Synthetic Image Generation with a Lens and Aperture Camera Model

MICHAEL POTMESIL
Rensselaer Polytechnic Institute
and
INDRANIL CHAKRAVARTY
Schlumberger-Doll Research

This paper extends the traditional pinhole camera projection geometry used in computer graphics to a more realistic camera model which approximates the effects of a lens and an aperture function of an actual camera. This model allows the generation of synthetic images which have a depth of field and can be focused on an arbitrary plane; it also permits selective modeling of certain optical characteristics of a lens. The model can be expanded to include motion blur and special-effect filters. These capabilities provide additional tools for highlighting important areas of a scene and for portraying certain physical characteristics of an object in an image.

Categories and Subject Descriptors: I.3.3 [**Computer Graphics**]: Picture/Image Generation; I.3.7 [**Computer Graphics**]: Three-Dimensional Graphics and Realism

General Terms: Algorithms

Additional Key Words and Phrases: camera model, lens and aperture, point-spread function, defocused optical system

1. INTRODUCTION

In the past few years several algorithms have been developed for realistic rendering of complex three-dimensional (3D) scenes on raster displays [3, 6, 12]. These algorithms have been generically called *hidden-surface* algorithms since they display only the visible surfaces from a given vantage point. All these algorithms, however, have continued to use the *pinhole camera* projection geometry which was developed for display of 3D line drawings on vector devices.

This paper was first presented at SIGGRAPH '81, the Eighth Annual Conference on Computer Graphics and Interactive Techniques, August 3–7, 1981, Dallas, Texas. An earlier version was first published in *Computer Graphics*, vol. 15, no. 3 (August 1981), pp. 297–305. Four papers from SIGGRAPH '81 were published in the first issue of TOG; two are being published in this second issue. The present version of this paper is an expansion and revision of the original.

This work was performed by the authors as part of their doctoral research in the Image Processing Laboratory, Rensselaer Polytechnic Institute. The research described here was supported by the National Science Foundation, Automation, Bioengineering, and Sensing System Program, under Grant ENG 79-04821.

Authors' addresses: M. Potmesil, Image Processing Laboratory, Rensselaer Polytechnic Institute, Troy, NY 12181; I. Chakravarty, Schlumberger-Doll Research, Old Quarry Road, Ridgefield, CT 06877.

The purpose of this paper is to develop a more complex camera model, which, although computationally more expensive, provides the means for generating more realistic synthetic images closely approximating a scene imaged by an actual camera. These synthesized images are suitable for display only on raster devices.

The purpose of generating such synthetic images, which in a sense incorporate the constraints of an optical system and the imaging medium, is twofold.

(1) It gives the ability to direct the viewers' attention to a particular segment of the image, that is, it allows selective highlighting either through focusing or some optical effects.
(2) It permits adaptation of many commonly used cinematographic techniques for animated sequences, such as fade in, fade out, uniform defocusing of a scene, depth of field, lens distortions, and filtering.

The objective in this paper is to model only those features which can be used to some advantage for special effects. No attempt is made to model flaws inherent in a lens (such as optical aberrations or the lens transfer function) or the characteristics of the imaging medium.

The image generation process described here consists of two stages. In the first stage, a hidden-surface processor generates point samples of intensity in the image using a pinhole camera model. Each sample consists of image plane coordinates, red, green, and blue intensities, depth distance, and identification of the visible surface. In the second stage, a postprocessor converts the sampled points into an actual raster image. Each sampled point is converted into a *circle of confusion* whose size and intensity distribution are determined by the depth of the visible surface and the characteristics of the lens and aperture model. The intensity of a pixel is computed by accumulating the intensity distributions of the overlapping circles of confusion for all sample points. A circle of confusion and its intensity distribution may be stretched in the image plane along the projected path of a moving surface to approximate motion blur caused by a finite exposure time. Special-effect filters, such as star or diffraction filters, may be convolved with the image at this stage.

2. THE CAMERA MODEL

2.1 Camera Geometry

The pinhole camera model provides a geometric transformation of a point in a 3D scene to a point in a two-dimensional (2D) image plane. This transformation is specified by a transformation matrix which is a function of several camera parameters. These parameters are the pan, tilt, and swing angles of the camera, the location of the center of projection, the focal length, and the size of the image frame (Figure 1). The camera transformation is a perspective mapping of a point from the global homogeneous coordinate system $O(x, y, z, w)$, in which a scene is defined, to the image-plane homogeneous coordinate system $O''(x'', y'', w'')$,

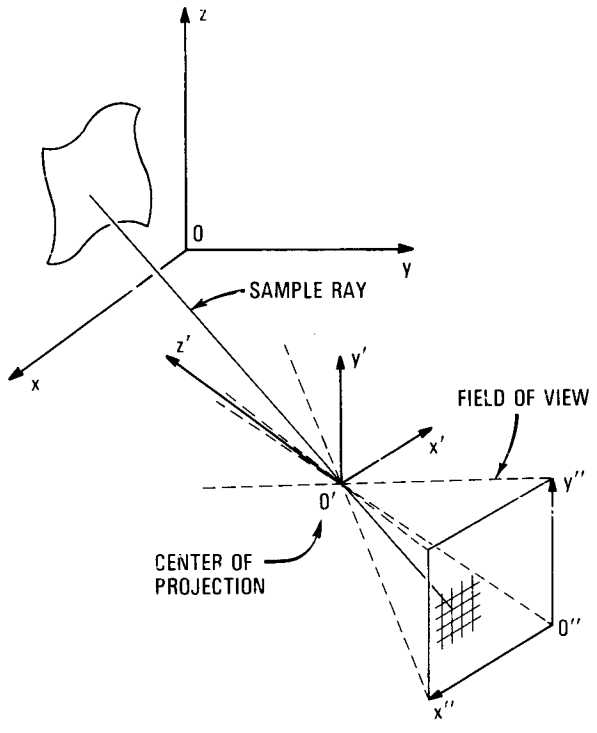


Fig. 1. Camera geometry.

and can be expressed by the following matrix [5, 10]:

$$\begin{bmatrix} x''w'' \\ y''w'' \\ w'' \end{bmatrix} = \begin{bmatrix} t_{11} & t_{12} & t_{13} & t_{14} \\ t_{21} & t_{22} & t_{23} & t_{24} \\ t_{31} & t_{32} & t_{33} & t_{34} \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}. \tag{1}$$

The pan, tilt, and swing angles specify the direction in which the camera is aimed, the center of projection is the vantage point relative to the $O(x, y, z, w)$ coordinate system, and the focal length along with the size of the image frame specify the viewing pyramid. Points in $O(x, y, z, w)$ which lie outside the viewing pyramid are clipped from the image plane.

2.2 The Finite Aperture Camera Model

The basic law governing image formation through a lens can be described by the lens formula used in geometrical optics:

$$\frac{1}{D} + \frac{1}{V} = \frac{1}{F} \tag{2}$$

where D is the object distance, V is the image distance, and F is the focal length of the lens, all measured along the optical axis SS' (Figure 2). We add to this

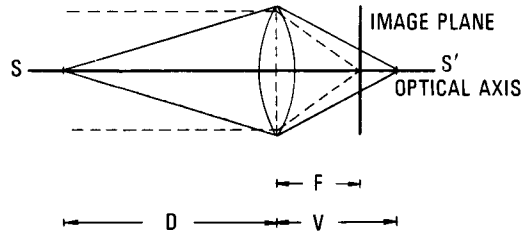


Fig. 2. Lens law.

basic lens model an aperture function, which limits the lens diameter and restricts the location of the image plane to the focal plane of the lens. The introduction of these two constraints allows the notion of focusing the lens (by moving it relative to the fixed image plane for object distances other than infinity) and, associated with it, a depth of field. It should be stated that the notion of depth of field (i.e., some objects appearing to be in focus while others are out of focus) is a function of both the optics and the resolving capability of the observer and the imaging medium.

We assume the accepted standard that for a viewing distance D the smallest area resolvable by the human eye has a diameter of $D/1000$. This means that anything larger than this diameter is viewed as a patch rather than as a point. Alternatively, one could say that the angular resolving power of the eye is 1 mil. On the basis of this criterion, the depth of field can be derived as follows. Assume that at distance D we have a patch diameter $D/1000$ which is in focus, that is, we can resolve it as a point. This is illustrated in Figure 3. Let the diameter of the patch be MN. Extending the patch in both directions, toward and away from the lens, we obtain triangles AMN and BMN. Note that LR is the effective lens diameter, defined as the focal length divided by the aperture number (F/n). The resulting extension of the patch MN, denoted by D^+ and D^-, on the optical axis (Figure 3), also remains in focus since their respective diameters are less than $D/1000$. Since ΔALR is similar to ΔAMN and ΔBLR is similar to ΔBMN, we obtain the following two equations:

$$\frac{D^+ - D}{MN} = \frac{D^+}{LR} \quad \text{and} \quad \frac{D - D^-}{MN} = \frac{D^-}{LR} \tag{3}$$

Solving for D^+ and D^-, and noting that the lens diameter $LR = F/n$, we obtain

$$D^+ = D \Big/ \left(1 - \frac{D}{(F/n)1000}\right) \tag{4}$$

and

$$D^- = D \Big/ \left(1 + \frac{D}{(F/n)1000}\right) \tag{5}$$

The following observations can now be made:

(1) If $D \rightarrow (F/n)\,1000$, then $D^+ \rightarrow \infty$ and $D^- \rightarrow (F/n)\,500$. The distance $(F/n)\,1000$ is a close approximation to the hyperfocal distance of a lens. A lens focused at this distance yields the maximum depth of field.

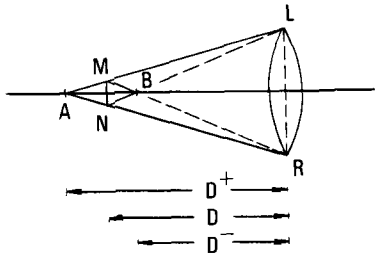


Fig. 3. Depth of field.

(2) As the effective lens diameter is decreased, by increasing the aperture number n, the hyperfocal distance becomes smaller in magnitude yielding greater depth of field.

Let us denote the hyperfocal distance as H. Then if we are focused at some plane D, the limits of the depth of field are given by

$$D^+ = \begin{cases} \dfrac{DH}{H-D}, & H > D \\ \infty & H \leq D \end{cases} \tag{6}$$

$$D^- = \frac{DH}{H+D}. \tag{7}$$

2.3 Image Focusing

In this section we develop a measure of how well an object point appears to be in focus on the image plane. Since the image plane is fixed at the focal length F, a point that is out of focus will converge on a plane away from F, projecting onto the image plane as a circle rather than a point. This circle is called the *circle of confusion* and is a measure of how defocused the image point is. When we are focused at some object distance D, our earlier calculations show that D^+ and D^- distances exist which also appear to be in focus. This implies that the corresponding image distances, V, V^+, and V^-, all form circles on the image plane whose diameters are less than or equal to $F/1000$. Thus, observed from a distance F, the images of these points are resolved by the eye as points and appear to be in focus.

The diameter of the circle of confusion for each point in the scene can be expressed as follows. Let V_D and V_P be the image distances corresponding to object distances D and P, respectively. V_P converges to a point on the image plane, whereas V_D projects into a circle as it converges a distance $|V_P - V_D|$ away. Following Figure 4, we note that ΔLRA and ΔEBA are similar. Thus we have

$$\frac{LR}{V_D} = \frac{EB}{V_D - V_P} \quad \text{where} \quad V_D = \frac{FD}{D-F} \quad D > F$$
$$V_P = \frac{FP}{P-F} \quad P > F. \tag{8}$$

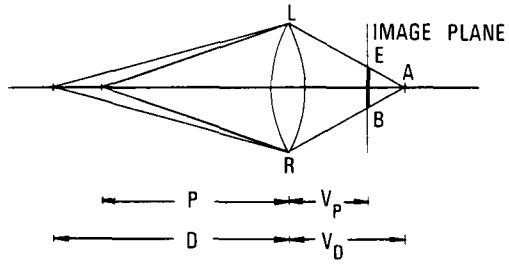


Fig. 4. Circle of confusion.

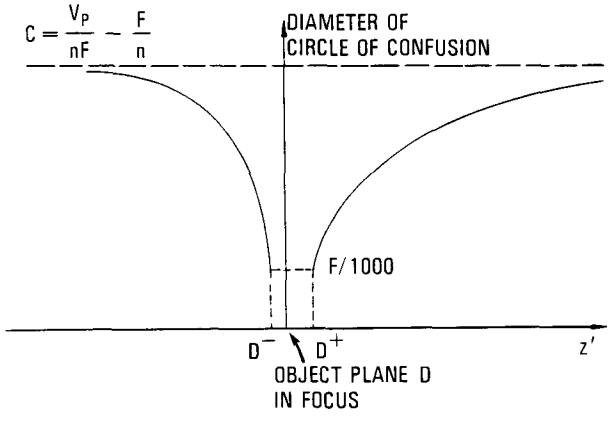


Fig. 5. Asymmetry of depth of field.

Since $LR = F/n$ and solving for EB, the diameter of the circle of confusion C, we obtain

$$C = | V_D - V_P | \frac{F}{n V_D}. \tag{9}$$

Note that as $D \rightarrow P$, the plane in focus, the diameter of the circle of confusion C approaches zero. Points at infinity approach a limiting diameter given by $(V_P/nF) - (F/n)$. The diameter of the circle of confusion is highly asymmetric as one moves away from the plane of focus in the two directions along the optical axis (Figure 5).

2.4 Diffraction Effects

In geometrical optics it is assumed that light rays always travel in straight-line paths. However, in examining the image of an aperture illuminated by a point source of light, one finds an intensity distribution pattern characterized by light and dark areas. This spreading of the image of a point source is due to diffraction; the shape, size, and intensity distribution of the pattern characterize the degradation that occurs due to the imaging system. Our objective here is to determine the intensity distribution within the circle of confusion for defocused points. We

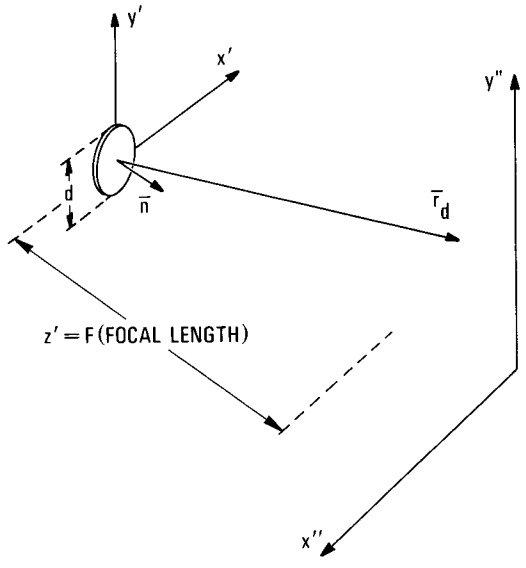


Fig. 6. Diffraction by a circular aperture.

present only some key results in formulating this distribution function; a complete derivation of optical imaging from principles of diffraction theory may be found in [4, 8].

Consider the generalized optical imaging system shown in Figure 6. The aperture consists of a circular opening in a planar opaque screen in the camera coordinate system $O'(x', y', z' = 0)$ and the image plane $O''(x'', y'')$, parallel to the screen, placed at a distance $z' = F$ (focal length) behind the aperture. Furthermore, we assume the existence of a thin lens placed behind the aperture, whose transfer function is simply a phase transformation and does not affect the transmittance. Consider the field distribution for a monochromatic wave $\xi(P, t) = \mathrm{Re}[\mathbf{U}(P)e^{-j\omega t}]$ where $\mathbf{U}(P)$ is a complex function of position given by

$$\mathbf{U}(P) = A(P)e^{-j\phi(P)} \tag{10}$$

and $A(P)$, $\phi(P)$ are the amplitude and phase of the wave at position P, respectively, and Re is shorthand notation meaning *the real part of*. Based on the Huygens-Fresnel principle, the propagation of the wave over the distance F, whose extent is limited by the aperture, can be described in terms of a linear space-invariant system. Thus the field amplitude $\mathbf{U}''$ on the image plane (x'', y'') can be expressed by a 2D convolution of the input wave field distribution $\mathbf{U}'$ and the impulse response of the system h evaluated over the area of the aperture:

$$\mathbf{U}''(x'', y'') = \iint_{\mathrm{circ}(r'(d/2))} h(x'', y''; x', y')\mathbf{U}'(x', y')\, dx'\, dy' \tag{11}$$

where the impulse response of the system is given by

$$h(x'', y''; x', y') = \frac{1}{j\lambda} \frac{e^{jk\bar{r}_d}}{\bar{r}_d} \cos(\bar{n}, \bar{r}_d)$$

and the wave number $k = 2\pi/\lambda$. Note also that $\bar{n}$ is a unit vector in the direction of the optical axis, and $\bar{r}_d$ is a vector from the center of the aperture to the observation point on the image plane whose magnitude is $|\bar{r}_d| = [(z')^2 + (x'' - x')^2 + (y'' - y')^2]^{1/2}$ (Figure 6). The integral (11) is evaluated over the extent of the aperture

$$\text{circ}\left(\frac{r'}{d/2}\right) = \begin{cases} 1 & r' \leq d/2 \\ 0 & \text{otherwise} \end{cases}$$

where $r' = [(x')^2 + (y')^2]^{1/2}$ is the radial distance in the $z' = 0$ plane of $O'(x', y', z', w')$. If the distance $|\bar{r}_d|$ is large compared to the diameter of the aperture and the region of observation is limited to around the optical axis (the Fresnel approximation [8]), then the obliquity factor $\cos(\bar{n}, \bar{r}_d) \approx 1$ and $\mathbf{U}''(x'', y'')$ reduces to the Fourier transformation of the portion of field subtended by the lens aperture. Assuming a unit-amplitude, normally incident plane wave, a circular aperture, and radius coordinates $r' = [(x')^2 + (y')^2]^{1/2}$ and $r'' = [(x'')^2 + (y'')^2]^{1/2}$, then

$$\mathbf{U}''(r'') = \frac{e^{jkF}}{j\lambda F} \exp\left[\frac{jk(r'')^2}{2F}\right] \mathbf{H}\{\mathbf{U}'(r')\} \tag{12}$$

where $\mathbf{H}$ is the Hankel transform (due to the radial symmetry) of the aperture circ $[r'/(d/2)]$. The light intensity (infinite time-average of the amplitude) for monochromatic light is given in this case by $I(P) = |\mathbf{U}(P)|^2$. Thus

$$I(r'') = \left(\frac{kd^2}{8F}\right)^2 \left[2\frac{J_1(kdr''/2F)}{kdr''/2F}\right]^2 \tag{13}$$

which is the well-known Airy pattern, where J_1 is the Bessel function of the first kind and the first order. The Airy pattern is shown in Figure 7a.

The effect of the thin lens in the imaging system causes the spherical waves to converge on the image plane. If the lens law (2) is satisfied, then it has been shown [8] that the impulse-response of the optical system with a circular aperture is simply the Airy pattern (13) centered at the observation point on the image plane. The impulse-response $h(x'', y''; x', y')$ is thus space invariant depending only upon coordinate differences $(x'' - x', y'' - y')$ and magnification of the optical system V/D. The impulse-response is called the *point-spread function* (PSF) and describes the form and egree of degradation in the imaging system [1, 8, 11].

The derivation of intensity distribution for defocused monochromatic images was first investigated by E. Lommel in 1885. Details of this derivation, based on the Huygens-Fresnel principle, may be found in [4]. The results derived by Lommel can be used directly with minor changes in interpretation. The Huygens-Fresnel integral used to determine the light intensity in the circle of confusion

cannot be obtained in a closed form and must be evaluated in terms of Lommel functions. The light intensity in the circle of confusion is given as follows:

$$I(u, v) = \left(\frac{2}{u}\right)^2 [U_1^2(u, v) + U_2^2(u, v)]\, I_0 \tag{14}$$

where

$$I_0 = \left[\frac{kd^2}{8F^2}\right]^2$$

and U_1 and U_2 are Lommel functions. These functions can be evaluated by a series approximation given by

$$U_m(u, v) = \sum_{s=0}^{\infty} (-1)^s \left(\frac{u}{v}\right)^{(m+2s)} J_{(m+2s)}(v) \tag{15}$$

where $J_{(m+2s)}$ is the Bessel function of the $m + 2s$ order. Asymptotic and recurrence expressions of Lommel functions are given in [9]. The variables u and v are dimensionless and specify the position of a point within the circle of confusion. They are expressed as follows:

$$u = k\left(\frac{d}{2F}\right)^2 z' \tag{16}$$

$$v = k\frac{d}{2F} r'' = k\frac{d}{2F}\sqrt{(x'')^2 + (y'')^2}. \tag{17}$$

As a result of the expression derived by Lommel, the following observations can be made:

(1) At $u = 0$, the distribution of light intensity of a point focused on the image plane is given by

$$\left[2\frac{J_1(v)}{v}\right]^2 I_0$$

which is the previously derived Airy pattern.

(2) Consider a plane at object distance P which is not in focus. This implies that the image of a point on this plane will converge at some image distance V_P. If the plane S is in focus, then $V_P - V_S$ is the distance by which the plane P is out of focus. To find the intensity distribution from the resulting circle of confusion, note that $z' = V_P - V_S$ or

$$u = k\left(\frac{d}{2F}\right)^2 (V_P - V_S), \tag{18}$$

and

$$v = k\frac{d}{2F} r'' \tag{19}$$

gives the position in the circle of confusion.

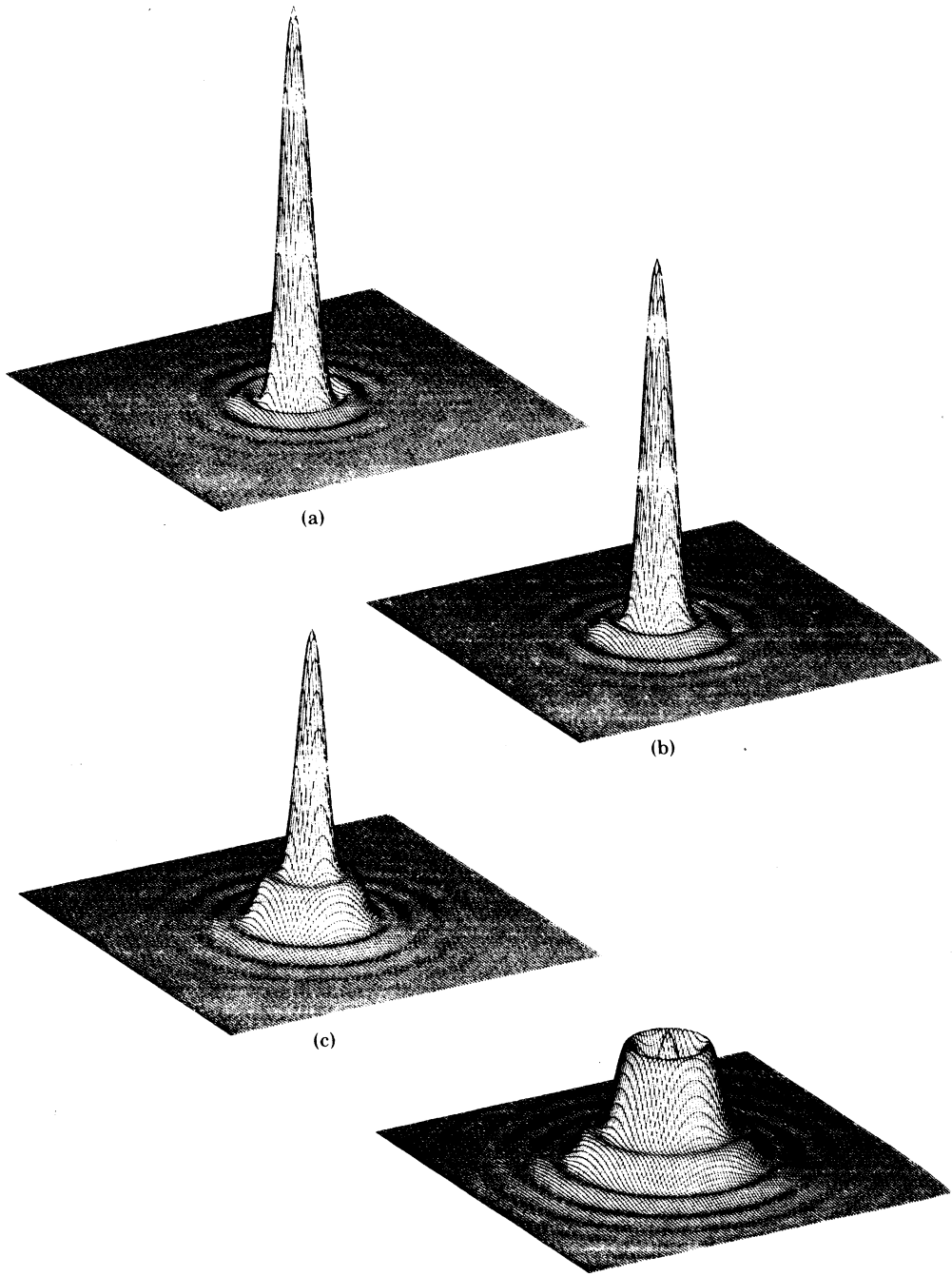
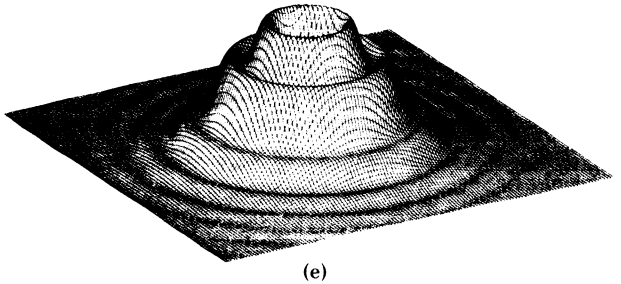
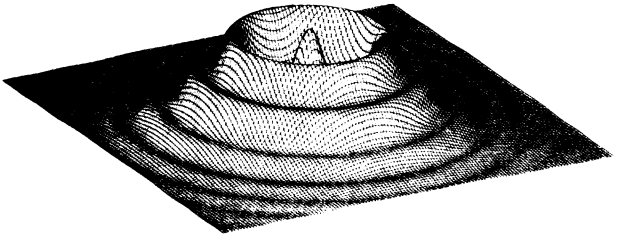


Fig. 7. The intensity-distribution function $I(u, v)$ plotted for parameter values (a) $u = 0$ (Airy pattern), (b) $u = \pm\pi$, (c) $u = \pm 2\pi$, (d) $u = \pm 3\,\pi$.

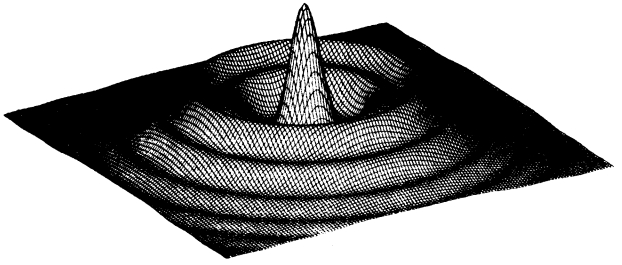
Figure 7 continued on next page.



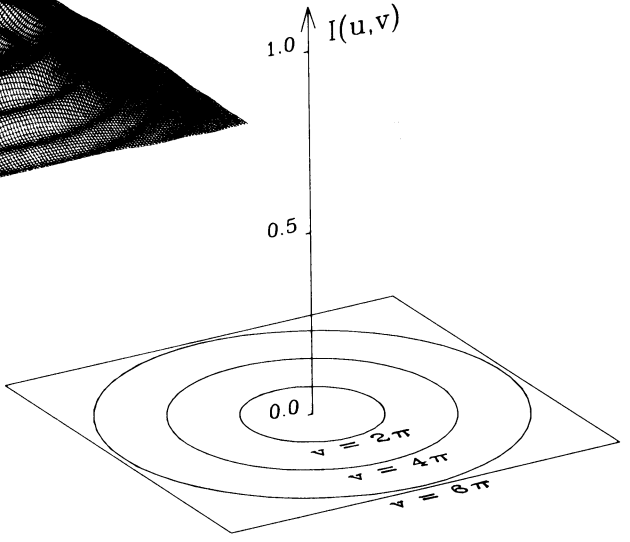
(e)



(f)



(g)



(h)

Fig. 7. The intensity-distribution function $I(u, v)$ plotted for parameter values (e) $u = \pm 4\pi$, (f) $u = \pm 5\pi$, (g) $u = \pm 6\pi$, and (a–g) $0 \leq v \leq 6\pi$; (h) scale of the plots.

(3) Note that for $v = 0$, the light distribution at the center of the circle of confusion, we obtain

$$I(u, 0) = \left[\frac{\sin(u/4)}{(u/4)}\right]^2 I_0. \tag{20}$$

This shows that the highly peaked light distribution of the Airy pattern obtained for points in focus reduces in amplitude until it actually becomes zero at $u/4 = \pm \pi$ or $z' = \pm 16 \pi F^2/kd^2$. Thus, the distribution changes significantly for points out of focus, resulting in a lack of contrast for defocused points.

Several intensity-distribution functions with different values of u are shown in Figure 7. Since the intensity-distribution functions are dependent upon the depth of the point, it is evident that we cannot model the PSF of the defocusing phenomenon to be space invariant. Thus h is a function of x'', y'', x', y', z'; the response of such a space-variant system must be calculated explicitly for each image point.

3. SYNTHETIC IMAGE GENERATION

3.1 Hidden-Surface Processor

The image formation process consists of two stages. In the first stage, a ray-tracing hidden-surface processor generates point samples of light intensity in the 3D scene within the field of view of the camera model. These samples can be directly converted by this processor into a raster image using the pinhole camera model, or they can be passed to a second processor which converts them into a raster image that is focused and has a depth of field. The hidden-surface processor uses Whitted's recursive illumination model [12], based on geometrical optics, to produce images which contain surface-to-surface reflections, refractions in transparent surfaces, illumination by point light sources, and shadows in complex 3D scenes. Planar, quadric, and bicubic surface elements can be rendered by the program at this time.

A raster image computed by this program, using the pinhole camera model, is considered to be an array of square pixels[1] lying in the image plane; that is, each pixel is a finite square area of the image. The 3D scene is sampled at the four corner points of each pixel. A 3D ray from a sample point in the image plane to the center of projection is extended into the scene and intersected with the nearest surface element (Figure 1). The intensity of the sample point is evaluated from visual properties of the intersected surface element, as well as from intensity information provided by additional rays which are bounced from the intersection point in the reflection, refraction, and light source directions. Information about each reflected and refracted ray is maintained in a node of a binary shading tree [12]. Finally, the intensity of a pixel is computed by averaging its four point samples. This is equivalent to fitting a bilinear function into the four sample values and integrating the function over the pixel area. Anti-aliasing is performed

[1] For simplicity we consider here only square pixels. The hidden surface and focus processors can handle rectangular pixels equally well.

by recursively subdividing a pixel, which has a large difference in the sampled values, into 2 × 2 square regions and repeating the sampling process at the four corners of each new region (Figure 8). The intensity of each region is again the average of its four samples and the final intensity of the pixel is the sum of all regional intensities, each weighted by its area. This subdivision is done only within pixels which contain sharp intensity changes, typically caused by edges, silhouettes, or textures.

In addition to generating the actual raster image with the pinhole camera model, the program can optionally save all the sampled values in a separate file. A sample i consists of the following information obtained from a node of a shading tree:

(1) the x_i'', y_i'' coordinates in O'', the 2D image plane coordinate system, of the sampled point,
(2) the red, green, and blue intensity values r_i, g_i, b_i, contributed by the currently intersected 3D surface point,
(3) the z_i' accumulated distance (along the camera's optical axis) in O', the 3D camera coordinate system, of the currently intersected 3D surface point, and
(4) the surface element identification number id_i, of the currently intersected 3D surface point (to be used in motion blur generation).

The parameters of the pinhole camera model are also saved in this file. While the hidden-surface processor produces the point samples to be used by the focus processor, the anti-aliasing method of subdividing selected pixels is overridden and the processor generates all samples on a uniform grid. If anti-aliasing is in effect, all pixels in the image are subdivided into square regions of the same area.

3.2 Focus Processor

The focus processor generates a raster image, which is focused and has a depth of field, from the image samples and the pinhole camera model supplied by the hidden-surface processor and from given lens and aperture parameters. The focus processor models the image formation as a *linear space-variant system* [1, 7, 8, 11] with its input and output related by

$$q(x'', y'') = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x'', y''; x_i'', y_i'', z_i')\, f(x_i'', y_i'')\, dx_i''\, dy_i'' \qquad (21)$$

where $q(x'', y'')$ is the output image intensity at point (x'', y'') due to all its input functions $f(x_i'', y_i'')$ and $h(x'', y''; x_i'', y_i'', z_i')$ is the response of the optical system (lens and aperture) at output coordinates (x'', y'') due to an input point function at (x_i'', y_i''). The impulse-response h is the PSF defined by eq. (14). The shape of this function depends on the z_i' depth of the input sample. Since the depth varies over the scene, the PSF is space variant and the input-output relationship cannot be considered a simple convolution. This means that eq. (21) has to be evaluated in the (x'', y'') spatial domain at a considerable computational cost rather than in the frequency domain.

In the discrete space, the focus processor models the output raster image as an array of square pixels. A pixel (j, k) covers image area $(x_j'', y_k'') \times (x_{j+1}'', y_{k+1}'')$ as

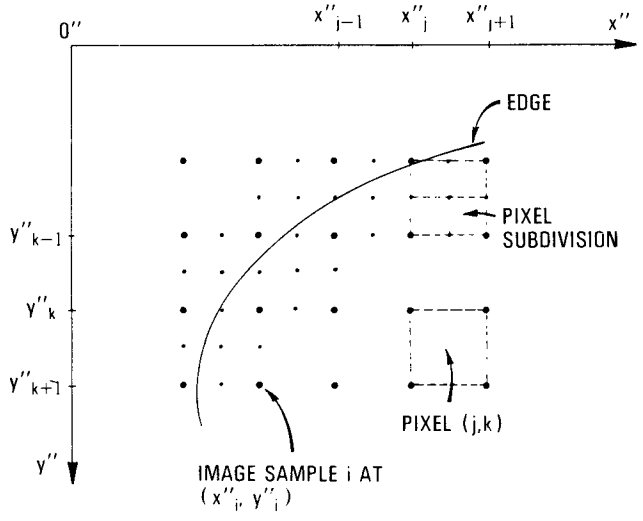


Fig. 8. Image sampling and pixel subdivision used for anti-aliasing.

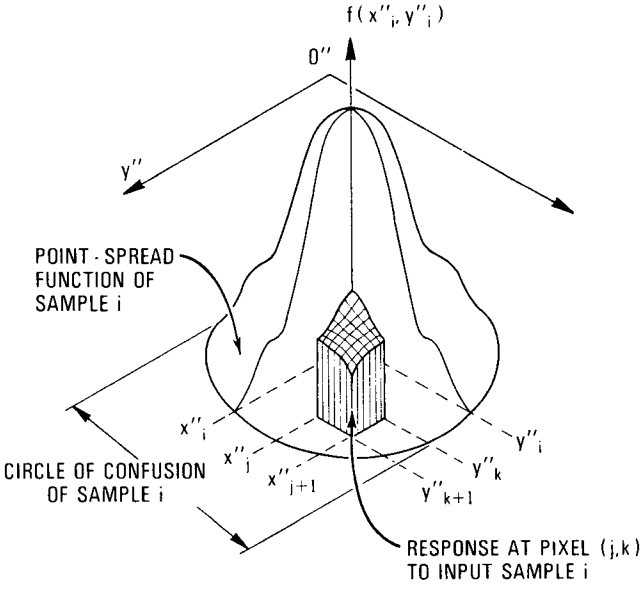


Fig. 9. System response at pixel (j, k) due to an input point sample $f(x_i'', y_i'')$.

shown in Figure 9. The input function $f(x_i'', y_i'')$ at point (x_i'', y_i'') is a delta function of magnitude equal to intensity r_i, g_i, b_i. The PSF $h(j, k; x_i'', y_i'', z_i')$ gives the response of the system over the area of pixel (j, k) due to the input function $f(x_i'', y_i'')$. This response is the integral of the normalized PSF centered at (x_i'', y_i'') over the area of the pixel (j, k). We assume that the PSF is zero

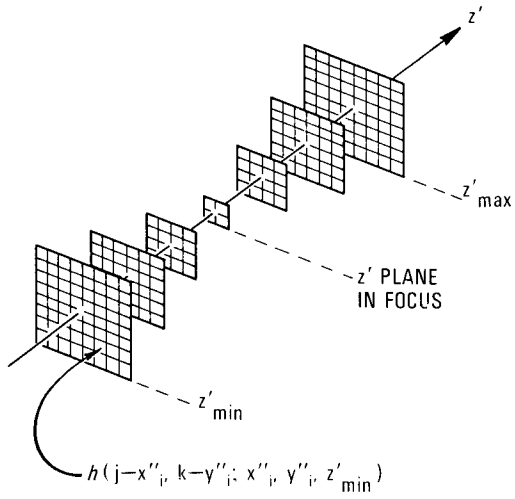


Fig. 10. Tables of point-spread functions in an image. (The number of entries in each table depends on the size of the circle of confusion at the given z' depth.)

outside the circle of confusion. The final intensity $q(j, k)$ of pixel (j, k) is computed separately for each primary color as the weighted average of all the input sample points:

$$q(j, k) = \frac{\sum_{i=1}^{N} f(x_i'', y_i'')h(j, k; x_i'', y_i'', z_i')}{\sum_{i=1}^{N} h(j, k; x_i'', y_i'', z_i')} \tag{22}$$

where N is the number of input samples in the image and h is the integral of the intensity-distribution function I of eq. (14) over pixel (j, k):

$$h(j, k; x_i'', y_i'', z_i') = \int_{x_j''}^{x_{j+1}''} \int_{y_k''}^{y_{k+1}''} I(z_i', \sqrt{(x'' - x_i'')^2 + (y'' - y_i'')^2})\, dx''\, dy'' \tag{23}$$

and the intensity-distribution function I is normalized so that

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} I(z_i', \sqrt{(x'' - x_i'')^2 + (y'' - y_i'')^2})\, dx''\, dy'' = 1. \tag{24}$$

The brightness of an image generated by the focus processor with eq. (22) is approximately the same as the brightness of the image generated by the hidden-surface processor. Brightness changes of the image caused by the film sensitivity, aperture setting, or exposure time are not modeled here, although they would be simple to add.

The focus processor proceeds in two steps. First, the program computes tables of the PSF's at equally spaced intervals along z' between the nearest and the farthest points in the image (Figure 10) for the given lens and aperture parameters. Each table contains values of the PSF, evaluated by eq. (23) at the z' coordinate of the table for all the pixels within the square area inscribing the circle of confusion. Since the PSFs are radially symmetrical, values in only one quadrant

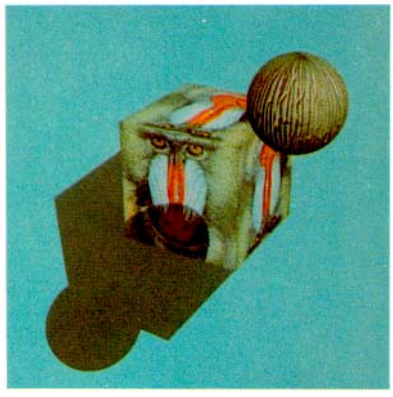


Fig. 11. Test image generated with the pinhole camera model.

are actually computed and stored. The spacing of these tables affects the program's memory requirements and execution speed. If the tables are computed at very small z' intervals, the values from the table nearest to the given sample can be used directly; otherwise the corresponding values from the two nearest tables need to be linearly interpolated.

In the second step, the program processes each input point sample by distributing its intensity to output pixels according to the values stored in the table nearest to the sample's z' depth. While computing the output image, the program maintains a block of four numbers for each pixel in the image. The first three numbers contain the numerator of eq. (22), one for each primary color, the fourth number is the denominator of eq. (22), common to all primary colors. After all the input samples have been processed, the final intensities of the three primary colors in each pixel are computed by dividing the three numerators by the common denominator.

The focus processor can perform anti-aliasing by evaluating the area intensity in eq. (22) at a resolution higher than the final pixel resolution, provided that the input point samples are also given at the higher resolution. The intensities of the subpixel regions are then simply averaged into the final pixel intensity. The execution time of the program is proportional to the number of input samples times the image area of the average circle of confusion in the image.

4. RESULTS

The focus processor has been used with different lens and aperture parameters to generate a number of versions of several test images.

The pinhole camera version of the first test image generated by the hidden surface processor is shown in Figure 11. The 3D scene consists of a 100 × 100 × 100-mm cube with an image of a mandrill mapped on its sides, a brown sphere of 20-mm radius with a texture added using Blinn's wrinkling technique [2], and a large blue background sphere which encloses the cube, the brown sphere, the camera model, and the light source. The distances of the brown sphere, the cube, and the background from the camera are approximately 290, 550, and 980 mm,

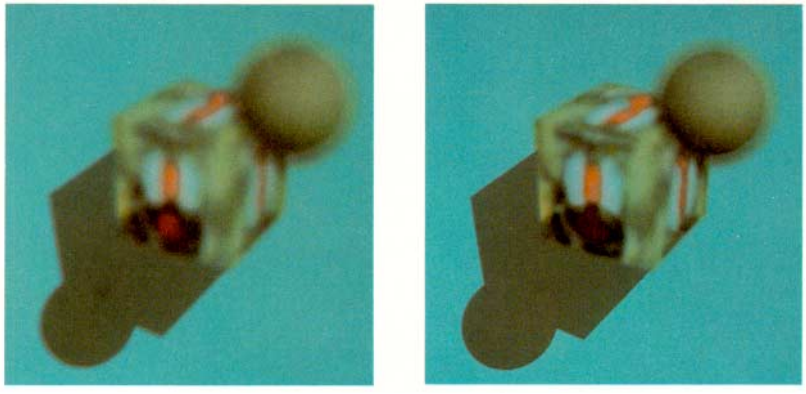
Table 1. Summary of Lens and Aperture Parameters Used to Generate Images in Figures 12–14

Figure	Aperture number n	Focused distance D (mm)	D^+ (mm)	D^- (mm)	Depth of field (mm)
12a	5.6	2000	2511	1662	849
12b	5.6	980	1089	891	198
12c	5.6	550	583	521	62
12d	5.6	290	299	282	17
13a	11	2000	3333	1429	1904
13b	11	980	1219	819	400
13c	11	550	618	495	123
13d	11	290	308	274	34
14a	22	2000	10000	1111	8889
14b	22	980	1612	704	908
14c	22	550	705	451	254
14d	22	290	328	260	68

Note: Focal length is 55 mm. Lens diameter is 9.82 mm at f/5.6, 5.00 mm at f/11, and 2.50 mm at f/22.

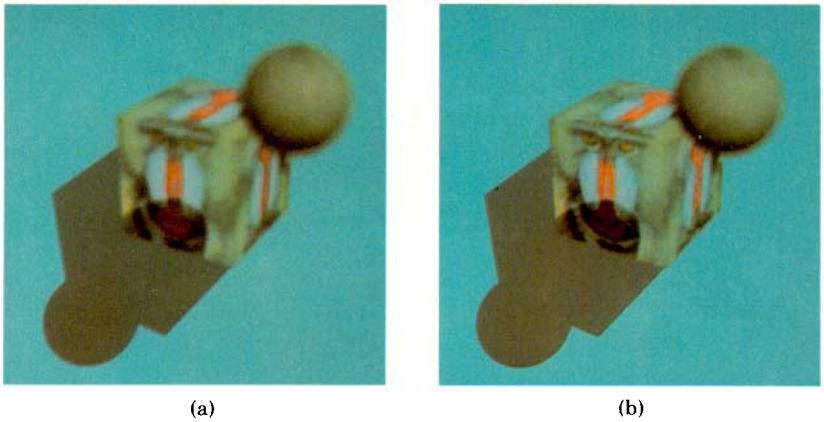
respectively. The camera model uses a 55-mm lens and a 30 × 30-mm image area. The focus processor generated the twelve images in Figures 12–14 from samples supplied by the hidden-surface processor. In all three figures, each image (a) is focused at 2000 mm (all objects in the image are out of focus), each image (b) is focused at 980 mm (the blue background), each image (c) is focused at 550 mm (the cube), and each image (d) is focused at 290 mm (the brown sphere). The four images in each figure were generated with the same aperture setting: f/5.6 in Figure 12, f/11 in Figure 13, and f/22 in Figure 14; that is, the aperture was increased by 2 and 4 stops (with corresponding "increases in exposure time" to maintain a constant image brightness) in Figures 13 and 14, respectively. Table 1 gives a summary of the focused distances, the aperture settings, and the depths of field in these images. From the table values and the associated images, one can observe an increase in the depth of field due to increasing aperture settings and a decrease in the depth of field due to shorter focused distances.

The last two images illustrate the effect of the lens and aperture camera model on reflected (Figure 15) and refracted (Figure 16) rays of light. Figure 15 displays an opaque carafe, modeled by 80 bicubic patches, standing on a tiled floor and illuminated by two light sources. The camera model has a 40-mm lens and a 30 × 30-mm image area. Figure 15a was computed using the pinhole camera model and Figure 15b was generated by the focus processor. The camera model is focused on the carafe with an aperture setting of f/1.4. The depth of field at this aperture is 110 mm; each tile measures 50 × 50 mm. In this image, while the carafe is in focus, the reflection of the floor on the surface of the carafe near its base becomes gradually defocused as the ray distance (along the camera's optical axis) of the reflected tiles increases. Similarly, the reflection of the carafe on the floor remains approximately in focus, whereas the floor itself in the vicinity of this reflection becomes defocused with decreasing distance to the camera. Figures 16a and 16b are identical in content to Figures 15a and 15b, respectively, except for the surface of the carafe, which has been modeled as transparent glass. Figure 16b shows the propagation of light rays through multiple transparent surfaces and the resulting defocusing of the tiled floor when viewed through the transpar-



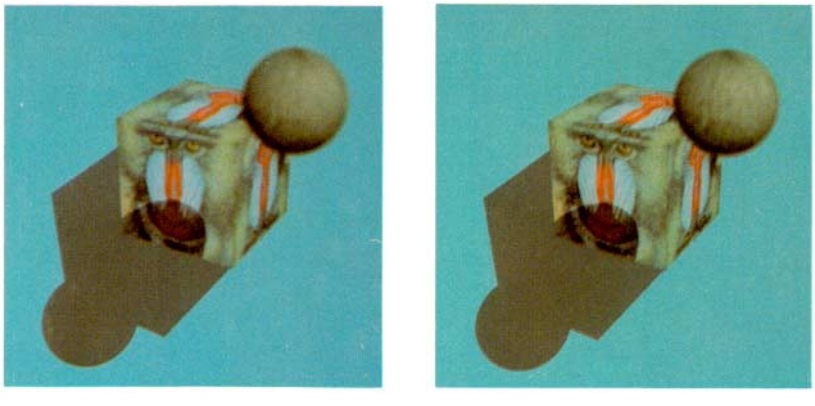
(a) (b)

Fig. 12. (a) $D = 2000$ mm, $n = f/5.6$, (b) $D = 980$ mm, $n = f/5.6$.



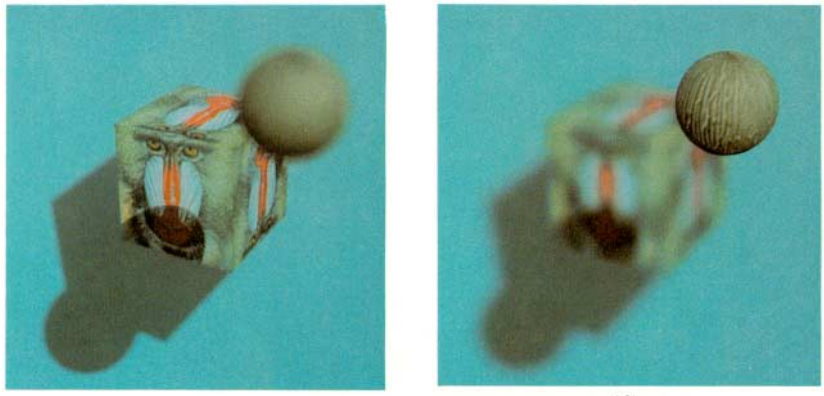
(a) (b)

Fig. 13. (a) $D = 2000$ mm, $n = f/11$, (b) $D = 980$ mm, $n = f/11$.



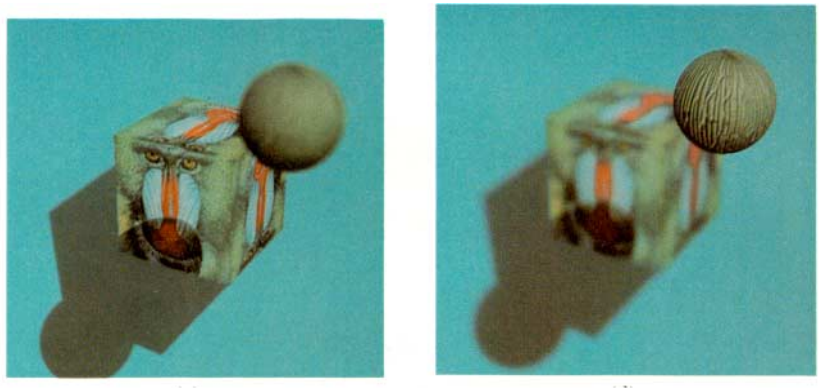
(a) (b)

Fig. 14. (a) $D = 2000$ mm, $n = f/22$, (b) $D = 980$ mm, $n = f/22$.



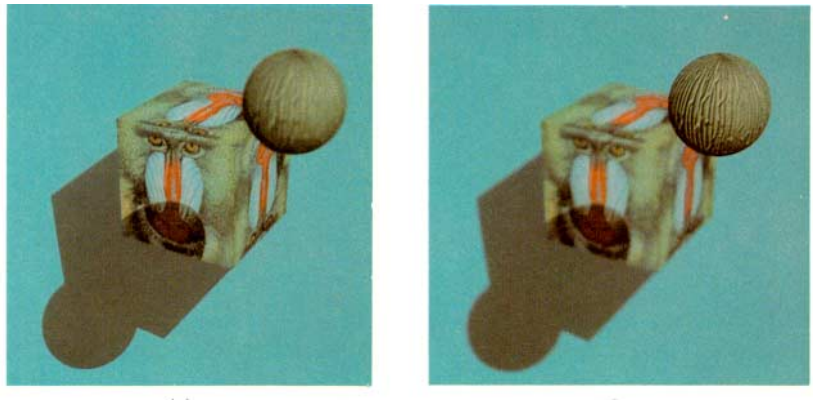
(c) (d)

Fig. 12. (c) $D = 550$ mm, $n = f/5.6$, (d) $D = 290$ mm, $n = f/5.6$.



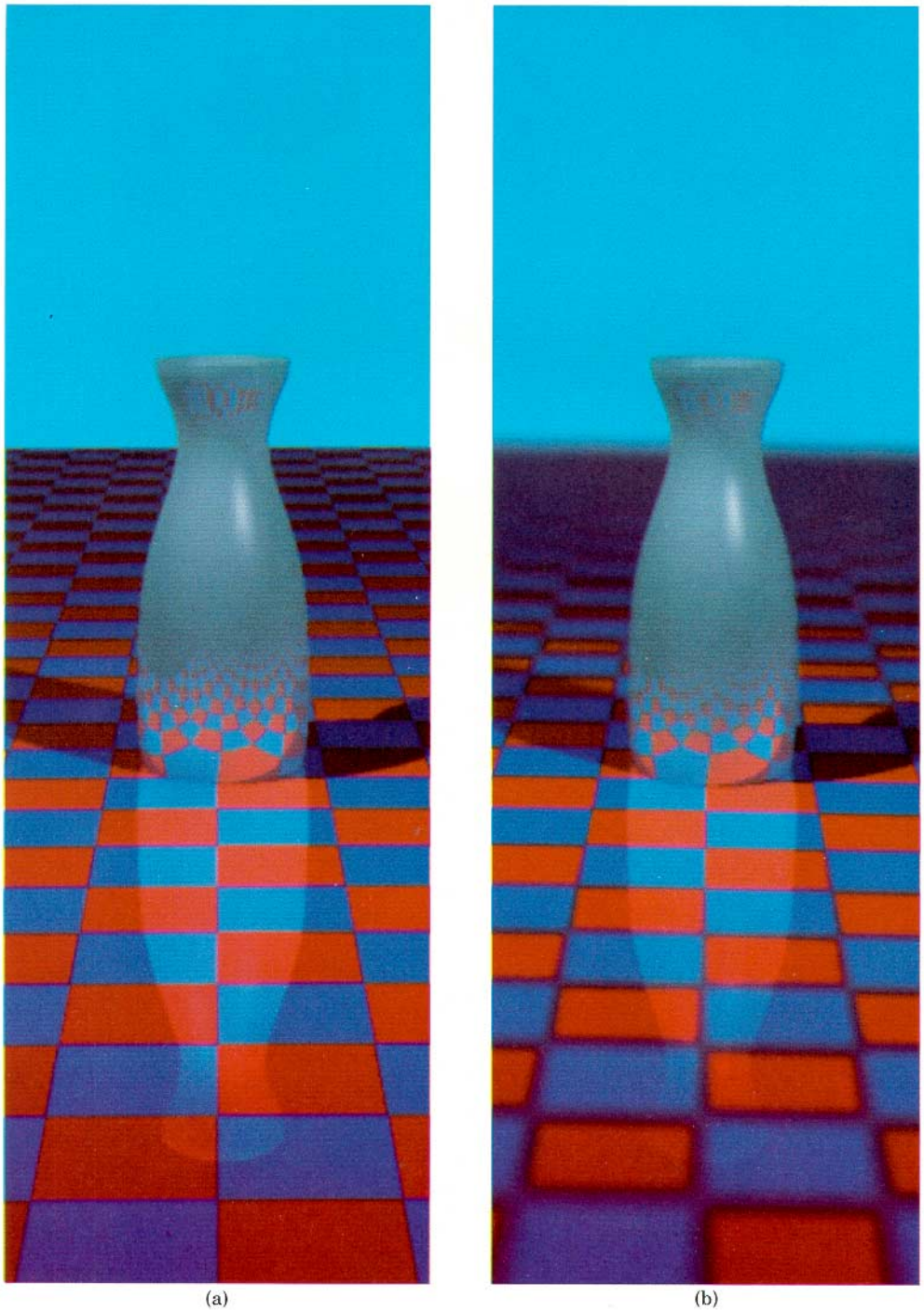
(c) (d)

Fig. 13. (c) $D = 550$ mm, $n = f/11$, (d) $D = 290$ mm, $n = f/11$.



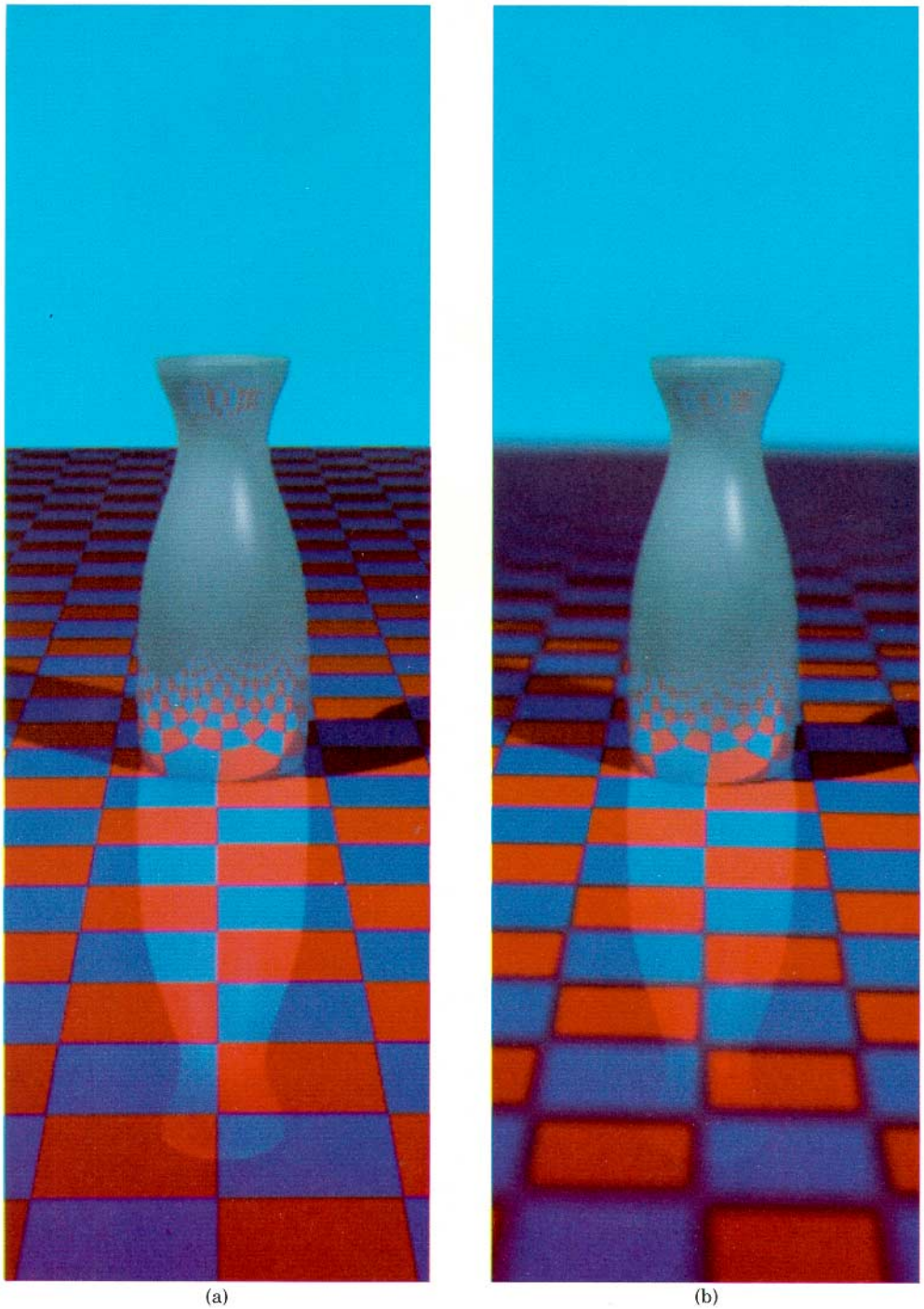
(c) (d)

Fig. 14. (c) $D = 550$ mm, $n = f/22$, (d) $D = 290$ mm, $n = f/22$.



(a) (b)

Fig. 15. Image of an opaque carafe generated with (a) the pinhole camera model and (b) the lens and aperture camera model. (The images shown are sections of the original images computed at 512 × 512 resolution.)



(a) (b)

Fig. 16. Image of a transparent carafe generated with (a) the pinhole camera model and (b) the lens and aperture camera model. (The images shown are sections of the original images computed at 512 × 512 resolution.)

Table 2. Summary of Program Performance for Image in Figure 15

Aperture number n	CPU time (minutes)
≥16	2.75
11	3
8	6
5.6	8
4	15
2.8	33
2	63
1.4	125

ent carafe. Notice that the defocusing and depth of field are caused entirely by the parameters of the camera's lens; any lens effect that may be caused by the glass carafe itself is not being modeled.

All images shown here were computed at 512 × 512 pixel resolution with 24 bits of intensity per pixel. Images produced by both processors were anti-aliased. In the hidden-surface processor, the anti-aliasing mechanism normally subdivides each pixel whose intensity is more than 10 percent different from any of its samples into 2 × 2 square regions and resamples each region. In the pinhole version of the first test image (Figure 11), the hidden-surface processor computed approximately 350,000 samples, which is about 100,000 samples more than would be required at this resolution without anti-aliasing. The image samples supplied by the hidden-surface processor to the focus processor were on a uniform square grid of 1025 × 1025 points. The focus processor evaluated image intensity in 1024 × 1024 square regions which were then averaged into 512 × 512 pixels. Tables of the PSFs were computed at 1-mm intervals along z' between the nearest and farthest points in an image. Values from the table closest to a given sample were directly used in eq. (22), rather than values interpolated between the two closest tables.

Both processors have been implemented in FORTRAN77 and run on a Prime 750 computer. The Prime 750 is similar in processing power to a VAX-11/780. Table 2 lists execution times of the focus processor for a series of aperture settings used on the image in Figure 15. The times are given for 512 × 512 pixel images without anti-aliasing. As expected, the execution time approximately doubles with each aperture step as the area of the aperture opening as well as the areas of the circle of confusion and the sizes of the tables of PSFs double.

5. SUMMARY

This paper has developed a lens and aperture model suitable for more realistic display of complex 3D scenes on raster devices. The purpose of this model is to provide new techniques for selectively implementing the optical effects of a lens and aperture on synthetically generated imagery. The model has been applied to the output of a ray-tracing hidden-surface program to generate images focused on various planes with associated depths of field as a function of the camera

parameters. This model could also be applied to a similar output of a depth-buffer hidden-surface algorithm. Ray tracing, however, allows focusing of rays which are reflected or refracted by multiple surfaces. The lens and aperture model has been described in terms of generalized imaging-system parameters. Further extensions of this work should include the generation and description of PSFs for modeling motion blur and special-effect camera filters.

There are, however, many limitations in the current approach. Ray tracing, which models light propagation as infinitely thin rays traveling in straight-line paths, cannot efficiently simulate effects such as light dispersion and scattering (attributed to the wave nature of light) or illumination by other than point light sources. Many of these effects are essential for realistic image generation. This paper has attempted to include the effects of diffraction in a ray-tracing model; the other effects would be difficult to add. The computational cost of using the model described in this paper is high. However, it is possible to develop special hardware and take advantage of the inherent parallelism in many of the computations involved. It may also be possible to approximate the effects of a lens and aperture with a PSF that is space invariant, in which case many of the computations could be done in the frequency domain.

GLOSSARY

Symbol	*Description*
$A(P)$	Amplitude of field distribution at position P
C	Diameter of circle of confusion
d	Diameter of aperture
D	Object distance
D^+	Farthest object distance in focus
D^-	Nearest object distance in focus
$f(x_i'', y_i'')$	Input image intensity r_i, g_i, b_i at point (x_i'', y_i'')
F	Focal length of lens
h	Impulse response of an optical system (PSF)
H	Hyperfocal distance of a lens
$\mathbf{H}(\cdot)$	Hankel transform
$I(u, v)$	Intensity-distribution function
$J_m(v)$	Bessel function of the first kind
k	Wave number
n	Aperture number
$\bar{n}$	Unit vector in the direction of the optical axis
$O(x, y, z, w)$	3D global coordinate system
$O'(x', y', z', w')$	3D camera coordinate system
$O''(x'', y'', w'')$	2D image plane coordinate system
$q(x'', y'')$	Output image intensity at point (x'', y'')
$\bar{r}_d$	Vector from aperture to the point of observation
r_i, g_i, b_i	Red, green, and blue intensity components of sample point i
r'	Radial distance in the $z' = 0$ plane of $O'(x', y', z', w')$
r''	Radial distance in $O''(x'', y'', w'')$

$[t_{ij}]$	Camera transformation matrix
$\mathbf{U}(P)$	Field distribution of wave at position P
$\mathbf{U}'$	Input wave field distribution
$\mathbf{U}''$	Image plane field distribution
$U_m(u, v)$	Lommel function of two variables
V	Image distance
z'	Depth of a 3D point
$\phi(P)$	Phase of field distribution at position P
λ	Wavelength of light
$\xi(P, t)$	Field distribution of a monochromatic wave at position P and time t

ACKNOWLEDGMENTS

The authors would like to thank Professor Herbert Freeman for his encouragement and patience, and for providing all the machine cycles needed to experiment with fuzzy images. The authors have also benefited from conversations with Turner Whitted of Bell Laboratories and Alan Barr of the Center for Interactive Computer Graphics at RPI.

REFERENCES

1. ANDREWS, H.C., AND HUNT, B.R. *Digital Image Restoration.* Prentice-Hall, Englewood Cliffs, N.J., 1977.
2. BLINN, J.F. Simulation of wrinkled surfaces. *Computer Gr. 12,* 3 (Aug. 1978), 286–292.
3. BLINN, J.F., CARPENTER, L.C., LANE, J.M., AND WHITTED, T. Scan line methods for displaying parametrically defined surfaces. *Commun. ACM 23,* 1 (Jan. 1980), 23–34.
4. BORN, M., AND WOLF, E. *Principles of Optics.* 3rd ed., rev. Pergamon, London, 1965, Chap. 8.
5. CARLBOM, I., AND PACIOREK, J. Planar geometric projections and viewing transformations. *Comput. Surv. 10,* 4 (Dec. 1978), 465–502.
6. CATMULL, E. A subdivision algorithm for computer display of curved surfaces. Tech. rep. UTEC-CSc-74-133, Comput. Sci. Dep., Univ. of Utah, 1974.
7. DAINTY, J.C., AND SHAW, R. *Image Science.* Academic Press, New York, 1974.
8. GOODMAN, J.W. *Introduction to Fourier Optics.* McGraw-Hill, New York, 1968, Chaps. 4, 5.
9. GRAY, A., MATTHEWS, G.B., AND MACROBERT, T.M. *A Treatise on Bessel Functions and Their Applications to Physics.* 2nd ed., Macmillan, London, 1922. Chap. 14.
10. NEWMAN, W.M., AND SPROULL, R.F. *Principles of Interactive Computer Graphics.* McGraw-Hill, New York, 1973.
11. ROSENFELD, A., AND KAK, C. *Digital Picture Processing.* Academic Press, New York, 1978.
12. WHITTED, T. An improved illumination model for shaded display. *Commun. ACM 23,* 6 (June 1980), 343–349.

Received August 1981; revised October and November 1981; accepted November 1981