Gaze-Directed Adaptive Rendering for Interacting with Virtual Space

Toshikazu Ohshima, Hiroyuki Yamamoto, and Hideyuki Tamura
Media Technology Laboratory, Canon Inc.
890-12, Kashimada, Saiwai-ku, Kawasaki, Kanagawa 211, Japan
e-mail : {ohshima, yumi, tamura}@cis.canon.co.jp

Abstract

This paper presents a new method of rendering for interaction with 3D virtual space with the use of gaze detection devices. In this method, hierarchical geometric models of graphic objects are constructed prior to the rendering process. The rendering process first calculates the visual acuity, which represents the importance of a graphic object for a human operator, from the gaze position of the operator. Second, the process selects a level from the set of hierarchical geometric models depending on the value of visual acuity. That is, a simpler level of detail is selected where the visual acuity is lower, and a more complicated level is used where it is higher. Then, the selected graphic models are rendered on the display. This paper examines three visual characteristics to calculate the visual acuity: the central / peripheral vision, the kinetic vision, and the fusional vision. The actual implementation and our testbed system are described, as well as the details of the visual acuity model.

1. Introduction

The rapid progress of 3-D computer graphics technology makes it possible to synthesize highly realistic images. For example, the radiosity method is often used to generate images that we can hardly distinguish from the real scenes which they depict. However, current virtual reality (VR) systems generate images with coarse geometric information and inelegant photometric effects. This is because VR applications require real-time interaction between the virtual space and human operator(s). Even the current high-end graphic machines cannot afford to render sufficiently realistic images at such a fast speed. In other words, current VR systems limit the number of rendered polygons and sacrifice the photo-reality of rendered images to insure the image-generation time.

In order to solve this problem, the two techniques categorized here, both of which are used to reduce the number of polygons to be rendered, have been proposed [1-10]. A short survey of the techniques is found in [6].

(1) **Visibility determination method**: This method culls polygons that do not affect the contents of images, such as the ones occluded from the viewpoint of the observer [1,2]. In many cases, this can reduce a significant number of polygons. If very complex objects are always observed, however, the method is not sufficient.

(2) **Detail elision method**: This method elides the details of each visible object, *i.e.*, the number of polygons that construct an object shape, adaptively at each frame. Practical implementations represent each object as a geometric model with multiple levels of detail (*LOD*s), - that is, a hierarchical geometric model with different numbers of polygons at each level, - and elide detail by selecting simpler LODs for each model [3-10]. LOD selection algorithms based on the size of an object on the screen and the distance from an object to the viewpoint are reported [7-10]. In addition, a technique that selects LOD by the number of pixels covered by a polygon is also reported in [5]. The techniques are effective when visible objects occupy very small areas on the screen.

These methods do not assume any condition about the observation such as the number of observers, the positions of observers, and the configuration of display devices. Therefore, the methods work on a standard graphic workstation.

On the other hand, this paper focuses on VR systems with special devices and assumes the following extra conditions. To begin with, it is assumed that there is only one operator, and that his/her viewpoint and gaze positions are tracked with special devices. Second, it is assumed that

0-8186-7295-1/96 $5.00 © 1996 IEEE
Proceedings of VRAIS '96

the operator sees stereoscopic images with a wide field of view [11]. Third, we focus on a technique not to stabilize the frame rate, but to achieve as fast a rendering time as possible.

Under these assumptions, a new method of gaze-directed adaptive control of LOD is presented in this paper. In this method, areas on the display that the operator does not focus on are drawn using simplified shape models. The approach reduces the burden of the drawing process significantly and maintains the quality of the generated image for the operator. The discussion starts in Section 2 with the basics of adaptive display strategy controlled by an operator's eye movement. Then, the implementation of adaptive display algorithm is discussed in Section 3. A testbed of display system and results are shown in Section 4.

2. Adaptive display strategy using gaze tracking

For VR applications, a gaze tracking device [12] can be built into a pair of liquid crystal shutter glasses or a pair of head mounted displays which shows stereoscopic images. Thus, it is realistic to assume that gaze tracking can be used for VR purpose. In this section, we presents the descriptions of our new technique based on the gaze information. Actual implementations of the technique are described in the following section.

Human eyes have various visual characteristics. For example, they subtend a wide field-of-view and provide high resolution detail at the center of gaze (*fovea*) with decreasing resolution in the remaining area (*periphery*). This suggests that most of our computing power can be concentrated to render the small foveal area [13]. That is, limited computing resources are sufficient to draw low-resolution images on the periphery. With carefully chosen parameters, the operator is likely to have great difficulty in distinguishing a loss of image quality in comparison to totally high-resolution images. This is especially true for a wide screen display where the operator observes a wide field-of-view and looks around to gather the visual data of the scene.

In order to realize this strategy, the importance of each graphic object for the operator should be calculated in advance. The measure of importance considers human visual characteristics and is called *visual acuity* in this paper. The LOD of the geometric model is controlled based on the visual acuity so that the images with detailed shape information in the model are rendered for the object with high visual acuity. Conversely, a simplified geometric model is used for the objects with low visual acuity.

Among the various characteristics of the human visual sensations, the following types are chosen.

Central / peripheral vision

As described above, the visual acuity of a human being is not uniform over the whole visual field. When a man gazes at an object, he recognizes the details of the object, but the objects around it are blurred. Thus, high visual acuity is allocated to the objects in the fovea.

Kinetic vision

We do not easily recognize the details of objects that move in the visual field such as objects that rotate or move across the field. This phenomenon is believed to be attributable to the decline of visual acuity against moving patterns. By taking this characteristic into account, the visual acuity of an object that moves with smaller angular velocity is set to a higher value, and the visual acuity of an object that moves with larger angular velocity is set to a lower value.

Fusional vision

The phenomenon discussed here is premised on the binocular vision and is illustrated in Figure 1.

Let assume that an operator looks at an object binocularly, his/her eyes moving to synchronize the two images of the object (Object A in Figure 1) seen by both eyes. This eye movement is called *convergence*. By this movement, he/she recognizes the focused object as a single stereoscopic image. Incidentally, if there is another object positioned in front of or behind the fixation point (Object B in Figure 1), the two images of the second object show a parallax that is different from the one at the fixation point. Therefore the second object is recognized as a "double image" with less reality. The phenomenon is called *fusion* and we experience it on an unconscious level

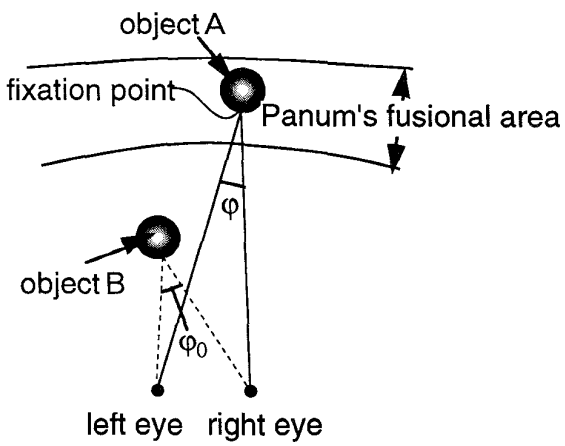


Figure 1. Fusional area.

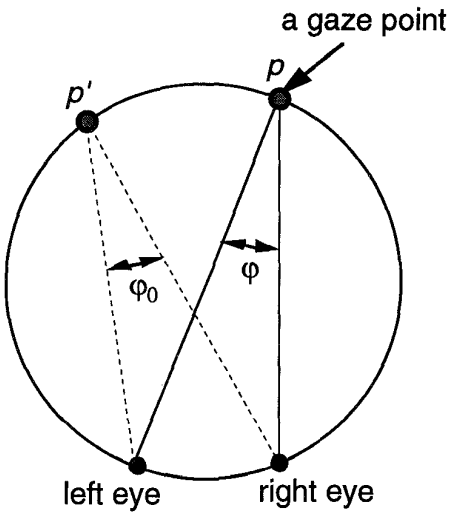


Figure 2. Geometry of Vieth-Muller's horopter.

every day.

In this sensation, the circumference that passes through the positions of observer's eyes and fixation point plays a great role. Since an object on the circumference is seen with the same parallax as that at the fixation point, the object is recognized as a single image. On the other hand, objects that are not on the circumference are seen as double images. The circumference is called *Vieth-Muller's horopter*. The geometry of the Vieth-Muller's horopter is illustrated in Figure 2. Accurately speaking, there exists some area around the horopter where the fusion is perceived. The area is called *Panum's fusional area* .

Taking these characteristics into account, the visual acuity is set with a low value if the object is placed out of the fusional area. Therefore, the objects in front of and behind the fixation point can be simplified.

3. Implementing adaptive rendering

3.1. Calculating visual acuity

In order to separate each factor of the visual sensations described above, the term *visual acuity factor* is used. That is, the visual acuity factor represents the importance of an object considering one visual sensation, whereas visual acuity means the total importance by examining all factors.

Figure 3 shows the geometric configuration to calculate the visual acuity factor k_1 of central / peripheral vision. Based on the figure, k_1 is modeled by

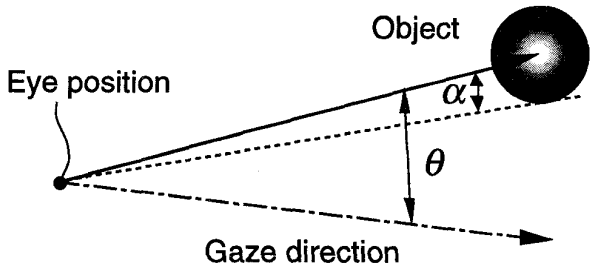


Figure 3. Viewing geometry considering central / peripheral vision.

$$k_1 = f(\theta) = \begin{cases} 1 & (0 \le \theta \le \alpha) \\ \exp\left(-\dfrac{\theta-\alpha}{c_1}\right) & (\alpha < \theta) \end{cases} \quad (1)$$

where θ is the angle between the visual axis and the vector from the viewpoint to the center of the bounding box of the object, and c_1 represents the parameter to adjust the decrement.

In the formula, the exponential function $f(\theta)$ is applied under the condition where the angle θ is greater than the value of α (see Figure 4). Here, the value of α represents the visual angle that the object occupies in the field of view; it is used to eliminate switching the levels of the shape models while the visual axis stays inside the object. In the following experiments, c_1 is set at 6.2°. With this value, the visual acuity factor is reduced to about 0.2 when the object is 10° apart from the visual axis.

The second visual acuity factor k_2, which takes the kinetic vision into consideration, is calculated by

$$k_2 = g(\Delta\phi) = \begin{cases} 1 - \dfrac{\Delta\phi}{c_2} & (0 \le \Delta\phi \le c_2) \\ 0 & (c_2 < \Delta\phi) \end{cases} \quad (2)$$

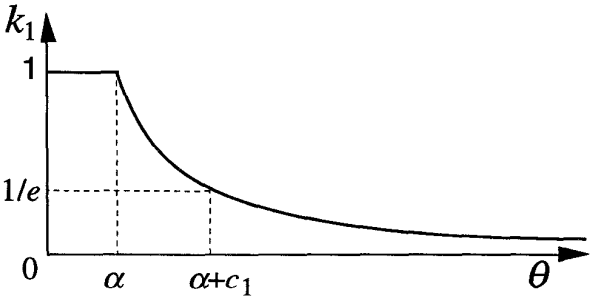


Figure 4. Visual acuity factor concerning central / peripheral vision.

where $\Delta\phi$ represents the angular velocity of a representative point of the moving object, and c_2, set at 180 degrees per second, is the parameter to adjust the decrement.

$\Delta\phi$ is obtained by the following algorithm. As illustrated in Figure 5, the graphic object is approximated as a sphere with the average radius of the object. The vector p is the position vector of the nearest position to observer's eyes, which is represented in the operator-centered coordinate system. Let assume that this vector p moves to p' after a certain time interval. The angular velocity $\Delta\phi$ is decided as the angle between p and p'.

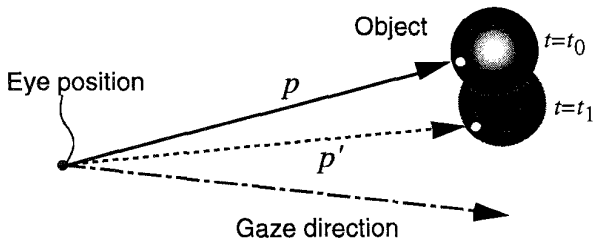


Figure 5. Motion geometry considering kinetic vision.

Third, the fusional area is modeled by the region expanded with certain width from the Vieth-Muller's geometric horopter. Thus, the third visual acuity factor k_3 is derived from

$$k_3 = h(\Delta\varphi) = \begin{cases} 1 & (0 \le \Delta\varphi \le b) \\ \exp\left(-\dfrac{\Delta\varphi - b}{c_3}\right) & (b < \Delta\varphi) \end{cases} \quad (3)$$

where $\Delta\varphi$ is the difference $\Delta\varphi = |\varphi - \varphi_0|$ between the angle φ_0 of convergence for the fixation point and the angle φ toward the object, b is the threshold value that decides the width of the fusional area, and c_3 is the parameter to adjust the decrement. This paper uses 0.62° as c_3 and 0° as b.

As explained above, the visual acuity factors are 1.0 at the best condition, and 0.0 for the object that the operator cannot see. Then, the three visual acuity factors are combined to obtain the visual acuity. There are several methods that can be used to calculate the visual acuity. The first possible method, one which was chosen in the experiments for its simplicity, is as follows:

$$a = \min(k_1, k_2, k_3) \times a_0 \quad (4)$$

where a represents the total visual acuity, and a_0 is the base visual acuity explained in Section 4.1 and the parameter for the system. Another possible method is to multiply the three visual acuity factors, *i.e.*,

$$a = k_1 k_2 k_3 a_0 \ . \quad (5)$$

3.2. Selecting a level from hierarchy

Once the visual acuity a is calculated, a level is selected from the hierarchy of the geometric model.

First, $r_{\min}$ is calculated as follows:

$$r_{\min} = 2D\tan\left(\frac{1}{2a}\right) \quad (6)$$

where D is the distance between the observer and the object. Then the highest LOD is selected from the hierarchy so that no polygon in the geometric model at the level is smaller than $r_{\min}$.

This is described as follows. Let assume that H_l represents the level l of the hierarchical geometric models of the object, p represents a polygon that is included in the geometric model, and $r(p)$ represents the length of the longest edge of all the edges in the polygon p. The biggest model H_l is selected from the set of $\{H_l | \forall p \in H_l, r(p) \ge r_{\min}\}$. Here the "biggest" model means the level of geometric model that has the most polygons.

4. Experimental results

4.1. Experimental methods

A 70-inch rear projection system that provides a diagonal visual angle of about 60° from the observer's position is employed as the display hardware. The distance between the observer and the display is about 1500 mm. The display images of 1280×1024 pixels are rendered by a graphic workstation (Silicon Graphics Inc., IRIS Crimson / RealityEngine). The images for each eye are time-multiplexed and displayed at a rate of 120 Hz. The operator looks at the images with liquid crystal shutter glasses (StereoGraphics, CrystalEyes) to experience a 3-D visual sensation.

Under the configuration, one pixel occupies an area of about 1.1mm×1.1mm, or 0.06° of the visual angle on the display. If, from the performance point of view, it is assumed that the system need not render polygon details under 5 pixels, 7.85mm, or 0.3° of visual angle on the display, the *base visual acuity factor* a_0 in Eqs. (4) and (5) is determined as follows:

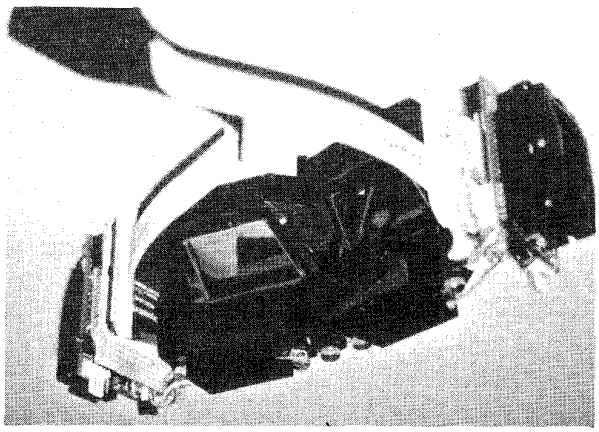


Figure 6. Gaze tracking eyewear.

$$a_0 = \frac{1}{2\tan^{-1}\left(\frac{l}{2D}\right)} = \frac{1}{2\tan^{-1}\left(\frac{7.85}{3000}\right)} \cong 3.3 \qquad (7)$$

where l is the minimum size of polygon details to be rendered, and D is the distance between the observer and the display.

In order to measure the gaze position, we are developing a stereoscopic eyewear with two eye trackers. Figure 6 shows the prototype of this eyewear. The body is composed of the CrystalEyes and two eye trackers each of which tracks the movement of each eye. The eye tracker is composed of two infrared (IR) light emitting diodes, a CCD image sensor unit including optical parts, and a processor unit. It decides the gaze direction by measuring reflected images of the IR light sources from the frontal surface of the cornea [12]. In addition, a Polhemus head tracker is mounted on the body to get the head position and attitude relative to the display device.

Since this eyewear is now under development, we use ultrasonic sensors built in the CrystalEyes to measure the head position and attitude. Then, the direction of the observer's head is used as the substitute of gaze direction of the dominant eye. The convergent angle is calculated by taking the intersection between the object and the gaze direction as the fixation point (see Figure 5).

In addition to the calculation of the convergence angle, the following measures are employed for the current display system.

Smoothing of gaze direction

As described above, the direction of the operator's head is used as the substitute of gaze direction. Therefore, frequent changes of the gaze direction caused by unconscious vibration of the operator's head, etc., are observed. If this information is directly applied into the calculation of the visual acuity, frequent switching of the shape models and unnecessary flickers are seen in the rendered images. Therefore, the gaze information is processed by smoothing detected movement and removing noise components to stabilize switching operation.

Fixation cursor

The actual gaze direction is independent of the head movement. Therefore, some measure is necessary to provide the fixation point to the operator. In the current configuration, a 3-D fixation cursor is displayed in the expectation that the fusional area works more effectively. By tracking the 3-D fixation cursor, the operator moves eyes so that his/her fixation point coincides with the gaze point used in the computation.

Saccade

Gaze movement whose angular velocity exceeds a certain value, - 180 degrees per second in these experiments, - is defined as saccadic movement. When movement of this kind is observed, the rendering process is skipped. In other words, the image is not updated until the angular velocity of the gaze direction falls back below the threshold. Thus, delays due to rapid movement are removed, and response time is improved.

4.2. Constructing hierarchical models

Hierarchical shape models are constructed from a *range image* of a real object (A range image is an image where each pixel contains 3-D information at that point.) The method to generate a hierarchical shape model from a radial range image has already been reported in [14].

This method converts a range image into polygons (triangular patches) with optimized structure. The sizes of the polygons are adaptively selected based on the local curvature to keep the original shape. Furthermore, the minimum size of polygon is specified to a larger value at a simpler LOD to construct a hierarchical model.

Since the method is based on the 3-D measurement of the object, the method can handle a complex object. Figure 7 shows the wire-frame representations of hierarchical triangle patches that are derived from a range image of a statue. Hierarchical geometric models used in the experiments are constructed in the same way.

4.3. Experimental results

Since it is difficult to present the effects of the kinetic visual sensation in a paper, the results are shown considering only the central/peripheral and the fusional visual factors. First, the results that are obtained by applying each visual factor separately are given. These

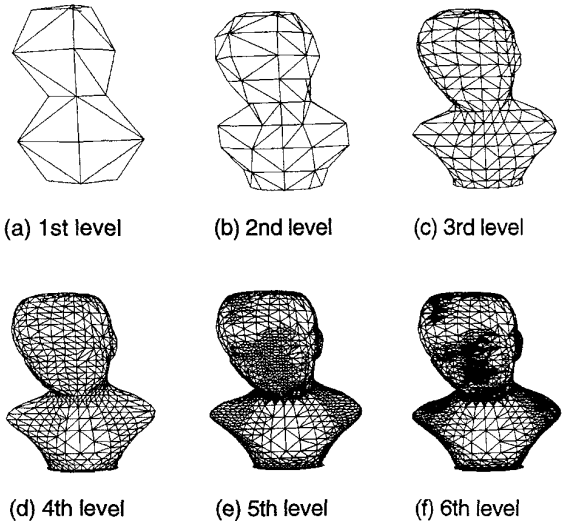


Figure 7. A hierarchical geometric model. The number of patches increases from 44 in the first level to 144, 534, 1612, 3562, and 6410 from the 2nd to 6th levels.

permit us to observe how each visual sensation contributes to the image quality and rendering speed.

In the following figures, graphic objects are rendered using a flat shading method to emphasize the results. In actual application, however, a smooth shading method, such as the Gouraud shading or Phong shading method, can be applied. Actually, our experiments have found that the degraded quality due to decrement of polygons is less recognizable when a smooth shading method and texture mapping are employed.

Central / peripheral vision

Here k_2 and k_3 in Eq. (4) are set to 1.0. Under this condition, an operator is asked to look around for a while and then to compare the quality of the degraded images derived from the proposed technique with the quality of the image with the highest resolution.

Figure 8 shows the displayed images while an operator is looking around the scene. In this figure, (a) represents the image when the fixation point of the operator is placed at the upper left object. In (b), the fixation point is moved to the lower right object. The results show that the shape models near to the fixation point are presented precisely, while LODs are lowered at the peripheral area.

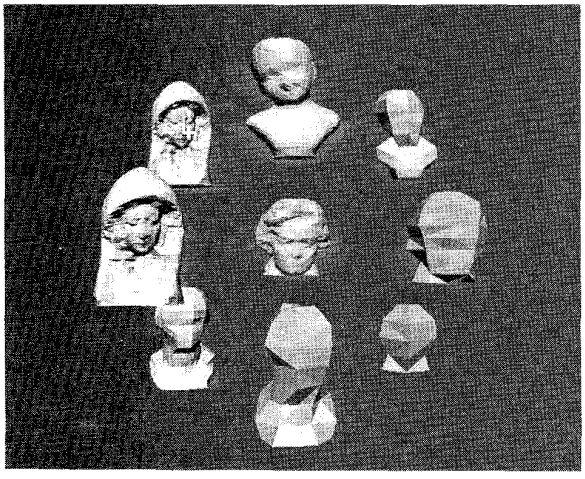
When the operator gazes at the fixation cursor, decrement in image quality at the peripheral area is only slightly recognized. Moreover, the display speed has been improved by a factor of five: the rendering time per frame is reduced from 230 msec per frame to 50 msec per frame.

Fusional vision

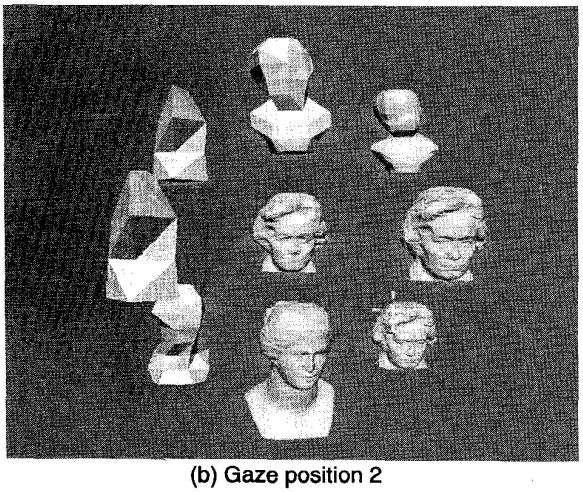
The k_1 and k_2 are set to 1.0. The operator performs the same task as in the previous experiment.

Figure 9 (a) shows the simulated result when the operator puts his gaze on the front object. The figure is composed of a pair of stereoscopic images. The image for the right eye is superimposed on the image for the left eye so that the fixation points in both images correspond to each other. The figure emulates a stereoscopic visual sensation in which we can see a single image around the fixation point and double images at the periphery. In Figure 9 (b), the operator moves his/her gaze to the object behind the fixation point. The results show that the single image around the fixation point is drawn precisely, while the double images are simplified.

The performance obtained through our experiments was



(a) Gaze position 1



(b) Gaze position 2

Figure 8. Rendered images considering central / peripheral visual factor.

similar to that obtained in our experiments on central / peripheral vision.

Total effects

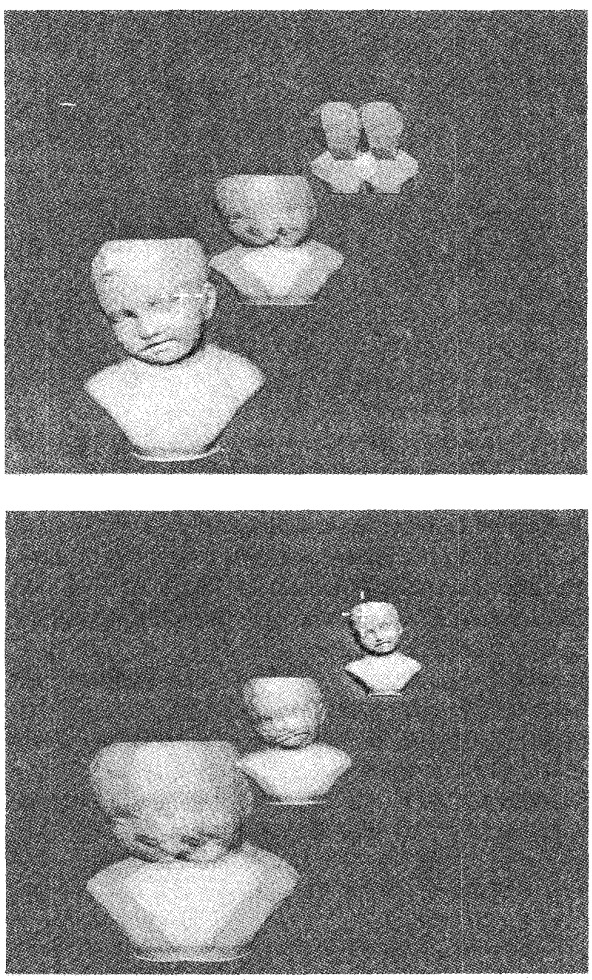
Eq. (4) or (5) can be used to combine the three visual sensations to obtain the visual acuity. Figure 10 shows the difference between the effects obtained by the two methods. Figure 10 (a) shows a result when Eq. (4) is employed, and (b) shows a result with Eq. (5).

Experimental results demonstrate that Eq. (4) shows better effect, that is, the operators recognize less degradation in image quality. This is because Eq. (5) tends to select lower LOD, and unnecessary degradation is often seen on the periphery.

5. Conclusions

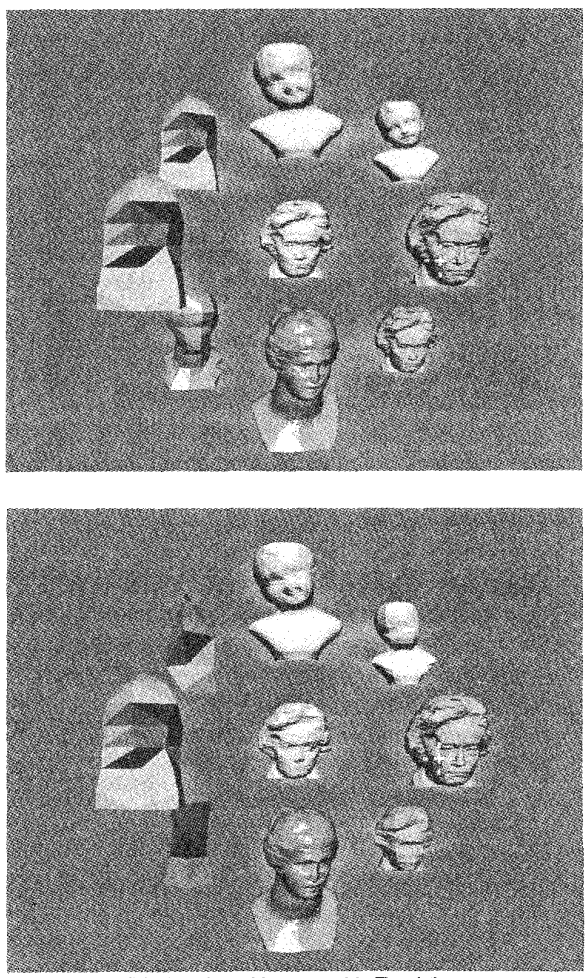
This paper proposes a new technique of adaptive rendering for virtual reality and describes its testbed system. The method adaptively selects a level from a set of hierarchical geometric models based on gaze information. Thus, simplified shape models are used to render unimportant areas that the operator is not focusing on. As a consequence, it reduces the burden of the rendering process and maintains the quality of generated images.

Our approach adopts three human visual characteristics: the central / peripheral vision, the kinetic vision, and the fusional vision. Based on the physiological backgrounds, they were modeled in Eqs. (1), (2), and (3). These computational models contains several parameters. The



(b) Gaze position 2

Figure 9. Rendered images considering fusional visual factor.



(b) Rendered image with Eq. (5)

Figure 10. Rendered images considering three visual factors.

parameter values were chosen carefully from human subjective observations. We confirmed that each of three models worked successfully as expected.

On the other hand, with regard to integration of three models, there is no theoretical basis. Thus, we examined two equations, Eqs. (4) and (5). In our experiments, Eq. (4) gave the better performance. It should be noted that further studies might find a much better solution.

The next step is to evaluate our approach quantitatively so that it could be applied to various practical tasks. For this purpose, it is necessary to define certain objective criteria on which the parameters and the integration method should be evaluated.

In the acquisition and utilization of gaze information, there are a couple of points to be improved. As the next stage of our research, we have the following plans:

(1) The current system uses head tracking as the substitute of gaze tracking. Under this situation, we obtained very promising results. In order to make the system better, especially to fully utilize the characteristic of fusional vision, the true gaze information is essential. Thus, we are developing the eyewear device with gaze tracking function shown in Figure 6.

(2) In the current system, the levels of hierarchical geometric models are switched object by object. However, there may be cases where it is more effective to control the details of the parts of one object separately. In such cases, it may be necessary to blend images to de-emphasize the border of resolution changes [13].

Acknowledgments

The authors would like to thank Mr. Shinji Uchiyama and Mr. Masakazu Fujiki for their cooperative works and useful discussions.

References

[1] J.M. Airey, J.H. Rohlf, and F.P. Brooks Jr., "Towards image realism with interactive update rates in complex virtual building environments," *Computer Graphics*, Vol. 24, No.3 (*Proc. SIGGRAPH'90*), pp.41-50 (1990).

[2] S.J. Teller and C.H. Sequin, "Visibility preprocessing for interactive walkthroughs," *ibid*, Vol. 25, No. 4 (*Proc. SIGGRAPH'91*), pp.61-69 (1991).

[3] J.H. Clark, "Hierarchical geometric models for visible surface algorithms," *Comm. ACM*, Vol.19, No.10, pp.547-554 (1976).

[4] J. Foley, A. van Dam, S. Feiner, and J. Hughes, *Computer Graphics - Principles and practice.* 2nd ed., Addison-Wesley (1990).

[5] T.A. Funkhouser, C.H. Sequin, and S.J. Teller, "Management of large amount of data in interactive building walkthroughs," *Computer Graphics* (*Special Issue on 1992 Symposium on Interactive 3D Graphics*), pp.11-20 (1992).

[6] T.A. Funkhouser and C.H. Sequin, "Adaptive display algorithm for interactive frame rates during visualization of complex virtual environments," *Computer Graphics* (*Proc. SIGGRAPH'93*), pp.247-254 (1993).

[7] E.H. Blake, "A metric for computing adaptive detail in animated scenes using object-oriented programming," in *Euro-graphics'87*, G. Marechal (Ed.), Elsievier Science Publishers, B.V., North-Holland, 1987.

[8] N. Kato and A. Okazaki, "A high-speed display method for 3-D object world based on optimal shape simplification," *IEICE Trans.*, Vol.J76-D-II, No.8, pp.1712-1721(1993) (in Japanese).

[9] K. Kitajima and Y. Yusa, "Study on saving display time for a real-time visual simulator based on the methods of grouping and multi-representation of shapes," *ibid*, Vol.J77-D-II, No.2, pp.311-320 (1994) (in Japanese).

[10] B.J. Schachter (Ed.), *Computer Image Generation.* John Wiley and Sons, New York (1983).

[11] R.S. Kalawsky, *The Science of Virtual Reality and Virtual Environments.* Addison-Wesley, (1994).

[12] L.R. Young and D. Sheena, "Survey of eye movement recording methods," *Behavior Research Methods, Instruments and Computers*, Vol.7, No.5, pp.397-429 (1975).

[13] M. Levoy and R. Whitaker, "Gaze-directed volume rendering," *Computer Graphics*, Vol. 24, No.2 (*Special Issue on 1990 Symposium on Interactive 3D Graphics*), pp.217-223 (1990).

[14] S. Uchiyama, H. Yamamoto, and H. Tamura, "Hierarchical shape representation with adaptive meshes from a range image," *IPSJ Trans.*, Vol.36, No.2, pp.351-361 (1995) (in Japanese).

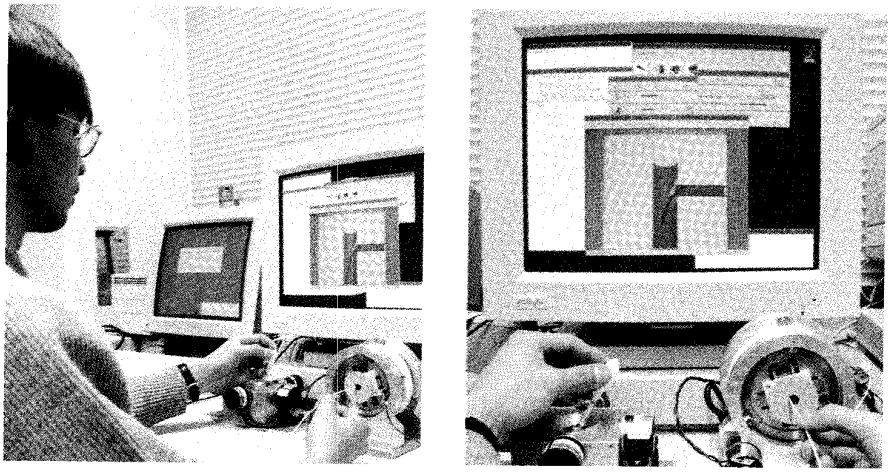


Plate 7: ***Distributed Virtual...*** – pg. 79; Fig. 12: Outline of the Simulator

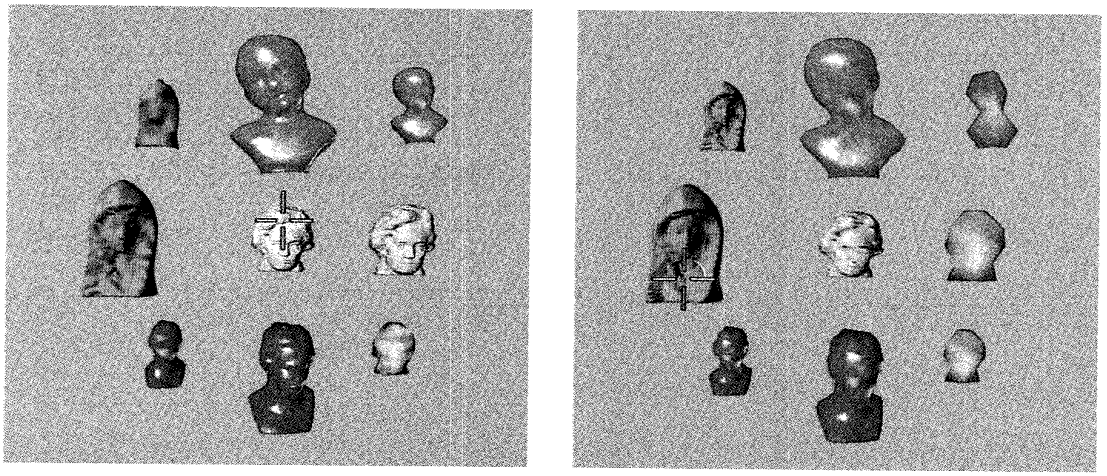


Plate 8: ***Gaze-Directed...*** – pg. 103; Figs.: Adaptively Rendered Images Using Two Gaze Directions

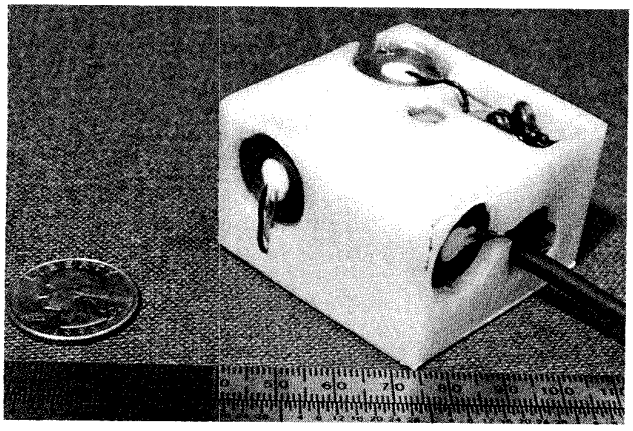


Plate 9: ***Inertial Head-Tracker...*** – pg. 185; Fig. 1: MIT Inertial Tracker 2nd Prototype