to 48 cameras. Systems of over three hundred cameras exist to try to reduce marker swap. Extra cameras are required for full coverage around the capture subject and multiple subjects.

Vendors have constraint software to reduce the problem of marker swapping since all passive markers appear identical. Unlike active marker systems and magnetic systems, passive systems do not require the user to wear wires or electronic equipment.Instead, hundreds of rubber balls are attached with reflective tape, which needs to be replaced periodically. The markers are usually attached directly to the skin (as in biomechanics), or they are velcroed to a performer wearing a full body spandex/lycra suit designed specifically for motion capture. This type of system can capture large numbers of markers at frame rates usually around 120 to 160 fps although by lowering the resolution and tracking a smaller region of interest they can track as high as 10000 fps.

Active marker

Active optical systems triangulate positions by illuminating one LED at a time very quickly or multiple LEDs with software to identify them by their relative positions, somewhat akin to celestial navigation. Rather than reflecting light back that is generated externally, the markers themselves are powered to emit their own light. Since inverse square law provides one quarter the power at two times the distance, this can increase the distances and volume for capture. This also enables high signal-to-noise ratio, resulting in very low marker jitter and a resulting high measurement resolution (often down to 0.1 mm within the calibrated volume).

The TV series *Stargate SG1* produced episodes using an active optical system for the VFX allowing the actor to walk around props that would make motion capture difficult for other non-active optical systems.[*citation needed*]

ILM used active markers in *Van Helsing* to allow capture of Dracula's flying brides on very large sets similar to Weta's use of active markers in *Rise of the Planet of the Apes*. The power to each marker can be provided sequentially in phase with the capture system providing a unique identification of each marker for a given capture frame at a cost to the resultant frame rate. The ability to identify each marker in this manner is useful in realtime applications. The alternative method of identifying markers is to do it algorithmically requiring extra processing of the data.

There are also possibilities to find the position by using coloured LED markers. In these systems, each colour is assigned to a specific point of the body.

One of the earliest active marker systems in the 1980s was a hybrid passive-active mocap system with rotating mirrors and colored glass reflective markers and which used masked linear array detectors.

Time modulated active marker

Active marker systems can further be refined by strobing one marker on at a time, or tracking multiple markers over time and modulating the amplitude or pulse width to provide marker ID. 12 megapixel spatial resolution modulated systems show more subtle movements than 4 megapixel optical systems by having both higher spatial and temporal resolution. can see the actors performance in real time, and watch the results on the motion capture driven CG character. The unique marker IDs reduce the turnaround, by eliminating marker swapping and providing much cleaner data than other technologies. LEDs with onboard processing and a radio synchronization allow motion capture outdoors in direct sunlight, while capturing at 120 to 960 frames per second due to a high speed electronic shutter. Computer processing of modulated IDs allows less hand cleanup or filtered results for lower operational costs. This higher accuracy and resolution requires more processing than passive technologies, but the additional processing is done at the camera to improve resolution via a subpixel or centroid processing, providing both high resolution and high speed. These motion capture systems are typically $20,000 for an eight camera, 12 megapixel spatial resolution 120 hertz system with one actor.

Semi-passive imperceptible marker

One can reverse the traditional approach based on high speed cameras. Systems such as Prakash use inexpensive multi-LED high speed projectors. The specially built multi-LED IR projectors optically encode the space. Instead of retro-reflective or active light emitting diode (LED) markers, the system uses photosensitive marker tags to decode the optical signals. By attaching tags with photo sensors to scene points, the tags can compute not only their own locations of each point, but also their own orientation, incident illumination, and reflectance.

These tracking tags work in natural lighting conditions and can be imperceptibly embedded in attire or other objects. The system supports an unlimited number of tags in a scene, with each tag uniquely identified to eliminate marker reacquisition issues. Since the system eliminates a high speed camera and the corresponding high-speed image stream, it requires significantly lower data bandwidth. The tags also provide incident illumination data which can be used to match scene lighting when inserting synthetic elements. The technique appears ideal for on-set motion capture or real-time broadcasting of virtual sets but has yet to be proven.

VIDEO BASED INPUT

Virtual reality (VR) is an exciting new medium with broad applications in entertainment, marketing, design, and more. Every bit as flexible and dynamic are the professionals who specialize in creating and adapting content for virtual reality. VR content creators use two main methods to create VR content: computer-generation, wherein every part of the world is synthetic, designed and integrated into an interactive experience using code; and 360-degree video, where video is taken using an omnidirectional camera and edited to create an immersive experience.

360-Degree Immersive Video

By far the least expensive and speediest method of creating virtual reality content is 360-degree video. Of course, that doesn't mean it's easy or cheap: video is shot using an omnidirectional camera, spliced together using special video stitching software, and edited further in order to be optimized for viewing on a head-mounted display (HMD). Frequently, content creators will fill more than one role on a single project, but VR content creators who work in this medium can take on a number of jobs, including:

- direction—conceiving and leading the project;
- filming—setting up tripods, guiding the Steadicam, or piloting a drone-mounted camera;
- audio capture—setting up lavaliermics or even sitting directly beneath the tripod; and postproduction.

As far as the projects themselves, 360-degree video can support fictional narratives and nonfiction documentaries, just like traditional video, but it's particularly effective when it comes to marketing and event management. The ability to live broadcast VR content allows sports games, concerts, conferences, and trade-shows to become immersive experiences for audiences at home, while brands take advantage of the eye-catching medium to capture viewers on sites such as Facebook and Youtube. Video-based VR also has cutting-edge applications in fields such as tourism, real estate, education, and surveillance—military and otherwise.

Interactive 3D Development

The "virtual" in virtual reality is why most people tend to picture this kind of content when they think of putting on a VR headset. If creating video-based VR content is like making a film, then creating interactive VR content is like developing a video game. Creators use 3D game development software—called "engines"—to build entire worlds from the ground up. This often means creating original 3D models and animations with software such as 3DS Max or Maya; using applications such as Photoshop or Zbrush to add color and texture; using the engine's built-in tools to add lighting to the scene; and designing original sound effects to bring the visual content to life. Perhaps most importantly, interactive VR content creators write code—using the advanced functionality provided by the game engine—to tie everything together.

While this method of creating VR content is much more expensive and time-consuming than shooting 360-degree video, it has the advantage of interactivity—thanks to the code—meaning that users can do much more than just look around. The ability to pick things up, move around, and edit virtual objects in an immersive environment makes interactive VR excellent for computer-aided design (CAD) as well as video game-style entertainment. Autodesk's VRED, for example, is a virtual prototyping software for automotive designers that allows them to see and edit their designs at scale without needing to build a model.

VIRTUAL REALITY MENUS

How do you design scalable Virtual Reality menus? How do you browse through the increasing number of 360 videos and photos or all those new VR apps you just downloaded with your headset on? While more and more content is pouring into VR, the question of how to explore it the best way in VR is far from solved. To get a feeling for the state of the art I embarked on a little design research journey through various VR applications to see what is out there in terms of menu design.

I like big picture overviews, so let's start with one. As I see it, there are three major design directions when designing menus for VR:

- Skeumorphic menus
- Flat menus mapped on (simple) geometry
- Real 3D interfaces

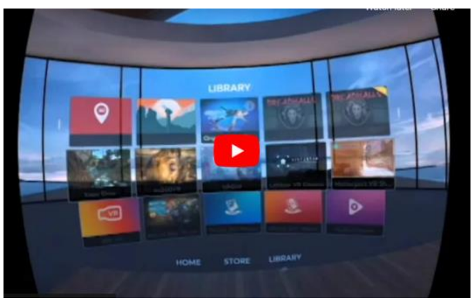
Skeumorphic Menus

Skeumorphic design is the idea of designing virtual things mimicking their real world counterparts. For example the note app on iOS looked like a real notebook with yellow pages and lines for writing.

In VR this could mean that a menu for a photo app looks actually like a shelf with plenty of photo albums. You can select an album, open it and start flipping pages. There is no need for a VR photo app to look like that since we do not have the same constraints as in the real world. No need to store things in shelves, we can just let them appear out of thin air. The big plus of skeumoprhisms: you enable the user to leverage the experiences with the real world to operate the new menu.

Flat Menus Mapped on Geometry

Flat menus are building upon interfaces we know from computers and especially tablets & phones. They rely on basic building blocks like cards & tiles and hierarchical folder structures. The apps on your smartphone for example are arranged in little tiles and you can group them together into a folder to create a hierarchy. Simple interfaces like that became very popular with the advent of the smartphone and one could argue that bringing them into VR already represents a skeumorphism of its own kind – but let's drift of into the philosophical here.



In VR those flat, 2D interfaces can be mapped onto simple geometric shapes like a curved screen or the inside of a sphere around the user, something we see in many of the currently available apps including Oculus Home on GearVR.

Real 3D interfaces

Before I summon the wrong pictures in your head: I'm not talking about crazy hand gestures and I'm most certainly not talking about the infamous Minority Report display! Real 3D interfaces so far are quite rare, although they seem like the most natural fit for VR. The problem: we have many years of experiences and a developed design language for flat interfaces, but not so much for 3D ones – especially when it comes to menus.

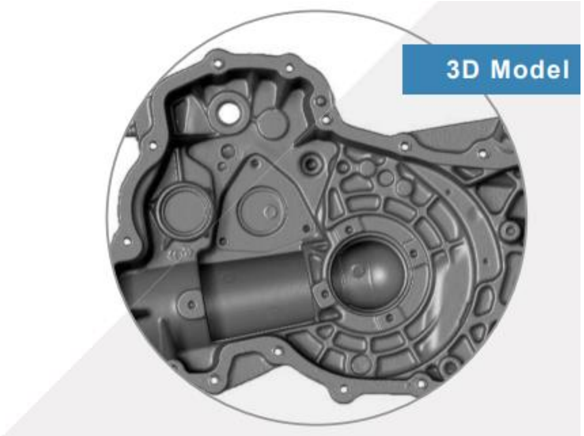
So why bother to dive into this when flat interfaces seem to work fine? Well first, an additional dimension, by definition, enables you to present richer information. While this does not scale infinitely on the human side of things, we are creatures that grew up in a three-dimensional world and thus well equipped to deal with three dimensions and using an environment that is more natural to us enables us make use of a lot of those very nifty subconscious co-routines we are equipped with.

We see a glimmer of what is possible with thoughtfully designed 3D interfaces in VR drawing apps like Tilt Brush. They put the menu around the controllers, making a large number of functions immediately accessible while requiring little space. Having all items present all the time in 3D space (continuity & physicality) keeps the interface very explorable and easy to learn – you will never search through a number of sub menus for that one particular item. Hierarchy is solved by proximity and location.

3D SCANNERS

3D scanners are tri-dimensional measurement devices used to capture real-world objects or environments so that they can be remodeled or analyzed in the digital world. The latest generation of 3D scanners do not require contact with the physical object being captured.

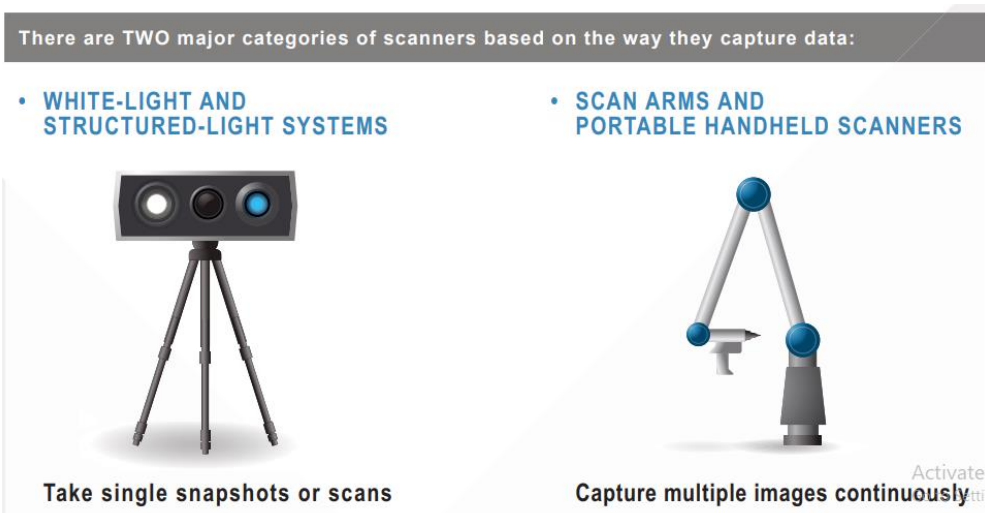
3D scanners can be used to get complete or partial 3D measurements of any physical object. The majority of these devices generate points or measures of extremely high density when compared to traditional “point-by-point” measurement devices.



OBJECTS ARE USUALLY SCANNED IN 3D FOR 2 PURPOSES:

1. Extracting dimensions to reconstruct a CAD reference file for reverse engineering or rapid prototyping.
2. Measuring the object for analysis and documentation. This is done for applications such computer aided inspection (CAI), digital archiving and computer aided engineering CAE analysis.

There are TWO major categories of scanners based on the way they capture data:



AUDITORY

The goal of 3-D sound system designer ought to involve the notion of complete control and manipulation of someone else's spatial auditory perception. In other words, by manipulating a machine—either in real time or algorithmically—the operator of a 3-D sound system will want to predict, within a certain level of allowable statistical predictability, the auditory spatial imagery of a listener using the 3-D sound system. This is referred to as spatial manipulation.

Aural Rendering Systems

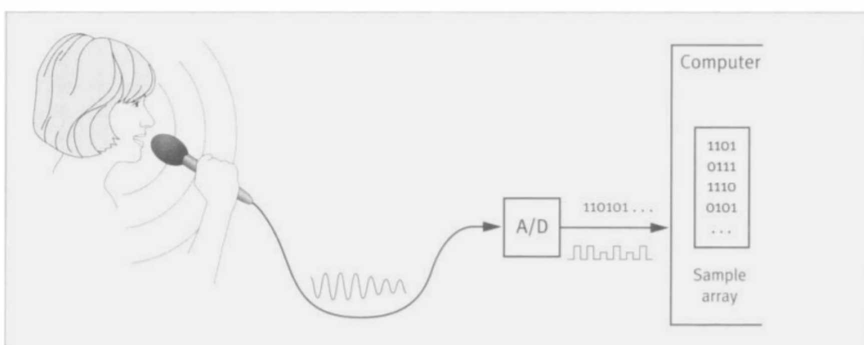
Computerized sound generation for virtual reality is also an active area of research. Information about how to create sounds via computer is less readily available than computer graphics information and is spread over many disciplines (including computer graphics, computer music, scientific visualization, and virtual reality). We'll begin our discussion with the basic ways sound can be rendered. From there, we'll cover sound generation hardware, aural software formats, and aural interface systems.

Methods of Aural Rendering

In this section, we discuss three basic methods of rendering aural signals: (i) the playback of recorded waveform samples, (2) synthesis (algorithmic rendering), and (3) the postprocessing of existing sound signals. Sounds are vibrations that travel through the air (or other media) in waves. The frequency of the waves, or how many cycles of the wave occur per unit of time, determines whether the sounds are highor low-pitched. The wave's amplitude, the other key feature of sound, determines the loudness of the sound.

Sampling

A common way of producing sounds is through the playback of digitally recorded samples of physical world sounds. An analog-to-digital (A/D) converter converts analog signals into digital signals (FIGURE 5-23). The recordings are created by sending the output of a microphone through an A/D converter, which measures the voltage at regular time intervals (sampling rate). The sampling rate and the number of digital bits used to encode the voltage are measures of the resolution of the sound sample. The frequency that the voltage measurements are taken at range from 8,000 Hz for telephone-quality spoken audio to 96,000 Hz for DVD-quality sound. The sound quality for CDs measures at 44,1oo Hz. The number of bits available for each measurement determines the dynamic range of the recorded signal.



will become bored or annoyed. Sound repetition can be easily detected by the participant (as can the repetition of patterns caused by visual texture maps). Multiple digitized sound samples can be altered and combined to create a richer, less repetitive environment. They can also be combined with other, algorithmically generated, sounds.

FIGURE 5-23 *Real-world sounds (such as a voice) can be recorded by doing analog-to-digital (A/D) conversion. The resulting digital signals can be processed, stored, or otherwise manipulated by the computer system.*

Synthesis

Synthesized sounds are those that are generated by executing an algorithm or some other component of the VR simulation that computes the waveform (FIGURE 5-2/4). This technique offers extreme flexibility to the experience developer, because creating any sound is possible, yet may require a significant computer engine or specialized synthesizer in order to render the samples in real time. Synthesizing complex, realistic broadband sounds is difficult. Soundgeneration algorithms range in complexity from producing simple sounds like sine waves to sounds created by modeling the properties of a sound-generating object. Methods of sound synthesis can be divided into three subcategories: spectral methods, physical models, and abstract synthesis.

Spectral Methods.

Spectral methods of sound synthesis involve observing a sound wave's frequency spectrum (the frequencies and amounts of each of the components that make up the sound) and re-creating that spectrum to mimic the original. Most of the sounds we hear in the real world are broadbanded; that is, they cover a wide variety of frequencies, so it is generally not feasible to create the entire spectrum ofsuch sounds in real time. Musical sounds, however, use only a small subset of frequencies.

Physical Models.

Physical models are created by computing sounds based on the physics of the object(s) generating the sound. For example, a physical sound model of a flute relies on computing the vibrations that occur as air flows through a tube of specific dimensions. That is, the underlying physical phenomena (air flow, vibration, etc.) are modeled, and the results of those computations are directly used to create the corresponding real-world sound. With this method, sounds can be made to seem more authentic and integral to the environment.

Abstract Synthesis.

Abstract synthesis of sounds creates sounds from streams of numbers generated by some system and maps the values to sound waveforms by some given function. Rather than re-creating a sound, this technique is used to create sounds that have never been heard before, based on some numerical system. For example, a risk analysis formula could reflect the change of commodity market instruments over time and allow an analyst to become familiar with patterns by listening to their fluctuations

HAPTIC RENDERING SYSTEMS

Of the major senses used in VR systems, the haptic sense is the most difficult to incorporate. One reason for this challenge is that the haptic sense results from direct contact with the environment and thus involves direct two-way communication between the participant and the world. Haptic devices are the only human-computer interfaces that both receive and provide stimuli. A malleable object is rendered so the user can feel its shape, texture, and temperature; in addition, if pushed on with sufficient force, the object's shape will be changed in accordance with its elasticity. It is challenging to create and maintain illusory displays for the haptic sense because of the direct contact required for the stimuliAnother challenge to successful haptic display is that the human haptic system includes both tactile (skin-based) and kinesthetic (muscle/joint-based) feedback. The two sensations are closely interrelated, but to date VR systems typically address only one or the other. So for practical purposes, our discussion on haptic rendering is divided into two basic techniques: skin-based rendering (temperature and surface texture) and muscle/joint-based rendering (surface texture and force).

Haptic Rendering Methods

In our discussion of displays in, we covered three types of common haptic displays: (I) tactile devices (attached to the skin), (2) end-effector displays (mechanical force applied to a stylus, finger-grip, etc.), and (3) robotically operated shape displays (mechanism for placing physical objects in the appropriate position). If an application designer is not motivated to design new forms of haptic display, they must use rendering methods that can be accommodated by these devices.

As we've said, haptic perception consists of both skin and kinesthetic (muscle and joint) sensations. Skin stimuli such as temperature, pressure, electric current, and surface texture can all be displayed using tactile display devices. Surface texture can also be rendered using small-scale end-effector (mechanical force) displays. Kinesthetic sensations allow people to determine such features as surface 267 I HAPTIC RENDERING SYSTEMS shape, surface rigidity, surface elasticity, object weight, and object mobility. Kinesthetic information can be rendered using either end-effector or robotically operated shape displays (ROSDs). The following sections provide an overview of the basic haptic rendering methods currently in use.

Thermal Rendering

Rendering temperature is fairly straightforward. All that is required is to transfer heat to and from an end-effector that is near or touching the skin. Fingertip transducers are end-effector devices used to provide the sensation that virtual objects are hot or cold. Of course, safety is a concern when applying heat or, indeed, extreme cold to the fingertips. The ambient air temperature may also be controlled via the activation of a heat lamp or other climate control device.

Pin-based Rendering

A tactile feedback array display renders tactile information by finger movement over small pins arranged to reproduce the sensation of texture. The movement of the pins is varied in accordance with finger motion in order to give the impression of rubbing a surface to feel texture. One rendering method uses pins arranged on a flat surface, rising and falling as the finger moves across the display. Another method uses pins mounted on a cylinder. In this instance the pins rise and fall as the cylinder rotates.

Kinesthetic Display

Object shapes can be rendered using kinesthetic rendering techniques such as a force display. By using robotics to control the movement of an end-effector, the user can "feel" where an object exists by their inability to penetrate the surface. They might also be able to sense qualities such as the object's elasticity and surface texture if that information is rendered. The texture tactile sensation of running a stylus over a surface to feel its texture can be rendered using a force display as illustrated in F~GURE 5-3O.

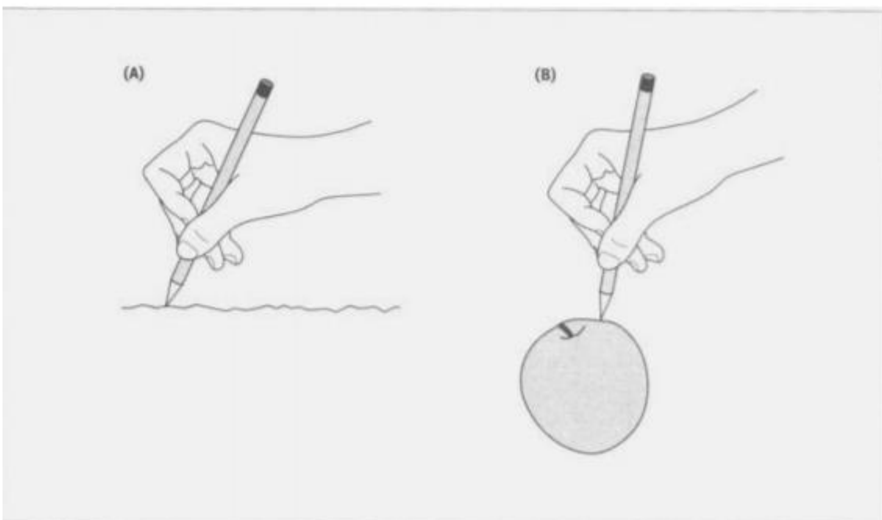


FIGURE 5-30 **(A)** *The sensation of a textured surface can be produced via a stylus that moves according to the virtual surface texture as a user drags the stylus across that virtual surface.* **(B)** *A stylus can be used to probe the surface characteristics of a virtual object.*

Robotically Operated Shape Display (ROSD)

Object surfaces can also be rendered using robotic displays as we saw in Chapter 4. A physical surface with an appropriate edge or surface angle is carefully placed in front of the user's finger or finger-surrogate (typically a stylus) to emulate the virtual object. As the user moves the probe, the surface being displayed changes appropriately.

Specific Robotically Operated Shape Display

Specific objects can also be rendered with specialized robotic displays equipped with sample items that serve to represent objects encountered in the virtual world, such as the many switches in an aircraft cockpit as implemented in a VR training application by Boeing [McNeely 1993]. These displays operate by having a robot place actual objects at the appropriate position before the user's finger or stylus arrives at that location. For example, an actual toggle switch can be placed by the robot at a location where a virtual switch is represented in a virtual world.

Physical Object Rendering (3D Hardcopy)

A final alternative is to "render in plastic." This is an intuitive, albeit nonmalleable, technique of actually creating a physical model of a virtual object, bringing it into the real world such that it can be held in the hand and directly experienced. One such technique is stereolithography.However, this technique does not offer interactive feedback.

Single Point of Contact

A primary form of force interface in VR experiences is at a single point of contact with an object. The force display provides stimuli to a fingertip or the tip of a handheid stylus, but no rotational (torque) information is provided. This type of display is often provided by basic end-effectors or robotically operated shape displays. A basic 3-DOF force display is all that is required.

Two Points of Contact (Pinching)

Pinching an object is another method for including torque interaction through force displays, in this case, two 3-DOF translational displays (FIGURE 5-33). By combining the two devices, torque is applied at the midpoint of the line between the two contact points. No torque can be driven around the line from one contact point to the other, which yields a 5-DOF display.

Multiple Points of Contact (Grasping)

Grasping an object means that the user and object have multiple points of contact. Objects not attached within the simulated world allow the user complete 6-DOF freedom in moving them. To provide enough haptic feedback to the user, a device that encompasses the entire hand (such as a glove) is generally used for grasping interfaces.

Haptic Rendering Techniques

There are several methods for simplifying the amount of information needed to pass between the world simulation and the haptic rendering:

1. Spring and dashpot
2. Point and plane
3. Multiple plane
4. Point to point
5. Multisprings
6. Inertial and resistant effects
7. Vibration
8. Error correction

Spring and Dashpot Model The spring and dashpot model allows the system to control direction, tension, and damping between a probe and surface in the virtual world. The makeup of a physical dashpot is that of a piston in a cylinder, with a small orifice in the end of the cylinder. As the piston moves back and forth within the cylinder, it pushes air in and out through the orifice. This results in viscous losses proportional to the rate of travel of the piston. A dashpot functions as a viscous damping device much like the shock absorbers of an automobile. Thus, some physical interactions can be approximated in the virtual world by using the equations that describe how springs and dashpots interact in the physical world.

Point and Plane and Multiple Plane Models The point and plane mode/represents the interaction between a probe stylus and a surface by placing a virtual planar surface tangential to (the nearest surface of) the probe's tip. As the probe traces the surface shape, the plane moves along the tangent of the surface to simulate the shape of the virtual object (FIGURE 5-34).

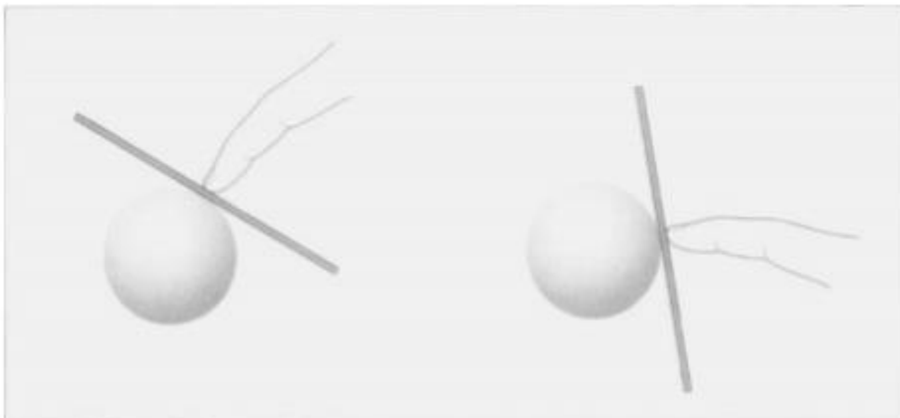


FIGURE 5-34 *The point and plane model uses the metaphor of providing a plane at the point of intersection with a probe. For example, as the tip of a finger or finger surrogate moves over a haptically rendered virtual sphere, the haptic computational model continually moves the plane to be tangent to the sphere at the point of contact to provide the sensation of touching an actual sphere.*

However, it is very difficult to simulate the corners of objects and movement through highly viscous fluids using this model. This model can be implemented using the spring and dashpot physical mode[. The multiple plane model is an extension of the point and plane model. With the addition of more virtual plane surfaces, this model provides a simplified method of rendering discontinuous corners in the virtual world. As the probe moves toward a corner, additional planes represent the complexity of the shape being depicted (FIGURE 5-35).

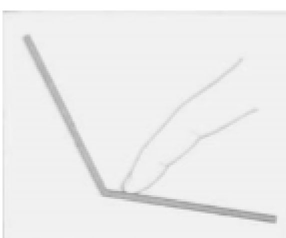


FIGURE 5-35 *The multiple plane representation allows discontinuous surfaces to be rendered.*

Point to Point Model The point to point model uses a basic spring model comprised of the equations that describe the forces of how a spring behaves when it is stretched and compressed between two points. Point to point is often used not as a general model, but as a transitory model to maintain stability in the midst of highly fluctuating forces gener- ated by complicated calculations. For a simulation that may be providing widely divergent forces, perhaps too drastic to be accurately rendered, a spring can be modeled to act between the simulated probe point and the tip of the physical probe. As an analogy, if you were hanging onto the end of an elastic cord, with the other end being tugged by someone else moving in a very erratic fashion, you would expe- rience only a dampened amount of the erratic motion from the other end.

The multispring model adds torque simulation. Torque rotation cannot be simulated if contact occurs at only one point. By arranging multiple springs in a pattern around the tip of the display, forces can vary on each side of the tip, resulting in rotational simulation (FIGURE 5-36). A force display must be capable of rendering torque for this model to be effective.

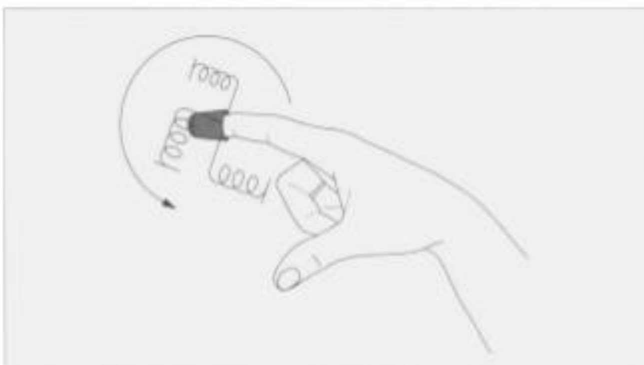


FIGURE 5-36 *Multispring models can be used to add torque to any of the other haptic repre- sentations [Adapted from Mark et al. 1996].*

Inertial and Resistant Effects Models Inertial and resistant effects models add characteristics of friction and viscosity (both resistive forces) and momentum (an inertial force) to the display. Friction is resist- ance caused by rubbing between surfaces (such as the tip of a stylus and an objecO. Viscosity is resistance to movement in a fluid (e.g., water, aiO-not necessarily on a surface. Surfaces define a shape, but when touched, we also get a sense of features like smoothness, compliance (how it conforms to our hand), and friction. These features are orthogonal. For example, a smooth surface can feel very different depend- ing on the material of which it's composed. Polished marble is smooth to the touch, with low surface friction, and thus has a different feel than a surface made of a high- friction, compliant material like rubber. Some simulated worlds emulate inertial effects. When a user contacts an object with high mass, the user feels resistance against slowing it down or getting it moving

Vibration Models Vibration can be modeled on a force display, or it can be rendered directly on a vibrating end-effector. For specialized devices, a signal indicates when the display should vibrate and at what frequency and amplitude. The use of specialized vibrator devices can allow vibration at frequencies greater than those of force display devices, which commonly operate at about 1,ooo Hz.

Error Correction Models Error correction models come into play when the haptic display violates the laws of a virtual world. This situation can occur because the frame rate of the simulation is significantly slower than both the frame rate of the haptic display and the user's speed. It is not uncommon for such errors to occur. The error correction model inter- venes to reconcile the discrepancies.

HAPTIC RENDERING ENGINES

There are not many generally available tactile rendering engines. In fact, not many tactile-based displays have even been developed, merely a modest selection of temperature, vibrator, inflatable bladder, and pin-based displays. Compared with tactile displays, more work has been done on force display and rendering systems. Some early work along these lines has been done at the University of North Carolina at Chapel Hill (UNC-CH).

Additional work has been done in the telepresence and robotics industries. Without specialized haptic rendering hardware to generate the signals needed to drive a haptic display device, the typical solution uses an additional computer system to generate the end-effector display or robotic movements.

A common solution off-loads the rendering to a separate, dedicated CPU, allowing the world simulation CPU to focus on higher level operations. In this respect, the computer system dedicated to haptic rendering becomes the haptic accelerator, analogous to graphical rendering engines.

VR HAPTIC INTERFACE SOFTWARE

- To this point, only a handful of haptic rendering interface software libraries have been produced.
- All the examples are for the control of force displays (which, again, can be used for rendering kinesthetic forces or surface texture).
- There are no widely available generalpurpose software libraries for temperature, tactile array, bladder, and other haptic displays.
- The University of North Carolina at Chapel Hill has created a software library for the control of force display devices called Armlib.

The key feature of Armlib is that it is a device independent, 6-DOF force display library freely available from U NC-CH, including the source code. The University's GRIP (GRaphical Interaction with Proteins) project developed several generations of Armlib for a variety of force devices [Mark et al. 1996].

Gesture interaction in virtual reality

Gesture is one of the most important communication methods of human beings, which can effectively express users’ demands. In the past few decades, gesture-based interaction has made significant progress.

The virtual environment can be a scenery conceived by a designer or a reproduction of the live scene. Professional equipment such as VR headsets and data gloves allow users to sense and control various virtual objects in real time, creating an experience that users cannot obtain in the real world and generating real response or feedback.

VR has three distinct characteristics: interaction, immersion, and imagination. Interaction refers to the natural interaction between the user and the virtual scene. It provides the users with the same feeling as the real world through feedback. Immersion means that the users feel that they are part of the virtual world in the VR scene, as if they are immersed; Imagination refers to the use of multi-dimensional perception information provided by VR scenes to acquire the same feelings as the real world while acquiring the feelings that are not available in the real world.

Hand gesture

In real life, people use their hands to perform operations such as grasp, move, and rotate. In the process of communication, people spontaneously attract the attention of others by hand movements. Gesture is the way people express their will under the influence of consciousness. Therefore, gesture interaction has two functions: one is the movement of the hand, and the other is the specific meaning of the gesture.

Gesture interaction technology in virtual reality

Human-computer interaction, as an important supporting technology of VR, provides a variety of interactive modes based on different functions and purposes, enabling people to obtain immersive feelings in 3D virtual scenes. As an emerging technology, VR is a new interactive human-computer interaction interface. Therefore, a new generation interface paradigm Post-WIMP and Non-WIMP has been defined, and the interaction also extends from a mouse- and keyboard-based 2D graphical user interface to a natural user interface or supernatural user interface. Unlike traditional WIMP, in a natural user interface or a supernatural user interface scenario, the user interacts in a 3D environment simply by VR simulation or interactive scenario design, which eliminates waste of resources generated by repetitive manufacturing and controls production costs. With the continuous development of interactive technology, VR has been widely used in many fields such as game entertainment, medical care, and education.

Gesture is a conscious or unconscious movement of the hand, arm, or limb. The gestures in the interaction process can be distinguished according to different spatiotemporal operation behaviors, different semantics, different interaction modes, and different interaction ranges. In the early interaction scenarios, the acquisition of gesture signals was mainly based on wearable sensor devices such as data gloves; with the development of mobile devices such as mobile phones and tablets, touch screens became popular; consequently, the acquisition of gesture signals has developed into visual signal acquisition from computer cameras. In recent years, information acquisition based on high-tech equipment such as electromyographic signal acquisition has gradually become a research focus. Owing to the variability and complexity of the gesture, how to use the existing technology to process the input signal collected by the device, how to determine the spatial position and posture of the hand, and how to obtain an accurate recognition result have great impact on the effect of subsequent gesture interactions. Gesture recognition technology is mainly categorized into gesture recognition based on wearable sensor devices, gesture recognition based on touch devices, and gesture recognition based on computer vision.

Gesture definition and gesture classification

Gesture is a posture or movement of the user's upper limbs. Through gestures, people can express their interaction intentions and send out corresponding interactive information. Generally, gestures convey information or intentions through physical movements of the face, limbs, or body. In the VR environment, the posture or movement of the users' body is used as input, and the posture information of the user through the interaction device in the physical world is used as system input. In the interaction process, users can not only generate gestures using hand, finger, and palm, but also generate gestures as an extension of the body through tools such as a mouse, pen, and glove. Users can send simple commands to the system through gestures, such as selecting, moving, and deleting, and can also express more complex intents, such as switching the current interactive scene, controlling virtual objects, and performing virtual actions. In different situations, the same gestures may have different meanings due to factors such as culture or geography.

2.1. Classification based on spatiotemporal status

According to the spatiotemporal operating state, gestures are generally classified into static gestures and dynamic gestures. A static gesture is a static spatial posture of a finger, palm, or arm at a certain moment. It usually only represents one interactive command and does not contain time series information. A dynamic gesture refers to the posture changes in a finger, palm, or arm over a period of time. Accordingly, it also contains time information. The difference between the two is that static gestures only include spatial gestures without causing changes in spatial position, whereas dynamic gestures are characterized by movements in spatial locations over time, such as waving gestures, which are typical dynamic gestures. Dynamic gestures also include conscious dynamic gestures and unconscious dynamic gestures, in which unconscious movements during physical activity are called unconscious dynamic gestures, and gestures for communication purposes are called conscious dynamic gestures.

2.2. Classification based on gesture semantics

Unintentional movements mean that the movement of the hand/arm does not convey any meaningful information, whereas intentional movement is referred to as the gesture. Gestures are divided into manipulative gestures and communicative gestures according to whether they contain communication characters. Manipulative gestures is the action of control the state of the objects through the hand/arm movement, such as moving and rotating objects through the movement of the arm. The communicative gestures can be customized gestures with a specific information function. In a natural interaction situation, they are usually accompanied by verbal communication, mainly including symbol gestures and action gestures. Symbol gestures have a fixed-constraint meaning and can be further divided into referential gestures and modalizing gestures. Action gestures include mimetic gestures that mimics the actual interaction behavior and an deictic gestures with a direction-indicating function.

2.3. Classification based on interaction mode

According to different interaction methods, gestures can be classified into media gestures, direct contact gestures, and non-direct contact gestures depending on the interaction mode. Media gestures are a single-point input device through physical contact, such as a mouse, a joystick, or a trackball, to transmit the obtained data series into the computing system. Direct contact gestures are gesture operations directly on the input device through part of the body or a physical object tool. When the user touches a device such as a screen, the gesture interaction is turned on, and the touch movement is an intermediate state of interaction. Direct contact gestures also include a single touch gesture and a multi-touch gesture. The single touch gesture is a gesture operation provided by the user through a single point of interest using a pointing input device (such as a mouse or pen) or a body part (such as a hand). A pen gesture is a typical single-touch gesture that records the movement trajectory of a pen and uses a symbol or logo for intention expression.

2.4. Classification based on the scope of interaction

According to the different scope of interaction, gestures can be classified into “stroke"gestures and “mid-air” gestures. The former is used mainly on a support surface based on a handwriting gesture of the upper part of the wrist. That is, the user directly uses the hand or tool to move on the support surface and communicate with the computer for information exchange, such as touch screen interactions. The mid-air gesture is based on the user’s limbs or interactive devices moving in a space without a supported surface. That is, the user’s limbs or interactive tools are not in contact with the computer while interacting in a 3D space.

3. Gesture interaction device in virtual reality

There are a variety of gesture interaction devices used for VR interaction, which enable users to interact more realistically with objects in the virtual world. For different devices, the gesture information input can be classified into wearable-sensor device-based input, touch device-based input, and computer vision interaction device-based input. Wearable sensor-based user interaction devices mainly include data gloves and accelerometers. Touch-based interactive devices mainly include capacitive touch-screen and resistive touch-screen. It is worth mentioning that some devices which have direct contact with the user could have a certain degree of impact on the user's health. For example, mechanical sensor materials may cause allergic symptoms in some users[15]. Computer vision device-based interaction mainly refers to input through the camera, including single camera, binocular camera, and depth camera. This method does not require direct contact, and is more user-friendly, but the configuration and processing method of the device is relatively complicated. There are also many gesture interaction output display devices in VR, such as computer monitors, VR glasses and VR helmets. Users can select appropriate input and output devices according to different interaction scenarios in order to build a gesture interaction system in a VR environment. The following introduces different types of gesture interactive input devices.[1]

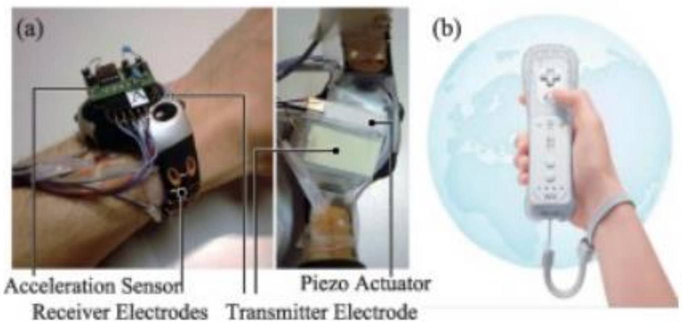
3.1. Wearable sensor-based device

1. Data gloves

As an important collection device for VR gesture interaction, data gloves can collect the posture and motion of the human hand in real time. The device directly detects the activity signal of each joint of the hand through various sensor arrays. It uses the coupling relationship between the data vector of each part of the hand, measured by the corresponding sensor, and the data vector of each bending angle of the hand, to obtain the degree of bending or stretching of each part, and to locate the 3D position of the hand. Data gloves can be classified into VR data gloves and force feedback data gloves. The VR data gloves can collect the posture and motion information of the hand, and force feedback data gloves add a force feedback function based on the VR gloves[16]. In the interactive system of VR, the data gloves can achieve two functions. On the one hand, the data glove acts as an input device, and can collect the gesture movements of the user in real time and convert the collected signal into a virtual hand motion. The user can observe the activity of the real hand in the virtual space through the movement of the virtual hand and can operate a virtual target by using various gestures. On the other hand, when using the feedback data glove in the virtual space, the output control of the feedback device enables the user to feel the physical property of the target during the operation and increases the realism for the user.

2. Inertial sensor

The micro-electro-mechanical system (MEMS) is an important branch of miniaturized inertial sensing. It is an independent intelligent system, mainly composed of three parts: sensor, actuator, and micro energy, which can effectively collect information about gestures. Common MEMS systems include MEMS accelerometers, MEMS gyroscope sensors, and MEMS pressure sensors. The sensors have high sensitivity and are usually fixed to the wrist, arm, or other positions to obtain gesture data, bringing new tools for gesture interaction. These devices are not affected by the external environment when collecting data and can be nicely applied to scene control in VR. MEMS accelerometers provide acceleration information of object motion, which is widely used in business and scientific research.



3. Myoelectricity (EMG) sensor

In recent years, myoelectricity is considered as one of the bioelectrical signals. Gesture recognition by collecting myoelectrical information has been widely studied, mainly through recognizing the biopotential changes of the myoelectric signal generated during the movement of the muscle tissue of the hand, to identify the current gesture of the user. This method is not easily affected by the external environment and is especially suitable for people with physical disabilities. In order to obtain accurate bioelectrical information, it is necessary to place at least two or three electrodes on the skin using conductive gels, which may stimulate allergic reactions in people with sensitive skin, and the skin condition, sex, age, and other factors changes affect on the signal generation, therefore, it is difficult to establish a unified measurement standard[21]. Currently, the identification device for myoelectricity detection of gestures is a wristband. In 2013, Thalmic Labs introduced the Myo smart wristband, which captures the bioelectrical change information of the user's arm muscle during exercise through a sensor in the wristband. It can identify 20 different interactive gestures after analyses to judge the intention of the wearer and can create visual control in the virtual environment to implement human-computer interaction. This method has better recognition even with slight movements and is a research focus in the future.

TOUCH DEVICE

1. Touch screen

The touch screen is composed of a touch detection component and a controller. The device detects the touch position of the user through the detection sensor, and then the controller converts the received information into contact coordinates and transmits them to the CPU. The touch screen provides an efficient connection between the human and the computer. Touch screens include capacitive touch screen, resistive touch screen, infrared technology touch screen, surface acoustic wave touch screen, and pressure touch screen[22].

2. Stylus pen

Stylus pen interaction is another common technique for operating a touch surface. The pen-type touch device mainly includes a touchpad and a stylus pen, and uses the stylus pen to interact on the touch surface of simulated paper to convert the collected data into horizontal coordinates. Currently, the pressure of the pen tip and the 3D posture information of the pen can also be collected by adding a sensor to the pen. The principle of stylus pen technology is mainly achieved by electric tracking technology, magnetic tracking technology, ultrasonic tracking technology, and ray tracing technology[23].

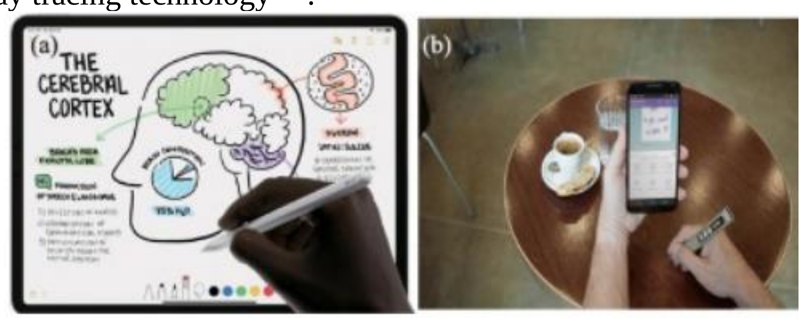


Figure 4. (a) Apple Pencil[6], (b) Phree[7].

GRASP REPRESENTATION

The transfer of object grasping performed in VR to virtual humans requires robustness against inaccurate sensor data from VR input devices, also due to missing tactile feedback with conventional data-gloves. To compensate for the vagueness of the input data, a knowledge-based approach involving an empirically founded grasp taxonomy as well as a collision-sensor enriched hand model have been developed.

Grasp Taxonomy

To ensure independence of hand and object geometry, a high level representation of the grasp is required. Grasp taxonomies which categorize different grasp types provide such high level representations. Several categorizations of grasps have been proposed in the literature. This began with research in the medical field where grasp sequences of humans have been studied by Schlesinger [Sch19]. His classification is based on the shape of the object to grasp and includes six different grasp types: “cylindrical grasp”, “tip grasp”, “hook grasp”, “palmar grasp”, “spherical grasp” and “lateral grasp”. Grasps can also be categorized by the stability of the grasp, as, e.g., in the work of Napier [Nap56] and Mishra and Silver [MS89]. This line of research differentiates between two basic grasp types, i.e., the power grip which holds an object firmly and the precision grip where the thumb andother fingers hold the object. Cutkosky developed a taxonomy on the basis of research about the work of mechanics to achieve optimal grasp operations in factories,distinguish static and dynamic grasps; in static grips the fingers remain unchanged, while in dynamic grips the finger positions vary to keep the grasped object stable.