

# Perception of 3D Spatial Relations for 3D Displays<sup>a</sup>

Paul Rosen<sup>1</sup>, Zygmunt Pizlo<sup>2</sup>, Christoph Hoffmann<sup>1</sup>, and Voicu Popescu<sup>1</sup>

<sup>1</sup>Computer Sciences, <sup>2</sup>Psychology

Purdue University, 250 N University Street, West Lafayette, IN 47907-2066

## ABSTRACT

We test perception of 3D spatial relations in 3D images rendered by a 3D display (Perspecta from Actuality Systems) and compare it to that of a high-resolution flat panel display. 3D images provide the observer with such depth cues as motion parallax and binocular disparity. Our 3D display is a device that renders a 3D image by displaying, in rapid succession, radial slices through the scene on a rotating screen. The image is contained in a glass globe and can be viewed from virtually any direction. In the psychophysical experiment several families of 3D objects are used as stimuli: primitive shapes (cylinders and cuboids), and complex objects (multi-story buildings, cars, and pieces of furniture). Each object has at least one plane of symmetry. On each trial an object or its “distorted” version is shown at an arbitrary orientation. The distortion is produced by stretching an object in a random direction by 40%. This distortion must eliminate the symmetry of an object. The subject’s task is to decide whether or not the presented object is distorted under several viewing conditions (monocular/binocular, with/without motion parallax, and near/far). The subject’s performance is measured by the discriminability  $d'$ , which is a conventional dependent variable in signal detection experiments.

## 1. INTRODUCTION

Traditionally, computer graphics systems project a 3D scene from the view desired by the user onto a rasterized plane. The 2D image so computed is then shown on a 2D display. This approach has three fundamental problems, for which there is no perfect solution yet. First, the system has to give the user exploring the scene an intuitive way of specifying novel desired views. Keyboards, joysticks, and trackers designed to tell the graphics system the user’s position and view direction are often non-intuitive, imprecise, and/or bulky and do not allow the user to navigate freely in the 3D scene. Second, the 2D image has to be recomputed from scratch for each novel view. In spite of the great advances in interactive rendering, rendering images that are mistaken for photographs still require several orders of magnitude more time than available in an interactive rendering system. Third, the user should be allowed to take advantage of binocular stereo vision. Existing head-mounted displays have very low resolution, are bulky and have a limited field of view. Active stereo glasses and polarized passive glasses are uncomfortable, and produce little and sometimes incorrect disparity.

One way of avoiding these problems is to display a sculpture of light that is the exact replica of the 3D scene to be rendered. Immersed in such a 3D image, the user requests novel views naturally by gaze, by head movement, and by walking. For static scenes at least, once the 3D image is computed, no per-view rendering is required. Lastly, the user will perceive correctly different images with each of his eyes. Such a 3D display technology does not yet exist, but several technologies have demonstrated promising results. In this paper we analyze the perception of 3D spatial relations in images generated by a rotating-screen 3D display. We briefly review the main 3D display technologies in the next section. Then we describe the psychophysical experiment we conducted. Section 0presents and discusses the results of the experiment and section 0gives possible directions of future work.

---

<sup>a</sup> Work supported in part by NSF grants DMS-0138098, EEC-0227828, EIA-0216131, and ACI-0325227. Hoffmann is also supported in part by an IBM Faculty Award.

## 2. THREE DIMENSIONAL DISPLAYS

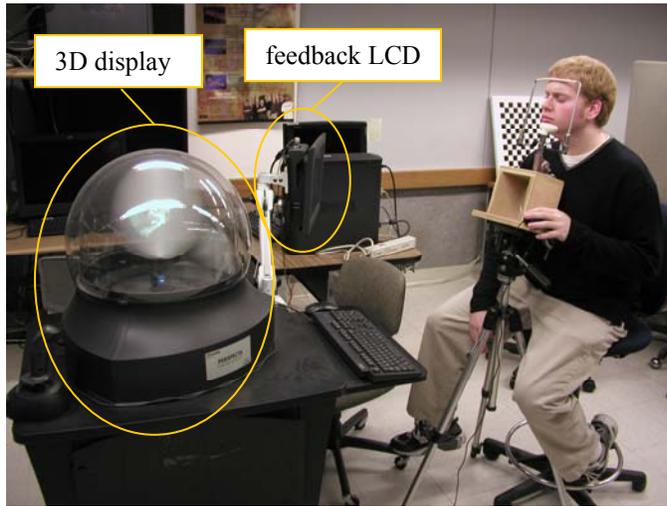


Figure 1 – Experiment setup. The lab lights were dimmed during actual experiment.

The Perspecta from Actuality Systems (Perspecta 2003) is a 3D display device that allows a volumetric image to be viewed by several observers simultaneously from any direction. A globular glass housing (Figure 1), approximately two feet in diameter, contains a rotating, semi-transparent screen of approximately 10" diameter on which 198 images are displayed by an internal DLP projector for each rotation of the screen. Each image has a resolution of 768 by 768 pixels and represents a slice of the 3D scene to be displayed. The eye of the observer combines these slices and perceives a complete volumetric image. The globe is free-standing, so the image is visible from all directions except from underneath (a  $360^\circ \times 270^\circ$  field of view). The volumetric image comprises approximately 100 million voxels, each with 64 possible colors. Thus the technology puts great demands on the bandwidth between the host computer and the display, which limits the ability of displaying moving scenes.

There are several alternatives to this technology for stereoscopic images. By far the most common stereoscopic device is based on presenting two separate images to each eye that account for parallax. In active systems, the images are presented alternately and a shuttered eye-glass device blocks one of the eyes accordingly. In passive systems (e.g. DTI 2003) the two images are color coded or presented with polarized light, and spectacles with suitable filters ensure that each image reaches the intended eye. Another passive technology is based on lenticular gratings. When the observer is positioned at the correct distance from the device, each eye sees a separate image on an LCD screen. The lenticular grating separates alternating image columns, so separating an interleaved stereoscopic pair of images. Such stereoscopic image pairs are a simpler technology to drive since the necessary data transmission rates are equivalent to normal 2D images.

An emerging technology is based on holography. At this time, there are no commercially available devices that create moving holographic imagery. There are, however, devices for creating holographic still images. As with the Perspecta, holographic moving images demand bandwidth and computational power that appears to exceed current technology.

Given the unique technological niche the Perspecta is situated in, we investigate the question whether the 3D images so generated confer a particular cognitive advantage when judging spatial relationships. If we consider images of familiar objects, such as houses, cars, and furniture, can an observer tell easily whether the object is correctly displayed or is distorted? To investigate this question, we conducted a series of experiments on the device. Briefly, in one experiment the object is displayed and the observer is encouraged to view it from several directions of his or her choosing before deciding whether the object is distorted. In another experiment, we fix the observer and ask whether a binocular view is superior to a monocular view, and at what distance. Finally, we run the experiment also using a nonstereoscopic flat panel display (the IBM T221 monitor with a pixel resolution of 3840 by 2400 on a 12" by 19" screen (IBM T221 2003)).

A generic limitation of the Perspecta is that it can only display illuminated voxels. Since opacity cannot be generated, no part of the image can truly obscure any other part. This can make the image of a familiar object seem unfamiliar, since the observer can see both front and back at the same time. Thus, the comparison with a traditional LCD screen makes the experimental evaluation of the Perspecta especially poignant.

### 3. PSYCHOPHYSICAL EXPERIMENT

In order to evaluate capabilities and limitations of Perspecta, we designed an experiment on the visual perception of spatial relations in a 3D scene. Perception of spatial relations in 3D scenes refers to a wide range of visual abilities: perception of distances among objects, distance from the observer, motion in depth, size and shape of an object. We decided to test the subject's ability to perceive 3D shapes, but this choice was somewhat arbitrary and one can easily argue that testing other abilities, like size or motion, would be equally, or even more informative. In order to minimize the role of mental processes that are not directly related to visual perception, we designed an experiment, in which the subject was presented with one stimulus on each trial and was asked to make a judgment about the shape of this stimulus. Other shape perception tasks like shape discrimination would require the subject to compare two or more shapes. Such comparison is likely to involve memory: the subject has to first look at one object, produce a mental representation of this object, and then compare this representation to the percept of the other object. Results of such a task are therefore likely to conflate the role of memory with the role of visual perception.

#### 3.1. Subjects

The authors were the subjects in this experiment. ZP is experienced as a subject in psychophysical experiments on shape, including binocular experiments. The other three subjects were inexperienced. PR received more experience with the stimuli due to the fact that he was directly involved in designing the stimuli.

Using a small number of subjects and including the authors among the subjects in psychophysical experiments is commonly accepted in the psychophysics community (the reader may verify this claim by consulting leading perception journals like *Vision Research* and *Perception & Psychophysics*). First, it is well established that all fundamental mechanisms underlying visual perception, including shape, are innate (Hochberg & Brooks, 1962; Slater, 1998). As a result, the magnitude of individual differences is extremely small. In other words, we all see things the same way, regardless of where we were born and raised. This means that results from just a few subjects who have normal vision is representative for the entire human population. Second, testing a subject who knows the hypotheses behind the experiment, not to mention, testing the person who formulated the hypotheses and designed the experiment in the first place, leads to valid data as long as a reliable psychophysical method is used. The main problem in studying perception is related to the fact that the subject's response is a result of two factors: the percept itself, and a decision, which mediates between the percept and the behavioral response. There are psychophysical methods, which allow measuring independently the percept and the decision. This method is called Signal Detection Experiment (SDE) (Green & Sweats, 1966). The main elements of SDE are described next.

#### 3.2. Signal Detection Experiment

In a signal detection experiment two types of stimuli are used:  $S_1$ , called *noise* and  $S_2$ , called *signal* (where  $S_2 > S_1$ ). The subject produces one of two responses,  $R_1$  and  $R_2$ , respectively. Each of the two stimuli is presented 50% of the time in a random order: the subject does not know which of the two stimuli is presented on a given trial. Response  $R_1$  is the correct response when  $S_1$  is presented, and response  $R_2$  is the correct response when  $S_2$  is presented. The percept of a stimulus is represented by a random variable  $X$ . When  $S_1$  is presented,  $X$  is subject to a normal distribution with the probability density function  $N(\mu_1, \sigma^2)$ , and when  $S_2$  is presented,  $X$  is subject to a normal distribution with the probability density function  $N(\mu_2, \sigma^2)$  ( $\mu_2 > \mu_1$ ). The subject's ability to detect the signal is related to  $(\mu_2 - \mu_1)/\sigma$ . This ratio is called detectability and is denoted by  $d'$ . For a given  $\sigma$ , when the difference  $(\mu_2 - \mu_1)$  is greater, it is easier to tell the two stimuli apart. Similarly, for a given difference  $(\mu_2 - \mu_1)$ , when  $\sigma$  is smaller, it is easier to tell the two stimuli apart. It is important to note that  $d'$  is a measure of the percept unconfounded with the subject's bias towards either of the two responses  $R_1$  or  $R_2$ . This is obvious because  $d'$  is a function of the parameters that characterizes the perceptual representation of the stimuli ( $\mu_1, \mu_2, \sigma$ ), but not the actual responses.

Now the main challenge is to show how to estimate  $d'$  from the responses  $R_1$  and  $R_2$ . Let  $h = P(R_2|S_2)$  be the *hit rate* and  $f = P(R_2|S_1)$  be the *false alarm* rate. It is assumed that the subject produces responses  $R_1$  and  $R_2$  based on a subjectively (and arbitrarily) adopted criterion  $k$  for the magnitude of the percept  $X$ . The decision criterion is as follows: if  $X > k$ , respond  $R_2$ , otherwise, respond  $R_1$ . Let  $\Phi(z)$  be the cumulative function of the standard normal distribution, and  $\phi(z)$  its density function. Let  $z_p = \Phi^{-1}(p)$ , be the inverse of the cumulative distribution function. It is easy to show that (Green & Sweats, 1966):

$$d' = z_h - z_f \quad (1)$$

In practice, one does not know  $h$  and  $f$ , but only their estimates. So,  $z_h$  and  $z_f$  must be computed from estimated  $h$  and  $f$ . To keep the notation simple, we will use symbols  $h$ ,  $f$ , and  $d'$  to represent not only the parameters, but also their estimators. This should not produce confusion. Hit rate and false alarm rate are then computed as follows.

Let  $N_1$  be the number of trials in which  $S_1$  was presented and  $N_2$  the number of trials in which  $S_2$  was presented. Let  $N_h$  be the number of trials where  $S_2$  was presented and the subject responded  $R_2$ . Similarly, let  $N_f$  be the number of trials where  $S_1$  was presented and the subject responded  $R_2$ . Then:

$$h = N_h/N_2, \quad f = N_f/N_1 \quad (2)$$

Now  $d'$  is estimated from (1) using hit and false alarm rates as estimated from (2). Interestingly, *even though the actual hit and false alarm rates strongly depend on response criterion  $k$ , and  $d'$  is computed from hit and false alarm rates,  $d'$  itself does not depend on the response criterion  $k$* . The remaining task is to estimate the standard deviation of the estimated  $d'$  (Wickens, 2002). This standard deviation, called standard error of estimated  $d'$  will be denoted by  $se$ :

$$se = [\text{var}(h)/\phi^2(z_h) + \text{var}(f)/\phi^2(z_f)]^{1/2} \quad (3)$$

where:

$$\text{var}(h) = h(1-h)/N_2; \quad \text{var}(f) = f(1-f)/N_1 \quad (4)$$

The standard error as expressed by (3) is a lower bound of the actual standard error; its value is obtained by assuming that the only source of variability in  $h$  and  $f$  is the sampling error. In reality, there are also other sources of variability, like subject's attention and subject's criterion  $k$  for producing responses  $R_2$  vs.  $R_1$ . In order to obtain a more realistic estimate of the standard error, one may use a different design of SDE, in which confidence rating is involved (see Macmillan & Creelman, 1991). In this study, we estimated the standard error by using formula (3). Finally, it must be pointed out that the formula (1) provides a good estimator of detectability only if the response criterion  $k$  is stable throughout the session. Otherwise, the estimated  $d'$  will be an underestimation of the true  $d'$ . So, even though, theoretically,  $d'$  is a measure of the percept that is unconfounded by a response bias, in practice this may not be true if the response criterion  $k$  is not stable. The best way to assure that the response criterion is stable is to use experienced subjects who are familiar with the stimuli and the experimental setup.

### 3.2.1. Stimuli

For each trial the subject is shown an image rendered from one of 18 complex objects, including automobiles, buildings and pieces of

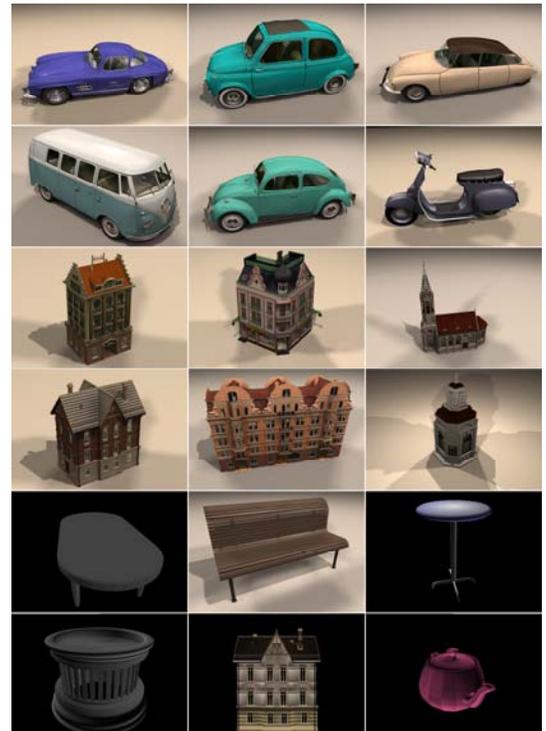


Figure 2. Objects used to generate stimuli

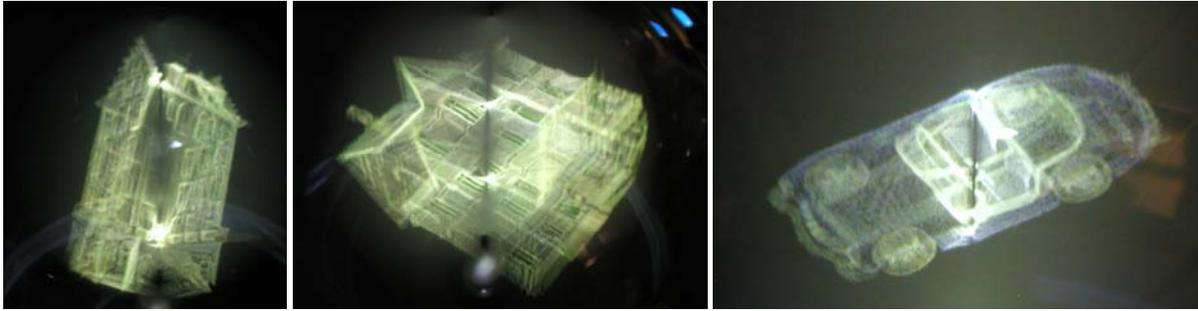


Figure 4. Examples of 3D images generated by the Perspecta display

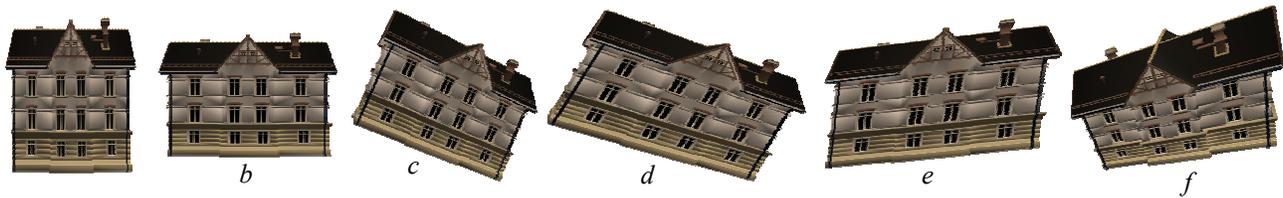


Figure 3. Distorted trial image generation

furniture (Figure 2). For about half of the images the objects are distorted, while for the other half they are not. The procedure to generate the images is shown in Figure 3. The original object (*a*) is stretched a random amount along all three major axes (*b*). This step ensures that although the same object is used for several tests, the object has a unique normal undistorted appearance for each of the tests. If an object is to be distorted, it is rotated about an arbitrary axis (*c*), stretched 40% (*d*), and then rotated back (*e*). Then, regardless of whether it has been distorted or not, the object is rotated about an arbitrary axis to generate a new random viewing angle (*f*). Figure 4 shows photographs of the stimuli as seen on the Perspecta.

### 3.2.2. Procedure

Five experimental conditions were used, one session per condition. In four conditions, the stimuli so generated were shown on the Perspecta (Figure 1). In the fifth condition, the stimuli were shown on a LCD screen (Figure 5). In the case of Perspecta, the stimuli were viewed binocularly from the viewing distances of 1m, and 2m, and monocularly from the viewing distance of 1m. In these three conditions, the head was supported by a chin-forehead rest. In the fourth condition, the subject was instructed to walk around Perspecta while making judgments. In the case of LCD, the viewing was binocular from a distance of about 1m, which was the simulated distance used in computing perspective images. In this case, however, binocular disparity was absent. Therefore, this condition is called monoscopic, as opposed to stereoscopic.



Figure 5. LCD experiment setup

Each session started with 20 practice trials. Then, 200 experimental trials followed. On each trial, an object was shown and the subject's task was to judge whether the object was symmetric. Exposure duration was 15 sec. The subject had the option to respond at anytime during the 15 sec period, or after that. Each subject received the same set of images, but the order in which the images were displayed was randomized with each subject and with each session.

The subject was provided with a display of the current test number, the currently selected answer, and whether the previous answer was correct or not. The verdict on the previous answer was given to help the subject maintain focus. The feedback was given on a nearby LCD (Figure 1) that had a black background so as not to interfere with the experiment.

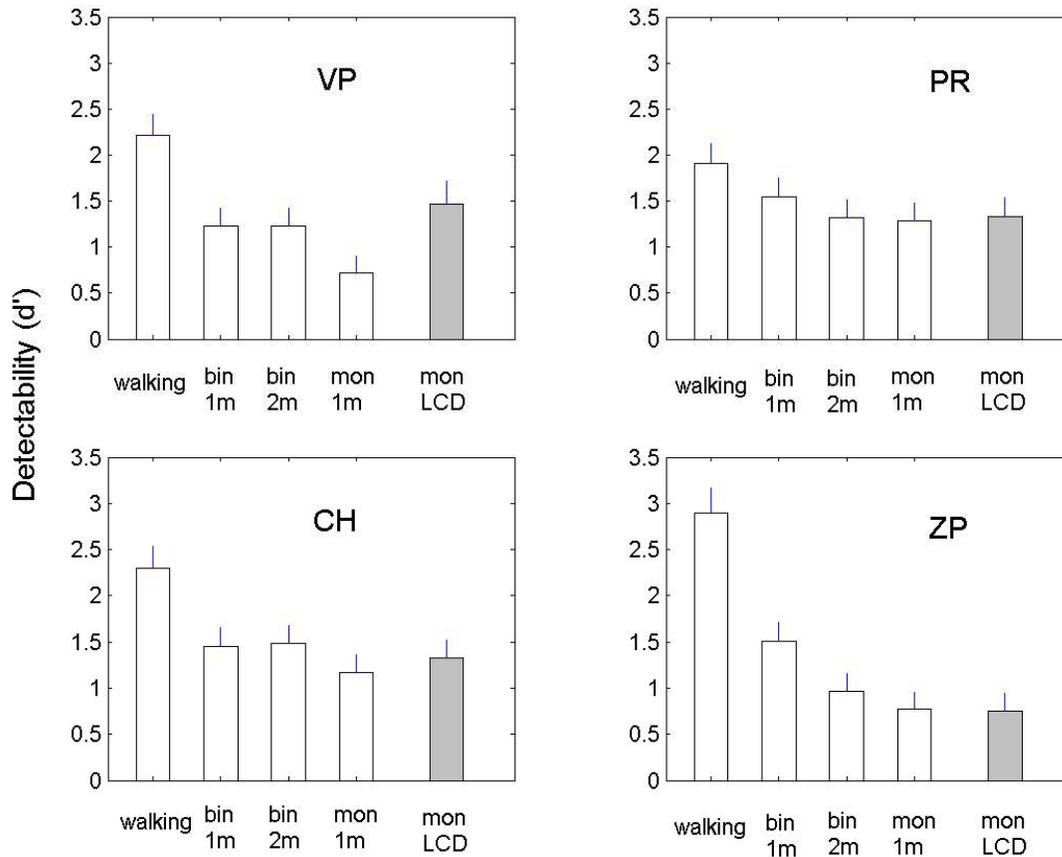


Figure 6. Results of the psychophysical experiment. White bars show performance in the case of Perspecta and the gray bar shows performance in monoscopic presentation of stimuli on an LCD monitor. Each panel shows results of one subject.

#### 4. RESULTS AND DISCUSSION

Performance in the case of walking around Perspecta is quite high in the case of all four subjects. All four subjects achieved  $d'$  around 2 or more. Note that  $d'=2$  corresponds to 84% correct responses (when  $h=1-f$ ), and  $d'=3$  corresponds to 93% correct responses. Binocular viewing of Perspecta leads to performance similar to monoscopic viewing of an LCD, in three subjects. This result is probably related to the fact that the images rendered on Perspecta have rather low contrast. Similarly, monocular viewing of Perspecta leads to performance similar to monoscopic LCD in three subjects. Finally, in the case of all four subjects, binocular viewing of Perspecta from a distance of 1m leads to somewhat better performance than monocular viewing of Perspecta. The fact that binocular viewing would lead to better performance than monocular viewing is obvious and was easy to predict. It is less obvious that the difference in performance is not very large. Apparently, a single perspective image contains enough information so that the judgments of symmetry of a 3D object are fairly reliable and adding binocular disparity does not improve the reliability much. That binocular shape perception is only slightly more reliable than monocular shape perception is not new and is not restricted to symmetry judgments (e.g. Chan et al., 1999). It is known that shapes of structured objects allow the human visual system to use strong priors, and the presence of the priors is often more important than the presence of depth cues (Pizlo, 2001).

## 5. FUTURE WORK

Our Perspecta display is a second generation device. Cost trade-offs in the manufacture of the device have resulted in an image that is slightly wobbly. A more costly, high-precision mechanical fabrication would lead to a more stable image. Furthermore, the current pixel resolution is expected to increase in future generations of the device. Both factors suggest that the full potential of the device has not yet been realized. Moreover, the bandwidth of the connection between computer and device undoubtedly will improve in the future. Today, moving complex images cannot be so displayed. When this has changed, new experiments should be conducted to assess the effect of these improvements.

We are pursuing an alternative path to improving the stability of the 3D image by modeling the wobbling pattern. We speculate that the pattern does not change considerably over time and that it can be almost eliminated by better calibration. The model parameters will be tuned manually by displaying a calibration scene and searching for a combination that minimizes the wobbling. We will also investigate the possibility of using a camera to provide the feedback needed for calibration and completely automate the procedure.

There are image applications that are traditionally difficult to do on conventional displays. They include point clouds and point/line arrangements in 3-space. Such images are easy to display and comprehend on the Perspecta, and are difficult to render well on conventional displays. A familiar strategy on conventional displays is to add as depth cue a reduced size and/or a fainter intensity of more distant parts of the arrangement. It would be interesting to compare the effectiveness of these display strategies with a straightforward rendering on the Perspecta.

A fundamental limitation of the 3D display technology employed by the Perspecta is its inability to display view dependent effects such as occlusions and reflections. For example a bright back surface cannot be hidden by a dark front surface, or a specular highlight's position is ambiguous when more than one view is considered. We will investigate how to reduce the artifacts resulting from these fundamental limitations. An a priori knowledge of the set of desired view locations could be employed to first eliminate the surfaces that are not visible from any or most of the desired views. Work in image-based rendering has shown that stationary specular highlights are preferable to eliminating all the highlights and treating all surfaces as diffuse. It is our experience that users generally do not notice that the specular highlights do not change with the desired view, formal user studies are in order to establish which applications could tolerate this approximation. When the correct highlight is important, tracking the user appears to be the only solution. For multiple users, each will have his own highlight but they will also see the highlight rendered for the other users, which will probably give the impression of a moving light.

3D displays are a very promising technology; however, the technology is still at its infancy. Level-of-detail and occlusion culling algorithms, schemes for parallel and distributed rendering, antialiasing algorithms (another view dependent effect), interfaces (for pointing, selecting, navigation, etc.) are yet to be developed and constitute interesting and potentially very fruitful avenues for future research.

## ACKNOWLEDGEMENTS

We would like to thank all the members of our computer graphics and visualization laboratory that put up with the dim lighting conditions required to run the numerous experiments. This research has been supported in part by NSF grants DMS-0138098, EEC-0227828, EIA-0216131, and ACI-0325227. Hoffmann is also supported in part by an IBM Faculty Award.

## REFERENCES

Perspecta 2003, Acutality Systems Inc., URL: <http://www.actuality-systems.com/>

DTI 2D/3D 2003, Dimension Technologies Inc. URL: <http://www.dti3d.com/>

IBM T221 2003, IBM Corporation, URL: <http://www.ibm.com/>

Chan M.W., Pizlo Z. & Chelberg D. (1999) Binocular shape reconstruction: psychological plausibility of the 8 point algorithm. *Computer Vision & Vision Understanding*, 74, 121-37.

- Green D.M. & Swets J.W. (1966) Signal Detection Theory and Psychophysics. NY: Wiley.
- Hochberg J. & Brooks V. (1962) Pictorial recognition as an unlearned ability: a study of one child's performance. *American Journal of Psychology*, 75, 624-8.
- Macmillan N.A. & Creelman C.D. (1991) *Detection Theory: A User's Guide*. Cambridge University Press.
- Pizlo Z. (2001) Perception viewed as an inverse problem. *Vision Research*, 41, 3145-61.
- Slater A. (1998) Visual organization and perceptual constancies in early infancy. In: Walsh V. & Kulikowski J. (Eds.), *Perceptual Constancy*, Cambridge University Press (pp. 6-30).
- Wickens T.D. (2002) *Elementary Signal Detection Theory*. Oxford University Press.