Chapter 1

Introduction

'*Geometry is not reality. Interactivity is reality.*'

(Myron Krueger)

1.1 Motivation

Virtual reality (VR): although the term was only introduced in around 1985, the possibility of being able to immerse and interact within a realistic computer-generated world has fascinated computer graphics researchers for decades. In his seminal paper, 'The Ultimate Display', Ivan Sutherland encapsulated his vision of the future when he speculated:

> '*The ultimate display would, of course, be a room within which the computer can control the existence of matter. A chair displayed in such a room would be good enough to sit in. Handcuffs displayed in such a room would be confining, and a bullet displayed in such a room would be fatal.*' (Sutherland, 1965)

Clearly we are far from attaining this considerable and provocative goal. Nevertheless, the field of computer graphics has advanced considerably over the past three decades. This has seen a progression from simple, monochrome, wire-frame figures to full-colour, lit, shaded, and textured representations. However, in the perpetual strive towards more efficient photo-realistic displays we are now starting to reach a fundamental limit: that of processing power.

VR is an extremely computational demanding paradigm, requiring the simulation and display of a virtual environment (VE) at interactive frame rates. Even with the use of powerful graphics workstations, a moderately complex VE can involve a vast amount of computation, inducing a noticeable lag into the system. This lag can detrimentally affect the user and may therefore severely compromise the diffusion of VR technology. It is widely considered that this problem of lag is the most pressing which currently faces the VR community (NSF, 1992; van Dam, 1993; Kalawsky, 1993).

This thesis presents the work which has been performed in an attempt to reduce these lags. The method which has been adopted is to trade visual complexity for interactivity. This is not a new concept in itself: the notion of level of detail (LOD) is well founded in the field of computer graphics. However this exposition presents a formal and extensive treatment of LOD which has not been attempted elsewhere to date.

Specifically, we will investigate the efficacy of the human visual system in an attempt to quantify the degree of detail which a user can perceive at any instant. Using this information we can then select the optimal LOD for each object in the VE without the user being aware of any degradation in image quality. In other words, we wish to investigate the fundamental limit of visual detail which a real-time computer graphics system need display; and hence optimise the performance of the system.

In summary, if the ambitions of this research programme could be distilled into a single sentence, then they could be encapsulated as follows:

> **To attempt to reduce the lags inherent in current virtual reality (VR) systems by seamlessly degrading the level of detail (LOD) of objects in the virtual environment (VE) based upon principled models of visual perception.**

The remainder of this first chapter will review the relevant background material to enable a complete understanding of the various issues that this statement involves, and to appreciate the decision processes which ultimately led to this objective.

1.2 Lag in Virtual Reality Systems

Every VR system suffers from an inherent lag to one degree or another. The minimisation of this lag is crucial to the development of effective VR systems, and is the key theme of this thesis. Accordingly, the following sections will detail precisely what we mean by the term lag. Then the effects of lag on the user will be assessed. And finally, a number of solutions will be presented which have been employed to reduce these effects.

1.2.1 A Definition of Lag

Lag is the delay which the user experiences when using a VR system. More precisely, it is the period of time between some change being effected on the VE, and that change being reflected on the display device (Bryson and Fisher, 1990). This can be broken down into two principal types of temporal degradation (Bryson, 1993):

1. **<u>Update Rate</u>** : the rate at which the visual display is refreshed. This is also often referred to as the frame rate of an application. A low update rate implies a high degree of lag.
2. **<u>Transport Delay</u>** : the period of time between a user's input and the effect of that input being represented on the display device, e.g. the lag between a user panning their head and the display being refreshed with the new perspective of the VE.

We shall refer to the cumulative effect of both of these delays as **System Lag**. There are a number of contributing factors which can affect the magnitude of the System Lag. These can be divided, in terms of the application program, into three fundamental components (Atkin, 1993):

1. **<u>Sensor Delay</u>** : the delay incurred by the sensing of the real environment, e.g. the tracking of a user's body part such as their head or hand. This constitutes the delay between the user moving a tracked body part and the computer system receiving notification of that movement.

2. **Processing Delay** : the delay incurred by the simulation of the VE, i.e. all of the calculations which must be performed on each object in order to update its state. Examples of this include: collision detection, implementing object behaviours, incorporating dynamics rules, updating representations of users' body parts, etc.

3. **Rendering Delay** : the delay incurred by the display of the VE. Depending upon the specific algorithm which is employed, this could involve various operations such as: clipping non-visible objects, projecting coordinates, back-face removal, depth sorting, accounting for light sources, applying textures, etc.

Normally, only one of these will prove to be the bottleneck for the entire system. However, which of these this is will depend upon the particular system being used as well as the nature and complexity of the VE being modeled.

For a single-processor model of a VR system, System Lag is simply defined as the linear combination of the above three components, i.e.

$$\textit{System Lag} = \textit{Sensor Delay} + \textit{Processing Delay} + \textit{Rendering Delay}.$$

For a multi-processor model, the relationship between these delays can become more complex and will obviously depend upon the degree and type of parallelism employed. As we are only concerned with the general issues here, we will not complicate the discussion at this stage with definitions of lag in multi-processor systems. Suffice to say however that the same components are relevant to both classes of VR systems; although normally a multi-processor system will incur an additional Synchronisation Delay to coordinate the activity of each processor (Wloka, 1995).

1.2.2 Effects of Lag

The degree of lag which a VR system exhibits can manifestly affect the user in a number of adverse ways. The following list attempts to summarise the potential consequences of high lag systems which have been reported to date:

- **Degraded User Performance** : humans are extremely sensitive to visual lags, and as a result our efficiency diminishes when we are forced to work in a high lag environment. For example, lags of between 30–120 ms have been shown to degrade user performance, depending upon the application (Held and Durlach, 1993). Singhal and Cheriton (1995) state that humans can detect latencies of around 100 ms, and will only tolerate maximum inconsistencies in the order of 200 ms. Gregory (1990) also reports that a lag of around 500 ms can seriously degrade hand–eye coordination tasks such as drawing and writing.

Bryson (1993) investigated the effect of both Update Rate and Transport Delay on the user's ability to perform a particular tracking task. He found that the impact of both of these lags was quantitatively similar, and that the user's performance degraded linearly with respect to lag between 0 and 0.5 seconds.

- **Motion Sickness** : delays in the update of visual information have also been reported to cause effects of motion sickness, the symptoms of which include nausea, pallor, and cold sweating (Money, 1970). This is thought to be due principally to the conflict between the user's visual and vestibular senses (Oman, 1993), often referred to as the Vestibular Ocular Response (VOR). Frank *et al.* (1988) have shown that humans can detect head tracking lags of greater than 5 ms and that delays of more than 30 ms can produce effects of motion sickness. Regan (1995) investigated the side-effects of head-coupled immersive VR applications and found a high incidence of motion sickness (61% of subjects in a 20 minute experiment).

  On a similar note, Uliano *et al.* (1986) report that asynchronous visual delays are also a culprit of motion sickness. This would suggest that fixed frame rate graphics systems should be advocated over variable frame rate ones (Hawkes *et al.*, 1995; Helman, 1994).

- **Degraded Sense of Presence** : the term 'presence' (with respect to VEs) is used to define the degree to which a user feels a sense of being physically present within a computer-generated environment. This concept was first developed by Sheridan (1992) and much research has been done subsequently to investigate the various determinants of presence in VEs (e.g. Slater *et al.*, 1995; Witmer and Singer, 1994; Slater and Usoh, 1993).

  The notion of presence has been proposed as a metric to measure the effectiveness of a VR application, based upon whether the user can interact with it in a natural manner. With respect to this, it has been reported that reduced frame rates (less than 15 Hz) can significantly diminish a user's sense of presence within a VE (Barfield and Hendrix, 1995).

1.2.3 Techniques to Reduce Lag

Based upon the above discussion, it is clearly undesirable to produce VR systems with a high degree of lag. As a result, a number of techniques have been developed to reduce System Lag in VR systems. A number of these are discussed below. After each case, an indication is provided towards the component of System Lag which the technique aims to reduce. This is symbolised with one or more of the labels: SENSOR, PROCESSING, or RENDERING.

It should perhaps be noted at this stage that it is not possible to totally eradicate lag: we must perform some calculations in order to process and display a VE, and these will always involve a finite period of delay. However, we can attempt to minimise the magnitude of the lag in order to avoid the above undesirable side-effects. The exact value of what constitutes an acceptable degree of lag is uncertain, although it will undoubtedly depend upon the particular application in question. However, it is generally accepted that a 15 Hz frame rate (i.e. a lag of $\sim$66 ms) is a minimum requirement for interactive, real-time, three-dimensional (3D) graphics systems (e.g. Barfield and Hendrix, 1995; Falby *et al.*, 1993).

- **Motion Prediction** : motion prediction techniques attempt to overcome the sensor delay incurred by sampling the position and orientation of the user's body part, or parts. The underlying principle is to pre-empt the tracking system by calculating estimated intermediate values and then using these approximations until a new reading is available from the sensor. As Kalawsky (1993) notes, simple extrapolation schemes are generally unsatisfactory, so most successful systems have been based upon Kalman filtering techniques (e.g. Liang *et al.*, 1991). [SENSOR]

- **Update Rates** : the notion of update rates[1] can be used to define the rate at which an object's state is recomputed (Wloka, 1993b). That is, to decide upon the amount of processing time to devote to an object.

[1] N.B. update rates in this context is different from the singular term update rate, which was defined previously on page 3.

The premise of this work is that many objects are either slow-moving, relatively unimportant background objects, or static objects and therefore do not require recomputation at every frame. [PROCESSING]

- **Visibility Culling** : in the demanding domain of VR, it becomes necessary to design graphics systems that restrict computation to only those entities which require consideration. Objects which are not within the current viewing frustum should be immediately discarded from the rendering process. Additionally, it is often desirable to discard any polygons from a scene which are back-facing or obviously not visible. This can alleviate much of the unnecessary burden on the graphics system. [RENDERING]

- **Frameless Rendering** : standard double buffered graphics systems can only display a new frame once the entire image has been rendered. In contrast, frameless rendering is a technique which displays each new pixel as it is calculated. Obviously, if the scene were to be rendered in a systematic left→right, top→bottom fashion then the user would be aware of the progressive update of the screen. Instead, pixels are rendered in a pseudo-random order so that the effect is less visually distracting (Bishop *et al.*, 1994). The applicability of this technique to polygon renderers (as commonly used in VR applications) has been demonstrated (Wloka *et al.*, 1995) and shown to increase the frame rate of the graphics system. [RENDERING]

- **Galilean Antialiasing** : Costella (1993) has proposed a hardware-based motion extrapolation system which operates at a pixel level. Basically, this involves a 'smart' frame buffer which, in addition to holding standard information such as colour and z-buffer values, can also record the velocity and acceleration of each pixel. Using this information, the frame buffer can generate intermediate frames by extrapolating the position of moving pixels. This is performed in order to preserve the update rate of the system while the graphics system is generating the next frame. [RENDERING]

- **Level of Detail** : by storing a number of different representations of objects, each varying in complexity, the VR system can select a particular representation to use for each object based upon a number of possible criteria. Using a less complex level of detail (LOD) model will produce a coarser representation than a more complex one, but it will also induce a smaller lag. [PROCESSING, RENDERING]

- **World Subdivision** : the entire VE is segmented into a number of smaller, self-contained volumes. Only objects contained within those volumes which the user is occupying, or can see at any time, are considered for processing and rendering. This technique is only of use if the VE is amenable to this sort of segmentation. For example, it is particularly applicable to architectural walkthroughs where the boundary of each room can be used as the boundary for the subdivision (Airey *et al.*, 1990): but it would be of little use for large, open spaces. [PROCESSING, RENDERING]

- **Employ Parallelism** : as alluded to in Section 1.2.1, we can attempt to reduce lags by using multi-processor systems and parallelising the major processes. This of course does not actually reduce any of the component lags in a system; but it does reduce the combined effect of these, i.e. System Lag. This approach has been used in a number of successful systems such as Division Ltd.'s PROvision systems (Atkin, 1993), Silicon Graphics Inc.'s IRIS Performer (Rohlf and Helman, 1994), and the freely-available MR-Toolkit package (Shaw *et al.*, 1992).

1.2.3.1 Discussion of Solutions

Each of the above solutions offer various advantages and disadvantages. In the case of the motion prediction techniques, such as Kalman filtering and Galilean antialiasing, because we are dealing with a human-in-the-loop simulation, there is only a finite degree to which data can be extrapolated. For example, Liang *et al.* (1991) report that the reliability of prediction decreases as the length of the prediction increases; and that predictions beyond three steps ahead were not feasible.

Graphics optimisation techniques such as visibility culling, frameless rendering and world subdivision provide a means of reducing the burden on the graphics system. However, there is only a certain extent to which this can be performed for any VE, and so these techniques do not provide us with any flexible means to modulate the workload of the system.

With regards to employing parallelism and designing custom architectures to reduce lag, it is the author's opinion that if VR is to become a widespread and accessible technology, then it must be made available on low end, off-the-shelf architectures. For this to be possible, work is required into the development of algorithmic solutions to combat the various latencies currently experienced in VR systems, as opposed to resorting to custom hardware solutions. This view is held by a number of other researchers (e.g. Billyard, 1993).

Arguably, the only solutions presented above which enable the application program to balance the load of the system in real-time are: update rates and level of detail. All of the others offer a single (meaningful) reduction in lag for any application. Of these two solutions, we are going to concentrate on the notion of level of detail. The following section will therefore describe this technique in further detail.

1.3 Level of Detail

The complexity of any VE will influence the time required to process and display that environment. For example, a VE which is composed of a large number of polygons will take longer to process and render than a VE with fewer polygons (Deering, 1994). This is not a simple relationship because, for example, the size and geometry of polygons can affect performance, and the use of texture maps can add visual detail to a VE which would otherwise require a large number of small polygons. However, in general, we can state that the more visual complexity that we include in a VE, the greater the lag which will be subsequently induced.

The technique of level of detail takes advantage of this relationship by effectively modulating the visual complexity of each object in real-time. This is done by storing a number of different representations for an object, each varying in complexity, and allowing the graphics system to select the most appropriate model to display at any juncture. Figure 1.1 illustrates this concept with two different levels of detail for a single object.

In essence, LOD techniques attempt to trade spatial fidelity for temporal fidelity: a less complex model will be displayed more rapidly, but it will appear more coarse. This is a valid trade-off for a real-time, interactive graphics system because it has been shown that humans are far more tolerant to reductions in image quality than to delays in visual update (Smets and Overbeeke, 1995).

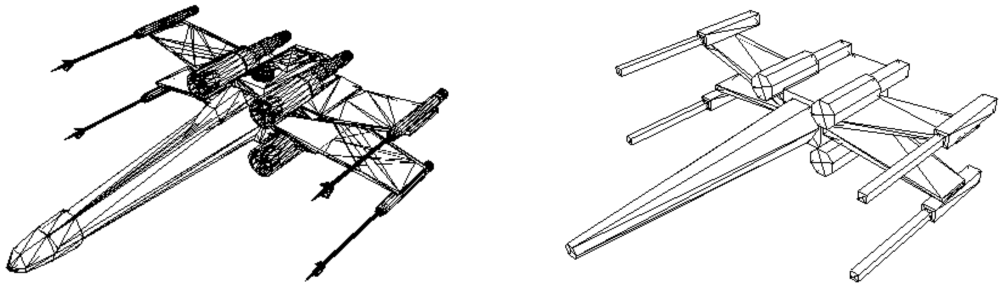

Figure 1.1: *An example of two different levels of detail for a single object, where (a) contains 6,048 polygons and (b) contains 278 polygons.*

For example, Swartz *et al.* (1992) have shown that frame rate is more important than resolution for target detection, recognition, designation, and tracking.

1.3.1 Generating Level of Detail

There are a number of ways in which we can vary the computational complexity of a polygonal model; and thus create different levels of detail for an object. The most common techniques are as follows:

- **Polygon Reduction** : the goal of this approach is to produce a new model with the same general form and genus as the original model, but containing fewer polygons. This normally implies the invocation of some polygon reduction or simplification process over the original object. A number of these will be discussed in the following chapter.
- **Texture Mapping** : texture mapping techniques can also be used to modulate the geometrical complexity of a model. Essentially, we can replace regions of high geometric detail with a single textured polygon. The polygon's texture is simply a rendered image of that section, from a certain viewpoint and distance. This optimisation can introduce visual artifacts if the model is viewed from a different viewpoint or distance. However, a number of solutions exist to combat this problem; including warping the texture image (McMillan and Bishop, 1995), warping the adjacent geometry (Aliaga, 1996), and smoothly transforming between geometry and texture (Maciel and Shirley, 1995).

- **<u>Illumination Models</u>** : the particular illumination model which is used to render an object can influence the display time. For example, lighting models such as Phong or Gouraud (smooth-shading) produce more realistically shaded results than the Lambertian (flat-shading) model, but at the price of higher computational complexity (Foley *et al.*, 1990). Therefore, we can use different lighting models to render an object to introduce different levels of detail (Funkhouser and Séquin, 1993).

1.3.2 A Typical Implementation

The normal way in which LOD is implemented is to base the model selection upon the distance of the object from the viewpoint. This means that we can use a high LOD model (e.g. many polygons) when an object is close to the observer, but substitute this with a cruder model (e.g. less polygons) as it progresses away from the observer. Thus we gain a computational advantage when processing distant objects without detrimentally affecting the fidelity of the display.

The basic premise behind this approach is that when a small object is rendered on any computer screen, it can only be displayed to a certain degree of accuracy and detail (due to the limited resolution of the display device). Therefore, it makes no sense to display a highly detailed object under these conditions because all of the sub-pixel detail will simply not be visible (see Figure 1.2).

The term level of detail is often synonymous with distance LOD (because this is the most prolific use of LOD). However, for the purpose of this thesis we should note that LOD is simply a technique to modulate the detail of an object: it does not imply any criteria for selecting an object's detail level, i.e. LOD does not imply distance LOD. As we will see, there are many other criteria which we can utilise to choose the most appropriate LOD at any point.

1.3.3 Identifying and Addressing Problems with LOD

The technique of LOD offers a powerful method to reduce the computational burden of a VR system. However, there are a number of problems with this technique as it stands. These include:

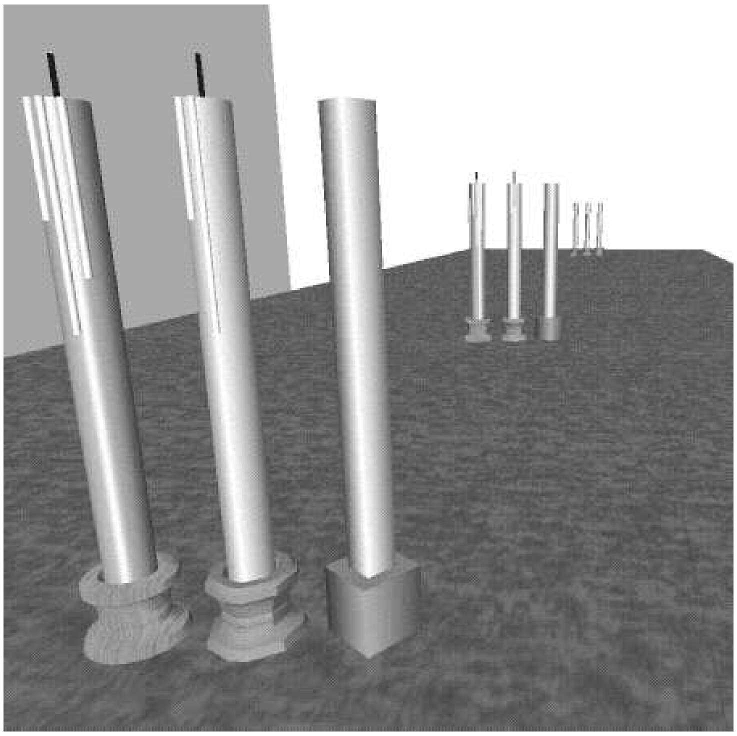

Figure 1.2: *An illustration of distance LOD for three different models of a candlestick. Reproduced with kind permission from Wernecke (1993). (c) Silicon Graphics Inc., 1994.*

1. There is currently no principled mechanism to select the optimal LOD at any point. As a result, most systems use trial and error judgements or *ad hoc* heuristics to decide this (Chrislip and Ehlert Jr., 1995). While this may be satisfactory for the occasional distance LOD implementation, a general LOD paradigm must be based upon a more formal and tractable solution.

2. As a direct consequence of there being no formal procedure for selecting LOD, we find that very often a noticeable flicker is incurred when switching between different levels of detail (the so called **popping effect**). In order to impose the minimum distraction on the user, we desire a system whereby the switch between levels of detail produces no visual artifacts.

3. Various selection criteria have been proposed to manage the modulation of LOD (in addition to distance LOD). These will be discussed later in Section 2.1. However, only a few simplistic attempts have been made to integrate all of these into a single, generic, and orthogonal framework.

4. There is very little support in currently available VR packages to actually generate different levels of detail for an arbitrary object (Reddy, 1995b).

Without the ability to produce various LOD models, the effectiveness of LOD is obviated.

This thesis will attempt to address each of these issues by investigating theories of visual perception. This is done in order to produce a model which quantifies the degree of visual detail that a user can perceive, and then to use this information to select the optimal LOD for each object without the user being aware of any change.

1.4 The Visual System

In order to modulate the LOD of an object based upon its perceptual content, one must first possess an appreciation for the way in which the human visual system is designed and how it is believed to function. The following sections describe the anatomy and physiology of the visual system, providing a basic foundation to help the reader understand and follow the subsequent perceptual material.

The visual system can be divided into three major processing sites (see Figure 1.3). These are the **eyes** (which act as the input to the visual system), the **visual pathways** (which transport and organise the visual information) and the **visual cortex** (the section within the brain which enables visual perception). Each of these sites performs a particular analytical process on the optic image data and will be described in detail in the subsequent sections. Following this, we will develop an appreciation for the spatial resolution of the visual system and how this varies systematically under a number of circumstances.

1.4.1 The Eye

The eye is our window onto the external world, being approximately spherical with a diameter of around 24 mm. The inner **vitreous chamber** is enclosed by three layers of tissue. These are the **sclera**, the **choroid**, and the **retina** (see Figure 1.4).

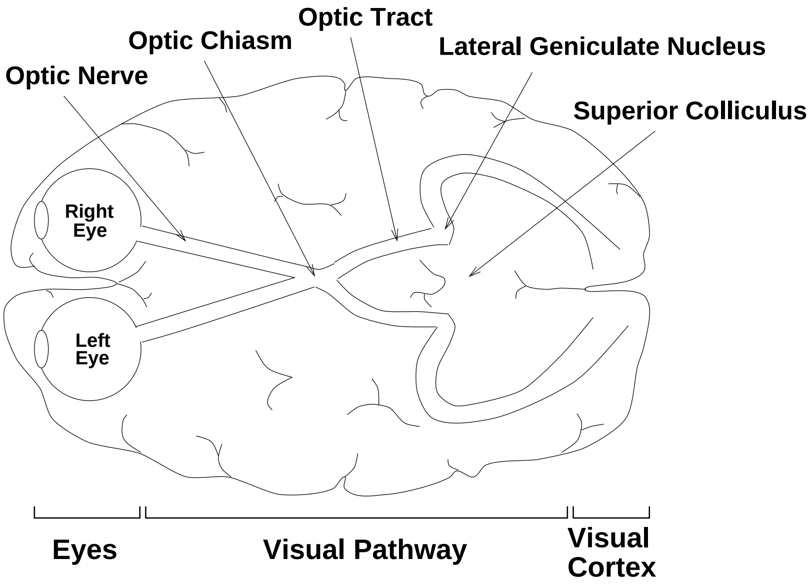

Figure 1.3: *Plan view of the brain showing the extents of the three major sites of processing in the human visual system.*

Of these, the retina is the most pertinent to our discussion because it is the point where light is first detected and processed. As such, its operation and efficiency characterise the raw data which are available to the ensuing vision processes and define the physical constraints and limitations of the human percept system. It would therefore be prudent to investigate the physiology of the retina further.

1.4.1.1 The Retina

Light enters the eye through the transparent cornea. It is focussed by the **lens** and passes through the vitreous chamber before reaching the retina at the back of the eye. Within the retina itself, light must pass through a number of layers of neurons before finally reaching the **photoreceptor cells**. These are responsible for converting the incident light energy into neural signals which are then filtered back through the network of neurons, consisting of the **collector cells** and the **retinal ganglion cells**. The axons of the retinal ganglion cells form the **optic nerve** which transports the neural signals to the visual cortex via the visual pathway. Figure 1.5 illustrates the elements involved in this process.

There are two principal classes of photoreceptor cells in the eye: these are the **rods** and the **cones**. Rods provide high sensitivity in dim light, whereas cones

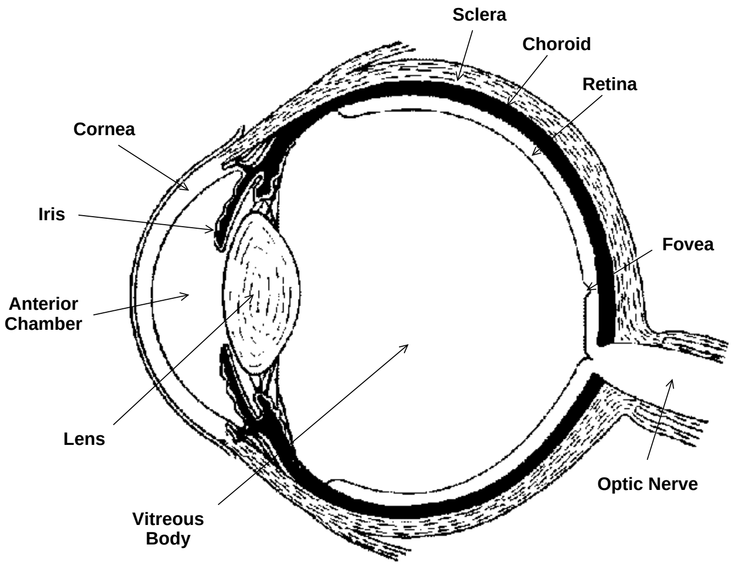

Figure 1.4: *Cross-section of a human eye illustrating its major layers and structures.*

offer high visual acuity in bright light. This duplex arrangement enables humans to see in a wide range of lighting conditions. Both receptor types contain a number of light-sensitive molecules, called **photopigments**. Each of these consist of a large protein called opsin and a vitamin A derivative known as retinal. When light strikes the photopigment, it initiates a reaction which results in the molecule splitting and the subsequent generation of an electric current.

The electrical signals which are generated in the photoreceptors are transmitted synaptically through the collector cells (incorporating the **horizontal**, **bipolar**, and **amacrine** cells) and the retinal ganglion cells. Each eye has roughly 130 million photoreceptors, but only around 1 million retinal ganglion cells. It is therefore apparent that the neural network reduces the retinal image into a more concise and manageable representation—extracting only the relevant parts of the image which are of particular interest. So which features of the retinal image is the eye interested in? In order to answer this question we must examine the visual stimuli which evoke a response from the retinal ganglion cells.

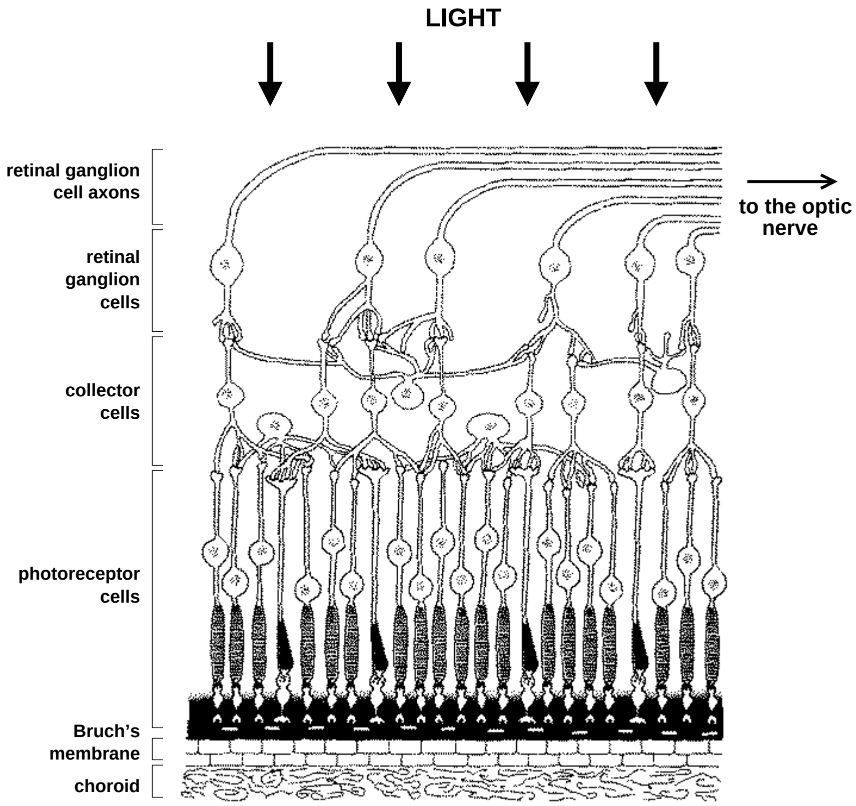

Figure 1.5: *A cross-section of the retina showing the various layers of photoreceptors and neurons which detect and filter incoming light. Adapted from Sekuler and Blake (1994). Note that light must travel through several layers of cells before reaching the photoreceptors.*

1.4.1.2 The Retinal Ganglion Cells

The inputs of the retinal ganglion cells are arranged in an antagonistic, concentric pattern composed of a centre and a surround region. The ganglion cell is continually emitting a background signal; however when light strikes the photoreceptors in one region, this stimulates an increased response from the retinal ganglion cell (a so-called **ON-response**). Whereas light falling on the other region will generate a reduced response, or **OFF-response**. This arrangement is illustrated in Figure 1.6. If the centre region is stimulated by an ON-response, then the retinal ganglion cell is referred to as an **ON-centre** cell. Conversely, if the centre region is stimulated by an OFF-response, then the cell is referred to as an **OFF-centre** cell.

The outputs from the ON-response and OFF-response regions are summed together to form the net response of the retinal ganglion cell. This means that

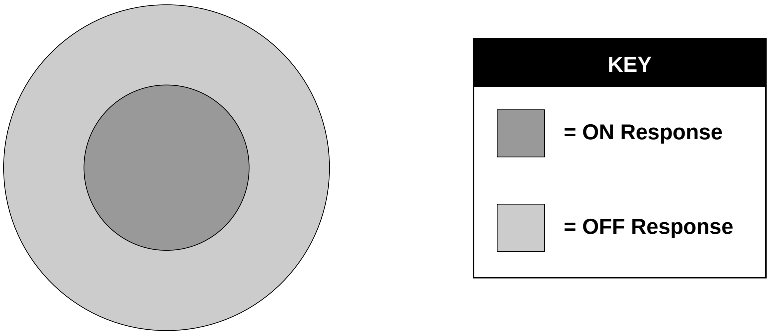

Figure 1.6: *Representation of an ON-centre retinal ganglion cell. Light falling in the ON-response region causes an increase in cell activity; whereas light falling in the OFF-response region causes a decrease in cell activity.*

if the same luminance is presented across the cell then it will elicit a weak response because of the antagonistic reaction between the centre and surround regions. However if an ON-response region receives light when the corresponding OFF-response region does not, this differential will cause a strong response. An illustration of this operation is presented in Figure 1.7.

Accounting for the above observations, we can define the following characteristics of retinal ganglion cells (and hence the first stage of processing which is performed on the retinal image).

- Retinal ganglion cells produce a marked response only when there is a contrast gradient across its **receptive field** (the area of the retina which the ganglion cell receives input from), i.e. they are sensitive to edges.
- A light stimulus which falls out with the cell's receptive field will have no effect on the cell's response.
- The size of the cell's receptive field defines the size of the light stimulus to which it is maximally sensitive.
- The orientation of a stimulus does not affect the cell's response (because the centre and surround regions are circular).

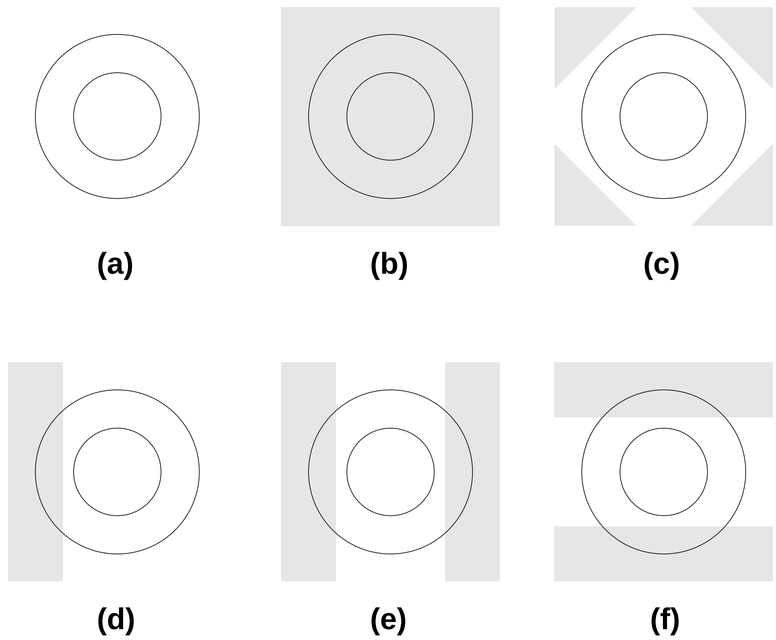

Figure 1.7: *Various stimulus arrangements for a single ON-centre retinal ganglion cell. Cells (a)–(c) all generate an equally weak (background) response whereas cells (d)–(f) generate a positive response, with cells (e) and (f) responding equivalently.*

1.4.2 The Visual Pathways

Once the image data leave the eye via the axons of the retinal ganglion cells, they are transported by the visual pathways to the higher vision centre of the brain. However, these pathways do not just passively transport the visual information: some degree of re-organisation and processing is also performed on the signals (although this has relatively little impact on the vision system as we are interested in it).

The main sites of processing on the visual pathways are the **superior colliculus** and the **lateral geniculate nucleus** (LGN). The superior colliculus is believed to be responsible for determining the location of a stimulus and also initiating and controlling eye movements (Carpenter, 1992). Of more interest to our discussion are the lateral geniculate nuclei (there are two—one on each side of the brain), which form a feedback loop with the visual cortex.

The cells of each LGN have antagonistic, concentric receptive fields much like those of the retinal ganglion cells. However, the surround region of an LGN cell exerts a stronger inhibitory response on the centre than its retinal counterpart. This effect induces a more powerful differential between adjacent retinal regions, but contributes very little new to the vision process.

There are two classes of cells in the LGN which remain largely segregated: the **parvocellular** (or P cells) and the **magnocellular** (or M cells). These have a certain correspondence with the cones and rods in the retina, respectively. P cells tend to have smaller receptive fields and so can resolve finer details. They are also sensitive to the colour of light which falls within their receptive fields. By comparison, M cells are sensitive to more coarse features and are stimulated by luminance, not spectral, variances. However, M cells respond vigorously to the rapid motion of a stimulus within the visual field. This provides evidence for the widely-held theory that the various components of vision—form, colour, movement, and depth—are transmitted independently via separate channels to the visual cortex (Livingstone and Hubel, 1988).

1.4.3 The Visual Cortex

The visual cortex (also referred to as the striate cortex, Area 17, and V1) is the major centre of vision. It is located in the occipital lobe, towards the rear of the brain.

As in the retina and LGN, the cells of the visual cortex have a receptive field which restricts the sensitivity of the cell to a certain region. Also, the cortical cells respond maximally to gradients of luminance across their receptive fields rather than ambient illumination levels. However, unlike the retinal and LGN cells, they are also selective on the orientation of a stimulus and the direction of moving stimuli (Blakemore and Campbell, 1969).

We can segregate the cortical cells into two classes (Hubel and Wiesel, 1962): the **simple cells**, which are orientation selective to stationary or slow moving stimuli, and the **complex cells**, which respond maximally to moving stimuli of a particular orientation. As may be envisaged, the receptive fields of cortical cells are not concentrically circular as in the retina and LGN. For example, the receptive field of a simple cell is an elongated shape with discrete excitatory (ON-response) and inhibitory (OFF-response) zones. Figure 1.8 illustrates some examples of how these zones are arranged in order to achieve their orientation selective nature. For example, Figure 1.8(b) will be maximally sensitive to a vertical edge and least sensitive to a horizontal edge. In general, a deviation of around 15° from a cell's preferred orientation is sufficient to render a feature undetectable to that cell (Sekuler and Blake, 1994).

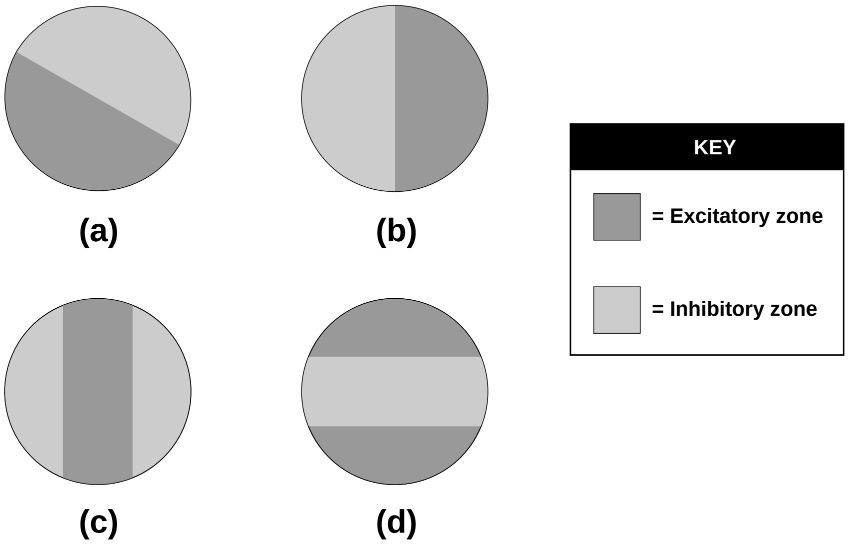

Figure 1.8: *Example receptive field layouts of four simple cortical cells, illustrating the orientation selective nature of these cells.*

Similarly, complex cells are also sensitive to the orientation of a contrast gradient. However, the position of the edge within its receptive field is not as important as it is for simple cells—edges of the preferred orientation can be detected anywhere within the cell's receptive field. In addition, complex cells respond strongly to the presence of a rapidly moving edge. Often this response is selective for a particular direction of movement through the cell's receptive field.

1.4.4 Sensitivity to Visual Detail

We have described the basic physiology of the human visual system. Now let us take a more detailed look at the implications of this design and the effect that these have on the degree of spatial detail that we can perceive.

1.4.4.1 Spatial Resolution

The size of a cell's receptive field (be it a retinal, LGN, or cortical cell) determines the size of stimulus to which it is optimally sensitive. Throughout all three of the vision processing sites we find collections of cells which exhibit a range of receptive field sizes; thus providing sensitivity to a range of stimulus sizes.

The resolution of the human visual system must therefore clearly possess a finite limit which is ultimately determined by the smallest receptive field size. For example, the spacing and pooling of photoreceptors in the retina (which form the inputs to the ganglion cells) will impose the primary limit on how much detail we can perceive. In the most densely packed region of the retina, photoreceptors subtend around 0.5 min of arc. Not surprisingly therefore, we find that the eye can detect detail down to a size of about 0.5 min of arc (Humphreys and Bruce, 1991)[2].

1.4.4.2 Retinal Inhomegenity

The eye's sensitivity to the size of a stimulus is not uniform across the entire retina. Instead we find that a very small part of the retina, known as the **fovea** (see Figure 1.4), has the ability to resolve the smallest features. However this ability degrades in proportion to **retinal eccentricity** (angular distance from the centre of the retina) such that in the peripheral field, the retina has very poor discrimination of fine detail.

This phenomenon means that whenever we wish to focus our attention onto an object, we must perform a combination of head and eye rotations so that the light reflected from that object is projected onto the foveae of our retinae. This ensures that we see the object in the highest detail. Although the peripheral regions of the retina are not as sensitive to visual detail, they are far more sensitive to movement. This provides humans with a highly adapted balance between acuity and sensitivity.

There are a number of physiological features of the visual system that vary with retinal eccentricity. These include:

- The concentration of cells varies dramatically across the retina. In the fovea, only cones are found; and the concentration of these drops off rapidly towards the periphery. Rods appear in large numbers just beyond the fovea, but these also reduce substantially in number towards the periphery. The density of ganglion cells across the retina is roughly proportional to

[2]We will be referring to measures of visual arc quite frequently throughout the text. As these units are often difficult to visualise, it may be illustrative to provide a simple rule of thumb: 1 cm at 57cm distance subtends 1 deg of arc. This roughly means that the width of your thumb in centimeters is its angular subtense in degrees at arm's length. (60 minutes of arc = 1 degree.)

the cone density (Hallett, 1991).

- The receptive field size of retinal ganglion cells increases linearly with eccentricity (Kelly, 1984). This is because the degree of photoreceptor pooling for ganglion cells varies with eccentricity. At the fovea, there is a 1:1 correspondence between cones and ganglion cells. This increases to a 7:1 mapping beyond the fovea (Cowan and Ware, 1985).

- There is a disproportionate number of cells in the visual cortex devoted to the foveal region of the retina. It has been estimated that 80% of all cortical cells are devoted to the central 10 degrees of the visual field (Drasdo, 1977).

The result of these characteristics is that our vision is maximally sensitive within a central region of approximately 5 deg of arc, and drops off smoothly towards the periphery (Zeki, 1993). This reduction in visual acuity across the retina is significant, with around a 35-fold difference existing between the fovea and the periphery (Nakayama, 1990).

1.4.4.3 Temporal Sensitivity

The human vision system cannot resolve as much detail in an object which is moving across the retina as it can in an object which is stabilised on the fovea. This causes the familiar effect of objects blurring as they move past our point of fixation, or as we pan our head to fixate on another target.

The reason for this effect is thought to be due, in part, to the eye's inability to track rapidly moving targets accurately, thus causing a slippage in the retinal image (Murphy, 1978). However, based upon the more recent studies of Tyler (1985), it has been suggested that the photoreceptors themselves are also responsible for limiting our temporal frequency sensitivity (Nakayama, 1990). That is, the process of detecting motion implies an integration of a moving object's stimulus energy over time. Therefore the visual information for precise features in that object are destroyed by this integration process.

1.5 Thesis Content

We have now completed the presentation of required background knowledge. The remaining chapters of this thesis are organised as follows:

Chapter 2 (Background) : the next chapter will present a review of scholarly literature which is relevant to this thesis. This will involve analysing the use and generation of level of detail as well as introducing contemporary models and measures of visual perception.

Chapter 3 (Development) : using the results from Chapter 2, this chapter will develop the base tools and theory which will be required for our model. This includes an appropriate method to generate level of detail, a measure for perceived detail within a computer-generated image, and a model for assessing the degree of detail which a user can perceive.

Chapter 4 (Implementation) : this chapter will describe how the material which has been formulated thus far can be integrated to form a prototype implementation. For example, how the visual metrics can be incorporated efficiently into the graphics system and how each object's level of detail can be calculated from this.

Chapter 5 (Results) : the antepenultimate chapter of the thesis is devoted to a systematic evaluation of the prototype implementation. This involves a suite of empirical studies to assess the merit of a perceptually modulated level of detail system.

Chapter 6 (Discussion) : leading on from the evaluation, we will present an analysis of the results which were obtained and embark upon a discussion of these results; assessing their implications and validity in the wider context.

Chapter 7 (Conclusions) : in the final chapter, we will summarise the result of the research programme, discuss the original contribution to knowledge which has been made, and identify further research activities which have been suggested through the current study.

Due to the cross-disciplinary nature of the thesis content, the initial chapters will necessarily present the background material from the fields of computer graphics and visual psychology as disparate discussions. However as the text develops, this boundary distinction will be eroded until, by Chapter 4, we will see an integrated formulation, where both disciplines merge to form a single, principled solution.

The reader's attention is also drawn to the Glossary of Terms chapter, located immediately after the appendices. This should provide a useful and convenient resource for locating concise definitions of various technical terms used within this dissertation.

The following typographic conventions are adopted throughout the thesis:

Italics is used for figure and table captions, Latin terms, and to emphasise a key word in a sentence.

Slanted Sans Serif is used for any quotations which are replicated verbatim from another source. An attribution will normally be supplied to indicate the source.

Boldface is used in the main text to introduce a new term. Many of these terms can be found in the glossary section towards the end of the thesis.

Boldface Underlined is used for item headings where a list of points is presented.

`Typewriter Font` is used to present any algorithms or sections of C++ source code.

1.6 Summary

Upon reading this exordial chapter, we have gained an awareness of the problems caused by introducing a high degree of lag into a VR system (i.e. reduced user performance, symptoms of motion sickness, and degraded sense of presence). After looking at a number of common techniques to reduce this lag, we have decided to channel our interest towards the notion of level of detail. This is a process whereby a number of different representations for each object are made available to the system, each varying in complexity. The system can then select the most appropriate model to display at any point based upon a particular selection criterion, or criteria.

Following from our examination of the human visual system, we can see that our vision is designed to detect the size and orientation of discontinuities in the retinal image. However, our visual acuity is not uniform under all situations. We have seen that our ability to resolve spatial detail degrades with respect to the retinal eccentricity and angular velocity of a stimulus. This provides us with perceptual justification to base the LOD for an object in a VE upon the degree to which it exists in the user's peripheral field, and the relative velocity with which it is travelling (in addition to the standard approach of distance-based modulation).