Obtaining Images

D.A. Forsyth,

University of Illinois at Urbana Champaign

Some slides adapted from
Svetlana Lazebnik, who adapted from Alyosha Efros, Derek Hoiem

Model of imaging

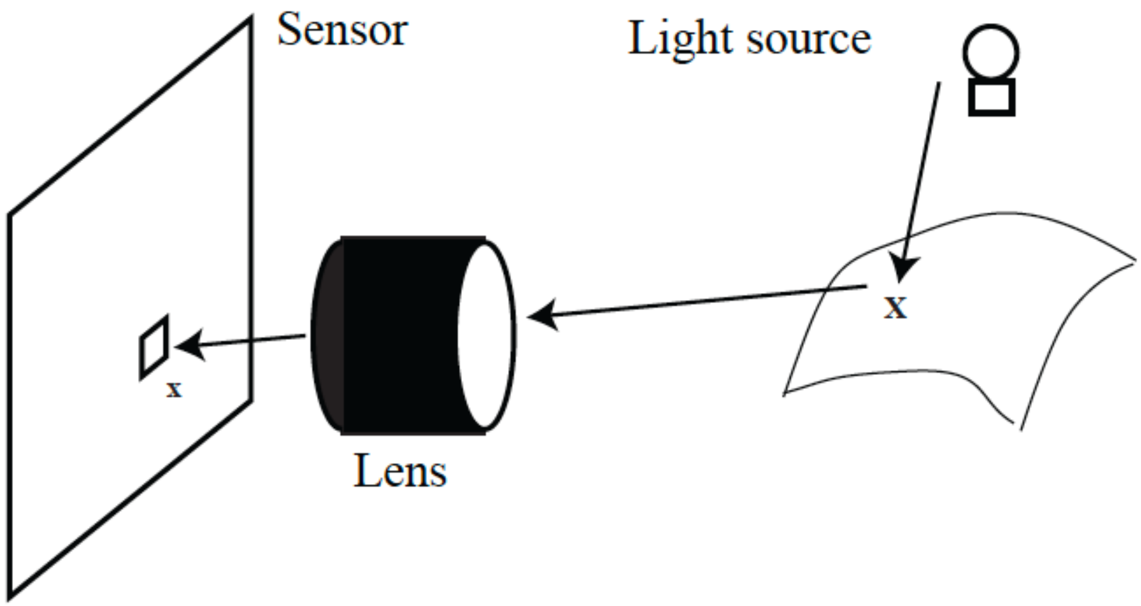


FIGURE 2.1: *A high-level model of imaging. Light leaves light sources and reflects from surfaces. Eventually, some light arrives at a camera and enters a lens system. Some of that light arrives at a photosensor inside the camera.*

A Pinhole Camera

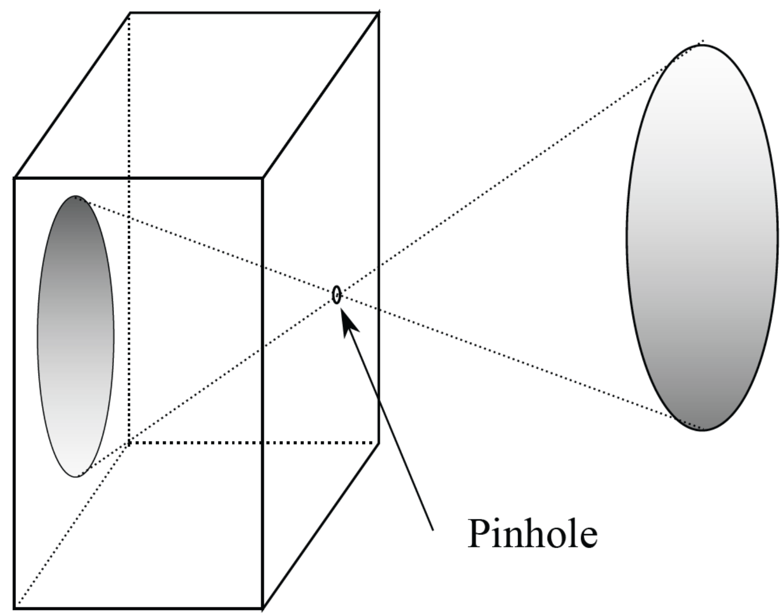


FIGURE 2.3: *In the pinhole imaging model, a light-tight box with a pinhole in it views an object. The only light that a point on the back of the box sees comes through the very small pinhole, so that an inverted image is formed on the back face of the box.*

The Pinhole Camera

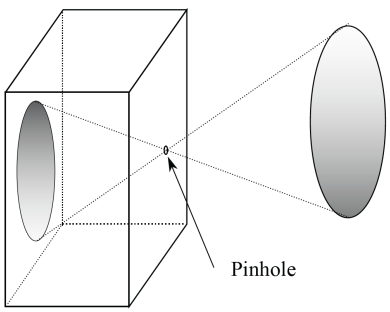


FIGURE 2.3: *In the pinhole imaging model, a light-tight box with a pinhole in it views an object. The only light that a point on the back of the box sees comes through the very small pinhole, so that an inverted image is formed on the back face of the box.*

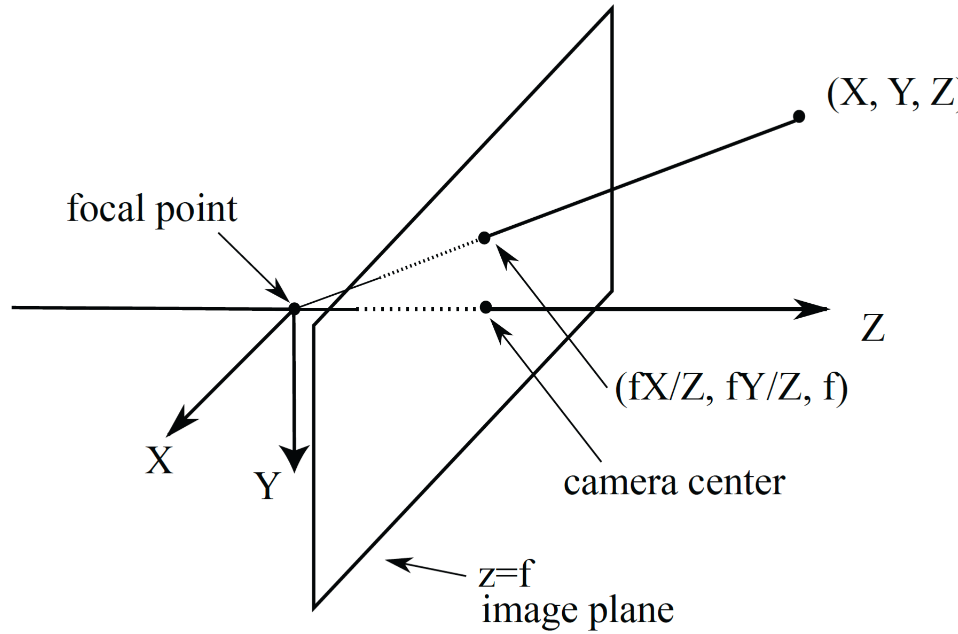


FIGURE 2.4: *The usual geometric abstraction of the pinhole model. The box doesn't affect the geometry, and is omitted. The pinhole has been moved to the back of the box, so that the image is no longer inverted. The image is formed on the plane* $z = f$*, by convention. Notice the y-axis goes down in the image. This allows me to use a* right handed *coordinate system and also have* z increase *as one moves into the image.*

The lens arranges that light arriving at $\mathbf{x}$ on the sensor all arrived from one point ($\mathbf{X}$ in Figure 2.1) on a surface in 3D, assuming it is in focus. At $\mathbf{x}$, the sensor collects power P for some period Δt, then passes the result on to the camera electronics. The sensor responds to energy $P\Delta t$, so collecting more power for a shorter period or less power for a longer period will result in indistinguishable results. The value of the pixel at i, j on the grid is a *sample* of a function of position (Figure 2.2).

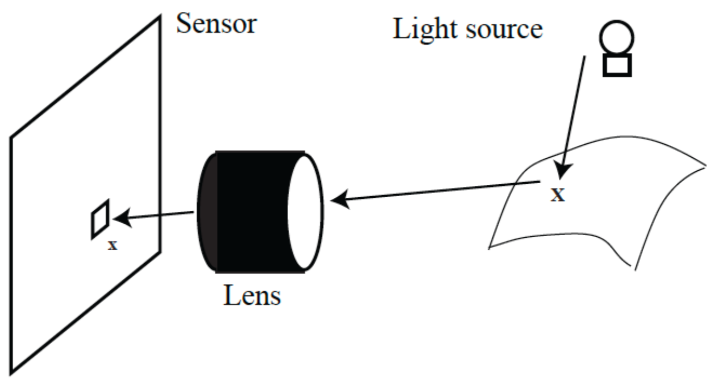
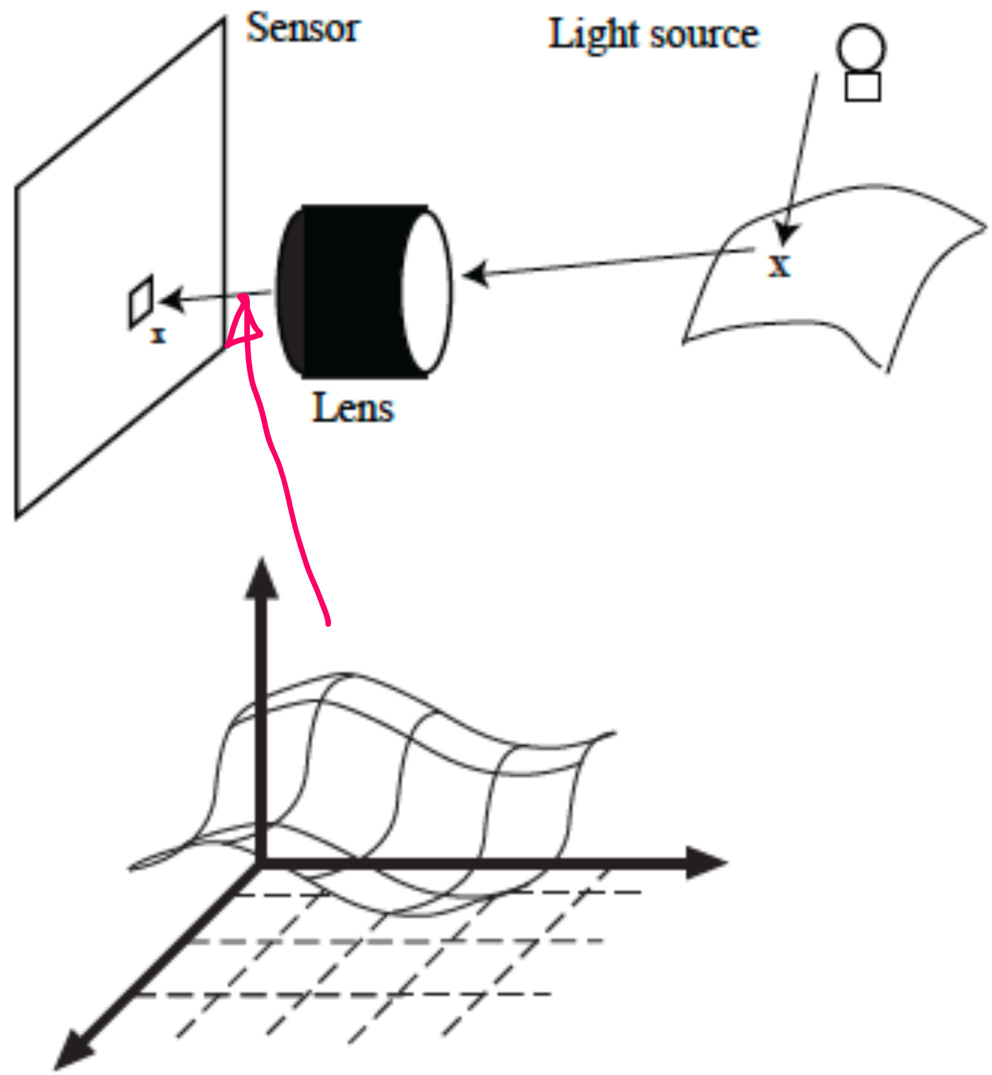


FIGURE 2.1: *A high-level model of imaging. Light leaves light sources and reflects from surfaces. Eventually, some light arrives at a camera and enters a lens system. Some of that light arrives at a photosensor inside the camera.*


Sensor
Light source
x
X
Lens

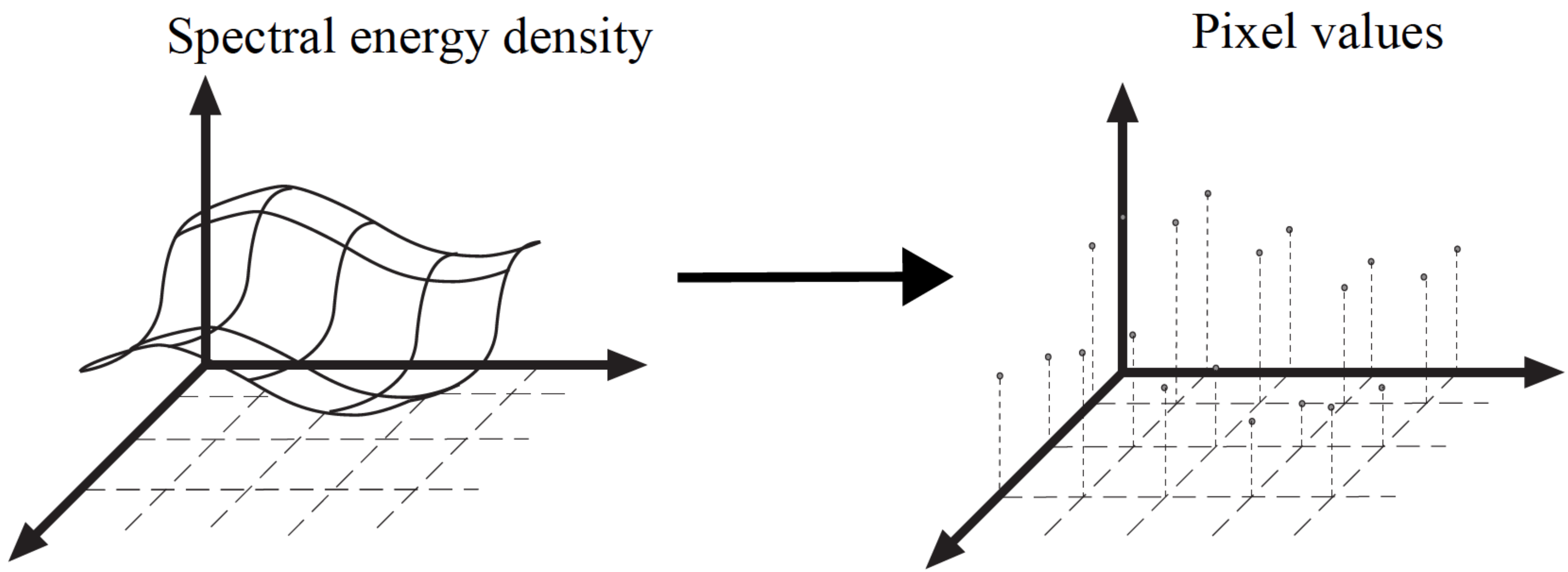
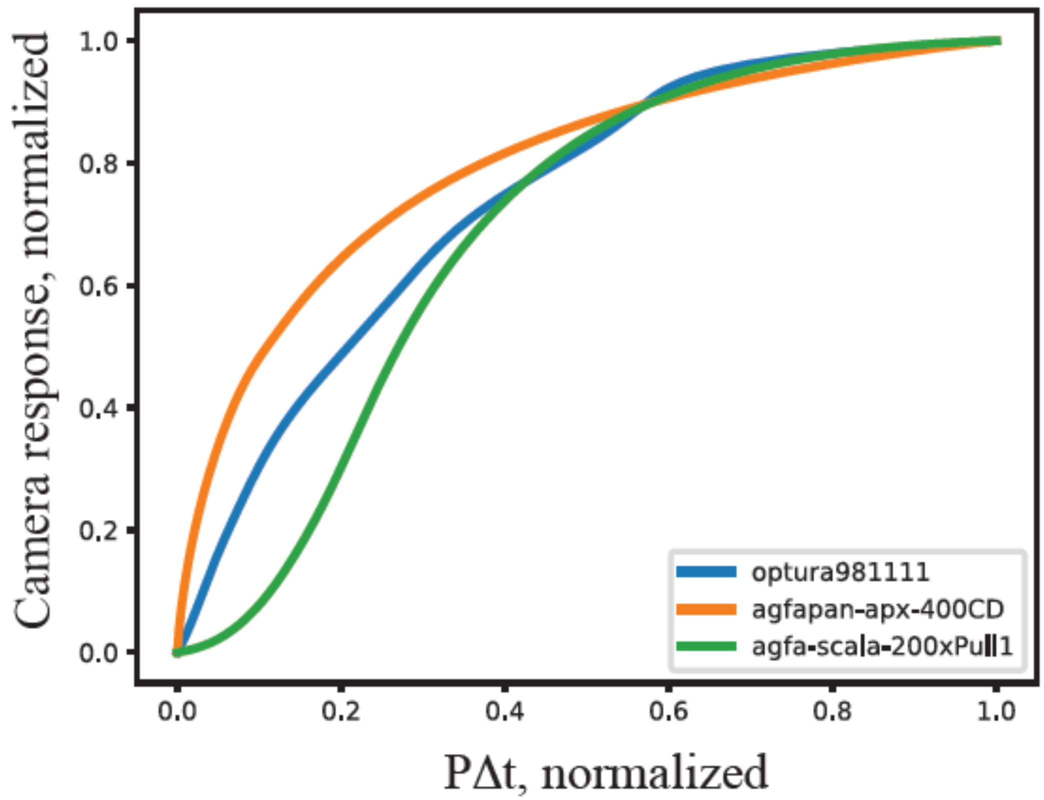


FIGURE 2.2: *Because each pixel in the sensor averages over a small range of directions and positions, the process mapping the input spectral energy distribution to pixel values can be thought of as sampling. On the* **left**, *is a representation of the energy distribution as a continuous function of position. The value reported at each pixel is the value of this function at the location of the pixel (***right***).*

Cameras aren't linear in input energy

- This is deliberate, and usually a property of camera electronics.
  - Helps imitate film
  - Increases dynamic range
- It's not the sensor (which is usually linear)
  - you can usually get linear pix from cameras with enough work
- Different cameras have different nonlinearities

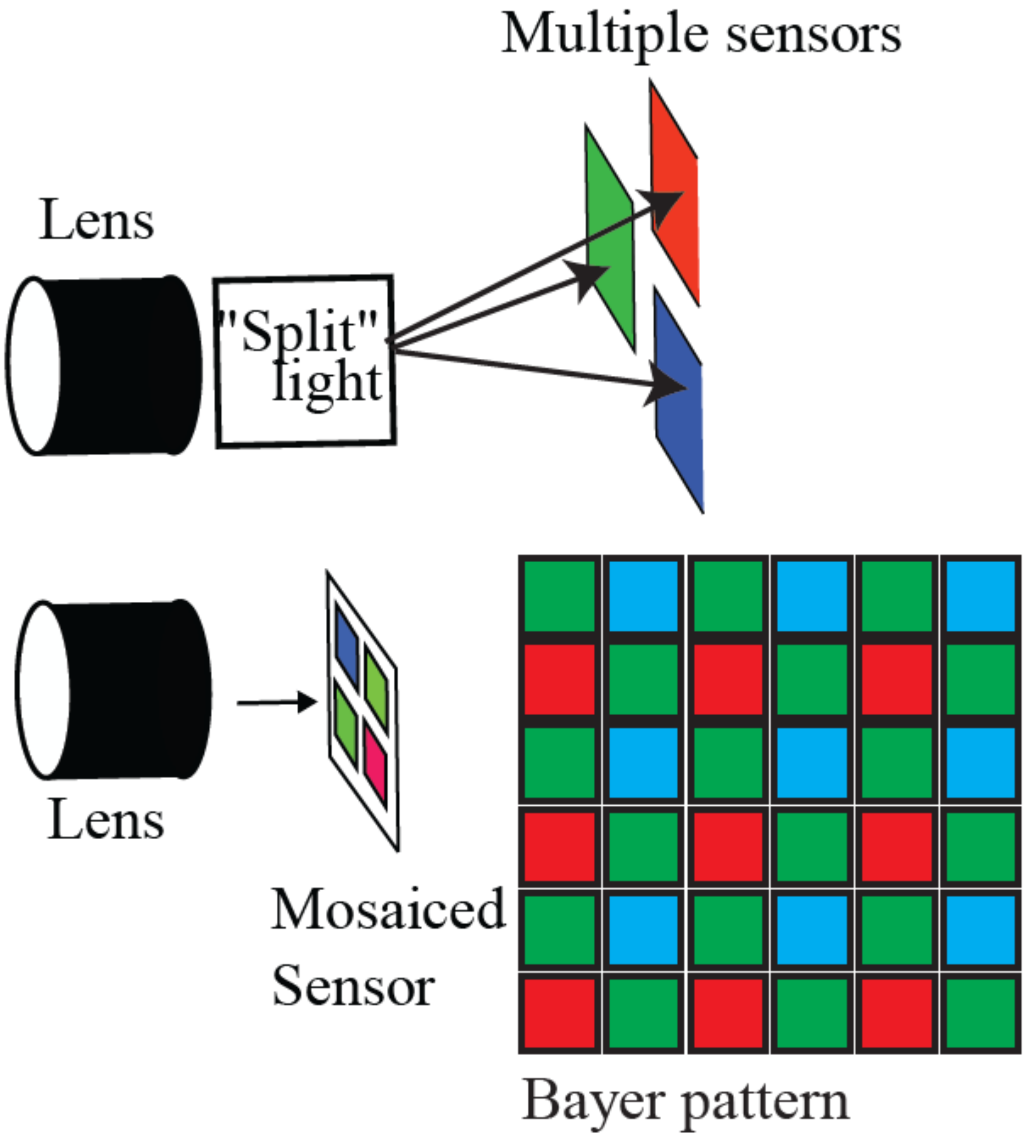


From wikipedia, https://commons.wikimedia.org/wiki/File:Raw_vs_jpg.jpg

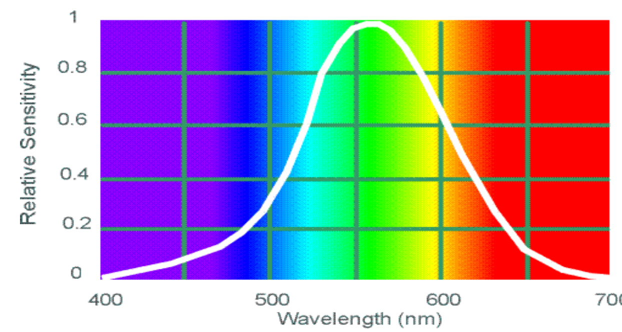
https://commons.wikimedia.org/wiki/File:Raw_vs_jpg.jpg

https://commons.wikimedia.org/wiki/File:Raw_vs_jpg.jpg

Digital color image



Why more green?

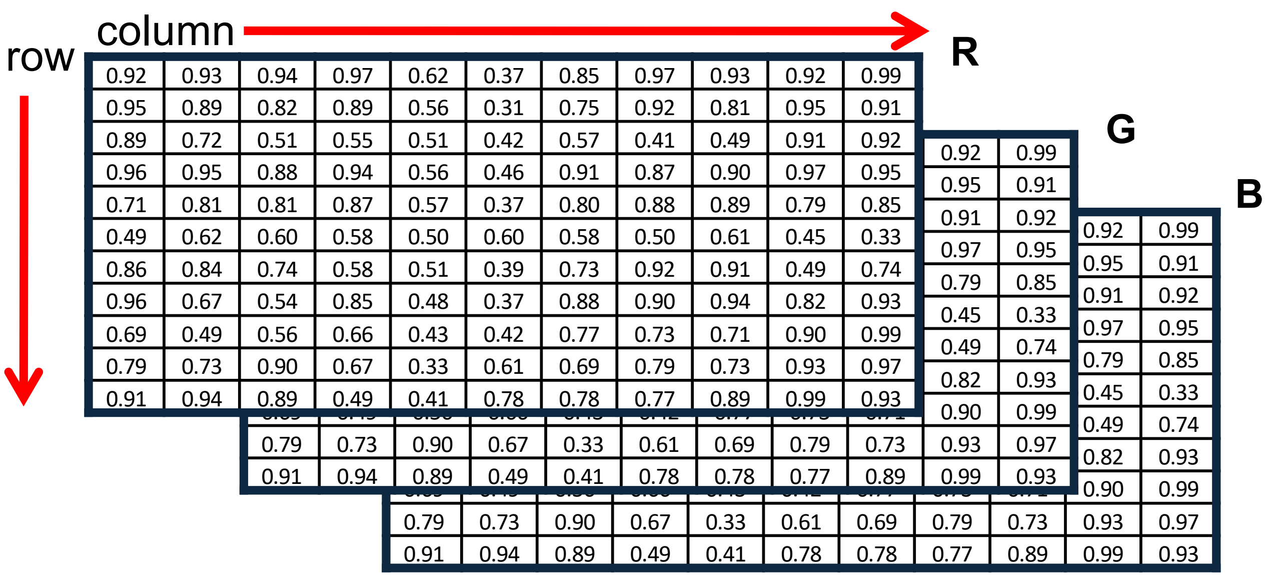


Human Luminance Sensitivity Function

Images in Python

```
im = cv2.imread(filename)                    # read image
im = cv2.cvtColor(im, cv2.COLOR_BGR2RGB) # order channels as RGB
im = im / 255                                # values range from 0 to 1
```

- RGB image `im` is a H x W x 3 matrix (numpy.ndarray)
- `im[0,0,0]` is the top-left pixel value in R-channel
- `im[y, x, c]` is the value y+1 pixels down, x+1 pixels to right in the c[th] channel
- `im[H-1, W-1, 2]` is the bottom-right pixel in B-channel



Pointwise image transformations

- Apply the same function to each pixel value
  - perhaps different for R, G, B

- Simple applications
  - remove camera non-linearity
  - simulate different camera
  - increase/reduce contrast

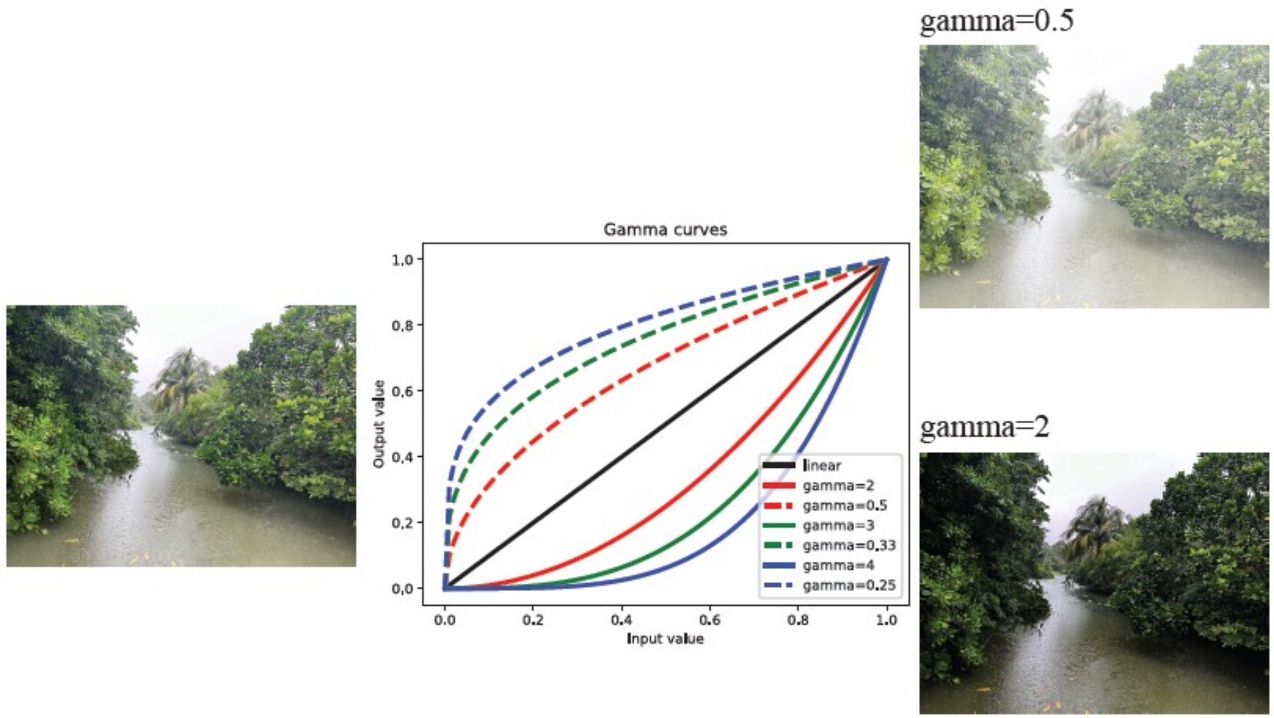


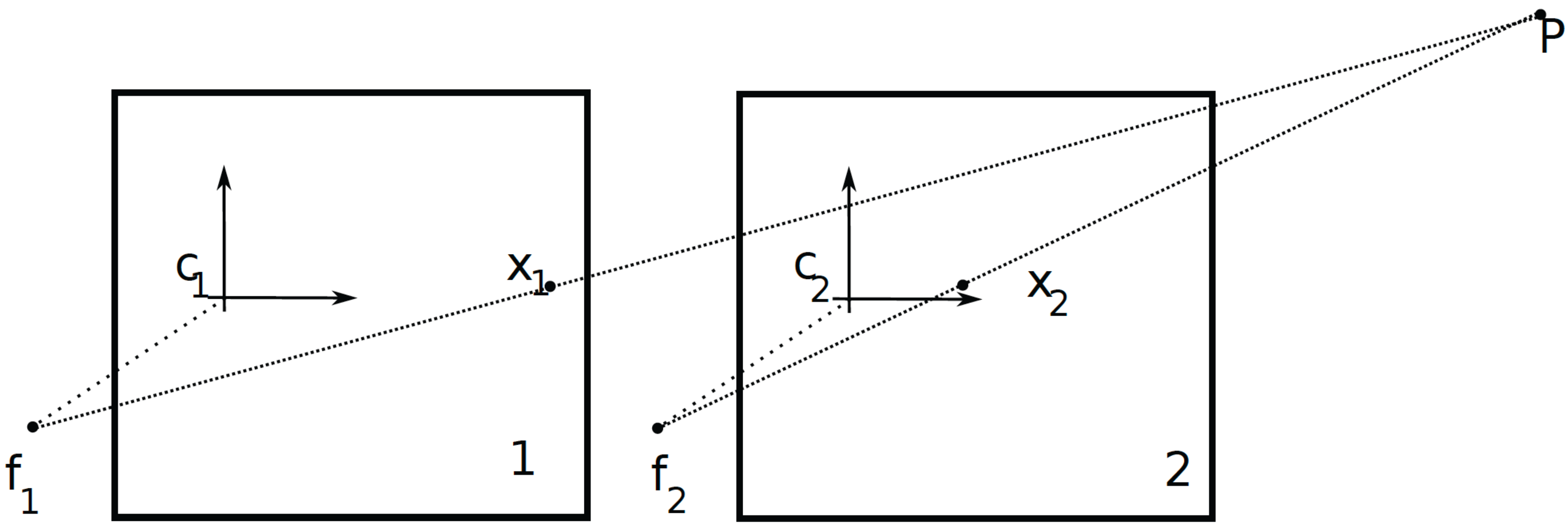
FIGURE 2.7: *Many imaging and rendering devices have a response that is a power of the input, so that* $output = C\,input^{\gamma}$*, where* γ *is a parameter of the device. One can simulate this effect by applying a transform like those shown in the* **center** *(curves for several values of* γ*). Note that you can remove the effect of such a transform –* gamma correct *the image – by applying another such transform with an appropriately chosen* γ*. The image on the* **left** *is transformed to the two examples on the* **right** *with different* γ *values.* Image credit: *Figure shows my photograph of a river in Singapore.*

Other sensors…

- Basically, light cameras with extra tricks
  - IR cameras
    - light cameras at different wavelengths
      - but you might have to cool sensors
  - Polarization cameras
    - specialized sensors to measure intensity at particular polarizations
  - Multispectral cameras
    - various arrangements to image with more sensors than R, G, and B

- Really different
  - Depth cameras

Two cameras yield depth



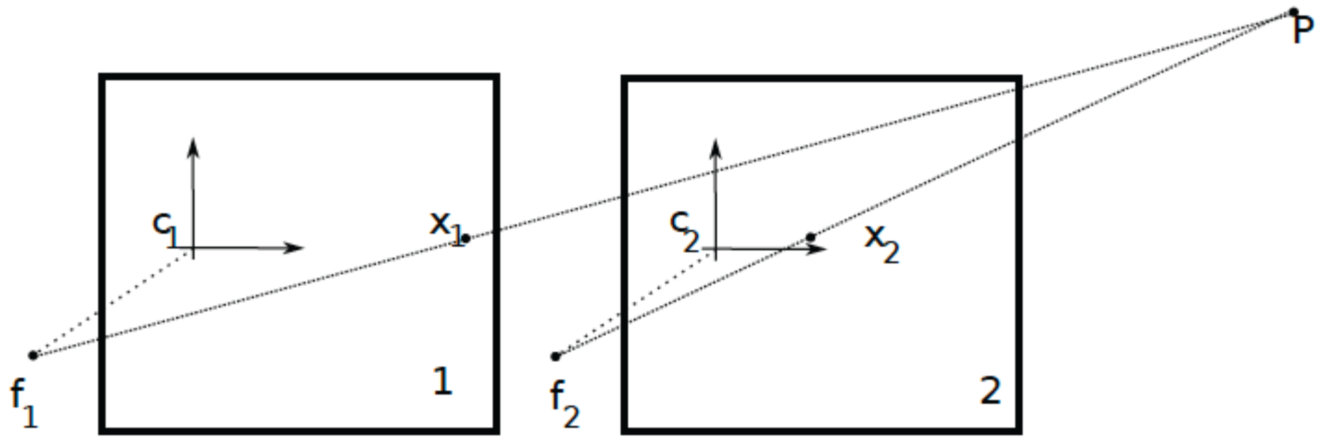


FIGURE 2.8: *When two pinhole cameras view a point, the 3D coordinates of the point can be reconstructed from the two images of that point. This applies for almost every configurations of the cameras. It is an elementary exercise in trigonometry (***exercises** *) to determine* **P** *from the positions of the two focal points, the locations of the point in the two images, and the distance between the focal points. Considerable work can be required to find appropriate matching points, but the procedures required are now extremely well understood (Chapters 35). One can now buy camera systems that use this approach to report 3D point locations (often known as RGBD cameras). Here we show a specialized camera geometry, chosen to simplify notation. The second camera is translated with respect to the first, along a direction parallel to the image plane. The second camera is a copy of the first camera, so the image planes are parallel. In this geometry, the point being viewed shifts somewhat to the left in the right camera.*

If's, and's and but's…

But there are limits to stereopsis. Measuring large depths with two cameras that are close together requires highly accurate estimates of point positions in images. Figure 2.8 shows a simple geometry that illustrates the problem. The point $\mathbf{P}$ projects to $\mathbf{x}_1$ in camera 1, and to $\mathbf{x}_2$ in camera 2. Notice because of the carefully chosen camera geometry, the y-coordinates of $\mathbf{x}_1$ and $\mathbf{x}_2$ are the same; only the x-coordinates differ. Write x_1 for the x-coordinate of $\mathbf{x}_1$; X for the x-coordinate of P, and so on. From the triangles in that figure, we have

$$d = x_2 - x_1 = f\frac{(X - B) - X}{Z} = -f\frac{B}{Z}$$

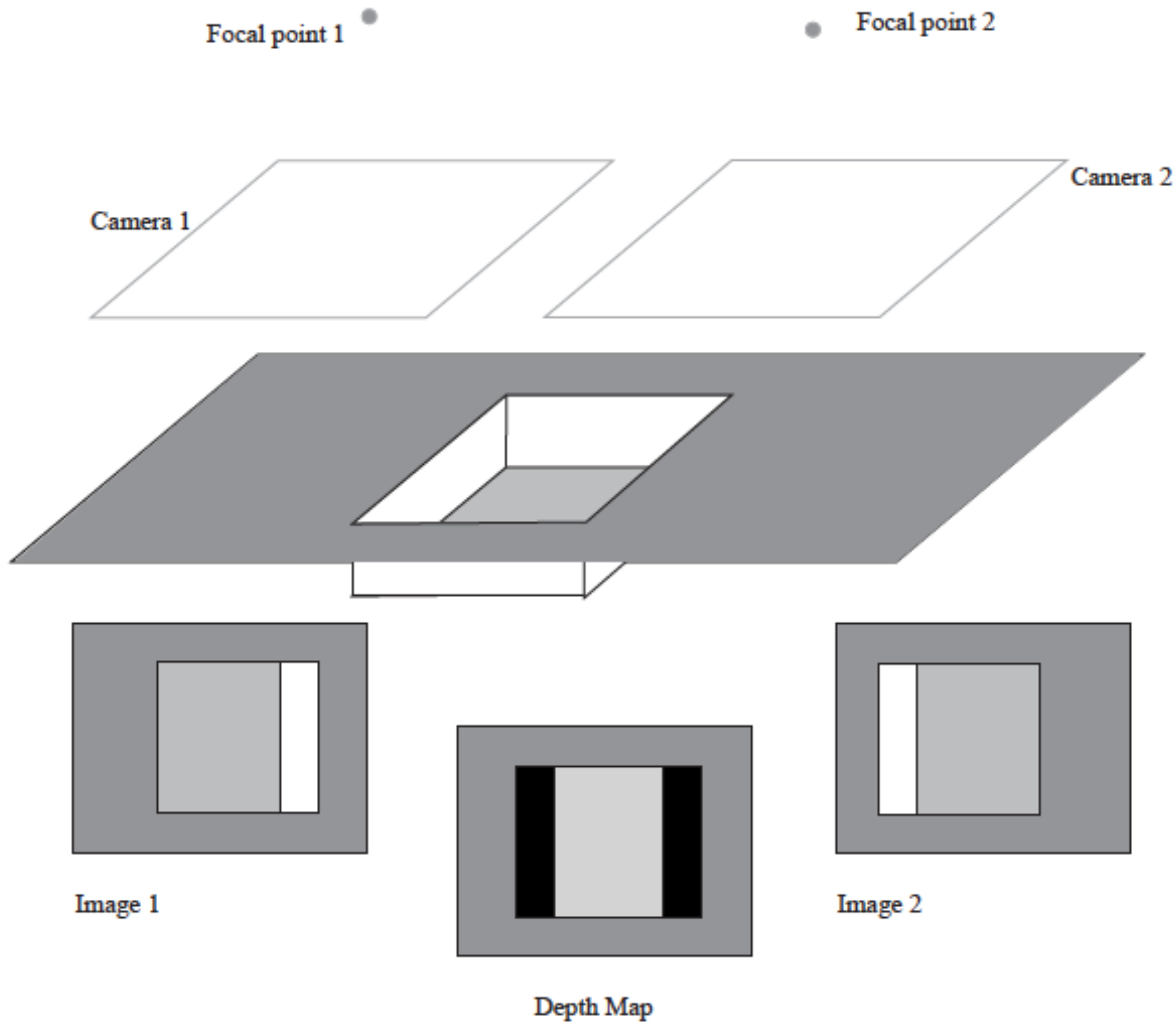
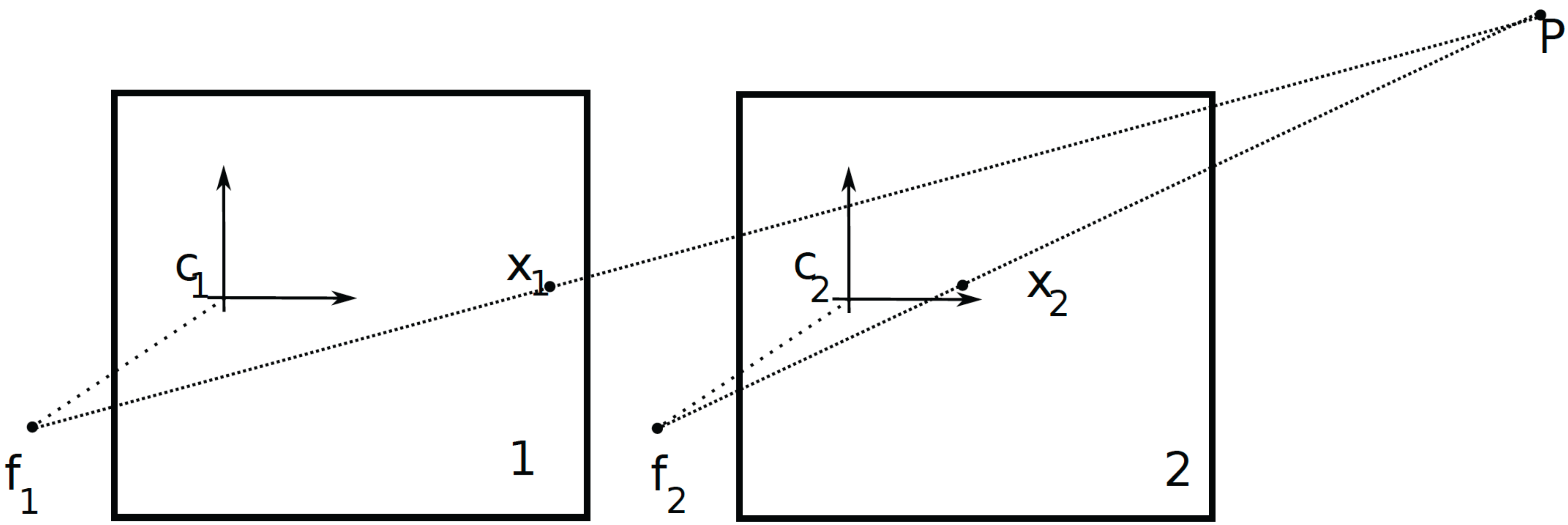


FIGURE 2.9: **Top** *shows two pinhole cameras viewing a rectangular depression in a flat surface. As the images show, camera on the left can see the right wall, and that on the right can see the left wall. This means that these walls cannot be reconstructed directly using trigonometry, and so the depth map will have holes in it. The depth map here is shown with a fairly common convention, where nearer surfaces are lighter, farther surfaces are darker, and holes are "infinitely far away".*

Camera – projector stereo



Structured light

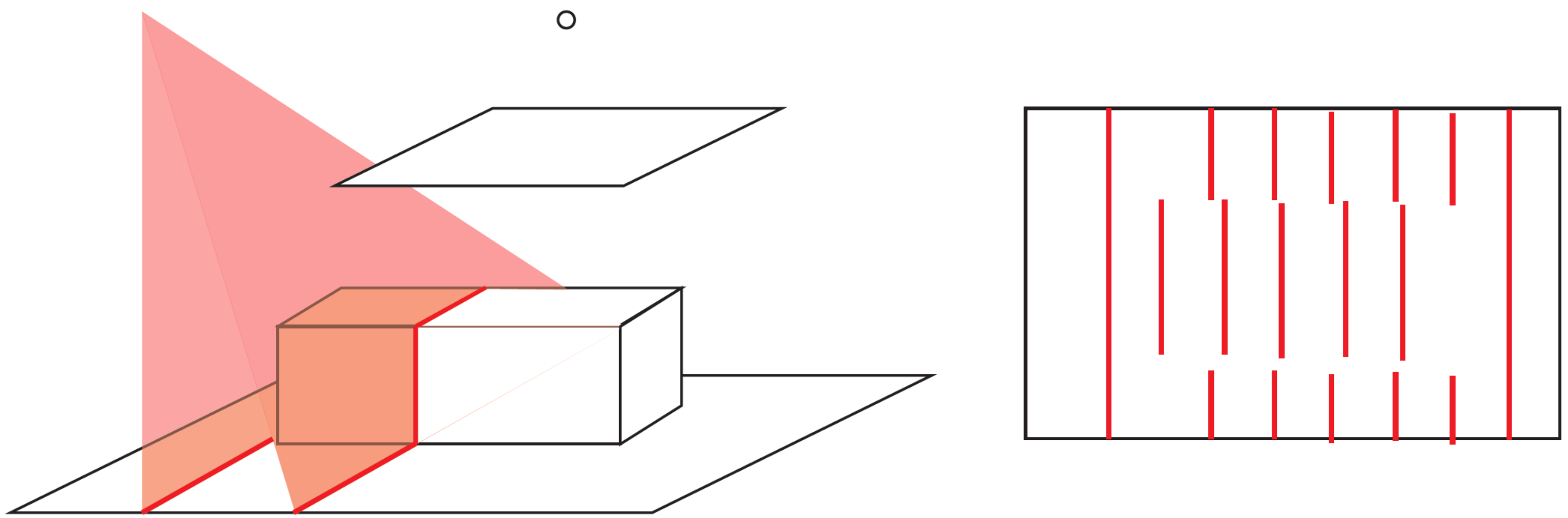


FIGURE 2.10: *The projector on the* **left** *casts planes of light into the scene. These are viewed by a camera. Their shape in the image (***right***) is a cue to the depth in the scene.*

LIDAR

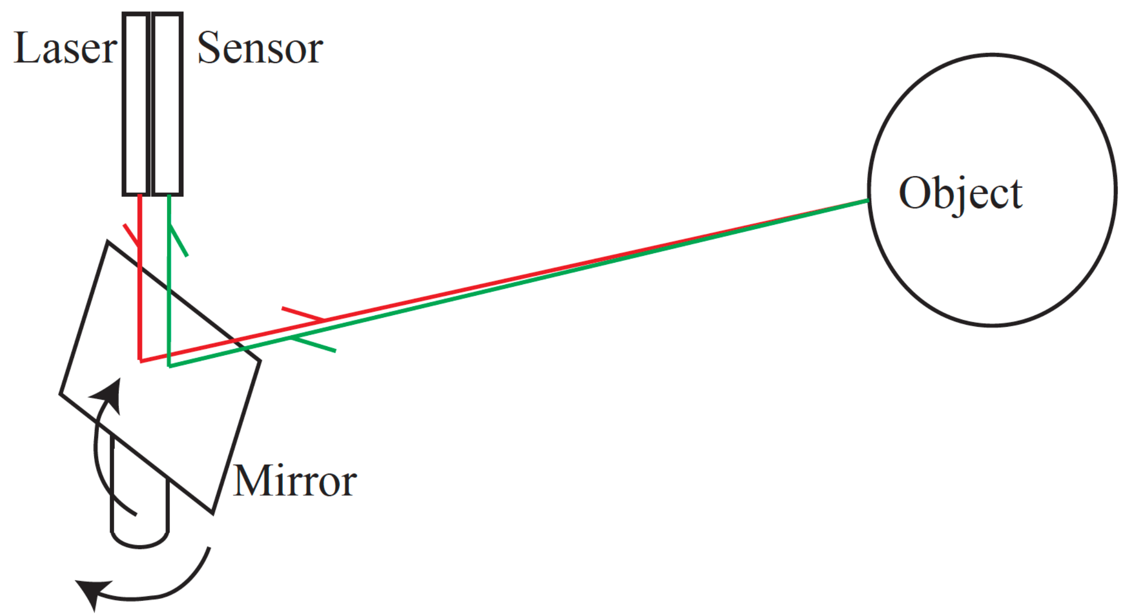


FIGURE 2.11: *A laser flashes a brief pulse of light onto a mirror. The light travels into a scene, is reflected from an object and returns to the mirror, and is reflected into a sensor. The time from flash to sensing reveals the distance tot he object. The mirror can be tilted or rotated to flash the light in different directions, and so to measure depth along different rays.*

2.1. As the pinhole in a pinhole camera gets larger, the image on the image plane (a) gets brighter and (b) is less focused. Why?

2.2. For a sufficiently small pinhole, diffraction effects cause the image to be defocused. Explain.

2.3. Why do cameras have lenses?

2.4. For smaller Δt, images will be darker; for larger Δt, moving objects may have blurred outlines. Why?

2.5. A common form of camera response function is Intensity$(P\Delta t) = Ce^{\gamma P \Delta t}$. Do you expect γ to be positive or negative? Why?

2.6. Why are mosaiced color CCD cameras more common than multiccd cameras?

2.7. You want to increase the contrast in an image by a pointwise image transformation that (a) makes dark pixels darker *and* (b) makes bright pixels brighter. Why will $f(I) = Ce^{\gamma I}$ not achieve this? Sketch a transformation that will.

2.8. Why is it hard to build a stereo camera that is (a) small and (b) good at measuring large depths?

2.9. What happens in a scanning time of flight sensor if the mirror moves slowly and objects in the scene move quickly?

2.10. An object is 30m away from a scanning time of flight sensor. How long does it take for the pulse to travel from camera to object to sensor? You can take the speed of light to be 3×10^8 meters per second.