

*IN 1000*  
*7/10/95*  
*1000*

**PERCEPTION-BASED TECHNIQUES FOR  
IMPROVING COMPUTER-GENERATED IMAGERY**

**ANNUAL STATUS REPORT**

**NASA AMES COOPERATIVE AGREEMENT NCC2-925**

**Dennis R. Proffitt  
Randy Pausch  
University of Virginia**

**Period of Performance: Year 1 (4/1/95 - 3/31/96)**

**Technical Officer: Mary K. Kaiser, NASA Ames Research Center**

## **PERCEPTION-BASED TECHNIQUES FOR IMPROVING COMPUTER-GENERATED IMAGERY**

Visual simulation plays a critical role in human-machine interfaces and other space systems. Advanced computer generated imagery (CGI) systems are used to create compelling visual displays for navigation/control systems, vehicle/system simulation, telerobotics, and scientific visualization applications. Inevitably, the realism of these displays is constrained by limitations in CGI hardware and software, especially if images need to be generated in real-time. Despite rapid advanced in image generation technology, human operators desire more realistic, higher fidelity displays; it is likely that such a demand for improved fidelity will continue for the foreseeable future (Padmos & Milders, 1992).

The current research program is focussed upon the development and evaluation of techniques that reduced the computational resources required to achieve specific levels of graphical fidelity in CGI systems. These techniques are based upon perceptual principles and allow a given level of apparent realism to be achieved at a reduced computational cost. The basic idea underlying all of these techniques is the following: **Since the human perceptual system neither uses nor requires all of the optical information available in a scene when forming spatial perceptions, efficiency gains can be realized in CGI systems by requiring that they render only information that is of perceptual utility.**

In our proposal, there were three primary areas of research. The first area examines the feasibility of exploiting aspects of visual fusion processes in order to increase the apparent resolution of images at a lower computational load than is required by current techniques. The second area examines the use of perception-based techniques to automate the modulation of level of detail in time-critical rendering. The third area focuses on developing techniques to streamline efficiency of animation systems by exploiting the fact that the perceptual system uses only a subset of the available optical information when deriving 3-dimensional

(3D) form from motion.

A fourth area of research developed over this first year of funding. Generalizing the principles of visual fusion to audition, we developed a technique for decreasing the computational load required to generate and transmit stereo auditory information.

### Enhanced Resolution through Visual Spatio-Temporal Fusion

The techniques examined in this program make use of the following principle of perceptual processing: **The fusion of two images can result in an apparent level of detail that is greater than what is actually presented in both image.** Using this principle, we developed and evaluated efficient techniques for creating an increase in apparent resolution at reduced computational costs in both stereo and non-stereo renderings of static and animated scenes.

#### Stereo Fusion

It currently takes twice the computational resources to present or store stereo images as opposed to single image displays. A pair of stereo images displays a scene from two different perspectives corresponding to the displaced viewing positions of each eye. The human visual system exploits the difference in the two images to recover depth.

We have developed a technique whereby stereo displays can be created or stored with a minimal increase in the computational resources required for single image displays. Called **hi-lo stereo fusion**, these techniques present a fully rendered image of the scene to one eye and a reduced resolution rendering of the scene to the other eye. When the two images are fused, depth is recovered from the stereo disparities available in the two images, and the details from the high resolution image are fused into the percept such that the loss of resolution in the second image is not apparent.

Differential resolution displays can be of two sorts. First, both displays can present the same surface boundaries; however, only one display presents surface texture. Such a pair of displays is presented in Figure 1. When these two images are fused in stereo viewing conditions, a

compelling depth impression of a three-dimensional solid is observed based upon the boundary information available in the two images. Moreover, the surfaces of the apparent object manifest the textures that are provided by only one of the images.

The second application of differential resolution displays is depicted in Figure 2. To the left is a complex high-resolution image and to the right is a stereo-appropriate rendering of the same object but at a lower resolution. The image to the right consists of 1/16th the number of polygons as that on the left. In computer graphics, resolution is defined by the size of the smallest polygons used to render the contrasts in the image. The higher the resolution, the greater is the number of polygons need to render the image, and thus, the greater is the computational resources needed to generate or store the image. Compared to the high-resolution image to the left, the image to the right has only about 6% the number of polygons, and thus, requires only about 6% of the computational resources. When the two images are fused, a three-dimensional object is seen based upon the low resolution information; moreover the percept has a high-resolution appearance. That is, the high-resolution details are fused onto the three-dimensional perception formed on the basis of the low-resolution information.

There are essentially two reasons why differential resolution stereo displays evoke high-resolution stereo depth percepts. First, the visual processes responsible for stereo-depth vision are driven primarily by low spatial frequency information corresponding to the low resolution components in both images (Tyler, 1983). Second, binocular rivalry is not evoked by differences in the high-resolution information (Liu, Tyler, & Schor, 1992). Instead, the high-resolution components of one image fuse onto the stereo-depth percept derived from the low-resolution components available in both images.

A number of hi-lo stereo fusion displays have been created and we are beginning to assess their effectiveness with user studies. Gossweiler (1995) looked first at an extreme application in which a sphere was presented at different resolutions to the two eyes. In this case, the extreme difference in level of detail resulted in a perception of translucency in which both the smooth edges of the sphere and the jagged

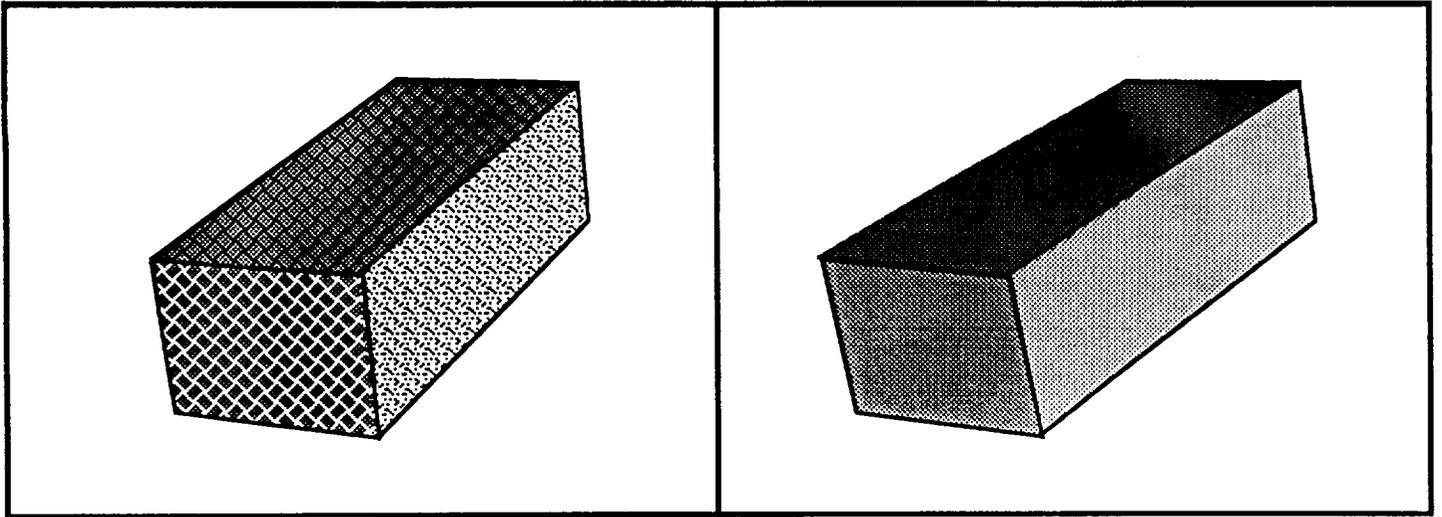


Figure 1

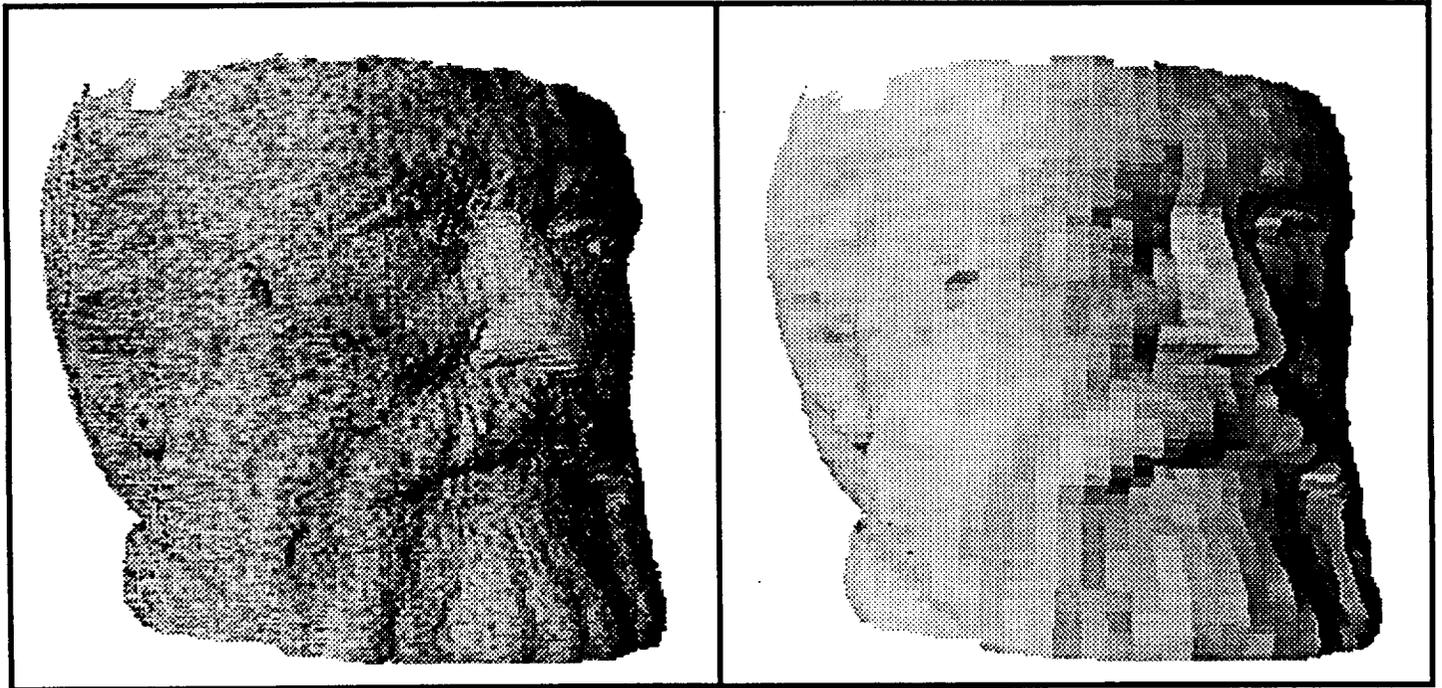


Figure 2

## Differential Resolution Stereograms

edges of the low resolution image were seen overlaid upon each other.

Since the facets of the object's external boundary were interfering with fusion, a rendering option was developed in which the lighting effects in one eye were turned off. This produced flat shaded coloring effects in one eye, and Gouraud shading in the other eye. An analysis was made of the impact of this technique on rendering speed. It was found that rendering speed was significantly reduced. As an example, for a 901-polygon object, by turning off lighting in one eye, the rendering speed dropped to be the same as if the object had been reduced to a 501-polygon object.

An informal user study was conducted to assess how distracting the effect was to naive observers. Seven observers wore a Virtual Research Flight Helmet (a binocular Head Mounted Display) and viewed a simulated room containing several items and textures. Observers were asked whether they could notice any rendering reduction techniques and they all reported that the scene appeared normal. They were then asked to first close one eye and then the other, thereby enabling them to see the differential resolution that was being presented to their two eyes. They all noticed the difference. They were then asked to view the scene with both eyes open and asked whether they could notice the differential resolution that they now knew to be present. All seven stated that they could not notice the reduced lighting when both eyes were open. This finding is remarkable, given that in the far periphery of one of the eyes, there was no lighting and the other eye could not provide fusion. Observers were asked to slowly turn their heads and look for the "sweeping" effect on the textures. When directed to look for the effect, all seven observers could identify the effect in the periphery.

We were concerned about the possibility that hi-lo stereo displays might tax the user's visual system with prolonged exposure. We conducted a user study that assessed the effect of viewing hi-lo stereo images on people's ability to fuse stereo images. Subjects were taught how to play a computer game in which they attempted to aim and shoot a simulated canon at targets displaced in depth. Stereo depth was made possible by alternating stereo-appropriate images at 120 Hz on an SGI Indigo 2 computer viewed with CrystalEyes stereo-shutter glasses. The task was

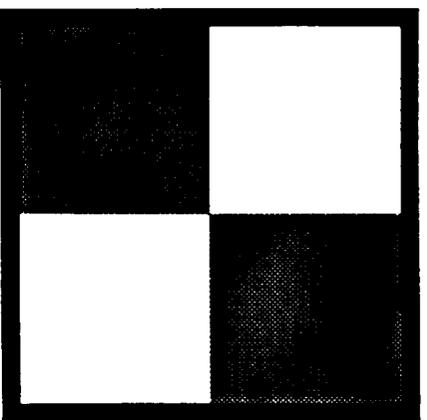
relatively easy when viewed in stereo, but quite difficult when viewed biocularly. Number of hits was recorded over a ten minute task period. There were three between-subjects conditions defined by whether viewing was biocular, hi-lo stereo, or normal hi-hi stereo. Prior to and after the task, subjects were assessed on their ability to fuse random-dot stereo images. Within each image pair, a number was presented that could be read only after the images had been fused. The subjects' task was to report each number after which a new stereogram was presented. The number of correct reports were recorded for a one minute test period.

As expected, it was found that the number of target hits was far greater for both stereo conditions relative to the biocular one. There was a trend toward slightly better performance in the hi-hi condition relative to the hi-lo one; however, all of the data has not been collected. Surprisingly, subjects' performance on the random-dot stereo fusion task was improved in the hi-lo viewing condition relative to the hi-hi one. We are currently looking at the relevant literature in order to determine if there is a precedent for this unanticipated finding. For our current purposes, however, the primary finding indicates that h-lo stereo viewing does not impair subsequent visual processing of normal stereo information.

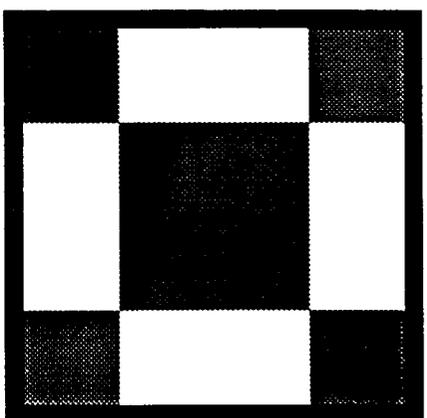
### Static Images

Modifications of the above technique can also be used to create non-stereo images. One method involves **temporal fusion** in which high and low resolution images are alternated in time. We have found that if two images, such those depicted in Figure 2 are alternated at frame-rates of 60 Hz or greater, then the apparent resolution that is observed is that of the high-resolution image. This finding led to a second technique, termed **resolution phase-shifting**.

This technique requires that two images of a scene be created that have the same resolution; however, the sampling of the resolution for one image is shifted by 1/4 of a cycle. The top panel of Figure 3 shows an apertured view of two images that have the same sized polygons, although in the second image their position has been shifted 1/4 cycle in the vertical and horizontal direction. When these two images are superimposed, the resulting resolution is 4 times greater than that of the images of which it is comprised. The bottom panel of Figure 3 shows this application for a complex object. If each of the low resolution images has n



+



=

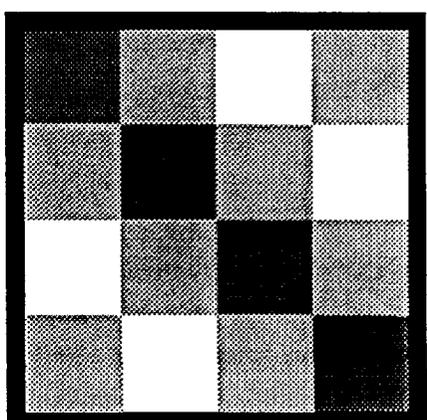


Figure 3

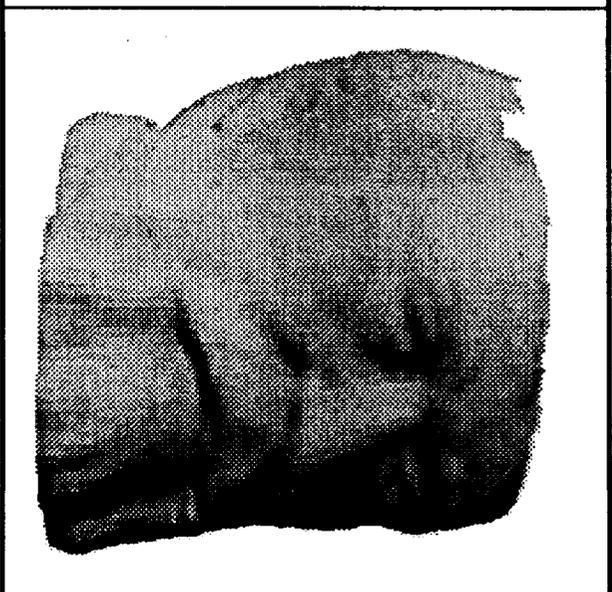
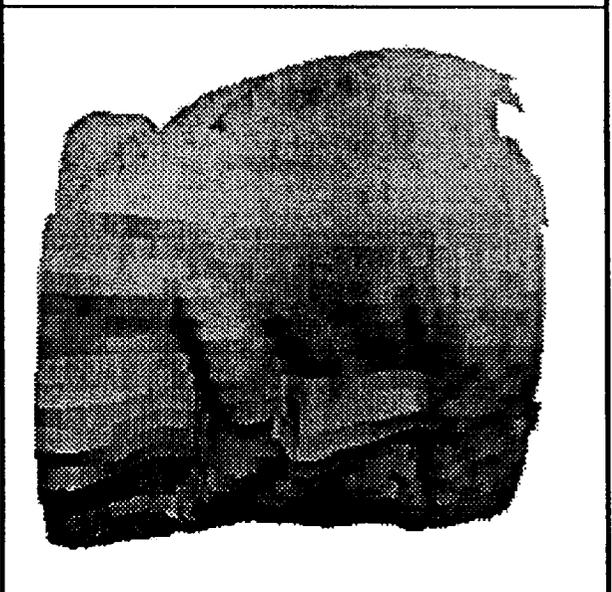
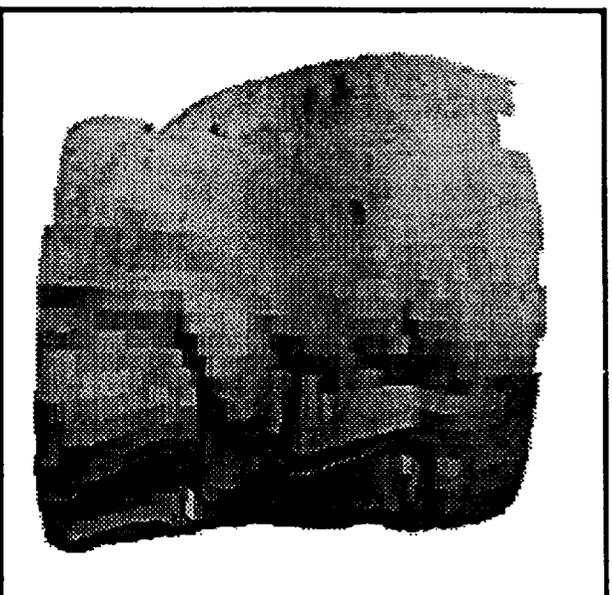


Figure 4

# Resolution Phase-Shifting

polygons, then the resulting image has  $4n$  polygons.

Resolution phase-shifting can be produced in two ways. At frame rates of 60 Hz or greater, phase shifted images can be alternated. The resulting apparent resolution is produced at a computational cost that is 1/4 of what would be required to create a single image of the same resolution. At lower frame rates, flicker artifacts are apparent using this technique; however, phase shifting can still be employed to good effect by rendering each frame with two superimposed phase-shifted versions of the scene. This is shown in Figure 4. In this application, the luminance in the scene is divided between two phase shifted images that are superimposed on each other. This technique yields an apparent resolution at 1/2 the computational cost.

A third technique in this set is the Gossweiler-Proffitt Image Reduction (**GPIR**) technique which reduces the amount of information in an image by segregating the image into two different component images; the first component represents an intensity map, and the second represents color. The intensity component image represents the greyscale or luminance values, and may be reduced along the dimension of the number of bits used to represent the intensity. At the extreme end, one bit of information can be used to represent whether the pixel is black or white. The color component image maintains a very low resolution representation of the color in an image. This is produced by taking the original image and sub-sampling it -- for example scaling the original image down by a factor of ten (e.g. storing only every tenth pixel).

When the image is presented, the color-component is re-scaled to the original size, resulting in a very blocky image. This is alpha-blended with the intensity component to create the final, composite image. Alpha-blending is a well-known process where the amount of color from one image is combined with the color of the second image. Numerical weighting is used to control how much of the first image's color is combined with the second. For example, if we represent color as an R,G,B triple [e.g. (255,0,0) is all red and (145,200,30) is another color], then we can combine two color images using a 90-10 combination with the formula:

$$\begin{array}{l} \text{New Color} \quad \text{color 1} \quad \text{color 2} \\ (\text{R,G,B}) = 0.90 \times (\text{R,G,B}) + 0.10 \times (\text{R,G,B}) \end{array}$$

To combine color and intensity, the intensity value is converted to a color by repeating it in each of the red, green and blue values. For example, an intensity component of 40 is (40,40,40).

This technique has a savings over a raw image of approximately 19.35 times. That is, if the original image is a 1,000 x 1,000, 24-bit color image, it would consume 3,000,000 bytes of memory. Using the black-and-white component (1,000 x 1,000 x 1 or 125,000 bytes) and the tenth-scaled color image (100 x 100 x 24, or 30,000 bytes) the GPIR image consumes only 155,000 bytes.

The reason that GPIR results in an image having very little apparent loss in image quality is that the technique takes advantage of the differential processing of luminance and color by the human perceptual system. The visual system extracts edges from an image by using only luminance contrasts. It extracts color information separately and then locates these colors within the edges that were defined by luminance. Thus, the colors appear to adhere to the luminance-defined edges, not to the low resolution edges that were rendered in the color component.

### Animations

The temporal fusion and resolution phase-shifting techniques described above for static images can also be applied to moving ones. In addition, the phase-shifting technique can be modified in animations to yield, not only a 4-fold computational savings, but also effective temporal anti-aliasing.

The essence of this technique involves creating each frame from two phase-shifted images. On each subsequent frame, only one of the two phase-shifted images is updated, and alternation occurs in image update across the sequence of the animation. This lag in one of the images used to render each frame produces a slight blur that corresponds quite well with naturally occurring motion blur. This technique produces temporal anti-aliasing at 1/4 the computational cost that would be required to produce an aliased sequence of comparable resolution.

### Emergence of Detail in Time-Critical Rendering

When a scene has too much detail to be rendered in a real-time CGI application, a common practice is to reduce the detail within the image in order to increase frame-rate to a desired value. Within an interactive CGI application, the amount of detail that can be rendered at any time depends not only on scene complexity but also on how much the scene is changing; thus, modulations in detail must occur in real-time. The practice of modulating level of detail in accord with the computational load on the system is termed time-critical rendering.

Previous systems using time-critical rendering techniques relied on application-specific information to improve frame-rates. This restricts the domain of applications for which this technique is applicable. The developer must determine whether the given technique is appropriate, and then structure the application to support the required data-constructs. This must be performed manually for each new application (and often, each new database).

Gossweiler (1995) developed a rendering system which performs degradation automatically, during run-time, as part of the rendering process. This system transparently separates the application semantics from the rendering process. The application-independent rendering engine uses a time-driven rendering scheduler which employs a combination of different perception-based degradation techniques. Perception-based degradation mechanisms are used because they are based on characteristics of the human, not on characteristics of the application. Since the human operator characteristics are constant across all interactive applications, this rendering system is application-independent.

Perception-based rendering techniques capitalize on the capabilities and limitations of the human visual system. For example, consider a scene with several objects placed at different distances from the user. These distances change during run-time based on the non-deterministic behavior of the user. Objects which are too distant to be perceived may be removed. From the user's standpoint, this image-degradation is not apparent, even though from an image processing standpoint pixels may

have changed. On the other hand, the reduction in detail at nearer distances within the scene will cause the user to perceive a noticeable degradation. Perception-based degradation techniques capitalize on the human-centered criteria for image fidelity to achieve increased rendering speed. The following are properties of the human visual system which may be exploited when performing degradation:

Field-Of-View -- within this space, objects may be rendered at different resolutions matched to visual acuity. If an object is near the center (foveal FOV), where visual acuity is higher, then it should be rendered with higher levels of detail than objects in the periphery.

Distance -- as objects are more distant, their projections becomes smaller, and there is a reduction in the level of detail that can be perceived. Level of detail can be modulated with distance.

Motion -- if an object is in motion, then the object will appear blurred. Moving objects can thus be rendered with reduced detail.

Attention factor -- also termed tunnel vision, the effective field-of-view narrows when the user is extremely attentive to a specific task or object. Peripheral objects can be rendered with reduced detail as a function of how long the user's input devices maintain the central FOV within a limited area in the scene.

Within the rendering engine, a rendering scheduler adaptively executes time-critical rendering algorithms based on the frame-rate and scene-complexity. When degradation in scene complexity is required to maintain some minimal frame-rate, then the loss in detail is determined by perception-based criteria. Since these decisions are made at run-time, and since a non-deterministic user may abruptly change the complexity of the scene through the manipulation of an input device, the scheduler must be reactive, seeking feasible, rather than optimal solutions. This differs from other time-critical rendering techniques which pre-process the 3D model and construct data-structures to be used at run-time (Airey, Rohlf, & Brooks, 1990; Funkhouser, Sequin & Teller, 1992; Yan, 1985). The run-time dynamics allow the rendering scheduler to orchestrate different degradation mechanisms, degrading for overloaded scenes, and for only a

few objects, rather than for all objects over the entire period of the application.

Montegut (1995) investigated the visual system's tolerance for various methods of continuous detail modulation (e.g., form-morphing, fade level of detail, spatial frequency modulation) which might better emulate veridical emergence of detail without overly compromising CGI system load. While reducing level of detail decreases computational load, it can introduce an annoying visual artifact at transition points; details can appear to "pop" in and out. Following preliminary psychophysical studies demonstrating that emergence of visual detail is a continuous perceptual process, several methods of modulating detail were examined. These included: morphing, fade level of detail, deblurring, and a combination of fade level of detail and morphing. Results indicated that the morphing technique of level of detail modulation most closely emulated the natural, continuous emergence of level of detail.

#### Simulating Object Rotations and Motion Parallax

The perception of depth from motion information occurs in two situations. First, depth is perceived when one views an object that is rotating around some axis other than the line of sight. Second, depth is seen when a viewer moves past stationary objects, or conversely when objects translate by a stationary observer on some path other than the line of sight. The depth evoking optical transformations that occur in this latter situation are termed motion parallax.

Within a range of object rotations and motion parallax, our perceptions of depth are based upon only a subset of the information available (Proffitt, Rock, Hecht, & Schubert, 1992; Caudek & Proffitt, 1993). Interestingly, the information that is extracted provides only an affine specification of depth, meaning that ordinal depth is specified but not metric depth. It has been found that the visual system scales the affine structure extracted with inherent biases that relate to 2D aspects of the scene.

These findings have clear implications for CGI application. For example, Kaiser and Proffitt (1992) showed how topographic contour maps

and air traffic control displays can be animated so as to evoke accurate depth perceptions at a reduced computational cost. These displays were animated using simple 2D motion algorithms instead of the far more complex 3D algorithms that would be required to render the slight depth rotations that were apparent in these displays.

We extended these techniques to the domain of visual flight simulators. Techniques already exist whereby 3D terrain objects are rendered as 2D representations, "billboards", that are kept normal to the line of sight. By applying simple 2D transformations to features on these billboards, far more realistic and veridical depth impressions can be evoked. Kaiser, Montegut, and Proffitt (in press) demonstrated the effectiveness of this technique. In a series of studies, we examined the perceptual efficacy of billboarding. Rendering objects in this manner greatly reduces their computational complexity, but critical object properties are lost. Specifically, the object always reveals the same side (or face) to the observer. Thus, the natural revealing of detail and observer-relative rotation is absent in such displays. Our experiments demonstrated that the absence of these rotations is fairly difficult for observers to notice (thresholds for translational anomalies were an order of magnitude lower). Further, we found the billboarding technique to be unobtrusive for several classes of object structure. Heretofore, graphics programmer had assumed the techniques would only be effective with radially-symmetric objects.

### Psychoacoustically Compressed Sound

Hi-lo stereo images exploit the fact that the human visual system fuses stereo images on the basis of the low spatial frequency content of each image. In a similar vein, it is known that the auditory system fuses binaural information on the basis of low frequency information. This led us to attempt to develop a sound compression technique that exploits this property of audition. This technique was developed in collaboration with Michael Kubovy (Faculty) and Steve Boker (Graduate Student) who are experts in auditory processing at the University of Virginia.

Called Psychoacoustically Compressed Sound (PACSound), this technique presents stereo-appropriate auditory information in both channels only for the low frequency content of the sound. The high

frequency information is identical in both channels. Since most of the costs of generating and transmitting sound are attributable to high frequencies, this technique results in a substantial efficiency gain.

Perceptually, PACSound is almost indistinguishable from full stereo. This is because the perceptual localization of sound is driven by low frequency. Moreover, high frequency sounds are perceived to be located in the vicinity of the low frequency sounds with which they share harmonic structure.

## Accomplishments

### Patent Application

Proffitt, D.R. and Kaiser, M.K. "Spatial-Temporal Resolution Process for Computationally Efficient Displays," (NASA Case No. ARC 12080-1). Application Date 6/28/95. (Dennis R. Proffitt and Mary K. Kaiser, Inventors)]

### Disclosure Statements Filed

Gossweiler, R. and Proffitt, D.R. "Gossweiler-Proffitt Image Reduction (GPIR), Submitted to University of Virginia administrators, 5/17/95.

Kubovy, M.K., Boker, S., & Proffitt, D.R. " Psychoacoustically Compressed Sound (PACSound), Filed with the University of Virginia Patent Foundation, 9/15/95.

### Conference Presentations

Kaiser, M.K. and Proffitt, D.R. (April, 1995). Participated in a panel on human factors issues in virtual reality. ACM 3-D Graphics Conference.

Proffitt, D.R. (August, 1995). Human Factors in Virtual Reality Development. Tutorial given at the annual meeting of SIGGRAPH.

Proffitt, D.R. (October, 1995). Invited attendee at ARPA Workshop on Human System Integration.

### Doctoral Dissertations

Gossweiler, R. (August, 1995). Perception-based Time Critical Rendering, University of Virginia.

Montegut, M. J. (1995). The emergence of visual detail, University of California, Santa Cruz.

Publications

Kaiser, M. K., Montegut, M. J., & Proffitt, D. R. (1995). Rotational and translational components of motion parallax: Observers' sensitivity and implications for 3-D computer graphics. Journal of Experimental Psychology: Applied, 1, 321-331.

Durgin, F.H., & Proffitt, D.R. (in press), Visual learning in the perception of texture: Simple and contingent aftereffects of texture density. Spatial vision.

Proffitt, D.R., Bhalla, M., Gossweiler, R., & Midgett, J. (1995). Perceiving geographical slant. Psychonomic Bulletin & Review, 2, 409-428.

## References

- Airey, J. M., Rohlf, J.H., & Brooks, F., Jr. (1990). Towards image realism with interactive update rates in complex virtual building environments, Computer Graphics, 24(2), 41-50.
- Caudek, C. & Proffitt, D.R. (1992). Depth perception in motion parallax and stereokinesis. Journal of Experimental Psychology: Human Perception and Performance, 19, 32-47.
- Funkhouser, T.A., Sequin, C.H., & Teller, S.J., Teller (April, 1992). Management of large amounts of data in interactive building walkthrough, Symposium on Interactive 3D Graphics, ACM SIGGRAPH, Cambridge Mass, 11-20.
- Gossweiler, R. (August, 1995). Perception-based Time Critical Rendering, University of Virginia.
- Kaiser, M. K., Montegut, M. J., & Proffitt, D. R. (1995). Rotational and translational components of motion parallax: Observers' sensitivity and implications for 3-D computer graphics. Journal of Experimental Psychology: Applied, 1, 321-331.
- Kaiser, M.K. & Proffitt, D.R. (1992). Using the stereokinetic effect to convey depth: Computationally efficient depth-from-motion displays. Human Factors, 34, 571-581.
- Liu, L., Tyler, C.W., & Schor, C.M. (1992). Failure of rivalry at low contrast: Evidence of a suprathreshold binocular summation process. Vision Research, 32, 1471-1479.
- Montegut, M. J. (1995). The emergence of visual detail, University of California, Santa Cruz.
- Padmos, P. & Milders, M.V. (1992). Quality criteria for simulator images: A literature review. Human Factors, 34, 727-748.

Proffitt, D.R., Rock, I., Hecht, H., & Schubert, J. (1992). Stereokinetic effect and its relation to the kinetic depth effect. Journal of Experimental Psychology: Human Perception and Performance, 18, 321.

Tyler, C.W. (1983). Sensory processing of binocular disparity. In C. Schor & K. Ciuffreda (Eds.), Vergence eye movements. Boston: Butterworths.

Yan, J.K. (1985). Advances in computer-generated imagery for flight simulation, IEEE Transactions on Computer Graphics and Applications, 5(8), 37-51.