NASA-CR-203974

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 (4/1/96 - 3/31/97)
Technical Officer: Mary K. Kaiser, NASA Ames Research Center

PERCEPTION-BASED TECHNIQUES FOR IMPROVING COMPUTER-GENERATED IMAGERY

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 computer generated imagery (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.**

Three primary areas of research were investigated during this project period. The first area examined 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 examined the use of perception-based techniques to automate the modulation of level of detail in time-critical rendering. The third investigated techniques for reducing the computational requirements of animated displays.

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 an efficient technique for creating an increase in apparent resolution at reduced computational costs.

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 in 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. 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. 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

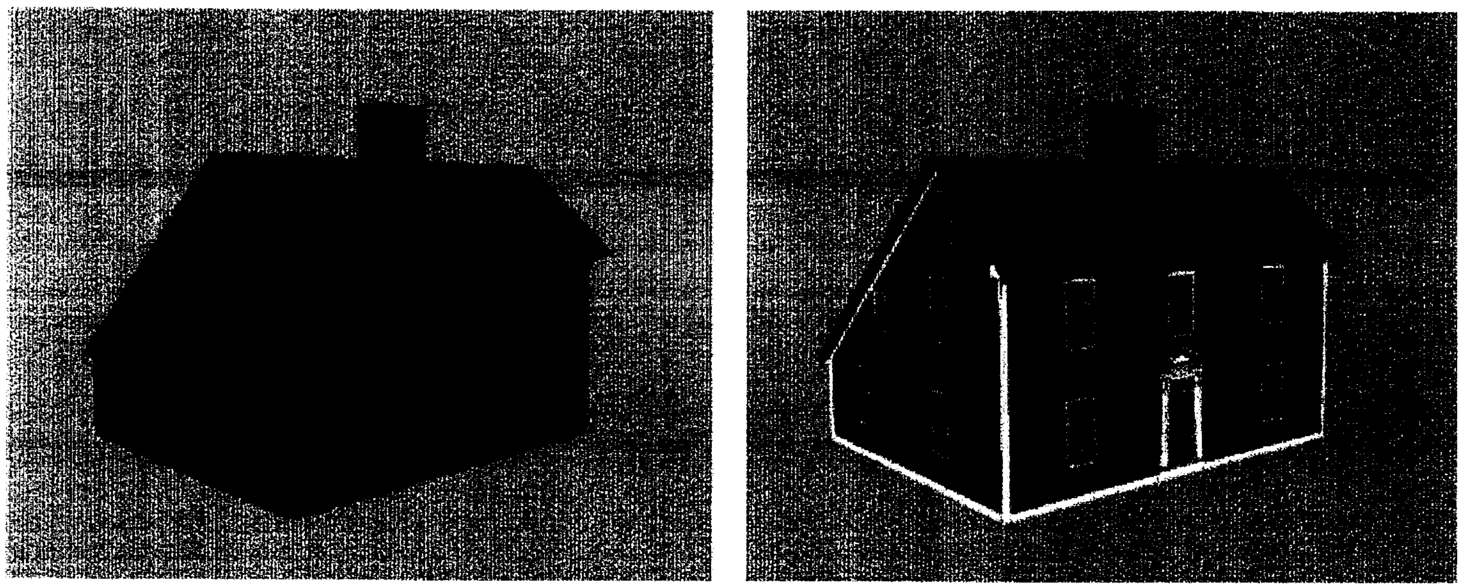

Figure 1

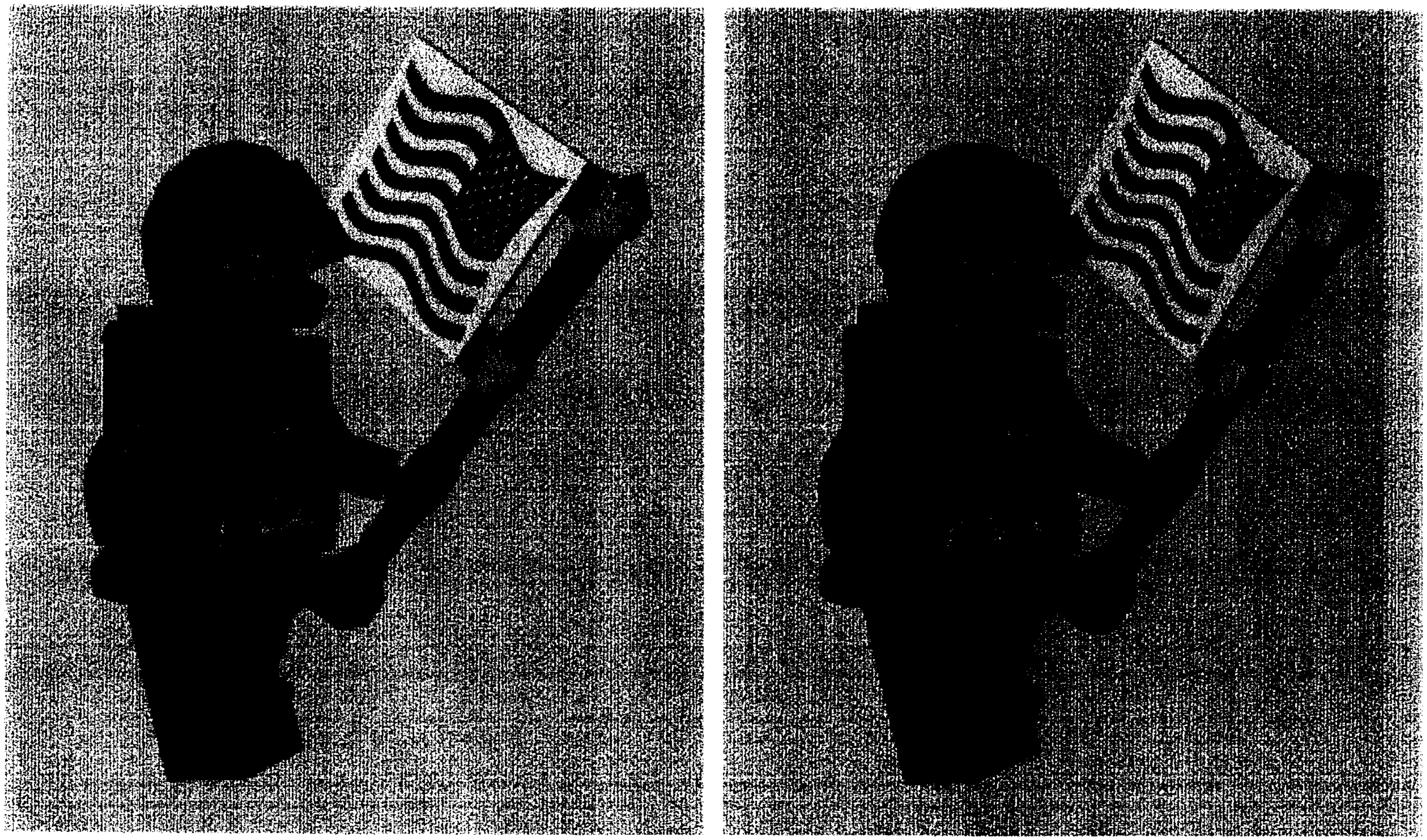

Figure 2

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.

We were concerned about the possibility that hi-lo stereo displays might tax the user's visual system with prolonged exposer. 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 four between-subjects conditions defined by whether viewing was (1) biocular hi-hi (BHH), (2) biocular hi-lo (BHL), (3) stereo hi-hi (SHH), or (4) stereo hi-lo (SHL). 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.

The results for both the cannon game and the random-dot stereogram task are presented in Figure 3. 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, but it was not a large difference. 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 president 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

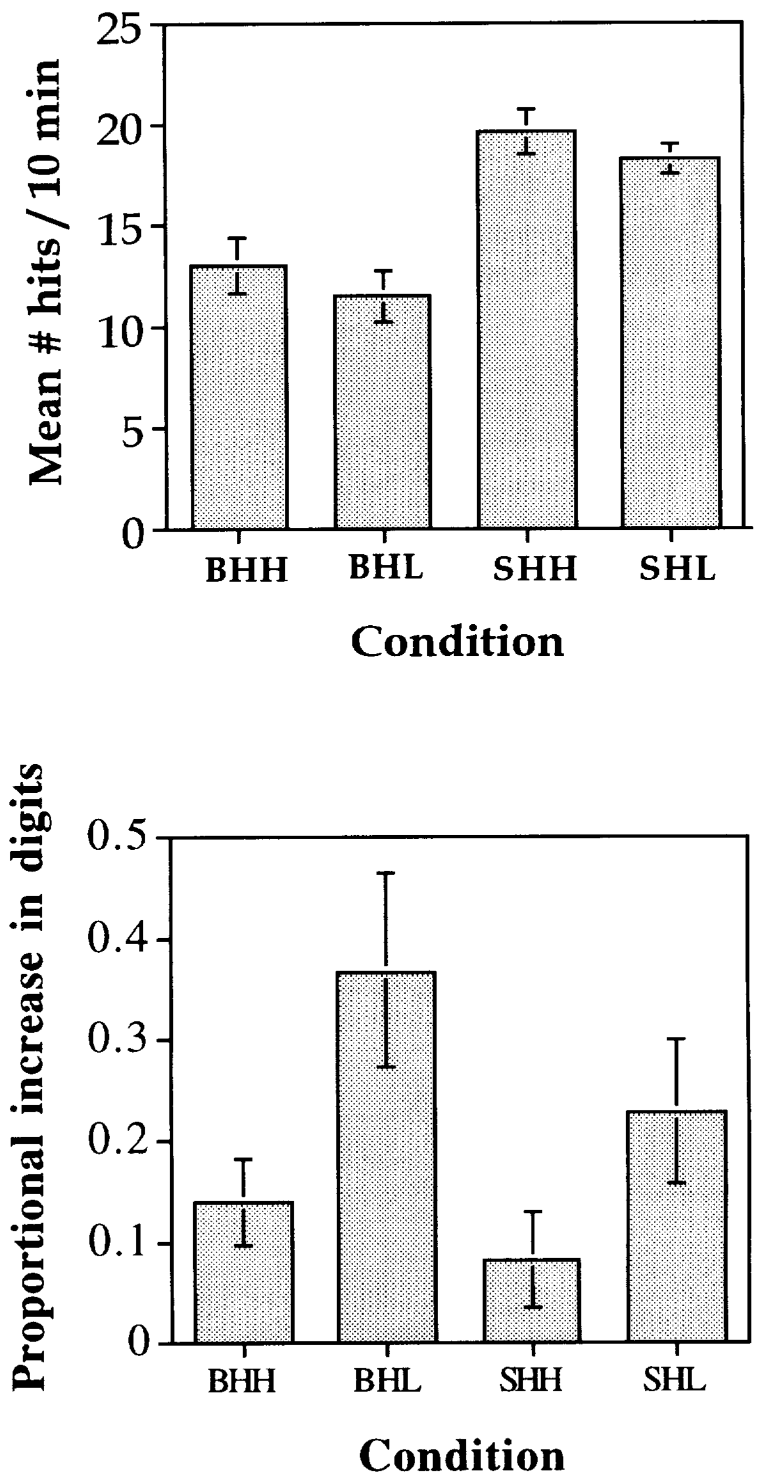
25
20
15
10
5
0
Mean # hits / 10 min
BHH
BHL
SHH
SHL
Condition
0.5
0.4
0.3
0.2
0.1
0
Proportional increase in digits
BHH
BHL
SHH
SHL
Condition

Figure 3

information.

The hi-lo stereo fusion technique and the user study were reported at the 1996 SIGGRAPH meeting (Proffitt & Kaiser, 1996).

A new technique that we have developed this period applies the hi-lo stereo process only to the least salient objects in a simulated scene. For example, more distant and small objects can be rendered in hi-lo stereo, whereas nearer and larger objects are rendered in normal hi-hi stereo. This technique allows a designer to realize a hi-lo saving without creating a loss of image quality in those parts of a scene that are most important. This technique can be incorporated into time-critical rendering methods.

Emergence of Detail in Time-Critical Rendering

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. As alluded to above, we are currently investigating hi-lo stereo in time-critical rendering. The notion is that an object's texture can emerge into the scene one eye at a time.

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 finding 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 require 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 (1995) 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.

We are continuing to investigate the perception of rigidity and 3D form in motion displays. Psychophysical research has shown that people are very sensitive to the first order derivatives of the instantaneous field of image displacements, whereas they show little sensitive to second order temporal derivatives. These latter temporal derivatives are essential for the canonical recovery of depth from motion. Moreover, it has been found that, when shown displays providing only first-order information, observers provide consistent judgements about 3D properties even though these properties are not uniquely specified by the information inherent in these displays. Finally, it has been found that the recovery of 3D form from motion is relatively accurate in certain situations but quite inaccurate in others. These aspects of perceptual performance cannot be accounted for by current models of the recover of 3D structure from motion.

We proposed a model based on the first order spatial derivatives of the image velocity field. This model can account for both veridical performance and systematic misperceptions in the human derivation of 3D form from motion. This work has been accepted for publication (Domini, Caudek, & Proffitt, in press).

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)]

Conference Presentations

Proffitt, D.R., & Kaiser, M.K. (1996). Hi-lo stereo fusion. Annual meeting of SIGGRAPH.

Publications

Proffitt, D.R. & Kaiser, M.K. (1996). Hi-lo stereo fusion. SIGGRAPH Visual Proceedings, 146.

Domini, F., Caudek, C., & Proffitt, D.R. (in press). Misperceptions of angular velocities influence the perception of rigidity in the kinetic depth effect. Journal of Experimental Psychology: Human Perception and Performance.

Durgin, F.H., & Proffitt, D.R. (1996). Visual learning in the perception of texture: Simple and contingent aftereffects of texture density. Spatial Vision, 9, 423-474.

References

Caudek, C. & Proffitt, D.R. (1992). Depth perception in motion parallax and stereokinesis. Journal of Experimental Psychology: Human Perception and Performance, 19, 32-47.

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.

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.