Chapter 2

Input and Output Devices

This chapter provides an overview of a number of the display devices that a rendering system commonly may have to drive. In addition it describes some of the other interface devices often connected to VR systems - for example to track head position. These devices are important since they strongly influence the images generated by the rendering system.

2.1 Input

2.1.1 General position and orientation detection

The most common input requirement in virtual reality systems is the detection of the position and orientation of the user's head and hands. The position and orientation tracker used in the Advanced Interfaces Group, is the Polhemus 3Space Fastrak™ which uses A.C. electromagnetic tracking. It consists of a transmitter unit containing three perpendicular coils, a number of receiver units containing a similar set of coils, and a control unit. Each of the three coils in the transmitter is excited in turn, each excitation induces a signal in all three of the receiver coils. These signals are measured and provide nine values. A large amount of processing is then required to convert these raw values into a position and orientation and to filter the results before presenting them to the host computer.

The following list (an extension of a similar list in **[Bhat93]**, which describes these terms in more detail) provides an overview of the important requirements for position sensors and how well the Polhemus system meets these requirements:

- Accuracy, Resolution and Vulnerability to Noise

  Electromagnetic trackers such as the Polhemus claim to be accurate to within a tenth of an inch and have resolutions better than one hundredth of an inch **[Bhat93].** They are subject to noise caused by electromagnetic emissions from display devices and other equipment. This exhibits itself as oscillations or sudden jumps in reported position, which are extremely annoying to the user. This becomes more problematic as the user moves further from the transmitter; in practice this reduces the useful working range of the system.

- Update rate

  The Polhemus Fastrak™ control unit can perform 120 readings per second. With four sensors connected to it, the position of each sensor is measured only thirty times per second.

- Lag/Latency

  The latency of the measurement system (i.e. the time from movement of the sensor to the new position being reported) is limited by the update rate. With only thirty updates per second, the worst case latency of the system is at least one thirtieth of a second (33ms)[1]. Although the update rate is not the sole factor in determining latency, at these low update rates it is the predominate one.

- Working range

  In general the working range of most magnetic trackers is small, for example well under three feet for the Polhemus Fastrak™. Electromagnetic trackers

1. Bhatnagar quotes Polhemus as specifying a latency of 4ms; this appears to be inconsistent with the quoted update rate **[Bhat93]**.

such as the Polhemus can often operate with a larger working range at reduced accuracy.

- Comfort

  The receivers used with the Polhemus are small and lightweight and are thus reasonably comfortable to hold and carry.

- Multiple trackers

  In most situations more than one tracker is required; for example a head tracker and one or two hand trackers; the Polhemus can accommodate up to four trackers from one control box, albeit with a reduced update rate. In some situations it is desirable to have more than one tracking system; this is common in systems which allow a number of users to cooperate. The Polhemus can be configured to use one of a number of four separate carrier frequencies; this allows a number of Polhemus systems to work in close proximity to each other.

2.1.2 General position sensing: Summary

The low working volume of electromagnetic trackers (such as the Polhemus) and the noise problems, change the way in which the user works; instead of walking naturally in the model, he tends to use the position sensing to look around but move through a model by using buttons on hand controllers to indicate forward or backward motion. This is not a very natural behaviour and takes a reasonable amount of time to get used to. The Optical tracker from the University of North Carolina (UNC) **[WABG92]**, resolves many of these problems, and in particular it provides a large working volume allowing the user to walk round a room. However this system has a higher latency and requires the user to wear heavy equipment which compromises the comfort of the user.

Position sensing systems exist which individually have a large working area, low latency, high accuracy and resolution, are comfortable to wear and are immune to environmental influences; however there is as yet no one single system which has all these properties. Using the magnetic tracking systems currently available (such as the Polhemus Fastrak™) is difficult and annoying. Unless the user is very close to the transmitter unit, the display constantly ‘jitters’ about making it very difficult to examine one object in detail. Walking about a room is only possible with the use of ‘walk forward’ and ‘walk backward’ buttons and not by actually walking around the room since this would leave the user too far from the transmitter. The poor characteristics of current tracking systems make large scale projection displays more attractive, since they are often used without head tracking systems and hence do not suffer from the lack of accuracy and ‘jitter’. A survey of many different position trackers is contained in **[Bhat93]**.

2.1.3 Other input technologies

A number of different input technologies exist for use in different applications:

- Finger position sensing

  Tracking only the position of the user’s hand is somewhat limiting. The VPL DataGlove is a glove with a position sensor mounted on it and optical fibre sensors to measure the position and flexion of the fingers. The flexion measurement allows the host to detect if a user grasps an object and to provide input to automatic gesture recognition software which can be used to control the host. Gesture input is a very flexible and reasonably natural way of communicating - the alternative is often selecting items off a menu which (in the authors experience of the Division dVs system) is difficult and unnatural. A description of the theory of operation of the DataGlove is given in **[Kala93]**.

- Voice Recognition

  Another way of issuing commands while immersed in the virtual environment is voice recognition. Crude voice recognition systems have existed for many decades but the technology has just reached the point where it is practical to incorporate it in commercial systems such as the IBM Personal Dictation System **[IBM94]**. One important advantage over gesture input is that the users hands are left available for other uses.

- Eye Tracking

  The resolution of the human eye is greatest at the centre of its field of view. Some display systems attempt to increase the image quality by overlaying a small higher resolution patch over the main image at the spot the user is looking at. To be able to do this it is necessary to track the direction of view of each eye. Various techniques exist for determining the direction of view of the eye. One technique is to observe the eye using a small camera. The image from the camera is then used to find the pupil and thus determine the direction of view.

2.1.4 Connecting input devices to the host

Many peripherals are connected to the host by means of low speed serial (RS-232) links designed for terminals, since this is one of the few interfaces which can be expected to be present on almost all host computers. The Silicon Graphics Crimson in the Advanced Interfaces Group is connected to the Polhemus Fastrak via a 9600 baud RS-232 link, higher speeds were tried but proved to be unreliable. One character takes about 1ms to be transmitted along such a link. In the most compact (and inaccurate form) the Fastrak sends 15 bytes for each sensor being tracked. With two sensors being tracked (head position and one hand position) a total of 30 bytes need to be sent taking just over 31ms. This gives a direct increase in the latency due to the tracking system, which without the communication delay would be under 5ms. The Polhemus also supports a fast IEEE-488 bus capable of about 100 kilobyte/second **[Polh92]**, unfortunately very few computers now include IEEE-488 as standard.

Another issue is whether the peripheral continually sends records to the host or whether the host has to ask for each one. The communication link may not have the required bandwidth to communicate all the samples to the host - indeed the Fastrak supplies no less than 1800 bytes of data per second which is too high for the slow serial link - thus samples are lost further increasing latency. The load of processing a continuous stream of information from the position sensors may also be a problem for a host which is already busy handling many other tasks. The alternative, requesting individual readings, can increase the latency dramatically. In this scheme the host must send a command to the peripheral controller and wait for the result. This incurs the time for the sending of the command, progression through many layers of buffering inside the operating system, and the subsequent processing in the peripheral controller. As noted above the serial communication lines on most hosts were designed for terminals and communications equipment. Similarly the driver software in the operating system often does not treat the serial ports with any level of importance. Some groups have found that this can contribute to many tens of milliseconds of latency **[CoOl94]**.

2.2 Visual output

The following section discusses various types of displays which are presently used or are proposed for use in virtual reality environments. The capabilities and limitations of the particular display used, influence the design of the rendering system which must drive them, and the applications which use the rendering system. The description given here is only intended to show how the properties of the device influence its suitability to VR applications and is not intended as a full description of the device operation. A survey of available device technology giving a more detailed description of the operation of each device is contained in **[Kala93]**.

Some of the requirements for displays in VR systems are:

- A wide field of view

  The instantaneous field of view of each eye is quoted in **[Kala93]** as 120° vertically and 150° horizontally. This figure increases if the movement of the eye is allowed for. This figure is one factor in determining the maximum useful field of view for each of the two displays in a stereo head mounted display (HMD). For displays which are not head mounted this figure increases considerably due to head movement and the combination of the field of view of both eyes. Most of the small display devices available for HMDs are not capable of providing a wide field of view; there are currently two solutions to this. The first is to ignore the problem and present the user with a narrow field of view - this is the approach taken by the Virtual Research Eyegen 3 HMD. Using an HMD with a small field of view is difficult because common actions which would have been very easy (such as looking down at your feet to check that you are not about to walk into something) now involve moving your head. The other solution is to use optics to stretch the image to fill a wider viewing angle. This produces a distorted image (which is difficult to correct in the rendering and display process). In stereo systems the two images need to overlap partially to provide binocular vision, thus increasing the required field of view.

- High resolution

  In general the highest display resolution available is desirable since this reduces aliasing problems and allows the user to examine a more detailed scene. Increasing the resolution does however increase the amount of processing required by the rendering system.

  The limit of the resolution required of a head set is determined either by the application or by the limits of the human eye. **[Davs63]** states that under optimal conditions the eye can distinguish between two objects separated by about 30 seconds of arc. If the eye were capable of this resolution over the whole field of view, this would indicate that to match the resolution of the eye a horizontal resolution of the order of 18000 pixels would be required.[1]

The eye does not have a constant resolution across its entire field of view however. The high resolution area is limited to a small section near the centre of the view called the 'fovea'. In systems which can track the direction of view of the eye, it is possible to create a display consisting of a low resolution display with a small high resolution section which follows the centre of the user's vision. Even without eye tracking it would still be possible to place a high resolution insert in the centre of the display. This technique is used in the CAE FOHMD (see **[HoLa93]**).

- Lightweight

  A head set which is in use for any length of time must be comfortable; therefore display devices (and any electronics needed to control of them) must be as light as possible.

- Small size

  Similarly to maintain a reasonable level of comfort the devices must be as small as possible (without of course reducing the field of view). Some of the early head sets (such as the one used in the VCASS system **[Kala93]** p.34) were much larger than the head of the user.

- Capable of colour

  Not all display devices are capable of providing a full colour display. Colour can help group different sets of information together (or differentiate between two sets of information on the same display) and it can also help emphasise patterns in data (for instance when colour coding ranges of statistical values on a map). Since colour can help so much in many applications it is usually considered essential.

1. The figure of 18000 pixels comes from the following approximation. An angular resolution of 30 seconds of arc corresponds to two pixels per minute or 120 pixels per degree, with a field of view of about 150˚ this corresponds to 120 x 150 = 18000 pixels.

2.2.1 The Cathode Ray Tube (CRT)

The raster scan Cathode Ray Tube (CRT) commonly used in conventional monitors is used in many VR display systems. The CRTs used in conventional monitors range in size up to about 35 inches (diagonally) - these large CRTs suffer from being extremely heavy and deep, however they offer high resolution and good colour.

At the other end of the range of CRT sizes are the small CRTs used in HMDs - these are typically around 1 inch in size. Monochrome CRTs are available that offer a useful resolution and are small and light, unfortunately the increased complexity of colour CRTs means that small, high resolution colour CRTs are not available. Another problem with CRTs is that they require extremely high operating voltages, which needs to be carefully shielded for safety reasons.

The Eyegen 3 HMD by Virtual Research uses monochrome CRTs and a colour wheel to produce a colour display. The image is separated into three separate red, green and blue images which are then displayed sequentially on the monochrome CRT at a much higher refresh rate. In front of the CRT is a colour wheel with red, green and blue filters which rotates and is synchronised to the display of the separate component images. Under most circumstances the eye and brain combine the images together so that the three separate colours are not noticed, however in fast moving scenes with high contrast a streaking effect is sometimes noticed where the trailing edge of a moving object appears to split into its component colours. One important disadvantage with the colour sequential display process, such as that used in the Eyegen, is that the process increases the latency of the system by a small amount. In the Eyegen 3 HMD the latency is increased by 11 ms **[Yeam94]**.

The fibre coupled display (**[Kala93]** p. 124-125) is another solution to the problem of the lack of small high resolution colour CRTs. In these systems a large high resolution CRT is coupled to the head set via a fibre-optic bundle. The headset can be reasonably light (especially if the fibre-optic bundles are supported externally) and have a high resolution.

High resolution (of the order of 1024 x 1024) in monochrome/grey-scale CRT based HMDs is available at a price, **[HoLa93]** lists a number of high resolution CRT based head sets starting at $74 000.

It is possible to use a conventional monitor to display stereo images, these are used on some very simple systems on the borderline of VR. This technique, *Frame sequential stereo*, displays the image for each eye alternately on the same monitor. The user wears a pair of glasses which have an LCD shutter over each eye. The operation of the LCD shutters is synchronised with the image display so that the appropriate shutter is transparent when the image for that eye is being displayed. These displays have high resolution and can produce a good stereo effect, however the feeling of presence is small since the display only covers a small part of the field of view and the rest of the physical world is still visible.

One useful benefit of the monochrome CRT is that it is one of the few display devices which does not have any type of visible pattern on the display device itself. Often the colour groups on CRTs and the individual elements in LCD based systems can be highly disturbing when they are placed a short distance from the eye or highly magnified (as is the case in projection systems).

2.2.2 Liquid Crystal Display (LCD)

LCDs offer the advantages of being small, light weight and easy to control electrically (no high voltages). However a serious problem with LCD technology is that the resolution is relatively small - for example 360 horizontal by 240 vertical monochrome pixels in the Virtual Research Flight Helmet displays. The Virtual Research Flight Helmet provides a colour display (in common with many other LCD based headsets) by using colour filters placed over the LCDs to convert groups of pixels into groups of primary coloured pixels. Many manufacturers quote the resolution of the displays in their head set in individual pixels - before the effects of colour grouping has been taken account of; thus an LCD quoted as having a 300 x 200 resolution may have 60 000 LCD cells making up 20 000 colour pixels.

2.2.3 Projection

Another use for LCD panels is as a component in projection systems. With suitable optics and a very strong light source, it is possible to project the image onto a screen many feet across, thus providing a large field of view. The technique of *back projection*, where the projector is placed behind the screen is used in flight and driving simulators. In these the windows of the mock-up shell of the aircraft act as the projection screens. Multiple projectors (when carefully aligned) can be used to increase the field of view still further.

There are a number of advantages to using LCD panels in projection systems. Instead of using one LCD panel with filters over it is possible to use three coloured LCD panels and combine the results optically. This means that the effective resolution is the same as the monochrome panels (as opposed to being reduced by the colour techniques discussed above) and there is less of a problem due to seeing the 'groups' of colours.

Projectors are not just constructed using LCD technology. High brightness CRTs have been used for many years to provide high resolution projection systems.

One system to make extensive usage of large scale projection systems is '*the CAVE*' **[CNSD93]** which consists of a system with four projectors (each driven by its own rendering system) which project onto three walls and up onto the floor. This creates a display which fills the participants field of view and provides a high degree of immersion. In '*the CAVE*' they use LCD shutter glasses to generate a stereo view on each wall.

Large scale projection systems benefit in a number of other ways:

- A number of users can view the system at the same time without an increase in the rendering requirements.
- Users can see their own bodies - this helps to judge distance.
- There is no latency due to movement
- Much more comfortable - no HMD.
- Users with glasses may find it difficult to use HMDs.

The technique of using LCD shutter glasses for stereographic display has the disadvantage that each participant has to wear expensive shutter glasses and that the display devices must refresh at twice the rate to ensure that no flicker is perceived. An alternative technique is to use two projectors and separate the images using polarisation. Each projector has a polarizing filter placed in front of it - the polarisation on the two projectors being separated by 90°. The user can then wear low cost passive polarizing glasses to see the stereo effect. This technique has been in use in the Advanced Interfaces Group for a number of months and has been found to be very useful. It suffers from two small disadvantages:

- If the user tilts his head the stereo effect is lost completely.
- The polarisation of the light produced by the particular projectors in use is different for each of the primary colours. It is necessary to use extra polarisation filters to ensure that the colour balance is not lost.
- The filtering process reduces the intensity of the projected image, often restricting the use of these systems to darkened rooms.

2.2.4 Other display techniques and considerations

[VRNe94] reports that the US government is funding development of a very high resolution (expected to be around 2000 x 2000 pixels per eye), large field of view (200° or more horizontal x 100° or more vertical) HMD. Although these technologies will probably be initially very expensive, the potential for the availability of high resolution, wide field of view headsets means that many applications which are presently impractical will become feasible. This will lead to a much greater load for the rendering systems both in terms of the number of pixels in the display and the scene complexity. The technologies which may be used in that headset include the tiling of multiple display panels together to increase the total resolution. A number of other display technologies which are still in the early stages of research are described in **[Kala93]**.

Many headsets are available which offer a 'see-through' mode, allowing the user to see both the real and virtual worlds simultaneously. This presents at least two problems:

- Registration between the real and virtual worlds is now very critical - any discrepancy will easily be noticed.
- It is possible for the user to place a virtual object behind an object in the real world. The display system needs some way of knowing about the real world so that it can hide the virtual object from view.

All of the display devices which have been described in this section require the image to be transferred to the display device by a scanning process one pixel at a time. All these devices also require constant refreshing even when the contents of the display do not change. The combination of these requirements together produce a requirement for a very high bandwidth data stream from the frame buffer to the display. Another consequence of this scheme is that the entire display is not refreshed at the same time; this causes *temporal distortion*, which is described in more detail in section **A.8.1**. In some devices (notably CRTs), there is a small break in the scanning process between the end of the last line on one frame and the first line on the following frame; this is known as *flyback time*. Some systems perform all rendering in the flyback time; this removes the need for double buffering on the frame buffer and thus reduces latency.

2.3 Other output devices

Although the visual display is the primary display device in VR systems, a number of other output devices are appearing:

2.3.1 Tactile Feedback

In one example virtual reality system that the author has used, the user has to touch a door handle to be able to go through a door. Most people who used the system found this to be very difficult. One possible aid would be tactile feedback to tell the user when he had touched the handle (or another part of the door). Similarly it would be very difficult for the user to catch a ball without the sensation of the ball hitting his hand. A number of devices now provide tactile feedback to aid in these and other tasks. Tactile feedback (and the similar *Haptic feedback*) allow the user to feel a simulated force or texture. Some examples of these devices are:

- The '*Teletact Glove*'

  The *Teletact Glove* developed at the Advanced Robotics Research Centre (see **[Kala93]** p. 40 for a brief description) consists of a pair of gloves - one is an input glove and the other an output glove. The input glove contains a number of force sensitive resistors to detect forces against the glove. The output glove contains a number of air pockets which can be inflated under the control of a computer which is monitoring the input glove. In this way the force on the input glove is transmitted to the output glove.

- The Argonne Remote Manipulator (ARM)

  The '*Argonne Remote Manipulator*' (ARM), consists of a mechanical arm with motors providing force feedback. The ARM is used extensively by the University of North Carolina in their *Grope* project **[Broo90]**. In their latest system

(GROPE-IIIB) the ARM is used to aid research chemists during a complex molecular docking task by simulating chemical forces on the ARM.

- Hydraulic Motion Platform

  One other form of force feedback is the motion platform. These are used to control the position of cabin bodies in driving and flight simulators and simulate some of the forces experienced when driving or flying in the simulated vehicle.

2.3.2 Acoustic output

Sound can enhance a virtual environment greatly. In some ways the lack of sound is a problem since users expect objects to make a noise when they are dropped or to hear the noise of an engine in a driving simulator; the lack of these cues can be disconcerting. In a virtual environment it is not only important to produce realistic acoustic cues, but to make the sound appear to come from the direction of the object producing it. One example of a technique to do this is described in **[WeFo90]**.