Contents

1 Imaging

1.1 Light interacting with surfaces

Visible light is electromagnetic radiation, exhibiting wave effects like diffraction. For many imaging models, however, it is helpful to introduce the abstraction of a *light ray*, describing the light radiation heading in a particular direction from a particular location in space. A light ray has a position, direction, intensity as a function of wavelength, and polarization.

Light sources, like the sun or artificial lights, flood our world with light rays. These reflect off surfaces, generating a field of light rays heading in all directions through space, as in Fig. 1.1. As the light rays reflect from the surfaces, they generally change in some attributes–their brightness or their color. It is those changes upon surface reflection that let us interpret what we see in the world. In this chapter, we describe how light interacts with surfaces and how those interactions are recorded with a camera.

1.1.1 Models of surface reflection

Let an incident light ray be of direction $\hat{p}$ and of power, $I_{\text{in}}(\lambda)$, as a function of spectral wavelength λ. The power of the outgoing light, reflected in the direction, $\hat{q}$, is determined by what is called the bidirectional reflection distribution function (BRDF), F, of the surface. If the surface normal is $\hat{n}$, then outgoing power is some function, F,

$$I_{\text{out}} = F(I_{\text{in}}, \hat{n}, \lambda, \hat{p}, \hat{q}) \tag{1.1}$$

In general, BRDF's can be quite complicated (see Matusik et al.), describing both diffuse and specular components of reflection. But surfaces with completely diffuse reflections, called Lambertian surfaces, have a particularly simple BRDF, which we denote F_L:

$$I_{\text{out}} = F_L(I_{\text{in}}, \hat{n}, \hat{p}) = A I_{\text{in}}(\lambda) \cos(\hat{n} \cdot \hat{p}), \tag{1.2}$$

where A is the surface reflectance, or albedo, $\hat{n}$ is the surface normal vector, and $\hat{p}$ is the direction of the incident light. Note that the brightness of the outgoing light ray depends on the orientation of the surface relative to the incident ray, as well as the reflectance, A, of the surface. For a Lambertian surface, the intensity of the reflected light is a function of

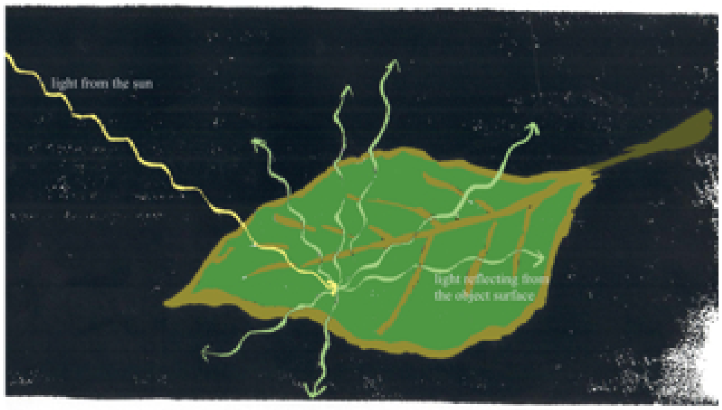


Figure 1.1
A light ray from the sun strikes a surface and generates outgoing rays of intensity and color depending on the angles of the incoming and outgoing rays relative to the surface orientation.

the direction of the incoming light ray, but not a function of the outgoing direction of the ray, $\hat{q}$. In general, surface reflection behaves linearly: the reflection from the sum of two rays is the sum of the reflections from the two individual rays.

To analyze the surfaces that reflect light, we need to know which light rays came from which direction in space. That lets us create an image, which we discuss next.

1.2 The Pinhole Camera and Image formation

One might wonder, when we look at a blank wall, why don't we see an image of the scene facing that wall? The light reflected from the wall integrates light from every reflecting surface in the room, so the reflected intensities are an average of light intensities from many different directions and many different sources. Mathematically, integrating the equation for Lambertian reflections, Eq. (1.2), over all possible incoming light directions $\hat{p}$, we have for the intensity reflecting off a Lambertian surface, I_L:

$$I_L = \int_{\hat{n}} AI(\hat{n}) \cos(\hat{n} \cdot \hat{p}) \tag{1.3}$$

I_L tells us very little about the light intensity $I(\hat{n})$ from any given direction $\hat{n}$. To learn about $I(\hat{n})$, we need to form an image. Forming an image involves identifying which rays came from which directions. The role of a camera is to organize those rays, to convert from the cacophony of light rays going everywhere to a set of measurements of intensities coming from different surfaces.

Perhaps the simplest camera is a pinhole camera. A pinhole camera has a light-tight enclosure, a small hole that lets light pass, and a projection surface. Figure 1.2 (a) shows the the geometry of a scene, the pinhole, and a projection surface. For any given point on the projection surface, the light that falls there comes from only from one direction, along the straightline joining the surface position and the pinhole. This creates an image of what's in the world on the projection plane, Figure 1.2 (b).

Figure 1.3 shows two different ways to make a pinhole camera. The first is simply a box with a hole in it, with a side removed for viewing. The second is a paper bag, padded to be opaque, with a hole in it, inside of which one sticks his head. We encourage readers to make their own pinhole camera designs. The needed elements are an aperture to let light through, mechanisms to block stray light, projection screen, and some method to view or record the image on the projection screen.

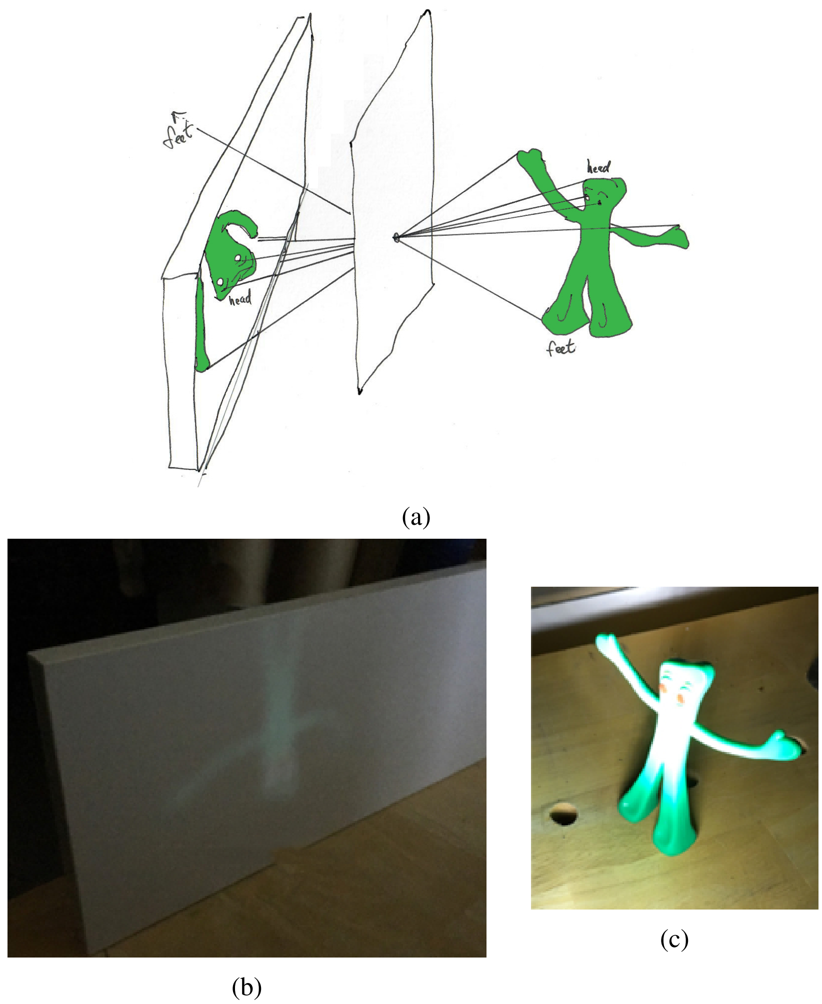
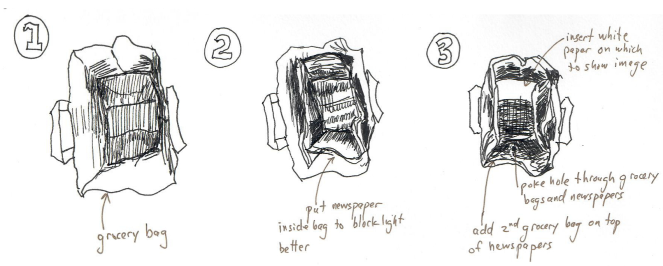


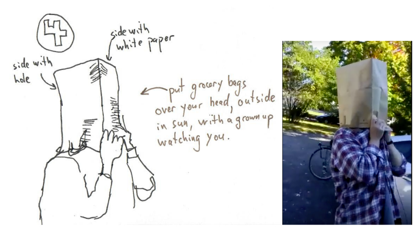
Figure 1.2
(a) Pinhole camera geometry, showing some light rays passing through the pinhole aperture to the projection plane. (b) Image projected onto the sensing plane creates an image of the subject, (c)

(a)

(b)



(c)



(d)

Figure 1.3
(a) Outdoor scene. (b) View inside pinhole camera imaging the scene, (a). Note pinhole at bottom face and inverted image on upper face. (c) We can also use a paper bag to make a light-tight pinhole camera, with the viewer inside. Newspapers can be added between two layers of paper bags to make a light-tight enclosure. (d) Use of the paper bag pinhole camera.

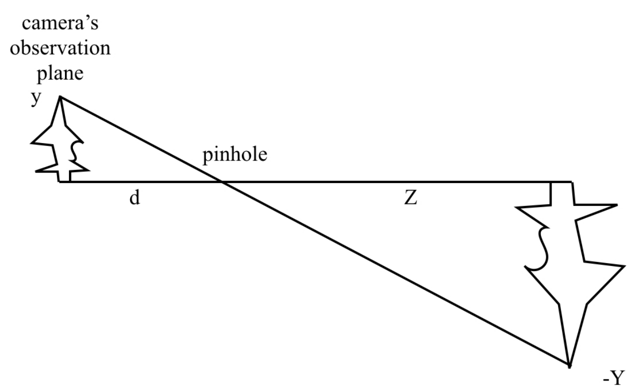


Figure 1.4
Perspective projection equations derived geometrically. From similar triangles, we have $y = -\frac{d}{Z}Y$.

1.2.1 Image formation by perspective projection

A pinhole camera projects 3d coordinates in the world to 2d positions on the projection plane of the camera through the straightline path of each light ray through the pinhole. The simple geometry of the camera lets us define the projection.

Let the origin of a Cartesian coordinate system be the camera's pinhole. The coordinates of 3d position of a surface in the world will be X, Y, Z, where the Z axis is perpendicular to the camera's sensing plane. Let the coordinates in the camera projection plane be x, y, parallel to the coordinate axes X and Y, respectively. If the distance from the sensing plane to the pinhole is d (see Fig. (1.4)) then similar triangles gives us the following relations,

$$x = -d\frac{X}{Z} \tag{1.4}$$

$$y = -d\frac{Y}{Z} \tag{1.5}$$

Eqs. (1.5) are called the *perspective projection equations*. Under perspective projection, distant objects become smaller, through the inverse scaling by Z. The perspective projection equations apply not just to pinhole cameras but to most cameras, and human vision as well.

1.2.2 Image formation by orthographic projection

Perspective projection is not the only feasible projection from 3d coordinates of a scene down to the 2d coordinates of the sensor plane. Different camera geometries can lead to other projections. One alternative to perspective projection is orthographic projection, where the size of objects is independent of the distance to the camera, scaled by a constant

(a)

(b)

Figure 1.5
Straw camera example (a) The subject: a hand in sunlight. (b) Showing the straw camera and the resulting image.

factor, k:

$$x = kX \tag{1.6}$$

$$y = kY \tag{1.7}$$

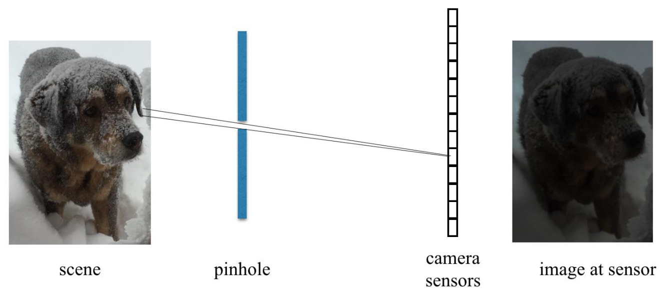
This is a good model for telephoto lenses, where the size of objects is roughly independent of their distance away. It's also a model for the "soda straw camera", shown in Fig. 1.5. A set of parallel straws allow parallel light rays to pass from the scene to the projection plane, but extinguish rays passing in all other directions. That camera doesn't invert the image, so there will be a sign difference in k for the two cameras.

1.3 Cameras with lenses

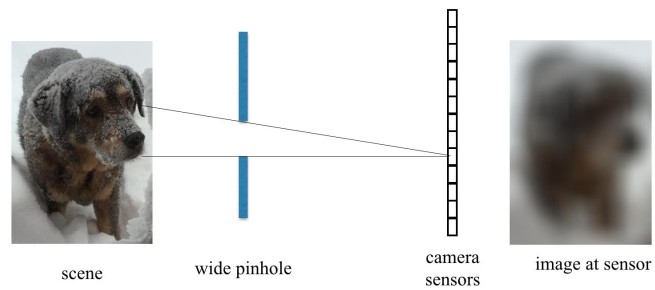
While pinhole cameras can form good images, they suffer from a serious draw back: the images are very dim, because not much light passes through the small pinhole to the sensing plane of the pinhole camera.

As shown in Figs. 1.6 and 1.7, one can try to let in more light by making the pinhole aperture bigger. But that allows light from many different positions to land on the sensor plane, resulting in a bright, but blurry, image.

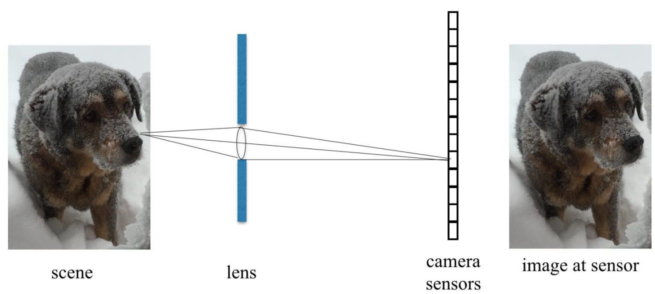
Putting a lens in the larger aperture can give the best of both worlds, capturing more light, while providing a focussed image on the sensor plane.



(a) Image from small pinhole

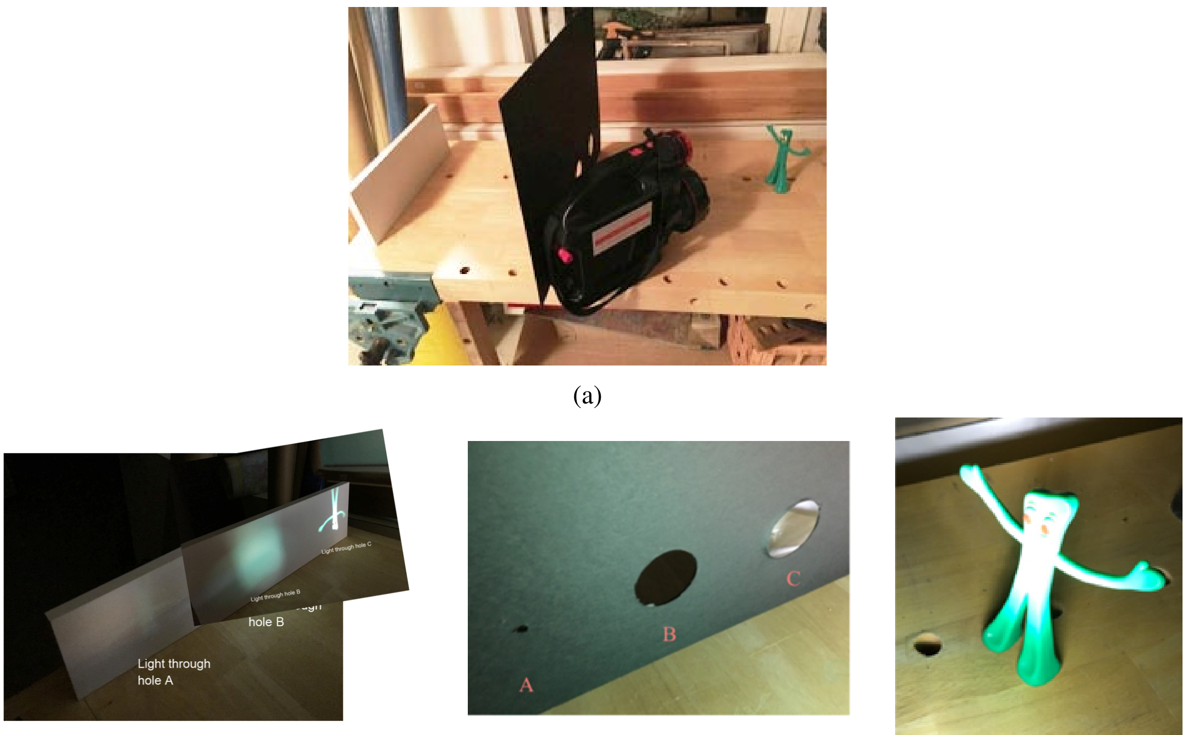


(b) Image from large pinhole



(c) Image from lens within the large pinhole

Figure 1.6
Brightness/sharpness tradeoffs in image formation. (a) A small pinhole will create a sharp image, but lets in little light, so the image may appear dark, for a given exposure time. (b) A larger pinhole lets in more light, generating a bright image. But each sensor element records light from many different image positions, creating a blurry image. (c) A lens can collect light reflected over many different angles from a single point, allowing a bright, sharp image.



(a) (b) (c) (d)

Figure 1.7
(a) Right to left: Gumby subject, illumination light, barrier with the three apertures and white projection screen. (b) Images formed by light through the three apertures. (c) the three apertures (small pinhole, large pinhole, and lens). (d) The subject.

1.3.1 Lensmaker's formula

Light slows down as it passes through glass. Under general conditions, that change in speed at the air-glass interface will bend the light.

A lens, positioned correctly with respect to a surface point and a sensor, has a remarkable property: every light ray from the surface point which passes through the lens is focussed to the same position at the sensor, no matter what part of the lens the ray hits.

Fig. **??** shows a demonstration of that property. The hand is flicking a laser pointer back and forth, sending light rays in many directions from a central point. All the the light rays which strike the lens are focussed to the same red dot. outside the lens, the light rays continue in their straight-line paths.

This property allows for high-brightness images to be formed in cameras. The question remaining is how to shape the lens to obtain this property. We'll present the correct shape, and then show that a lens with that shape will have the desired properties.

Consider a lens composed of two spherical surfaces, joined at a circular seam. Let there be two planes parallel to the plane of the lens, at distances $\frac{1}{d_1}$ and $\frac{1}{d_2}$ from the plane of the lens. Let the radii of the two spherical surfaces be the same: R. Referring to Fig. **??**: if a ray from a surface point, p_1 on plane 1, passing through the center of the lens, arrives at a point p_2 on plane 2, then every ray leaving p_1 and passing though the lens will also arrive at point p_2 on plane 2.

The distance relationship over which this focusing condition is met is called "the lens-maker's formula". The relationship, derived in Appendix **??**, is:

$$\frac{1}{f} = \frac{1}{d_1} + \frac{1}{d_2} \tag{1.8}$$

where f is a property of the lens called the focal length and d_1 and d_2 are the distances from the lens to the object and sensor planes.

[[figure to make: flick a laser pointer at one focal point of a lens. show that all the rays get focussed to one point (the nodal focal point). show that as blurred]]

1.4 Cameras as linear systems

Lens-based and pinhole cameras are special case devices where the light falling on the sensors or the photographic film form an image. In the more general case, the light recorded by a camera may look nothing like an image of the scene in the world.

Let the light intensities in the world be represented by a vector, $\vec{x}$. The value of the jth component of $\vec{x}$ gives the brightness of the light at position j, heading in the direction of the camera. If the camera sensors respond linearly to the light intensity at each sensor, their measurements, $\vec{y}$, will be some linear combination of the light brightnesses, given by the matrix, $\boldsymbol{A}$:

$$\vec{y} = \boldsymbol{A}\vec{x} \tag{1.9}$$

For the case of conventional cameras, where the observed intensities, $\vec{y}$ are an image of the reflected intensities in the scene, $\vec{x}$, then $\boldsymbol{A}$ is approximately an identity matrix.

For more general cameras, $\boldsymbol{A}$ may be very different from an identity matrix, and we will need to estimate $\vec{x}$ from $\vec{y}$. In the presence of noise, there may not be a solution $\vec{x}$ that exactly satisfies Eq. (1.9), so we often seek to satisfy it in a least squares sense. In most cases, $\boldsymbol{A}$ is either not invertable, or is poorly conditioned. It is often useful to introduce a regularizer, an additional term in the objective function to be minimized. If the regularization term favors small $\vec{x}$, then the objective term to minimize, E, could be

$$E = |\vec{y} - \boldsymbol{A}\vec{x}|^2 + \lambda|\vec{x}|^2 \tag{1.10}$$

Setting the derivative of Eq. (1.10) with respect to the elements of the vector $\vec{x}$ equal to zero, we have

$$\begin{aligned} 0 &= \nabla_x|\vec{y} - \boldsymbol{A}\vec{x}|^2 + \nabla_x\lambda|\vec{x}|^2 && (1.11)\\ &= \boldsymbol{A}^T\boldsymbol{A}\vec{x} - \boldsymbol{A}^T\vec{y} + \lambda\vec{x} && (1.12)\\ & && (1.13) \end{aligned}$$

or

$$\vec{x} = (\boldsymbol{A}^T\boldsymbol{A} + \lambda\boldsymbol{I})^{-1}\boldsymbol{A}^T\vec{y} \tag{1.14}$$

For the simplicity of visualization, let's consider a 1-dimensional pinhole camera: As shown in Fig. 1.6 the sensor is 1-dimensional and the scene lives in flatland. Let the sensor measurements be $\vec{y}$ and the unknown scene be $\vec{x}$. We represent the camera by the matrix, $\boldsymbol{A}$. For the case of a pinhole camera, and assuming 13 pixel sensor observations, the camera matrix is just an 13x13 identity matrix, depicted in Fig. 1.6.

Next, consider the case of a wide aperture pinhole camera, shown in Fig. (1.9). If a single pixel in the sensor plane covers exactly two positions of the scene intensities, then the geometry is as shown in Fig. (1.9) (a). The imaging matrix, $\boldsymbol{A}$, and its inverse, $\boldsymbol{A}^{-1}$, and shown in (b). (b) also shows the regularized inverse of the imaging matrix, which will usually give image reconstructions with fewer artifacts.

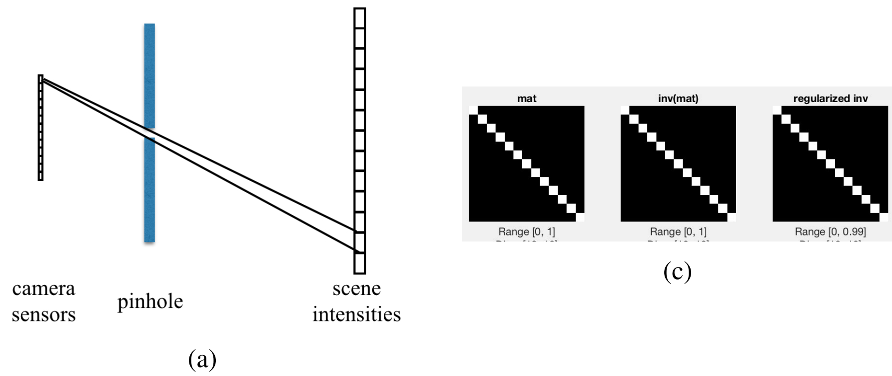


Figure 1.8
(a) Schematic drawing of a small-hole 1-d pinhole camera.(b) Visualization of imaging matrices: The imaging matrix relating scene intensities to sensor readings; the inverse of that matrix; the regularized inverse. For the small-pinhole imager, all three matrices are approximately identity matrices.

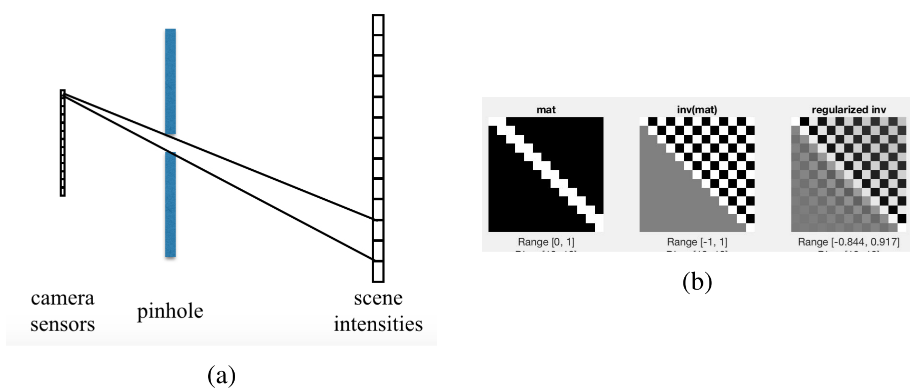


Figure 1.9
(a) Large-hole 1-d pinhole camera. (b) Visualization of imaging matrices: The imaging matrix relating scene intensities to sensor readings; the inverse of that matrix; the regularized inverse. For the small-pinhole imager, all three matrices are approximately identity matrices.

1.5 More general imagers

Many different optical systems can form cameras. Even a simple edge will do. Consider the example of Fig. 1.10. For the edge camera of (a), we have

$$\boldsymbol{A} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{1.15}$$

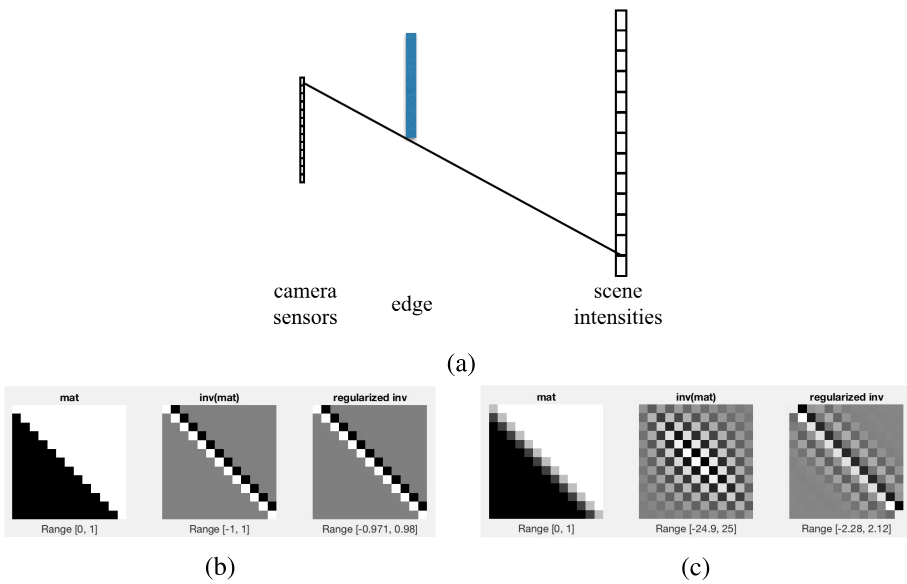


Figure 1.10

(a) An edge camera (b) Visualization of idealized imaging matrices: The imaging matrix relating scene intensities to sensor readings; the inverse of that matrix; the regularized inverse. (c) A blurred edge imaging matrix, and its inverse and regularized inverse

1.5.1 Corner camera

To show the general case of a camera imager, let us consider the “corner camera” Bouman et al. (2018). This is very similar to the edge camera of Eq. 1.15 and Fig. 1.10, but with slightly more complicated geometry. As shown in Fig. 1.11, a vertical edge partially blocks a scene from a ”projection plane”, creating intensity variations on the ground plane, observable by viewing those intensity variations from around the corner.

In practice, we will subtract a mean image from our observations of the ground plane, so in the rendering equation below, we will only consider components of the scene that may change over time, under the assumption that only a person behind the corner is moving. We will call these intensities $S(\phi, \theta)$ (”S” for the subject), where ϕ measures vertical inclination and θ measures azimuthal angle, relative to position where the vertical edge intersects the ground plane. The observed intensities on the ground will be $y(r, \theta)$, where the polar coordinates r and θ are measured with respect to the corner.

$$y(r, \theta) = \int_{\phi=0}^{\phi=\pi} \int_{\xi=0}^{\xi=\theta} \cos(\phi) S(\phi, \xi) \mathrm{d}\phi \mathrm{d}\xi, \tag{1.16}$$

where we have assumed a Lambertian diffuse reflection from the ground plane, which introduces the $\cos(\phi)$ term in Eq. (1.16).

The dependence of the observation, y, on vertical variations in $S(\phi, \theta)$ is very weak, just through the $\cos(\phi)$ term. We can integrate over ϕ first, to form the 1-d signal, $x(\xi)$:

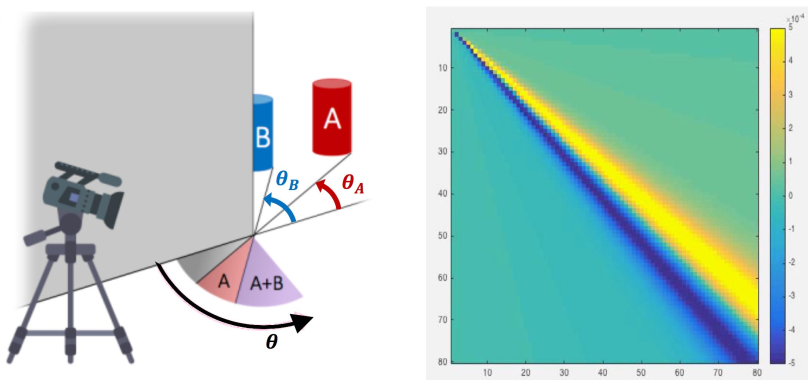
$$x(\xi) = \int_{\phi=0}^{\phi=\pi} \cos(\phi) S(\phi, \xi) \mathrm{d}\phi \tag{1.17}$$

Then Eq. (1.16) has the form,

$$y(r, \theta) = \int_{\xi=0}^{\xi=\theta} x(\xi) \mathrm{d}\xi, \tag{1.18}$$

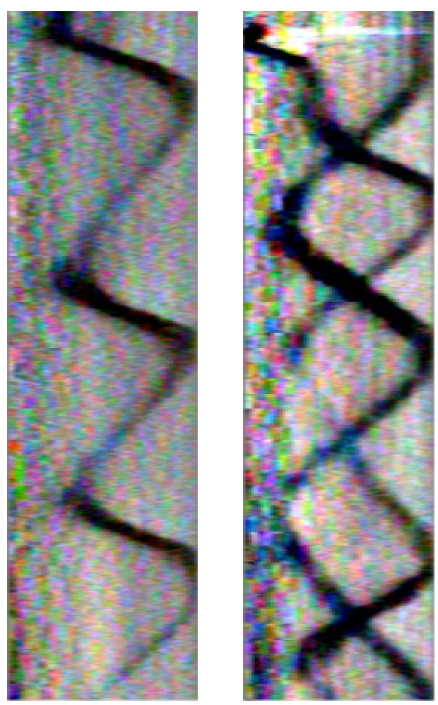
where $x(\xi)$ is a 1-dimensional image of the scene around the corner from the vertical edge.

If we sample Eq. (1.18) in its continuous variables, we can write it in the form y = Ax. Solving Eq. (1.14) for the multiplier to apply to y to estimate x yields the form show in Fig. 1.11 (b). We see that the way to ”read-out” the 1-d signal from the ground plane is to take a derivative with respect to angle. This makes intuitive sense, as the light intensities on the ground integrate all the light up to the angle of the vertical edge. To find the 1-d signal at the angle of the edge, we ask, ”what does one pie-shaped ray from the wall see that the pie-shaped ray next to it doesn’t see?”



(a) (b)

Figure 1.11
Corner camera geometry.



(a) (b)

Figure 1.12
(a) corner camera trace with one person moving (b) with two people moving.

Bibliography

Bouman, Katherine L., Vickie Ye, Adam B. Yedidia, Fredo Durand, Gregory W. Wornell, Antonio Torralba, and William T. Freeman. 2018. Turning corners into cameras: principles and methods. In *Proc. ieee international conference on computer vision*.

Matusik, W., H. Pfister, M. Brand, and L. McMillan. A data-driven reflectance model. *ACM Trans. Graphics* 22 (3): 759–769.