Camera Models and Optical Systems Used in Computer Graphics: Part I, Object-Based Techniques

Brian A. Barsky[1,2,3], Daniel R. Horn[1], Stanley A. Klein[2,3]
Jeffrey A. Pang[1], and Meng Yu[1]

[1] Computer Science Division
[2] School of Optometry
[3] Bioengineering Graduate Group
University of California, Berkeley, California, 94720-1776, USA

http://www.cs.berkeley.edu/optical

Contact author: barsky@cs.berkeley.edu

Abstract. Images rendered with traditional computer graphics techniques, such as scanline rendering and ray tracing, appear focused at all depths. However, there are advantages to having blur, such as adding realism to a scene or drawing attention to a particular place in a scene. In this paper we describe the optics underlying camera models that have been used in computer graphics, and present *object space* techniques for rendering with those models. In our companion paper [3], we survey *image space* techniques to simulate these models. These techniques vary in both speed and accuracy.

1 Introduction

The goal of increasing realism fuels progress in computer graphics. However, perfect realism is impossible to achieve simply due to the number of interactions that occur when sending light into a scene and collecting the results on a discrete plane. The human eye can discern many details; thus, humans can notice the surreal clarity of most computer-generated scenes. In particular, images often lack depictions of focus, blur, optical aberrations, and depth of field. This paper illustrates the concepts behind modeling computer graphics with a lens system.

The rendering of a 3D scene requires transforming from *object space* to *image space*. In object space, the entire geometry of the scene, with all the materials and surfaces, is present. This transformation is parameterized by a camera angle and location. There are various algorithms that can be used to translate the 3D geometry into a final, 2D image. When the scene is transformed into a 2D image, it often retains a depth map which contains distance measurements to the camera at any given pixel; this may be used for additional post-processing algorithms. Two ubiquitous methods to translate a 3D model in object space into a 2D image are *scanline rendering* and *ray tracing*. However, images generated with these methods are rendered in focus at all distances, as if photographed through a pinhole camera with an infinitely tiny aperture. Thus, these standard rendering techniques cannot convey focus nor represent arbitrary lens configurations.

The rendering of images with focal blur is an interesting and important field of research in computer graphics. Several techniques have been proposed to model camera lens systems to varying degrees of accuracy. They range from *object-based* modifications to ray tracing and scanline rendering to *image-based* techniques that convolve and otherwise distort an image after it has been rendered. We discuss object based techniques in this paper, and image based techniques in our companion paper [3], published later in these proceedings. Currently, these techniques offer varying realism and speed in rendering images with camera models. Before studying them, some background on optics will be reviewed.

2 Camera Models

Several camera models have been employed in computer graphics to approximate physical optical systems. Although each has distinctive qualities and produces characteristically different images, they all share some common traits.

A camera model simulates the capture of light from a three-dimensional scene in *object space* onto a two-dimensional image, or *image space*. Most models contain or approximate a system of parallel *lenses* such as that of a camera or the eye. An axis that passes through the geometric center of the system of lenses is called the *optical axis*. In computer graphics, the center of a single lens system is sometimes called the *center of projection* (COP).

Objects in the scene are projected through the lens system to form an *image* located on the opposite side of the system. Each lens has an *aperture* which defines the area through which light is allowed to pass to the image.

Although rendering models usually consider an *image plane* in front of the lens system when forming an image, camera models consider a *film plane* behind the lens system. Like physical optical systems, the image formed on the film plane is inverted. An appropriate image plane can be derived from the location of the film plane in some camera models. The field of view is parameterized by the size of the film plane.

In optics, the standard unit of lens power is the *diopter* [1], a unit which is measured in inverse meters. Although the amount blur is not linear in distance, it is approximately linear in diopters.

2.1 Camera Obscura (Pinhole Camera)

The standard rendering algorithms in computer graphics are equivalent to modeling a *camera obscura* or pinhole camera. In this model, all the light rays from the scene pass through a single point, called the center of projection, or through a lens with an infinitesimal aperture.

Since only a single light ray cast from each point in the scene can pass through the COP regardless of its location in space, each point will be rendered exactly

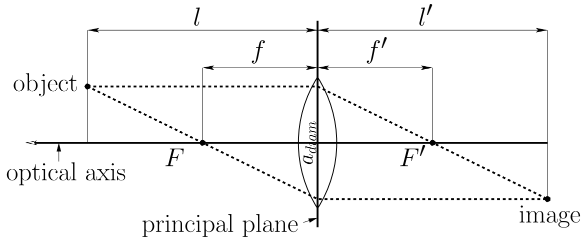

Fig. 1. Thin lens model.

once on the film plane. This causes all objects in the scene to appear in sharp focus on the image produced. The pan of the camera can be parameterized by the location of the film plane along the axis.

This model could not be realized physically since in reality, a point-sized aperture would produce an image on the film plane that would be too dim to observe. However, simply reversing this model yields a ray tracer, where a ray is projected from the COP through each pixel on the image plane and then the colors of the materials it intersects determine the color of the particular pixel. Analytic techniques have also been developed using this model to project objects to image space efficiently [7].

2.2 Thin Lens Approximation

In reality, all lenses have a finite aperture, and hence each point in the scene emits a cone of light which is visible to the lens. In geometric optics, when light rays encounter the boundary between two media, the angle through which light is refracted can be calculated using Snell's law. Since real lenses are composed of a refractive material that has a *finite* and non-constant thickness, Snell's law can be used to project light rays through a lens.

The *thin lens approximation* instead assumes that even though lenses have a finite aperture, they have an *infinitesimal* thickness. In general, thin lenses are also spherical in form, either convex or concave, and *ideal*. A lens is ideal if it has the property that the change in slope for a ray passing through the lens is proportional to the distance from the center of the lens to the point at which the ray encounters it [1]. The plane that is normal to the optical axis at the lens is called the *principal plane* (Figure 1).

Along the optical axis of the lens is the *focal point*, F. Formally, F is an axial point having the property that any ray emanating from it or proceeding toward it travels parallel to the axis after refraction by the lens [8]. A secondary focal point, F' exists on the opposite side of the lens. Analogously, this point is defined as an axial point having the property that any incident ray traveling parallel to the axis will proceed toward F' after refraction (Figure 1).

The distance between the center of the lens and a focal point is called the *focal length*. If the medium is same on both sides of the lens (hence the indices of refraction, n and n', are equal) and if the lens is symmetric, then these distances, denoted by f and f' are the same.

The law governing image formation through a thin lens in Gaussian form is:

$$\frac{1}{l} + \frac{1}{l'} = \frac{1}{f} \tag{1}$$

where l is the distance from the object to the lens, l' is the distance from the lens to the image, and f is the focal length (Figure 1). By solving for l' in equation (1), we can derive that the projection of an object at a distance l in front of the lens will converge at a distance

$$l' = \frac{fl}{l - f} \tag{2}$$

behind the lens.

The *f-stop* (or F number), which we denote as f_{stop}, is defined as the ratio of the focal length to the diameter of the aperture of the lens, denoted by a_{diam}:

$$f_{stop} = \frac{f}{a_{diam}} \ . \tag{3}$$

When focusing at a particular distance d_{focus} in front of the principal plane, there is a specific location d'_{focus} where the film plane must be located such that each point that is at distance d_{focus} forms its image on that plane (Figure 2). The plane perpendicular to the axis at d_{focus} is called the *focal plane*. To derive the distance d'_{focus} behind the lens to place the film plane, we substitute d_{focus} and d'_{focus} for l and l', respectively, in equation (2), yielding:

$$d'_{focus} = \frac{f d_{focus}}{d_{focus} - f} \ . \tag{4}$$

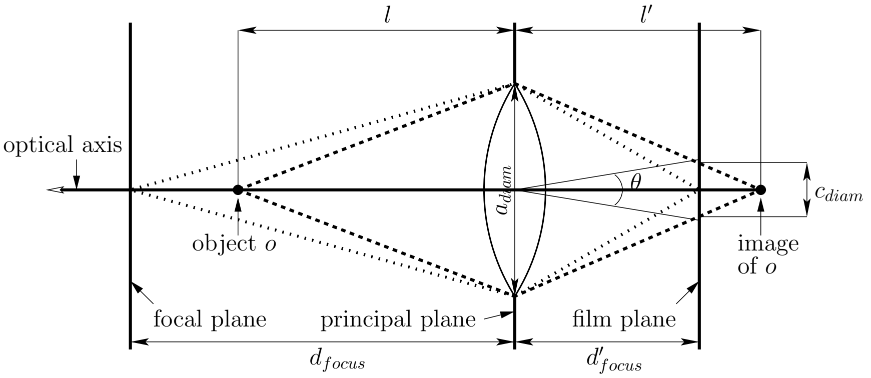

Fig. 2. The circle of confusion c_{diam} and blur angle θ in the thin lens model.

A point that is closer to the lens than the focal plane projects behind the film plane, whereas the projection of a point farther from the lens than the focal plane converges in front of the film plane. Thus, each of these projected points forms a circle rather than a point on the film plane. This is called the *circle of confusion* or blur disk, and it measures how defocused an image point is. Using similar triangles [12] in Figure 2, we can deduce from equations (2) and (4) that the diameter of the circle of confusion c_{diam} for a particular point is:

$$c_{diam} = a_{diam} \frac{l' - d'_{focus}}{l'} \ . \tag{5}$$

In optics, it is often simpler to measure blur in terms of the *blur angle* rather than the diameter of the circle of confusion. The blur angle, which we denote as θ, is the angular size, measured in radians, of a given point's blur [1]. From Figure 2, it can be seen that:

$$\tan \frac{\theta}{2} = \frac{c_{diam}/2}{d'_{focus}} \ . \tag{6}$$

When θ is small, as is usually the case for normal lenses, $\tan \frac{\theta}{2} \approx \frac{\theta}{2}$; thus equation (6) can be approximated as:

$$\theta = \frac{c_{diam}}{d'_{focus}} \ . \tag{7}$$

Substituting equation (5) for c_{diam} yields:

$$\theta = a_{diam} \frac{l' - d'_{focus}}{l' d'_{focus}} \ . \tag{8}$$

Rearranging,

$$\theta = a_{diam} \left(\frac{1}{d'_{focus}} - \frac{1}{l'} \right) \ . \tag{9}$$

The quantity in parentheses, $\frac{1}{d'_{focus}} - \frac{1}{l'}$, measures the change in curvature due to a change in the location of the object or a change in the location of the image. This is called *diopters of defocus* and is denoted by $D_{defocus}$ [1]. Using this notation, equation (9) can be written as:

$$\theta = a_{diam} D_{defocus} \ . \tag{10}$$

2.3 Thick Lens Approximation

The thin lens approximation is appropriate for the case where the thickness of the lens is small relative to the focal length. However, some lenses do not satisfy this assumption and most lens systems that comprise multiple "lens elements," like those in photographic systems, cannot be accurately approximated by a single thin lens. However, it is sometimes possible to treat these systems as ideal

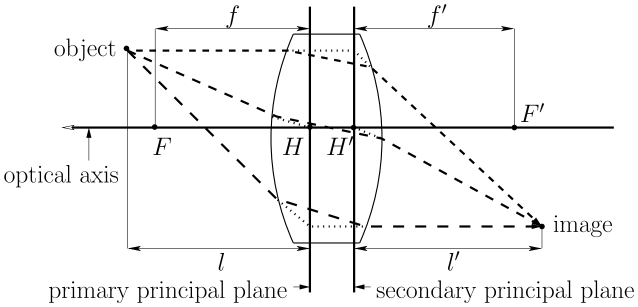

Fig. 3. Thick lens model. The dashed lines show the actual paths of rays passing through the lens. The dotted lines demonstrate that we can calculate exit directions for the rays in a similar manner as in the thin lens case with the addition of a translation between the principal planes.

thick lenses. A thick lens is usually described by a single homogeneous lens with two spherical surfaces separated by an appreciable distance.

A thick lens can be characterized by *primary* and *secondary principal planes*, H and H', and their corresponding *primary* and *secondary focal points*, F and F'. The two principal planes are defined in the same manner as was the single one in the case of the thin lens. The primary principal plane is the plane that is normal to the optical axis where any ray cast from the primary focal point F intersects the corresponding axis-parallel ray exiting the lens. The secondary principal plane is defined similarly with respect to the secondary focal point F'.

As with the thin lens, the primary and secondary focal lengths, f and f' are defined to be the distances from H to F, and from H' to F', respectively (Figure 3).

The principal planes are usually determined by tracing rays through the lens system, although they can also be calculated analytically using lens thickness formulas [10]. Once we know the location of the principal planes and focal lengths, we can determine image formation through the lens. Assuming that the medium is the same on both sides of the lens, the formula for image formation is the same as equation (1) [8]. However, since there are two principal planes instead of only one (as there is in the thin lens model), here l and l' are the distances from H to the object, and from H' to the image, respectively.

Thus, the circle of confusion for a point can be derived in a similar manner to the derivation for the thin lens model (Section 2.2). We merely have to translate each ray entering the lens from the primary to secondary principal planes (in parallel to the axis) before applying the thin lens equations to determine where the corresponding image forms.

2.4 Full Lens Systems

Kolb et al. [10] argued that the previous idealized models cannot accurately simulate the behavior of particular physical optical systems. They developed a

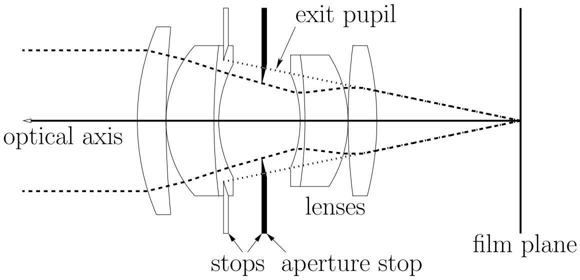

Fig. 4. Example of a full lens system. The black stop is the aperture stop for this particular point on the film plane since it most limits the incoming parallel rays of light. The exit pupil, shown by the dotted lines, defines the cone of rays from the point that pass unobstructed through the lens. (Courtesy of Craig E. Kolb. Adapted from [10] © 1995 ACM, Inc. Reprinted by permission.)

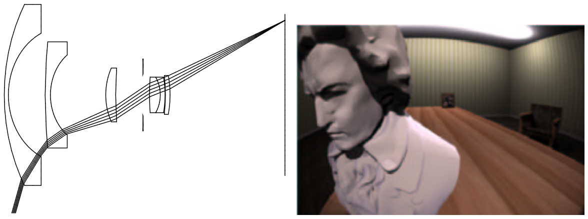

Fig. 5. Full lens simulations can exhibit aberrations that are not possible with thick lens simulations, such as appear in this image taken with a 16mm fish-eye lens. The image shows the lens' signature barrel distortion. (Courtesy of Craig E. Kolb [10] © 1995 ACM, Inc. Reprinted by permission.)

rendering technique based on a physically-based camera model that simulated a system of lenses (Figure 4).

A lens system usually comprises a series of spherical lenses and stops centered on the optical axis. A *stop* is an opaque element with a approximately circular opening that permits the passage of light. The stop that most limits the passage of rays through the system is termed the *aperture stop*. One or more of the lenses in the system usually moves relative to the film plane to change the focal point of the system. One advantage of using a lens system model is that moving a lens changes the field of view as in a physical system. The previous approximations assume that the film plane is always located at the secondary focal point and the lens can be focused at any arbitrary distance without change of configuration.

Full lens geometries can simulate an array of specific physical systems, such as that depicted in Figure 5.

Image formation through a lens system is described by Snell's law. Since the lens geometry is arbitrary, generalizations are not possible, although characteristics of the system such as the *exit pupil* can be derived to aid simulation algorithms.

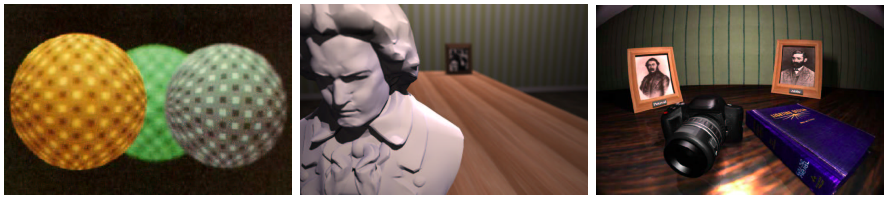

Fig. 6. Images synthesized using distributed ray tracing using a thin lens model (left), thick lens model (center), and a full lens system (right). The full lens simulation exhibits noticeable barrel distortion. (Left image courtesy of Robert L. Cook [5]. Center and right images courtesy of Craig E. Kolb [10]. All images © 1995 ACM, Inc. Reprinted by permission.)

3 Distributed Ray Tracing

Super-sampling or oversampling was originally employed as a mechanism for dampening aliasing artifacts in ray traced images by increasing the sampling rate to better approximate the analytic model of a scene. Cook et al. [5] demonstrated that distributing ray samples could be used to model other "fuzzy phenomena" as well, including depth of field and motion blur.

Actual lenses have a finite aperture and project cones of light from the scene to each point on the film plane, as described in Section 2. To approximate the cone using the thin lens approximation, the image plane is placed at the focal plane. In computer graphics, the traditional pinhole model traces a *single ray* from the COP to each pixel on the focal plane, whereas a distributed approach traces rays from *many points* on the principal plane to each pixel on the image/focal plane. Cook et al. showed that this model accurately captures the circle of confusion that was described in Section 2.2.

Although this simple model generates photographic-like effects, Kolb et al. [10] argued that the model is not precise enough to accurately approximate specific physical optical systems. They proposed a similar distributed ray tracing technique that traces rays directly through systems of lenses (a similar rendering model was subsequently described by Barsky et al. [2] for a broader set of applications). Their algorithm traces rays from each pixel on the film plane to the surface of the lens in front of it and computes new directions for rays using Snell's law. Rays travel through each lens in the system before exiting the system to then sample the scene. Their full lens model accounts for radiometry as well. The ray tracing algorithm captures the blocking of light by lens elements when rays pass through the system at large angles to the axis; thus, the images produced will correctly exhibit vignetting and other exposure effects.

Figure 6 compares images synthesized using distributed ray tracing with different lens models.

Distributed ray tracing methods for capturing optical effects such as depth of field have several benefits. First, since these approaches integrate these effects with shading and visible surface calculations, they more accurately solve the depth of field problem and do not have problems dealing with occlusion. Second, these methods are easily implemented by simply ray tracing a stochastically chosen set of rays and then averaging the results.

Nevertheless, there are several considerations when applying these methods. The most significant is the problem of deciding the number and nature of samples to be distributed over the lens to obtain an accurate value for each pixel on the image plane. Cook [4] addressed the problems that accompanied uniform sampling at regularly spaced sample points by introducing *stochastic sampling*, or sampling at random intervals. Dippe and Wold [6] analyzed two such sampling techniques: Poisson and jitter sampling. Lee et al. [11] developed a statistical model to probabilistically limit the error in sample variance and discussed stratified sampling techniques to limit the number of samples required for good approximations. Kajiya [9] discussed a Monte Carlo algorithm that he termed *uniform sequential sampling* for sampling the aperture of a lens. This method is compatible with stratified sampling techniques and converges to the analytic model of a scene. Kolb et al. [10] used a stratified strategy based on a mapping of concentric squares to concentric circles.

The sampling problem is magnified by the multi-dimensional nature of the samples. For example, in Kolb et al.'s algorithm, the origins of sample rays must be distributed on the film plane, and target points must be distributed on the nearest lens in the system. To capture motion blur, the dimension of time must be added to the sample space. Hence, the number of samples required for a good approximation can become prohibitive.

Kolb et al. found that 16 rays per pixel were required for their images. Although tracing rays through their lens systems only consumed 10% of the rendering time, a number of additional ray samples were required. For example, the center image in Figure 6 required about 90 minutes to compute on a Silicon Graphics Indigo2.

Furthermore, distributed ray tracing techniques are only applicable when the 3D geometry to a scene is available. In addition, even though these methods correctly model certain attributes of physical optical systems, they do not model arbitrary (position or time dependent) lens aberrations nor many wavelength dependent effects such as chromatic aberration. Finally, all such systems are assumed to be aberration-limited and thus ignore the effects of diffraction.

4 Summary

In this paper, we described the optics underlying camera models that have been employed in computer graphics, and presented *object space* techniques for rendering with those models. The models range from the common pinhole camera to completely specified lens system geometries. Rendering techniques that use these models trade off complexity and efficiency for accuracy and realism. In our companion paper [3], published later in these same proceedings, we survey several *image space* techniques to simulate these models, and address topics including linear filtering, ray distribution buffers, light fields, and simulation techniques for interactive applications.

Acknowledgments

The authors would like to thank Adam W. Bargteil for his helpful comments as well as Craig E. Kolb for providing figures for this paper.

References

1. David A. Atchison and George Smith. *Optics of the Human Eye*. Butterworth-Heinemann Ltd., Woburn, Mass., 2000.
2. Brian A. Barsky, Billy P. Chen, Alexander C. Berg, Maxence Moutet, Daniel D. Garcia, and Stanley A. Klein. Incorporating camera models, ocular models, and actual patient eye data for photo-realistic and vision-realistic rendering. Submitted for publication, 2003.
3. Brian A. Barsky, Daniel R. Horn, Stanley A. Klein, Jeffrey A. Pang, and Meng Yu. Camera models and optical systems used in computer graphics: Part II, Image based techniques. In *Proceedings of the 2003 International Conference on Computational Science and its Applications (ICCSA'03)*, Montréal, May 18–21 2003. Second International Workshop on Computer Graphics and Geometric Modeling (CGGM'2003), Springer-Verlag Lecture Notes in Computer Science (LNCS), Berlin/Heidelberg. (These proceedings).
4. Robert L. Cook. Stochastic sampling in computer graphics. *ACM Transactions on Graphics*, 5(1):51–72, January 1986.
5. Robert L. Cook, Thomas Porter, and Loren Carpenter. Distributed ray tracing. In Hank Christiansen, editor, *ACM SIGGRAPH 1984 Conference Proceedings*, pages 137–145, Minneapolis, July 23–27 1984.
6. Mark A. Z. Dippe and Erling H. Wold. Antialiasing through stochastic sampling. In Brian A. Barsky, editor, *ACM SIGGRAPH 1985 Conference Proceedings*, pages 69–78, San Francisco, July 22–26 1985.
7. James D. Foley, Andries van Dam, Steven K. Feiner, and John F. Hughes. *Computer Graphics: Principles and Practice, 2nd Edition*. Addison-Wesley Publishing Co., Reading, Mass., 1990.
8. Francis A. Jenkins and Harvey E. White. *Fundamentals of Optics*. McGraw-Hill, Inc., New York, 1976.
9. James T. Kajiya. The rendering equation. In *ACM SIGGRAPH 1986 Conference Proceedings*, pages 143–150, Dallas, 1986.
10. Craig Kolb, Don Mitchell, and Pat Hanrahan. A realistic camera model for computer graphics. In Robert L. Cook, editor, *ACM SIGGRAPH 1995 Conference Proceedings*, pages 317–324, Los Angeles, August 6–11 1995.
11. Mark E. Lee, Richard A. Redner, and Samuel P. Uselton. Statistically optimized sampling for distributed ray tracing. In Brian A. Barsky, editor, *ACM SIGGRAPH 1985 Conference Proceedings*, pages 61–67, San Francisco, July 22–26 1985.
12. Michael Potmesil and Indranil Chakravarty. Synthetic image generation with a lens and aperture camera model. *ACM Transactions on Graphics*, 1(2):85–108, April 1982. (Original version in ACM SIGGRAPH 1981 Conference Proceedings, Aug. 1981, pp. 297-305).