MANAGING LEVEL OF DETAIL IN VIRTUAL ENVIRONMENTS: A PERCEPTUAL FRAMEWORK

Abstract
In the companion paper, Watson et al. (1997), we demonstrated the effectiveness of using perceptual criteria to select the amount of detail that is displayed in an immersive virtual reality (VR) system. Based upon this determination, we will now attempt to develop a principled, perceptually oriented framework to automatically select the appropriate level of detail (LOD) for each object in a scene, taking into consideration the limitations of the human visual system. We apply knowledge and theories from the domain of visual perception to the field of VR, thus optimizing the visual information presented to the user based upon solid metrics of human vision. Through a series of contrast grating experiments a user's visual acuity may be assessed in terms of spatial frequency (c/deg) and contrast. The results of these tests can be modeled mathematically using a contrast sensitivity function (CSF). Therefore, we can use the CSF results to estimate how much visual detail the user can perceive in an object at any instant. Then, if we could describe this object in terms of its spatial frequencies, this would enable us to select the lowest LOD available without the user being able to perceive any visual change.

I Introduction
In any virtual reality (VR) system, the complexity of the virtual environment (VE) being
simulated influences the time required to process and display that environment. For example, a VE that is composed of a large number of polygons takes longer to process and render than a VE with fewer polygons. This is not a simple relationship because, for example, the size and geometry of polygons can affect performance; also, the use of texture maps can add visual detail to a VE that would otherwise require a large number of small polygons. However, in general, we can state that the more visual complexity we include in a VE, the greater the lag that will be subsequently induced.

The magnitude of this lag can have very real consequences for the user of a VR system. Visual delays in immersive applications have been reported to cause effects of motion sickness (Regan, 1995), the symptoms of which include nausea, pallor, and cold sweating (Money, 1970; Uliano et al., 1986). Systems that suffer from a noticeable lag can also affect the performance of the user, particularly for coordination and navigational tasks. For example, Gregory (1990) states that a lag of around 500 ms can seriously degrade hand-eye coordination tasks such as drawing and writing. In addition, Barfield and Hendrix (1995) have reported that reduced frame rates (less than 15 Hz) can diminish a user's sense of presence within a VE: the feeling that a user has of being in an environment other than where their body is (Slater and Usoh, 1993).

It is therefore clearly undesirable to produce VEs that are unnecessarily complex. As a result, a number of techniques are commonly employed to reduce the complexity of a VE at any given instant. These include the process of world subdivision, clipping nonvisible objects, using various levels of detail for certain objects (Astheimer and Poche, 1994) and supporting texture-mapped polygons. The technique that this paper will focus upon is level of detail (LOD). This technique involves holding a number of representations of an object, each varying in complexity (e.g., polygon count) and alternating between these during the simulation based upon certain selection criteria—thus allowing the complexity of the VE to be modulated in real-time.

In the first part of this series, Watson et al. (1997) suggested a paradigm for design of LOD management systems and tested a prototype of such a system. Their prototype reduced the amount of detail displayed in the periphery of the user's field of view. For the particular evaluation task employed, testing showed that a normal high-resolution display offered no significant performance advantage over a display with a low-resolution periphery. Although these results demonstrated the potential efficacy and utility of peripheral detail reduction, they did not suggest how the distribution and degree of level of detail loss might most appropriately be controlled.

Therefore, in this second part, we will characterize the human visual system, with the aim of finding a perceptually based method for controlling LOD. In the process, we will identify those elements that might best be exploited by designers of LOD management systems. We will conclude by suggesting a possible LOD management system design.

2 Limits of Vision 2.1 Visual Acuity
The resolution of the human visual system must clearly possess a finite threshold that is ultimately determined by the spacing and pooling of photoreceptors in the retina (the cells that detect incoming light). However, when a user dons a head-mounted display (HMD), the resolution of that device will normally be far lower than that of the individual's biological vision system, and so the resolution of the HMD will define the user's visual acuity while that person is immersed in the VR system. (The often-quoted example is that when using a modern HMD, the user's visual acuity is reduced to around 20/200, i.e., effectively the user can see from 20 feet what a "normal" person can see from 200 feet.) As a demonstration of this point, the eye can detect detail down to a size of about 0.5 min of arc (Humphreys and Bruce, 1991), whereas the angular resolution of a modern LCD-based HMD is in the order of 10 min of arc. (Based upon manufacturers' figures, the Virtual i-O i-glasses! offers an angular resolution of 6.8 min of arc,
the Forte Technologies VFX1: 10.4 min of arc, and the VictorMaxx CyberMaxx: 12.6 min of arc.)

### 2.2 Peripheral Vision

The eye's sensitivity to detail is not uniform across the entire visual field. Instead we find that our visual acuity is highest towards the center of the retina, at a point called the fovea. There are a number of physiological reasons for this difference. For example, the concentration of cone photoreceptors in the retina decreases rapidly with eccentricity (distance from the fovea). Also, the ratio of cortical cells in the brain that are devoted to the foveal region is notably disproportionate with around 80% of all cortical cells dedicated 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).

### 2.3 Motion Sensitivity

The human vision system cannot resolve as much detail in an object that is moving across the retina as it can in an object which is stabilized on the fovea. This fact 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 to the eye's inability to track rapidly moving targets accurately, thus causing a slippage in the retinal image (Murphy, 1978). Recent studies have suggested that this limiting of our temporal frequency sensitivity occurs as early as the photoreceptors in the retina (Nakayama, 1990).

### 3 Formulating the Framework

Based upon the preceding examination, we can see that there are three principal ways in which visual complexity can be optimized in a VE with respect to limitations of the display device and the human visual system. We can therefore describe these in terms of criteria for the LOD selection mechanism:

1. **Size/Distance:** When a complex object is positioned at a distance from the viewpoint, many of its detailed features will be projected onto an area less than the size of one pixel. As such, these features will make little or no contribution to the visual representation of the object at that distance. It is therefore possible to select a lower LOD model when the object exceeds a particular threshold distance from the viewpoint. Doing so will have little impact on the display, but will afford a certain performance advantage. An alternative method for implementing distance LOD is employed by the Open Inventor toolkit and the AVRIL graphics library. This approach bases the LOD selection upon the pixel area occupied by the bounding volume of an object after it is projected into screen coordinates. This is perhaps a more accurate technique because it is the size of a feature in screen space that determines whether it will be visible on the display device; also, it avoids the issue of choosing a single arbitrary point within the object's 3D volume to use for the distance calculation.

2. **Eccentricity:** As described previously, our visual system is limited to detailed resolution in only a very small region of the visual field (around 5 deg of arc). The field of view (FOV) of contemporary HMDs range from around 30 degrees to 120 degrees of horizontal arc (although most low-end HMD units offer a FOV of less than 60 deg of arc). There is, therefore, substantial opportunity for a highly detailed object to be present within the user's peripheral field. When this is the case, the object's visual complexity may be wasted because the eye cannot resolve features in the periphery to the same degree as in the fovea. In addition to the study described in the first part of this series, a number of graphics systems have been developed to take advantage of this phenomenon by degrading the detail of features in the user's peripheral field (e.g., Hurault, 1993; Levoy and Whitaker, 1990).
3. Motion: We have seen that an object moving rapidly across the retina cannot be perceived in as much detail as a stationary object. It would therefore be sensible to attempt to reduce an object's LOD in proportion to its motion relative to the user. Considering that an immersive VE is an inherently motion-rich experience (e.g., users' head movements generate motion flows as do independently moving objects within the VE), then we would expect that a motion-sensitive LOD mechanism will favorably affect the performance and interactivity of the system. Funkhouser and Sequin (1993) incorporated support for this feature into their architectural walk-through system; as did Hitchner and McGreevy (1993) in the Virtual Planetary Exploration (VPE) Testbed.

4. Modeling a User's Visual Acuity

Of the three criteria presented above, distance modulation is the most prolific LOD mechanism in use today (although surprisingly most VR toolkits do not explicitly support this criterion, and often leave the task of distance LOD selection to the application program). However, one problem commonly associated with LOD techniques is the often detectable flicker (or "popping") when different models are swapped in. This problem arises because the selection thresholds are normally arbitrary in nature and not based on any models of visual perception. If the user is not to be distracted by the changes in visual complexity, then the LOD selection thresholds must be based on a solid metric that accurately models what the user can and cannot see.

The remainder of this paper is, therefore, concerned with introducing such a metric from the domain of visual perception; investigating how this can be applied to an immersive VR system; and analyzing the pertinent issues that are involved in the implementation of this process.

4.1 A Metric for Visual Detail

Since the initial work of Schade (1956), the most common means of accurately measuring a subject's visual acuity is through the use of a pattern known as a contrast grating. This is simply a pattern where contrast is varied sinusoidally across the display, producing a series of alternating light and dark vertical bars [see Figure I(a)]. The spacing between bars is measured by a quantity called spatial frequency, defined in units of contrast cycles per degree of visual field (c/deg). For example, a high spatial frequency implies a short distance between adjacent bars and hence represents a stimulus of high detail.

The visibility of a grating is dependent upon its spatial frequency and its contrast (luminance difference between adjacent bars). A curve known as a contrast sensitivity function (CSF) can be plotted to record a subject's ability to resolve a grating based upon these two factors. For example, the CSF in Figure 1(b) states that for a contrast grating moving at 3 deg/s, the subject will be unable to resolve any spatial frequencies greater than about 8 c/deg.

The effect of increased velocity and/or eccentricity is to shift the CSF closer towards the y axis, thus reducing the threshold of visual acuity in both cases. In other words, a subject will be able to resolve fewer high spatial frequencies (regions of high detail) as the stimulus moves faster across the retina (Kelly, 1979) or is presented further towards the periphery (Virsu and Rovamo, 1979).

Spatial frequency has been used to analyze and describe the limitations of the human visual system for almost three decades. If this metric can be applied to the domain of real-time computer graphics, then we would have a wealth of literature at our disposal to determine the perceptual content of a computer display and the visibility of any region of detail therein.

5 Applying the Visual Metric

5.1 When and Where to Calculate Spatial Frequency

Spatial frequency is a measure of the detail that is presented to the visual system--i.e., in a computer graphics system it is a measure of the detail presented on the display device. In order
to accurately gauge the perceptual content of a VE? we must therefore apply our analysis to the rendered image that is transmitted to the HMD. We cannot accurately predict what will be displayed by simply looking at the geometry of the scene because the geometry can be displayed differently depending upon the shading model being applied, the level of lighting that is in effect, the use of texturing and atmospheric effects, and so on. We must therefore base our analysis of the scene upon the rendered image, not the geometry of the objects within the scene.

There are however a number of complications. If we take the isolated case of one object, then the spatial frequency content of that object will change as it rotates or moves away from the viewpoint, i.e., spatial frequency is viewpoint-dependent. Therefore, to accurately calculate the spatial frequency content of that object we must apply our analysis to the display every time the object or the viewer moves. This is obviously unacceptable in a time-critical application such as VR. However, more importantly, this is impractical because we are required to know the spatial frequency profile before the image is displayed--once the image has been rendered we will have already expended the computational resources that we wish to conserve. Therefore, we must endeavor to precalculate the spatial frequencies in an object. This calculation must be done from several viewpoints around the object in order to capture all of the object's features. We can then interpolate these values during the simulation in order to predict the spatial frequency content of any arbitrarily positioned object in real-time and subsequently select the most suitable LOD to utilize.

5.2 Calculating the Spatial Frequencies in a Rendered Image

Given a snapshot of an object from a particular viewpoint, we want to be able to find all of the relevant spatial frequencies (c/deg) that are component in that image. Specifically, we wish to locate each visual feature in an object and calculate its fundamental (lowest) frequency at a number of orientations (where a feature is loosely defined here as a region of similar color). A feature's fundamental frequency will be inversely related to its cross-sectional length at each orientation. A method for extracting all of the relevant spatial frequencies from a computer-generated image in this way is detailed in Reddy (1996).

Briefly, from the rendered snapshot of an object we can calculate the values for each feature's fundamental spatial frequency in units of cycles per pixel (c/pixel)--essentially a measure of how many pixels a feature extends over at each orientation. Then, once we know the FOV for the screen or HMD, we know how much of the visual field is subtended by a single pixel. We can therefore apply a suitable scaling factor to the c/pixel values to gain results in units of c/deg.

Consequently, for each object, we can find all of the spatial frequencies (and their orientations) at any instant. The orientation of the spatial frequencies is of little importance when considering a stationary object (except perhaps to compensate for the aspect ratio of the display device). However, because we wish to include motion optimizations in our framework, we should record the orientation of spatial frequencies because the visibility of a moving spatial frequency is dependent upon its alignment with the direction of motion.

5.3 Example Spatial Frequency Analysis

In order to illustrate this concept, Figure 2 presents three different levels of detail for a model of a die. Next to each LOD is a graph of the relevant spatial frequencies in that image. The x axis represents increasing spatial frequency (c/pixel), i.e., regions of high detail are coded to the right of the graph. The y axis denotes the number of features in the image with a particular spatial frequency.

From these results we can note that there appear to be three major groupings of spatial frequency. In the case of the lowest LOD, Figure 2(c), we see that there are only three low spatial frequencies. These will obviously represent the three facets of the die that are the only
visible features at that level. In the medium LOD, Figure 2(b), we can see that we still have these three low spatial frequencies, but we now also have a batch of higher spatial frequencies. These will represent the spots on the die that have been introduced at that level. Finally, in the highest LOD, Figure 2(a), we can see the same trends as in the other two instances, but we also have some even higher spatial frequencies. These will represent the added detail induced by the curvature of the die at its edges.

From this analysis we can see that we have a high LOD that contains a number of high spatial frequencies and that for each lower LOD we have fewer of these high spatial frequencies (less high detail). Therefore, if this die was presented to a user in such a situation where they could only perceive frequencies below 0.01 c/pixel, then we could select the lowest LOD model-Figure 2(c)—and the user would not be able to perceive any change.

There are a number of implementation questions that are raised by this process, and we will consider some of these presently. We do not propose to offer definitive answers here, merely to highlight the relevant issues and provide tentative suggestions towards possible solutions.

**Size of Test Snapshot:** How large should the object be for the test analysis? Obviously, if we display the object too small, then we will lose some fine detail. We therefore suggest that the frequency analysis be performed on an object that fully occupies the display resolution. This assumes that when an object is positioned to fill the screen, it will be rendered in full detail. This also means that if a number of different resolutions are to be supported, then the analysis should be performed at the highest resolution and scaled appropriately, or reiterated for each resolution.

**Factoring-in Object Size:** How do we account for objects that are displayed smaller than the test case (i.e., support distance/size LOD)? If we reduce the size of an object in screen space, then we can obviously see fewer frequencies. However, we can predict the highest visible frequency in an object of any size smaller than the test case by simply applying a scaling factor based upon the relative size of the displayed object to the test case. For example, if an object is displayed half the size of the test condition, then we will not be able to see frequencies of 0.5 c/pixel (i.e., 1 pixel detail).

**Orientation Sampling:** How many spatial frequency orientations should we sample for any feature? This will obviously depend upon the size of the particular feature being analyzed (there would be no point in analyzing 360 different orientations for a feature that is only one pixel large). We therefore envisage a relation of the form $a = \tan^{-1}(k/r)$, where $a$ is the orientation increment in radians, given; $r$, is the radius of the feature in pixels, and $k$ is some suitable scaling increment.

**Number of Snapshots:** How many snapshots should we take around an object? At the very least it would seem intuitive to take snapshots aligned with each face of the object's bounding box (i.e., six snapshots). However, some complex objects may require a greater fidelity to accurately encapsulate the spatial frequency content of the 3D shape. This relationship requires further investigation.

**6 Implementation**

Once we have a method for estimating the perceptual content of a display in terms of spatial frequency, we can then begin to look at integrating this into the LOD selection criteria. Obviously, for this integration to be effective, there must be some correlation between spatial frequency and LOD. Specifically, we might expect that because a lower LOD model is more coarse and has less detailed features in it, it should contain fewer high spatial frequencies. This correlation will generally be true, and we can observe this relationship in Figure 2. Based upon
the preceding discourse, an implementation for a perceptually based LOD system could be formulated as follows:

**Offline:** Before the simulation of the VE is initiated, each LOD for every degradable object is analyzed to discover its spatial frequency profile. This involves taking a number of snapshots of an object from different viewpoints. For each snapshot, all of the features within that image are located and their size calculated in units of c/pixel for a number of orientations. These values can then be scaled to units of c/degree based upon the FOV of the HMD being used.

**Online:** During the simulation, the LOD Scheduler will analyze the distance, eccentricity, and velocity of each degradable object (in units of m, deg, and deg/s, respectively). By applying the results from the various CSF curves it becomes possible to calculate the highest spatial frequency that the average person will be able to resolve in this situation. Then with the results from the offline process, we can estimate the spatial frequencies that will be contained in each LOD if it were displayed. This will enable us to select the most appropriate LOD to use at that instant. For example, if two different LOD models are expected to be perceived equally (i.e., they have essentially the same spatial frequency profile below the visibility threshold), then the less complex model will be chosen.

Figure 3 presents the standard sense-process-display loop of a typical VR application and illustrates how the LOD Scheduler fits into this scheme. The position of the Scheduler within the process stage is important in order to gain the optimal performance. It must follow any processing that updates the location and velocity of each object (including the viewpoint), but it should precede any subsequent processing (e.g., collision detection, dynamics equations, and so forth) so that this will be applied to the actual model about to be displayed.

**7 Conclusions**

In this paper we have looked at a paradigm for reducing the visual complexity of objects in a VE based upon various characteristics of the human visual system. This has been done with the expressed aim of improving the performance and interactivity of the VR system while using proven metrics and principled selection criteria.

There are two points that must be appreciated, however. First, the results for the CSFs are for a "standard observer." The visual acuity of any particular user could therefore differ slightly from this standard. This variation could be compensated for, and it is unlikely that it would pose a major problem. More importantly, it should be highlighted that any system that attempts to make judgments on the perceptual content of a display must, to be completely accurate, track the user's gaze. This is because the user's perception is based upon the image formed on the retina, which at any instant could be focused on any region of the display device. It would be acceptable, though less accurate, to assume that the user will always be looking towards the center of the display. This was the assumption made by Funkhouser and Sequin (1993) and also by Hurault (1993).

In this respect, it is the authors' opinion that the above approximation can be more strongly advocated in an immersive VR system than in a desktop VR system. The reason is that whenever the user evokes a large change in their point of fixation, there will normally be an associated head movement. As a result, the resting gaze will generally be quite closely related to the user's head orientation. Thus the differential between a user's point of fixation and the center of the display will be less pronounced in an immersive, head-tracked system. This point is also made by Hitchner and McGreevy (1993). There is some perceptual evidence to support this claim. For example, Bahill et al. (1975) report that most individuals will make a combined head and eye movement to focus on objects that are outside the central 10-15 degrees of their fixation field.
Consequently, if an appropriate eye-tracking technology is not employed, then one should be aware that the user may be able to perceive slight inconsistencies in the rendered image. This could, of course, also be compensated for by slackening the selection thresholds somewhat, or introducing a degree of hysteresis (a lag in the transition between different LODs). However, it must be borne in mind that most researchers in the VR field agree that interactive update rates are more important than display fidelity, and so some visual incongruities can be tolerated. This is aptly expressed by Wloka (1993) when he states that "Presentation quality is expendable: to be wrong and on time is more valuable than to be right and late." These same sentiments are encapsulated by Krueger when he makes the statement: "Geometry is not reality. Interactivity is reality" (Garassini, 1991).

In conclusion, the product of this series of two papers has been an evaluation of the effect of perceptually based detail degradation on the performance of a user. The results from this study were encouraging, and so subsequently we have embarked upon a general and principled framework to manage the selection of LOD based upon a model of human visual acuity. This provides a mechanism to automatically select the lowest LOD at any juncture without affecting the user’s perception of the scene and hence to optimize the performance of a VR system.

GRAPHS: Figure 1a. A sample contrast grating. The curve below the grating illustrates the sinusoidal variance of contrast across the image. If this grating was positioned to occupy 1 deg of visual arc, then it would have a spatial frequency of 4 c/deg I b. A contrast sensitivity curve illustrating detection thresholds for contrast gratings moving at 3 deg/s. All gratings represented by the region below the curve are considered to be detectable. Adapted from Kelly (1979).

GRAPHS: Figure 2. An example spatial frequency analysis of three different LOD models. In each graph, the x axis represents increasing spatial frequency (c/pixel).

DIAGRAM: Figure 3. Diagram illustrating how the proposed LOD Scheduler (right-hand flow diagram) fits into the typical main loop of an immersive VR system (left-hand flow diagram).

**References**


~~~~~~~~~