Chapter 8

Optics

This chapter covers the essentials of geometrical optics. Radiometry is covered in Chapter 9. Machine vision relies on the pinhole camera model, which models the geometry of perspective projection but omits the effects of depth of field, based on the fact that only those points within a certain depth range are in focus on the image plane. Perspective projection assumes that the view volume is an infinite pyramid, limited only by the top, bottom, and sides of the viewable rectangle on the image plane. Optical effects such as depth of field and light attenuation limit the view volume by introducing limits on the distances to viewable objects. Objects may not be seen clearly if they are too close or too far away.

The pinhole camera model assumes that the perspective rays pass through an infinitesimal aperture at the front of the camera. In practice, the aperture must be larger to admit more light. Lenses are placed in the aperture to focus the bundle of rays from each point in the scene onto the corresponding point in the image plane as indicated by the geometry of perspective projection. The scene point, the projected point on the image plane, and the center of the aperture are on the ray for the line of sight in perspective projection. A lens gathers more light, allowing the camera to work with less ambient illumination or with a faster shutter speed, but the depth of field is limited. With the exception of limited depth of field and the camera calibration problems discussed in Section 12.9, lenses do not introduce any effects that violate the assumptions of ideal perspective projection, and most machine vision algorithms do not depend on the design of the optical system for image formation.

It may be necessary to calculate the depth of field in order to select a camera lens for a particular machine vision application. This chapter will cover the lens equation and present the equations for calculating the depth of field.

8.1 Lens Equation

The lens equation relates the distance z' of the image plane from the center of the lens (optical origin), the distance z to the point in the scene, and the focal length f of the lens:

$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}. \tag{8.1}$$

When $z \to \infty$, the distance z' of the image plane from the optical origin is equal to the focal length f. The focal length is the distance of the image plane from the optical origin when parallel rays are focused to a single point in the image plane. In photogrammetry, the optical origin is called the center of projection and z' is called the camera constant (see Chapter 12 on camera calibration). When the lens is focused on points that are not at infinity (points at close range), $z' < f$, so using f to approximate z' overestimates the camera constant. The distance z' of the image plane from the optical origin approaches f as the camera focus is changed to accommodate points in the scene that are farther away.

8.2 Image Resolution

Spatial resolution is determined by pixel spacing, lens aberrations, diffraction, and depth of field. For most machine vision applications, lens distortions and diffraction are not the limiting effects. Spatial resolution is determined by the interplay between pixel spacing and depth of field.

Suppose that the spacing between pixels in the image plane is Δ. The resolution limit is 2Δ because that is the separation between features that allows two features to be perceived as distinct. In other words, perceiving some separation between features requires at least one imaging element between the features. Resolution is often listed as resolving power in units of lines per inch or lines per millimeter:

$$\text{RP} = \frac{1}{\text{RL}} \text{ lines/mm} = \frac{1}{2\Delta} \text{ lines/mm}. \tag{8.2}$$

In electronic cameras using photodiodes, the resolution of an image is limited by the spacing between the imaging elements (pixels) in a charge-coupled imaging device.

Photographic film has grains (crystals) of silver halide suspended in clear gelatin. When a ray of light hits a grain of silver halide, it changes to metallic silver. The unexposed grains are washed away during the developing process. The resolution of film is limited by the average spacing between grains of silver halide. A typical spacing between grains is 5 μm, with each grain about 0.5 μm^2 in cross section. The sensitivity (film speed) increases with grain size.

In the human vision system, the spatial resolution is determined by the spacing between cones in the fovea. A typical spacing is 30 seconds of arc or $1/120°$. This provides a resolution limit of one minute of arc, or $1/60°$, or 0.3×10^{-3} radians. Since this visual angle is small, the distance from the eye can be multiplied by the resolution limit in radians to calculate the spacing between features that corresponds to the resolution limit of human vision. For example, the length of your arm is around 40 cm. At this distance the resolution limit is

$$0.3 \times 10^{-3} \times 40 \text{ cm} = 120\ \mu\text{m}. \tag{8.3}$$

The spacing between dots placed on a page by a laser printer is 300 dots per inch or 85 μm. The human vision system can perceive seperations below the resolution limit by interpolating feature locations to subpixel resolution, so the jaggedness of printed text and line drawings is just barely noticeable when reading a page held at arm's length.

8.3 Depth of Field

The purpose of a lens is to allow a larger aperture to be used in the construction of the camera so more light can enter the camera, allowing the camera to function with less ambient illumination or a faster shutter speed. A larger aperture comes with the penalty of reduced depth of field. There is a trade-off between aperture size and depth of field: a smaller aperture provides more depth of field but admits less light; a larger aperture admits more light but reduces the depth of field.

For a particular setting of the distance z' of the image plane from the center of the lens, only points on a plane at distance z, obtained by solving the lens Equation 8.1, are in perfect focus. In practice, the depth of field is determined by the spatial resolution of the imaging device. Some amount of defocusing below the resolution of the imaging device can be tolerated. There is a range of image plane distances z' with an acceptable level of defocusing and a corresponding range of scene distances z, called the depth of field, with scene points that are in focus to an acceptable degree.

When a scene point is out of focus, it creates a circle of image intensity on the image plane instead of a single point. If the diameter of this circle is below the resolution of the imaging device, then the amount of defocusing is not significant. Suppose that the diameter of the circle is b, the diameter of the lens aperture is d, the focal length is f, and the correct setting of the image plane distance from the center of the lens is z'. If the image plane is moved closer to the lens, to a distance z_1', then the amount of blur is given by

$$b = \frac{d\,(z' - z_1')}{z'} \tag{8.4}$$

since the ratio of $b/2$ to $(z' - z_1')$ must be equal to the ratio of $d/2$ to z' by similar triangles. We can solve the lens Equation 8.1 for z' and z_1', corresponding to z and z_1, respectively, and substitute these expressions in Equation 8.4 to obtain an expression that relates the amount of blur to distances in the scene:

$$b = \frac{d\,f\,(z - z_1)}{z\,(f + z_1)}. \tag{8.5}$$

Suppose that b is the maximum diameter of the blur circle for acceptable defocusing. Solve Equation 8.5 for z_1 to obtain an expression for the distance to the near plane that bounds the view volume:

$$z_1 = \frac{f\,z\,(d - b)}{d\,f + b\,z}. \tag{8.6}$$

To calculate the distance z_2 to the far plane that bounds the view volume, let

$$b = \frac{d\,(z_2' - z')}{z'}, \tag{8.7}$$

where z_2' is the setting of the image plane distance (beyond the correct setting z') that corresponds to the maximum amount of blur. Solve the lens equation for z' and z_2' and substitute into the equation for the blur diameter:

$$b = \frac{d\,f\,(z_2 - z)}{z\,(f + z_2)}. \tag{8.8}$$

Solve this equation for the far plane distance:

$$z_2 = \frac{f\,z\,(d + b)}{d\,f\; - b\,z}. \tag{8.9}$$

These equations provide the positions of the near and far planes for a particular setting of the nominal plane of focus z and the aperture diameter d, focal length f, and maximum acceptable blur diameter b. There is a distance $z = d\,f/b$, called the hyperfocal distance, at which the far plane distance and the depth of field become infinite.

The depth of field D is the difference between the near and far plane distances, $z_2 - z_1$, given by

$$D = \frac{2\,b\,d\,f\,z\,(f + z)}{d^2\,f^2 - b^2\,z^2}. \tag{8.10}$$

8.4 View Volume

Consider the perspective model for image formation, introduced in Chapter 1, with the image plane in front of the center of projection. Let the center of projection be located at point (x_0, y_0, z_0) and the view vector from the center of projection to the origin of the image plane (the principal point) be (v_x, v_y, v_z). Assume that the image plane is perpendicular to the view vector. Note that the view vector defines both the direction in which the camera is pointed and the distance from the center of projection to the image plane (the camera constant). The view vector defines the orientation of the camera in space, except for the twist about the optical axis. Let the vector (u_x, u_y, u_z) from the origin in the image plane define the direction that corresponds to up. The projection of the up vector onto the image plane, normalized to unit length, defines the y' axis in the image plane. The x' axis is the unit vector perpendicular to y' and the view vector.

Assume that the viewable area on the image plane is a rectangle with height h and width w centered about the principal point. The corners of the rectangle and the location of the center of projection define four planes (top, bottom, left, and right) that bound the view volume. The near and far planes are perpendicular to the image plane at distances determined by the depth of field. By convention, the plane normals point to the outside of the view volume.

The six bounding planes can be calculated from the position and orientation of the camera, the dimensions of the imaging element, and the lens optics. The plane coordinates are in the same coordinate system used to define the position and orientation of the camera, typically the absolute coordinate system for the scene. The bounding planes can be used to determine if all scene points that should be within the view of the camera are inside the view volume. The position and orientation of the camera, the nominal focus, the aperture diameter, and the focal length can be adjusted to change the view volume until all scene points that should be viewable are within the view volume. The minumum aperture diameter is constrained by the ambient illumination and the shutter speed necessary to minimize motion blur.

8.5 Exposure

The amount of light collected by the camera depends on the intensity of light falling on the image plane (the image irradiance) and the duration of the exposure (shutter speed):

$$\mathcal{E} = Et. \tag{8.11}$$

Exposure is in units of joules (energy) per square meter while image irradiance is in units of watts (power) per square meter. Multiplying power by time gives energy, so the units work out correctly.

The F-number or F-stop is proportional to the ratio of the focal length f to the diameter d of the aperture:

$$\text{F-number} = \frac{f}{d}. \tag{8.12}$$

The lens aperture on cameras is designated in units of F-number because image intensity (or film exposure for constant shutter speed) is the same for different lenses at the same F-stop:

$$\text{F-number} \sim \frac{1}{E} \tag{8.13}$$

$$\sim \frac{1}{\mathcal{E}} \quad \text{for constant shutter speed } t. \tag{8.14}$$

In other words, the F-stop is the aperture diameter provided in a system of units that factors out the different light gathering capabilities of lenses with different focal lengths.

The F-number on camera lenses is marked with numbers that are multiples of $\sqrt{2}$ because doubling the aperture area is equivalent to increasing the aperture diameter by $\sqrt{2}$,

$$2 \times \text{aperture area} \sim \sqrt{2}d \approx 1.4d. \tag{8.15}$$

The F-stops are numbered 2.8, 4, 5.6, 8, 11, and so on, so each F-stop changes the aperture diameter by a factor of 1.4 and increases the amount of light on the image plane by a factor of 2.

Further Reading

There are several good books on optics, but most texts give short coverage to radiometry. An excellent text on optics which includes some coverage of radiometry and the optics of the eye is a paperback book by Young [259]. The vision text by Horn [109] contains a good discussion of radiometry and imaging. Horn was the pioneer in the use of optics and radiometry in machine vision. The text by Beynon and Lamb [31] contains a chapter on CCD image sensors.

Exercises

8.1 The geometry of the view volume was formulated in Section 8.4. Assume that the camera position and orientation are given in camera coordinates, so that the position is $(0, 0, 0)$, the view vector is $(0, 0, 1)$, and the up vector is $(0, 1, 0)$.

a. Calculate the position of the corners of the view rectangle on the image plane in terms of the position and orientation of the camera and the height and width of the imaging element.

b. Calculate the formulas for the coefficients of the top, bottom, left, and right bounding planes in terms of the camera position and orientation and the height and width of the imaging element.

c. Calculate the formulas for the coefficients of the near and far planes in terms of the focal length, aperture diameter, and nominal distance to a scene point.

d. Generalize these formulas to the camera position and orientation in absolute coordinates.

e. Outline an iterative algorithm for adjusting the camera position and orientation, nominal focus, aperture diameter, and focal length to accommodate a given list of scene points that should be in the view volume.

8.2 Consider a moving object with position (p_x, p_y, p_z) and velocity vector (v_x, v_y, v_z) in camera coordinates. It is desirable to adjust the shutter speed so that the blur due to motion is reduced to an acceptable level, defined for this exercise to be the minimum distance between imaging elements.

a. Calculate the projected velocity vector on the image plane assuming perspective projection and camera constant f.

b. Calculate the minimum acceptable shutter speed.

c. Suppose that there are no moving objects in the scene, but the camera moves with velocity vector (v_x, v_y, v_z). Repeat the calculation for the minimum shutter speed. The result will depend on the distance to points in the scene.

d. Suppose that the camera motion is due to vibration, modeled as movement in a uniformly distributed direction with vector magnitude v. Repeat the calculation for minimum shutter speed in terms of the maximum magnitude of the velocity vector due to vibration.