

10 Technological Background About VR

Mariano ALCANIZ, Ph.D.¹,
José Antonio LOZANO, M.S.¹, Beatriz REY, M.S.¹

¹ *MedICLab, Medical Image Computing Laboratory,
Universidad Politécnica de Valencia, Valencia, Spain*

Abstract. User interface (UI) design is a critical component of any virtual environment (VE) application, and especially for VE applied to medicine. User interfaces for VE are becoming more diverse. Mice, keyboards, windows, menus, and icons - the standard parts of traditional WIMP interfaces- are still prevalent, but nontraditional devices and interface components are proliferating rapidly. These include spatial input devices such as trackers, 3-D pointing devices, and whole-hand devices allowing gestural input. Three-dimensional, multisensory output technologies -such as stereoscopic projection displays, head-mounted displays (HMDs), spatial audio systems, and haptic devices- are also becoming more common. In this chapter we present a brief overview of 3-D interaction and user interfaces technologies for VE.

1. Introduction

User interfaces (UIs) for computer applications are becoming more diverse. Mice, keyboards, windows, menus, and icons - the standard parts of traditional WIMP interfaces- are still prevalent, but nontraditional devices and interface components are proliferating rapidly. These include spatial input devices such as trackers, 3-D pointing devices, and whole-hand devices allowing gestural input. Three-dimensional, multisensory output technologies -such as stereoscopic projection displays, head-mounted displays (HMDs), spatial audio systems, and haptic devices- are also becoming more common.

With this new technology, new problems have also been revealed. People often find it inherently difficult to understand 3-D spaces and to perform actions in free space.

Although we live and act in a 3-D world, the physical world contains many more cues for understanding and constraints and affordances for action that cannot currently be represented accurately in a computer simulation. Therefore, great care must go into the design of user interfaces and interaction techniques for 3-D applications, especially for Medicine. It is clear that simply adapting traditional WIMP interaction styles to three dimensions does not provide a complete solution to this problem. Rather, novel 3-D user interfaces, based on real-world interaction or some other metaphor, must be developed.

This chapter is a brief overview of the current state of the art in 3-D user interfaces and interaction available today.

2. The Virtual Experience: INPUT Interface

In a Virtual Reality (VR) System is very important that the user can “live” the Virtual Reality Experience of the most realistic possible way. But, how the user can to *participate* (*communicate*) with a virtual experience? What *technological methods* can to use this user in order to manipulate and travel through the virtual experience?

Before continuing is important to make this explanation. Basically there are two ways of input to a VR experience: *User Monitoring* (the real-time monitoring of the user’s actions) and *World Monitoring* (the real-time monitoring of the real world). Now, we will centre us in the first way because, possibly, it is the most interesting since it affect directly to the user and how he/she can to make anything in this VR experience.

The *User Monitoring* provide to the user a way in order to tell the virtual world what he/she want and track at least some part of their body. We will make a brief introduction to several ways of monitoring a user’s interaction with a virtual world, that is, *Tracker Systems* and other *Physical Input Devices* (*Physical Controls*).

2.1 Tracker Systems

In a VR experience it is very important to know (to track) continuously the user’s movements in the virtual world because this is what allows the VR system to render and display the virtual world from a user-centric perspective and, in this way, providing the effect of physical immersion.

This is possible thanks to the *Position Sensors*, that is, devices that provide its location and/or orientation to the computer. These devices are used in order to communicate the VR system where is located the user within a virtual reality world; the user head, the user hand and even both things.

There are several types of *Position Sensors*:

- *Electromagnetic*
- *Mechanical*
- *Optical*
- *Ultrasonic*
- *Inertial*

- *Electromagnetic:*

This type of position sensors uses a *Transmitter* (it is fixed at a know location and orientation) in order to generate a low-level magnetic field from three orthogonal coils within the unit. Each one these fields will generate current in another set of coils that it will be in a smaller *Receiver* unit (it is fixed on the user head, hand or in any other device manipulated by him). Thanks to the measures of the signal in each coil of receiver is possible to know its position and orientation relative to transmitter, that is, where the receiver is and how is oriented.

Limitations:

- Interferences because of metals in the environment
- Short range (3-8 feet) of the magnetic field generated

Advantages:

- No line of sight restriction (no problems with visual or sonic obstacles between transmitter and receiver)

- *Mechanical:*

In this type of position sensors the track may be carried out through mechanical systems. For example, the BOOM systems, mechanical articulated armlike with visual device in one of their ends. In these systems, the user can strap this visual device in their head or put their face in it and grasp it with some handles. The BOOM will follow the head user movements within a limited range and, thanks to each elbow joint and connecting link, it will can to measure and calculate the head user position and orientation. Moreover, it is possible to incorporate motors to the BOOM linkages. In this way, it will possible to simulate force (pressure and resistance) effects.

Limitations:

- User location restrictions

Advantages:

- Very fast and accurate system

- *Optical:*

This type of position sensors uses visual information in order to track the user or the objects. It is possible tanks to a video camera (it is in a fixed location) that acts as an electronic eye that it is “watching” the tracked user or object. Normally, this user or object will have placed a sensor device (light-sensing devices) that it will be watched. Using complex computer vision techniques and being based on what the camera sees, it will be possible to calculate the user or object position.

If the user or object have a single sensor device, their position will can be reported in only two dimensions but without depth information. Nevertheless, this problem can to solve if the user or object have multiple sensors. In this way, the system will can triangulate the location and/or orientation of the tracked entity, providing three-dimensional position information. Moreover, if the system use three visual input devices (three video cameras) in different locations, it will be possible to calculate a full 6 DOF (Degrees of Freedom –a particular way in which a entity may move in space) position.

Limitations:

- Line of sight restriction (problems if obstacles between the tracked person or object and the camera)
- Therefore, limitation in the participant’s range of movement

- *Ultrasonic:*

This type of position sensors uses high-pitch sounds emitted at timed intervals in order to calculate the distance between the transmitter (speaker) and the receiver (microphone) and, in this way, to track the user or the objects. In order to obtain the full 6 DOF position of a user or object is enough the use of three transmitters combined with three receivers.

Limitations:

- Interferences when operated in a noisy environment
- Short range of action
- Encumbrance (wires attached to the transmitter and receiver)

Advantages:

- Fairly inexpensive way of position tracking

- *Inertial*:

This type of position sensors uses electromechanical instruments (accelerometers, inclinometers) in order to detect the relative motion of sensors by measuring change in gyroscopic forces, acceleration and inclination [1].

An inertial device is a small sensor that is attached to the object being and connected to the computer via connecting wire (often there is an intermediate device or “black box” that it will convert the signals to appropriate levels and digital communication protocols) or with wireless transmission technology (radio). An inertial device has the same technique of operation by which the inner ear aids in knowing the head orientation, that is, a fluid tends to remain motionless while the surrounding structure rotates. In this way, inertial sensors relay information about the location of the structure relative to fluid.

Limitations:

- Not provide location information
- Limited to orientation-only measurement

Advantages:

- No range limitation
- User freedom movements
- Quality inexpensive
- Easy incorporation and use with HMDs
- Can be combined with other tracking systems in order to provide the best method

Any of these technologies are better than the other. All these types of *Position Sensors* have their advantages and disadvantages, their benefits and limitations (normally, these limitations arise from technology used to determine the relationship from some fixed origin and the sensor) . Besides cost, these are the questions to keep in mind:

- *Accuracy / Precision* and *Speed* of the reported sensor position
- *Interfering Media*, that is, metals, opaque objects, etc.
- *Encumbrance*, that is, wires, mechanical linkages, etc.

Therefore, when a designer and developer of VR systems are planning a VR experience, will must keep in mind all this and choose the most appropriate system. A possibility is to consider the ability of the system chosen in order to produce an acceptable experience, that is, an experience without noise, low accuracy and lag time, factor that decrease the realism and immersion of the virtual experience and that they can lead to nausea in the users.

2.2 *Physical Controls*

Physical Controls are another part of the interface between the user and the virtual world, that is, individual *buttons* (two positions: depressed or released), *switches* (two or more positions), and *valuators* (a range of continuous values - sliders and dials) that allow users to actively provide input directly into the virtual world. Several of these devices can be combined into a single input device that it will can be designed as generic devices (usable in multiple applications) or as specific device (usable in specific use). Also, they can be mounted on a handheld *prop* tracked by the VR system, on a *platform* used in a virtual reality experience, etc.

Now, we will make a brief introduction to the two new terms mentioned previously:

- *Props*
- *Platforms*

- *Props:*

These devices are physically objects used as an interface to a virtual world.

Normally, these objects are built for a specific purpose in an application and represent some object in a virtual world, that is, they may be embodied by a virtual object. These devices might have physical controllers (buttons, switches and/or valuator) mounted on it and, also, some tracker system. The props have physical properties (shape, weight, texture, centre of gravity and solidity) that provide some haptic information to the user and they suggest him its use in a virtual reality experience. There are several types. The most common are wands and 3D mouse (they can move in two directions and also they can report height information).

When a user manipulates a real object that it is simulated in the virtual world, he/she think that the virtual experience is more real and, also, the rest of the virtual world seem more real too. These feature, well-known like *transference of object permanence*, and the other features mentioned previously, convert this input interface in a natural and intuitive devices in order to interact with the virtual experience.

- *Platforms:*

These devices are physical structures in which the user is situated. Normally, they simulate a real world device found in the virtual world, but sometimes they are simply a place to sit or stand. Also, it is possible the user interact with the virtual world manipulating real world objects included in this platform. Moreover it is possible to incorporate motion systems (hydraulic or electric systems) in the base of these platforms (*motion platforms*) providing, in this way, a very natural interface with the virtual experience. There are several types. The most common are *ring*, *kiosk*, *ambulatory* and *vehicle* platforms. In any of them are limited the visual device that could use. There are examples in both HMDs and screen/projection devices.

3. The Virtual Experience: OUTPUT Interface

In a Virtual Reality (VR) System is very important that the user can “live” the Virtual Reality Experience of the most realistic possible way. But, how the user can to *perceive* a virtual experience? What *technological methods* can to use a virtual reality system in order to present information to any of the users’ senses?

Before continuing is important to make this explanation. The human perceptual system has five senses providing information to the brain but, only, we will study three of these senses (visual, aural and haptic) because the other two senses is not sufficiently developed from a technological point of view.

Now, we will make a brief introduction to several visual, aural and haptic VR output devices. We will discuss their properties, different modalities for each of them, their components and features.

It is very important keep in mind that, in order to select one of the visual, aural or haptic options there are not a rule that say us which is the best choice. Depending on parameters like available resources, venue constraints and/or requirements, number of users of the virtual experience, etc, the designer of the VR system must make this choice.

3.1 Visual Devices

In the real world, when we look around, we perceive information regarding the relative distance of objects. In this process, this information is given thanks to several indicators of distance, also known like *depth-cues* (*monoscopic, stereoscopic, motion and physiological*). Not all these *depth cues* have the same importance but, undoubtedly, the *stereoscopic* is the most eminent. The stereoscopic image depends on *parallax*, that is, the apparent displacement of objects viewed from different locations.

Nevertheless, besides these *depth-cues*, they are also important other properties of visual devices that it is important keep in mind when we want to choose the appropriate visual device. These properties are the *Visual Presentation Properties* (color, spatial resolution, contrast, brightness, number of display channels, focal distance, opacity, masking, FOV –Field of View-, FOR –Field of Regard-, Graphics Latency Tolerance and Frame Rate), that they are a very important factor in the overall quality of the virtual reality experience, and the *Logistic Properties* (Portability, Encumbrance, Safety, Cost, etc.).

At present, technically is possible to produce a virtual visual sensation similar to the reality, also with the stereoscopic effect, that it will help to increase the realism and immersion of the user in the virtual experience. But, what devices will can to use for this? Basically, there are two categories of visual devices:

- *Head Based Visual Devices:*

Occlusive HMDs

Nonocclusive HMDs

- *Stationary Visual Devices:*

Monitor Based VR

Projection VR

- *Head Based Visual Devices (Occlusive and Nonocclusive HMDs):*

The *Head Based Visual Devices*, also know like *Head Mounted Device* (HMD), is the device that most people associate with Virtual Reality. All people, when they think in “virtual reality”, they think in a HMD.

These devices have some screens that they move with the movements of the user head. These screens are small and lightweight since they are worn or held by the user. Most of the HMD allow stereoscopic image using, normally, a dual visual output (one for each eye) system.

Although these devices display the virtual world through the user’s viewpoint, in these devices it is possible to add some tracking method and, also, some could have it incorporate in the unit. Tracking the location and orientation of the user head, the HMD can to be the most intuitive visual interface because, if a user wants to see another side of an object he will have simply walking to this new side and looking the object.

Occlusive HMDs:

These devices isolate completely the user from the real world. Anything the user needs to see will must to be generated by the virtual world, including this own body if this was necessary.

Nonexclusive HMDs:

These devices not isolate completely the user from the real world. They are used in order to obtain an augmented copy of the physical world, mixing, with the view of this real world, the view of a virtual world. That is, mapping virtual information onto the real world view. It is know like Augmented Reality (AR), a very important subclass of virtual reality.

In order to mix a virtual and a real image there are two methods: optics or video.

The first method uses lenses, mirrors and half-silvered mirrors. In the second method, where the images of real world are generated by cameras mounted on the HMD, is used electronic systems in order to make this mix [2].

The main advantages of these devices are the following:

- Lower cost for lower resolution models
- Can occlude the real world (useful in some virtual reality experiences)
- Can be used for AR
- Complete FOR
- Greater portability

The main disadvantages of these devices are the following:

- Limited resolution
- Limited FOV
- Encumbrance

- *Stationary Visual Devices (Monitor Based and Projection VR):*

These devices are known as “stationary” because they are fixed in place and unlike previous systems, they can not be moved with the user head movements in the virtual experience. Basically there are categories:

- *Monitor Based VR*
- *Projection VR*

Monitor Based VR:

Normally, this visual device is an extension of a simple desktop computer (standard computer monitor) setup but it requires other additional components in order to be considered a monitor VR. Basically, a monitor VR should provide the user to move their head (side to side and up and down) in order to see around, over and under objects and to see the virtual world in a stereoscopic way.

In order to track the user’s head and, in this way, to render scene changes in response to the tracked head users movements, it is possible to use a monitor-top video camera that it will be useful in order to capture the images of the user head, to process these images and to calculate the location of their head. This system is an inexpensive system (sometimes it might already be part of the computer system), but it is possible to use other tracking technology too.

In order to obtain stereoscopic vision in the monitor VR is used a time interlacing system that it provides left and right view. The user will need to use a shutter glasses with liquid crystal display (LCD) lenses. Other possibility is to use some special filter over the monitor screen. This option avoids the use of special glasses.

The main advantages of these devices are the following:

- Inexpensive
- Easy to use
- Higher visual resolution

The main disadvantages of these devices are the following:

- Less immersive (than most other VR visual devices)

- Limited FOR

Projection VR:

This system uses large screens in order to visualize the virtual experience. This large visual “surface” can be created by setting several CRT monitors side by side but, normally, is used wide screens on which it is projected images by means of projectors like in the cinema. In this last option, the users are placed between the projector and the wide screen which can produce shadows of the users on the screen. An alternative to this is the use of rear-projection screens. In this option, the users are not between the projector and the wide screen avoids the annoying shadows. Now, most projection VR systems are rear-projected.

In order to track the user’s head and, in this way, to render scene changes in response to the tracked head users movements, it is possible to use some of the tracking systems well-known. Nevertheless, it important to keep in mind this question: if the screen of projection is not vertical (normal in some specific projection systems) and it is necessary to provide to the user a stereoscopic vision, the system will need a greater range of tracking, plus head-orientation data.

In order to obtain stereoscopic vision in the projection VR system is possible to use two methods: shutter or polarized glasses. The first option is easier to implement because most graphical displays already include the necessary hardware. The second option can need a projector (specialized) or two projectors (regulars) equipped with filters to polarize the two images (left and right) for the appropriate eye. In this case, the user will need a pair of glasses also equipped with polarized filters.

The main advantages of these devices are the following:

- Wider FOV
- Greater user mobility (fewer cables)
- Longer periods user permanence
- Lower safety risk
- Better for group viewing

The main disadvantages of these devices are the following:

- Necessary more maintenance and equipment

Therefore, because the visual sense is the first sense used in order to obtain the immersion of the user in a virtual reality experience, it is very important to choose the appropriate visual devices if we want to obtain the goal of our virtual environment.

3.2 *Aural Devices*

In the real world we hear sounds on all sides. Thanks to the *localization* phenomenon, the brain of the listener can determine the direction and distance from which a sound emanates and, therefore, the listener can to understand the nature and features of the environment he is in.

Normally, the virtual worlds represent 3D environments like the real world. It is possible to create in a virtual world a virtual illusion similar to the *localization* phenomenon of the real world? The term *spatialization* describes this possibility. The *spatialization* is the act of creating the illusion that a sound is emanating from a specific 3D location.

This virtual illusion is possible thanks to the *Transfer Functions*, mathematical transformations that can be applied to a signal to alter it in some specific desired way. In this way, if a virtual reality experience developer wants to make that a sound is coming from a specific place in a virtual world, he will can to use some specific transfer function (filter) in order to obtain that the user perceives that the sound came from this specific place.

At present, technically is possible to produce this real sensation that it will help to increase the realism and immersion of the user in the virtual experience. But, what devices will can to use for this? Basically, there are two categories of aural devices:

- *Head Based Aural Devices:*
Headphones

- *Stationary Aural Devices:*
Speakers

- *Head Based Aural Devices (Headphones):*

Headphones are for one person only and move with the user head. They can to isolate in the user the sounds of the virtual world from sounds of the real world (Close-ear Headphones) or to overlap the virtual and real sounds (Open-ear Headphones).

Generally, these devices are dual-channel displays located near each ear. Therefore is much easier to carry out the presentation of stereophonic and 3D spatialized sounds that using speakers. In order to obtain that the user of a 3D virtual world has the sensation that the sound in this virtual world appear to come from a particular location, it is very important to track this head position so that the spatialization information reflects the changing location of the listener's ears.

The main advantages of these devices are the following:

- Works well with Head Based Visual Displays
- Easier to implement spatialized 3D sounds
- Isolate sound of the real world
- Greater portability
- Private

- *Stationary Aural Devices (Speakers):*

With the speakers, the sounds can to be listened for several persons at the same time. To create spatialized sounds using this technology can be more difficult than with headphones.

Ambisronics, the term that define the presentation of 3D sound using several speakers, it is a line of investigation that will may result in a usable system in the future [3].

Nevertheless, because both ears can hear the sounds from each of the speakers, to obtain this it don't seem an easy task.

The main advantages of these devices are the following:

- Works well with Stationary Visual Displays
- Greater user mobility

Normally, when a designer and developer of VR systems are planning a VR experience, they give much importance to the visual aspect of this experience: the realism of the virtual environments (3D models and textures) and the visual device that it will used.

Nevertheless, sometimes, the addition of high-quality sound in a virtual world can help in creating a compelling experience, a more realistic and immersive virtual experience, even when the quality of the visual presentation is not enough or the visual device selected is not the more appropriate. Both high-fidelity audio devices, headphones and speakers, are not expensive and, moreover, their incorporation in a VR system is very easy.

Therefore, it is necessary to give to the audio devices as importance as the visual devices because, both are essentials in order to add realism and immersion to a virtual experience.

3.3 Haptic Devices

In the real world when we touch something (object) we comes to believing this object is real and, therefore, this object exist. The existence of an object is verified by touch sense, the “haptic perception”. It is very important to clarify that this perception involves two combined sensations: *kinaesthesia* and *taction*.

Kinaesthesia is the perception of movement or effort from within muscles, tendons and joints of the body. Normally, the term *force feedback* is used as a synonym of *kinaesthesia*.

Taction is the sense of touch that comes from the nerve sensors at the surface of the skin, that is, stimuli for temperature as well as pressure. In the human body, the reception of the *kinaesthesia* and *taction* information is produced at the same time. The separation of both processes is impossible.

Maybe, in a next future, it will be possible to combine both senses in an only device.

Nevertheless, at present it already is possible already obtain some of these sensations by means of haptic devices and, sometimes, their use in virtual worlds is extremely effective.

Therefore, technically is possible to produce a virtual haptic sensation that it will help to increase the realism and immersion of the user in the virtual experience. But, what devices will can to use for this? Basically, there are two categories of haptic devices:

- *Tactile Devices*

- *End-Effector Devices*

- *Tactile Devices:*

These devices use the ability of the skin in order to interpret stimuli as pressure, temperature, electricity and pain. They provide to the user information in response to touching, grasping, feeling surface textures or sensing the temperature of an object.

In order to present these stimuli to the user, several *actuators* are affixed to a wand that it will be grasped by the user hand because the tactile nerve sensors are located in the fingertips. These actuators will can to be *bladder actuators* (pockets that can be expanded

and contracted by controlling the flow of air or liquid into and out of them) or *vibrator actuators* (more robust and easier to control than previous). Both are used in order to translate the sensation of pressure.

These devices are not very advanced because the market for them is small and, therefore, less research is done on this field.

The main advantages of these devices are the following:

- Facilitates the fine manipulation of virtual objects
- Can be combined with end-effector displays in specific applications
- Often less expensive than other haptic devices
- Generally portable

- *End-Effector Devices:*

Normally, these devices act as input and output devices. Basically, they are mechanical devices that allow the user to perform a force by means of one of their extremities (hands and/or feet) and, moreover, to provide him a resistance to this force.

An end-effector device needs two components: a way of sensing the movements of the user and a way of providing resistance at the point of contact with him. In order to obtain this it is necessary using electric motors and hydraulics and pneumatic pressure systems.

Moreover, if we want to obtain a fast and accurate response it is possible to incorporate some mechanical user-tracking system in the mechanical devices in which are mounted these devices.

These are some examples of end-effector devices: *Argonne Remote Manipulator (ARM)* (multijointed hand grips), *PHANTOM* (desktop point controls), *Sarcos Unipoint System* (resistive push and rotate pedals) and *Rutgers Master I and II* (hand-motion limiters).

The main advantages of these devices are the following:

- Can be world- or body-grounded
- Fast and accurate tracking usually built into device

Therefore, at present it is possible to offer to user tactile or force stimuli incorporating haptic displays in the virtual reality system and, in this way, to increase the realism and immersion that supposes the inclusion of virtual objects in a virtual experience.

Nevertheless it is very important to say that it will be necessary to choose if we want to offer tactile or force stimuli because, most commercially available haptic displays provide either tactile or force stimuli but not both.

4. Input Devices .vs. Interaction

A distinction must be made between input devices and interaction techniques. Input devices are just the physical tools used to implement various interaction techniques. In general, many different interaction techniques can be mapped onto a given input device. The question is how naturally, efficiently, and appropriately a given input device will work with a given technique. Input devices are also governed by the degrees of freedom (DOFs) they

have. In general, an input device with a smaller number of DOFs can be used to emulate a device with more DOFs with the addition of buttons or modifier keys.

Input devices can be roughly categorized based on the types of events they generate.

Devices that generate one event at a time based on the user are considered discrete-input devices. A button press or other action signals an event, which is usually a Boolean (up/down) value. Pinch gloves, developed by Fakespace, are an example of a discrete-input device (the user pinches two or more fingers to signal an event). In contrast to discrete-input devices, continuous-input devices generate a stream of events. Two of the most common continuous-input devices are position/orientation trackers and datagloves, which transmit bend angles of the fingers. Devices that combine both discrete and continuous events to form single, more-flexible devices are called combination or hybrid-input devices.

Examples of hybrid devices include the Ring Mouse (a small device worn on the user's finger that combines ultrasonic tracking with two buttons) and pen-based tablets, which are becoming more popular in VE applications because they give users the ability to interact in two dimensions. Speech input is unique because the "device" is the human voice.

Speech provides a nice complement to other input devices and, as a result, it is a natural way to combine different modes of input (multimodal interaction). In general, speech input can be a valuable tool in 3-D user interfaces, especially when both the user's hands are occupied. A major myth of speech recognition is that having a good speech recognizer will solve all problems. Unfortunately, many other issues need to be considered when dealing with speech input. Letting the computer know when to listen to the user's voice is one such issue. A user may issue a speech command unintentionally when talking to another person. One of the best ways to avoid this problem is to use an implicit or invisible push-to-talk scheme. A push-to-talk interface lets users tell the application when they are speaking to it. In order to maintain the naturalism of the speech interface, one can embed the "push" into existing interaction techniques so the user does not have the burden of remembering the signal to the application that a voice command is about to be issued.

5. Interaction techniques

5.1 Navigation

The task of navigation is the most prevalent user action in most large-scale 3-D environments, and it presents challenges such as supporting spatial awareness, providing efficient and comfortable movement between distant locations, and making navigation lightweight so that users can focus on more-important tasks. We subdivide navigation into the motor component called *travel* and the cognitive component called *wayfinding*.

Navigation tasks can generally be classified into three categories. *Exploration* is navigation with no explicit target: the user is simply investigating the environment. *Search* tasks involve moving to a particular target location. Finally, *maneuvering* tasks are characterized by short-range, high-precision movements that are used to place the viewpoint at a more advantageous location for performing a particular task.

Travel is a conceptually simple task—the movement of the viewpoint from one location to another. Further, viewpoint orientation is usually handled in immersive VEs by head tracking, so only techniques for setting viewpoint position need be considered.

There are five common metaphors for travel interaction techniques, and most published interaction techniques for travel fit into one of these five categories:

- **Physical movement:** The user's body motion is used to travel through the environment. Examples include wide-area motion tracking, walking in place, and

locomotion devices such as treadmills or stationary bicycles, Such techniques are appropriate when an enhanced sense of presence is required or when the application requires the user to experience physical exertion when traveling.

- **Manual viewpoint manipulation:** The user's hand motions are used to effect travel. For example, the user "grabs the air" and pulls himself along as if with a virtual rope. Another type of technique uses a selected object as a center point around which user motion can be specified. These techniques can be efficient and easy to learn, but they can also cause fatigue.
- **Steering:** Steering is the continuous specification of the direction of motion. This is the most common travel metaphor and includes techniques such as gaze-directed steering (wherein the user's head orientation determines the direction of travel) or pointing (in which hand orientation is used). Steering techniques are general and efficient.
- **Target-based travel:** The user specifies the destination, and the system handles the actual movement. This may take the form of "teleportation," in which the user jumps immediately to the new location, or, preferably, the system may perform some transitional movement between the starting point and the destination. Target-based techniques are very simple from the user's point of view.
- **Route planning:** The user specifies the path that should be taken through the environment, and the system handles the actual movement. The user may manipulate icons or draw a path on a map of the space or in the actual environment in order to plan a route. These techniques allow the user to control travel while he retains the ability to perform other tasks during motion.

Besides choosing a metaphor, other design issues for travel techniques include velocity control and the use of constraints or guides to aid travel.

5.2 Selection and manipulation

Interaction techniques for 3-D manipulation in VEs should provide means to accomplish at least one of three basic tasks: object selection, object positioning, and object rotation.

Because direct hand manipulation is a major interaction modality not only in the 3-D virtual world but also in natural physical environments, the design of interaction techniques for object selection and manipulation has a profound effect on the quality of the entire VE user interface.

The classical approach to design manipulation techniques is to provide the user with a "virtual" hand – a 3-D cursor, often shaped like a human hand, whose movements correspond to the movements of the hand tracker. Selection and manipulation simply involve touching an object, then positioning and orienting this virtual hand within the VE.

The virtual hand technique is rather intuitive because it simulates a real-world interaction with objects, but only those objects within the area of reach can be picked up. A number of techniques have been suggested to overcome this problem. The Go-Go technique [4] allows the extension of the user's reach by using a nonlinear mapping applied to the user's hand extension. When the user extends the hand farther than a threshold distance D , the mapping becomes nonlinear and the virtual arm "grows". Different mapping functions can be used to achieve different control-display gain between real and virtual hands [5].

The other common way to select and manipulate objects in VEs is to point at them using a virtual ray emanating from the virtual hand. When the virtual ray intersects an object, it can be picked up and manipulated. Several variations of ray casting have been developed to help users in selecting very small or faraway objects. For example, the

spotlight technique [6] provides a conic selection volume, so that objects falling within the cone can be easily selected. However, when more than one object falls into the spotlight, further disambiguation of the target object is required. The aperture technique [7] uses a conic pointer whose direction is defined by the location of the user's eye (estimated from the head location) and a hand sensor. The user can control the size of the selection volume simply by bringing the hand sensor closer or moving it farther away. The image plane family of interaction techniques [8] develops and extends this idea.

All the techniques described above provide users with tools that allow them to select or reach further in the immersive virtual world. An alternative approach would be to allow the user to manipulate the relative scale of the virtual world. One of the earliest uses of this approach was in the 3DM immersive modeler [9], in which users could "grow" or "shrink" themselves to manipulate objects of different sizes. The World-in-Miniature (WIM) technique [10] provides a handheld model of the VE. The user can then indirectly manipulate virtual objects by interacting with their representations in the WIM.

Because all manipulation techniques have particular strengths and weaknesses, a number of attempts have been made to integrate and combine their best features. For example, the Virtual Tricorder [11] combines ray casting for selection and manipulation with techniques for navigation and level-of detail control within one universal tool.

5.3 System Control

System control refers to a task in which a command is applied to change either the state of the system or the mode of interaction. The issuing of a command always includes the selection of an element from a set. Thus, some similarities can be seen between system control and object selection techniques. In desktop applications, the use of commands has received much attention. Unfortunately, interaction styles used in desktop environments, like pulldown menus and command-line input, are not always usable within a VE. One of the basic problems of VE system control is that a normally one- or two-dimensional task becomes three-dimensional, which reduces the effectiveness of traditional techniques. For example, touching a menu item floating in space is much more difficult than selecting a menu item on the desktop, not only because the task has become 3-D, but also because the important constraint of the physical desk on which the mouse rests is missing. Evaluation results for system-control techniques are relatively sparse. Although many ad hoc implementations have been reported, system control has not been studied in a structured way.

We can categorize system-control techniques for immersive VEs into four groups, namely graphical menus (visual representations of commands), voice commands (menus accessed via voice), gestural interaction (command sets accessed via gesture), and tools (virtual objects with an implicit function or mode). Also, hybrid techniques exist that combine several of the types.

System control is often integrated within another universal interaction task. Due to this integration, one should avoid disturbing the flow of action of the main interaction task.

The user should stay focused on the task. "Modeless" interaction (in which the mode changes are very natural) is ideal. One way of supporting easy access to a system -control interface is to use a natural spatial reference, such as a fixed position relative to the user's head or body, for placement of the interface. This guideline is mostly applicable to graphical menus, but tools also benefit from a strong spatial reference. Another method to allow a more seamless integration of system control into a flow of action is to use a multimodal system control interface.

After accessing a system-control interface, one has to select a command. When the set of functions is large, one needs to structure the items. This might be achieved by

methods like context-sensitive menus, or by clearly communicating the hierarchy of items and (sub) menus.

Finally, the designer should try to prevent mode errors by providing the user with appropriate feedback during and after selection of a command. Mode errors can be highly disruptive to the flow of action in an application.

6. Two-Dimensional Interaction in Three-Dimensional Environments

A common misconception of 3-D user interface design is that, because the applications usually contain 3-D worlds in which users can create, select, and manipulate 3-D objects, the interaction design space should utilize only 3-D interaction. In reality, 2-D interaction offers a number of distinct advantages over 3-D interaction techniques for certain tasks. If haptic or tactile displays are not present, 2-D interaction on a physical surface provides a sense of feedback that is especially useful for creating objects, writing, and annotating.

Most efficient selection techniques are essentially 2-D, although further manipulation may require a 3-D interaction technique. By taking advantage of the benefits of both 2-D and 3-D interaction techniques, it is possible to create interfaces for 3-D applications that are easier to use and more intuitive for the user. Seamless integration of 2-D and 3-D interaction techniques is a critical design consideration from both a physical and logical perspective. Physical integration is important because we do not want to make it difficult for users to switch between 2-D and 3-D devices. Logical integration is also important because we want the devices in the application to know whether they are used for 2-D or 3-D interaction. This contextual application-based information helps to reduce the user's cognitive load.

The 2-D /3-D interfaces can be roughly classified into three categories. Note that, in all categories, some type of physical surface is required for 2-D input. The distinguishing characteristic of these interfaces is how the physical 2-D surfaces are utilized. The first category covers applications that use fully immersive displays such as HMDs, where the user cannot physically see the 2-D surface. Here, the 2-D surface is usually a piece of tracked plastic or pen-based tablet, and users must have a graphical representation of the surface in order to interact with it in the virtual world. Examples of this type are the Virtual Notepad, a system for writing and annotating in VR [12] and the Virtual Habitat [13].

The second category of 2-D/3-D interfaces cover applications that use semi-immersive displays such as workbenches. The physical 2-D interaction surface is usually either on top of the workbench display so users can directly interact with the display surface, or on a tracked, transparent tablet that users can hold in their hand. In the latter case, graphics are projected on the primary display but virtually appear as if they are on the surface of the tablet. Examples of this category are the ErgoDesk system [14], a modeling application using the physical display surface for a 2-D interface, and the Transparent Pad, a clear tablet that users can hold to interact with a landscaping design application [15]. The third category uses separate 2-D display surfaces, such as handheld computers and pen-based LCD tablets. An example of this type of interface is the use of a Palm Pilot in a CAVE-like device for camera, environment, and geometry controls [16]. In this area, there are many open research issues to explore, especially when dealing with wireless technology and the weight of these devices.

References

- [1] Foxlin, E.; "Inertial Head-Tracker Sensor Fusion by a Complementary Separate-Bias Kalman Filter", Proceedings of the IEEE 96 Virtual Reality Annual International Symposium (VRAIS), **1996**, pp. 185-194.
- [2] Rolland; Holloway, R.; Fuchs H.; "A Comparison of Optical and Video See-Trough Head-Mounted Displays", *Proceedings of SPIE Telemanipulator and Telepresence Technologies*, vol. 2351, **1994**.
- [3] Gerzon, M.A.; General Meta Theory of Auditory Localisation., Paper presented at the 92nd Audio Engineering Society Convention, Vienna, March, **1992**, (Preprint 3306).
- [4] Poupyrev, I., Billinghamurst, M., Weghorst, S., & Ichikawa, T., The Go-Go interaction technique: Non-linear mapping for direct manipulation in VR. *Proceedings of the ACM Symposium on User Interface Software and Technology*, **1996**, 79 -80.
- [5] Bowman, D., & Hodges, L., An evaluation of techniques for grabbing and manipulating remote objects in immersive virtual environments. *Proceedings of the ACM Symposium on Interactive 3D Graphics*, **1997**, 35-38.
- [6] Liang, J., & Green, M., JDCAD: A highly interactive 3D modeling system. *Computer & Graphics*, 4(18), **1994**, 499-506.
- [7] Forsberg, A., Herndon, K., & Zeleznik, R., Aperture based selection for immersive virtual environments. *Proceedings of the ACM Symposium on User Interface Software and Technology*, **1996**, 95-96.
- [8] Pierce, J., Forsberg, A., Conway, M., Hong, S., Zeleznik, R., & Mine, M., Image plane interaction techniques in 3D immersive environments. *Proceedings of the ACM Symposium on Interactive 3D Graphics*, **1997**, 39-44.
- [9] Butterworth, J., Davidson, A., Hench, S., & Olano, M., 3DM: A three dimensional modeler using a headmounted display. *Proceedings of the ACM*, **1992**.
- [10] Stoakley, R., Conway, M., & Pausch, R., Virtual reality on a WIM: Interactive worlds in miniature. *Proceedings of CHI: Human Factors in Computing Systems*, **1995**, 265-272.
- [11] Wloka, M., & Greenfield, E., The virtual tricorder: A unified interface for virtual reality. *Proceedings of the ACM Symposium on User Interface Software and Technology*, **1995**, 39-40.
- [12] Poupyrev, I., Tomokazu, N., & Weghorst, S., Virtual notepad: Handwriting in immersive VR. *Proceedings of the Virtual Reality Annual International Symposium*, **1998**, 126-132.
- [13] Bowman, D., Wineman, J., Hodges, L., & Allison, D., Designing animal habitats within an immersive VE. *IEEE Computer Graphics & Applications*, 18(5), **1998**, 9-13.
- [14] Forsberg, A., LaViola, J., & Zeleznik, R., ErgoDesk: framework for two and three dimensional interaction at the ActiveDesk. *Proceedings of the Second International Immersive Projection Technology Workshop.*, **1998**
- [15] Schmalsteig, D., Encarnacao, L., & Szalzvარი, Z., Using transparent props for interaction with the virtual table. *Proceedings of the ACM Symposium on Interactive 3D Graphics*, **1999**, 147-154.
- [16] Watsen, K., Darken, R., & Capps, M., A handheld computer as an interaction device to a virtual environment. *Proceedings of the third Immersive Projection Technology Workshop*, **1999**