Image sampling and resolution

The visual world around us contains an enormous amount of detail. We can see individual blades of grass at a distance of several meters, and we recognize faces at distances of tens of meters. The photoreceptor cells in our eyes *sample* the images, and our ability to see details depends in part on this sampling. It also depends on the defocus blur in the images, which we will discuss shortly.

What is the number of samples per degree of visual angle of our eyes or of a camera? To calculate this sampling rate, we need to know the distance s between samples on the image sensor, and the distance f of the image sensor from the center of the aperture. The angular distance between samples is then s/f radians. The sampling rate is the number of samples per angle which is the inverse of the distance between samples, i.e. f/s.

For example, consider some camera that has a given number of pixels on its sensor, say 3000×2000 and suppose the sensor area were $30mm \times 20mm$. One could calculate the distance between pixels from these values. If the distance from the sensor to a small aperture were given (and pretending there was no lens) then one could calculate the number of samples (pixels) per radian or per degree. See the Exercises for some examples.

Blur due to finite aperture

Up to now we have only considered eyes that are formed by having a concave surface. In the case of an extreme concavity, light enters only through a small aperture. The image that is formed will have limited sharpness because each sensor point will receive light from a cone of directions and these directions come from different 3D points in the world. The image sensor will average together the intensity values of the rays coming from these different 3D points.

Another way to think about blur is to note that each 3D point in the world will send rays of light to different points on the image sensor. That is, rather than thinking about each sensor receiving light from many different points in the 3D scene, we can think of each 3D scene point as sending light to many different sensor points. These two ways of thinking about blur just differ in what we are holding "fixed", either a single 2D sensor point or a single 3D scene point. Both are valid ways of thinking about blur. We will see this concept pop up several times as we study vision.

Thin lens model

Let's now consider lenses in the eye. A lens changes the direction of incoming light rays. Lenses allow some 3D scene points to produce a focussed image despite the eye having a finite size aperture. If you need a refresher on how lenses work, see
`https://www.khanacademy.org/science/physics/geometric-optics/`
If you want to learn more about how lenses evolved, see `http://www.youtube.com/watch?v=mb9_x1wgm7E` (Richard Dawkins video)

We restrict our discussion to the *thin lens* model. You may have seen the derivation of this model in your high school or your freshmen physics course. A key assumption of the thin lens model is that, for any 3D object point (X_o, Y_o, Z_o) in the world, the light rays that diverge from this point and that pass through the lens all will converge at some image point (X_i, Y_i, Z_i) behind the lens. Such points are called *conjugate pairs*. Using simple geometric arguments, one can derive relationships between the X, Y, Z variables of conjugate pairs. For example, you may recall that rays that are parallel to the optical axis (Z axis) and that pass through the lens will then pass through the optical axis at the point $Z = f$. The constant f is called the *focal length* of the thin lens. The focal length depends on the curvature of the two faces of the lens and on the material of the lens i.e. the "index of refraction" which has to do with whether it is made of water, glass, etc. The inverse of f, i.e. $1/f$ is called the *power* of the lens. [ASIDE: Note we have changed the definition of the variable f and the term "focal length". The definition from the last lecture (which is used in computer vision) was based on a pinhole camera model.]

The above property about parallel rays allows one to derive (details omitted) the following, called the *thin lens equation*:

$$\frac{1}{Z_o} + \frac{1}{Z_i} = \frac{1}{f}$$

which you should have seen in your Physics 1xx courses. The case of parallel rays is a special case in which the object is very far away from the lens, i.e. $Z_o \approx \infty$). Taking $Z_o = \infty$, and plugging into the thin lens formula gives $Z_i = f$. This holds for any set of incoming parallel rays, as long as the direction is not too far from the optical axis.

One way to think of the thin lens model is that if we have an object in the scene at some distance Z_o, then the image of that object will be at some distance Z_i behind the lens. But we can think of the thin lens equation in the opposite way too. Suppose we have an image sensor plane that is a distance Z_{sensor} from the center of the lens. The thin lens models that say points on the sensor plane have a set of conjugate points on a scene plane, called the *focal plane* which is at depth $Z_{focalplane}$ such that:

$$\frac{1}{Z_{focalplane}} + \frac{1}{Z_{sensor}} = \frac{1}{f}.$$

Example

Suppose your eye is focused on an object that is a distance of 10 m away and you hold up your finger at arm's length. Assume Z_{sensor} is 2 cm (the length of your eye) and suppose the aperture ("pupil") is 3 mm. What will be the blur width of your finger?

We apply the thin lens equation twice – once for the focal plane at 10 m :

$$\begin{aligned} \frac{1}{f} &= \frac{1}{Z_o} + \frac{1}{Z_{sensor}} \\ &= \frac{1}{10} + \frac{1}{.02} \end{aligned}$$

and once for the finger:

$$\frac{1}{f} = \frac{1}{.57} + \frac{1}{Z_i}$$

This gives $\frac{1}{Z_i} = 48.1$ and so $Z_i \approx .0207$. Thus the image of the finger is focussed slightly beyond the sensor, which causes blur on the sensor.

To compute the blur width w, we use similar triangles:

$$\frac{A}{Z_i} = \frac{w}{Z_i - Z_{sensor}}$$

so

$$\frac{.003}{.0207} = \frac{w}{.0007}$$

which gives $w \approx .0001$ m.

The blur width w spans some distance on the sensor surface. What is the visual angle covered by this blur width? It might not be clear what this question means, since visual angle was defined last lecture for pinhole cameras only and here we obviously don't have a pinhole camera. The way to think about it to ask what the visual angle *would be* for that distance if we *were* to have a pinhole camera. The answer is:

$$\text{blur width (radians)} = \frac{w}{Z_{sensor}} = \frac{.0001}{.02} \text{ radians} \approx \frac{1}{4}\text{degrees}$$

Recall that the angular width of your finger at arm's length is about 1 degree. So in this example, the blur width is about $\frac{1}{4}$ of the finger width. **[April 23. This seems like a lot! I will doublecheck with a camera. We can't always trust our eyes on these things.]**

In the exercises, I ask you to show that the blur width for a point at depth Z_o is

$$A(\mid \frac{1}{Z_o} - \frac{1}{Z_{focalplane}} \mid).$$

Note from this expression that the image is in perfect focus when $Z_o = Z_{sensor}$ and that the blur increases linearly with the distance in diopters from the focal plane.

Depth of Field

If a scene has a range of depths, then it is impossible for all points in the scene to be in perfect focus. Only one depth is in perfect focus. That said, our vision systems are limited in how well they can *detect* defocus blur, since we have a finite grid of photoreceptors. Some points that are out of focus will still appear perfectly focussed to us. So, a range of depths appears in perfect focus, and it is useful to give this range a name: the *depth of field* is the range of depths that are *perceived* to be in focus. The term is most often used in photography where one is describing a captured image, but it can be used in vision too. Note that this range of depths straddles the single depth that is in perfect focus. Some points closer than the focal plane and further from the focal plane appear to be in focus.

The typical depth of field that is quoted for human vision is 0.3 diopters (D). This means that the range of depths $[Z_{near}, Z_{far}]$ that *appears* in perfect focus at any one time typically satisfies about

$$\frac{1}{Z_{near}} - \frac{1}{Z_{far}} = 0.3.$$

The value 0.3 is only ballpark, however, and it depends on the pupil size, the individual person, , and the scene. Different scenes produce different image patterns, and for some of these patterns blur might be easier or harder to detect.

For example, the following depth intervals each have a difference of about 0.3 diopters (see slides for illustration):

- $[3.3m, \infty m]$ or $[.3D, 0D]$

- $[2m, 5m]$ or $[0.5D, 0.2D]$
- $[1m, 1.43m]$ or $[1D, 0.7D]$

Because blur increases linearly with diopter distance from the focal plane, the focal plane for these examples is at $0.85D, 0.35D$, and $0.15D$. Note that these focal plane distances are not the halfway point in the above interval in meters, but rather they are the halfway point in inverse meters (diopters).

Accommodation

The *power* of a lens is defined as the reciprocal of the focal distance f of that lens. When an optical system has a sequence of lenses that have focal lengths say f_1, f_2, etc, the combined power of the lenses is approximately the sum of the powers, $\frac{1}{f_1} + \frac{1}{f_2} + etc.$ In the eye, there are two refracting elements: the lens and the cornea. The cornea (the hard protective surface that interfaces with air) has more power than the lens. The high power of the cornea is due mainly to the large difference in index of refraction between the cornea and air. A typical power for the cornea is 40, i.e. a focal length 1/40.

The cornea is hard and doesn't change shape, and so it is lens that allows the focal plane (or power of the eye) to vary. There are muscles in the eye that can squeeze the lens, causing the lens curvature to increase, which decreases the focal length. So how does changing the lens power affect the focal plane? Consider the thin lens equation

$$\frac{1}{f_{cornea}} + \frac{1}{f_{lens}} = \frac{1}{f_{focalplane}} + \frac{1}{Z_{sensor}}$$

where the left hand side is the combined power of the cornea and lens. The power of the corner and the distance Z_{sensor} are fixed, and so there is only one varying element on the left and right side, namely by changing the power of the lens, you change the focal plane distance.

As we age, our lens becomes more rigid. The effect is huge. A ten year old child can change the power of the lens over a range of 15 diopters, but this range steadily decreases as one ages up to about 50 where the range is reduced to a mere 1 diopter. So beyond the age of 50, one still can accommodate a bit, but not much. This is a well known and universal effect of aging, and it is called *presbyopia*.

Another problem which many of you are more familiar with is short sightedness (*myopia)* versus long sightedness (*hyperopia*). Myopia means that the lens is too powerful relative to the size of the eye, and so rays coming from distant objects tend to be focussed in front of the sensor surface, and the rays then diverge before they reach the surface, creating blur. To counter myopia, one wears glasses that have negative power. For example, I am myopic and my prescription is about -3 diopters. When I wear my glasses, the power of my eye is,

$$\frac{1}{f_{cornea}} + \frac{1}{f_{lens}} + \frac{1}{-3}.$$

How does my optometrist decide what prescription I need? Since my problem is that I cannot see distant object clearly, I need glasses that correct my vision to allow me to see objects at a distance up to infinity. Note that this does not mean that I need to be able to focus at infinity. It is enough that I can focus up to about 0.15 diopters from infinity i.e. $1/0.15 \approx 7$ meters, since my depth of field will allow me to see clearly those last 0.15 diopters beyond 7 m.

People who are hyperopic have the opposite problem. They need optical corrections to see objects that are nearby. They need to add power, rather than subtract power. For example, if they want to see something clearly at a distance of say 20 cm or 5 D (diopters), they need the near end of the depth of field to be at that distance. Strictly speaking, this means that is enough for them to be corrected a distance of slightly greater than 20 cm since their depth of field will give them clear vision in an interval around 20 cm. However, note that at this distance, tiny changes in depth lead to relatively large changes in blur (for fixed focal length). For example, a 0.3 diopter range 'centered' at 20 cm is $[4.85D, 5.15D]$ which corresponds to a mere $[20.6cm, 19.4cm]$ only!

Open questions

We have reviewed some of the basic models and concepts of blurring that results from defocus. But we have just talked about image formation here, not about vision. Here are few questions that we would like to be able to answer about vision:

First, How does a visual system determine if an image is in focus ? When I take my glasses off while lecturing the scene in front of me is very blurry. This is obvious, and there is nothing I can do about it other than put my glasses on. But objects that are close to me might be only slightly out of focus at any given time. But how can I tell that? How can I characterize the image I am measuring as blurred? Similarly, when you look at a photograph, how can you decide if it is sharp or blurred? What properties of the image make it sharp?

Second, how does the visual system accommodate ? If a scene around us is slightly blurred, then we need to adjust the focal length of our lens, i.e. we need to accommodate. But how? Should we focus at a farther or closer distance? (That is, is the object we are looking at blurred because it is too close or too far from where we are focusing?)

Third, a related question: how does accommodation interaction with binocular vergence ? If I rotate my eyes in to look at something closer to me, then I should increase the power of my lens too, so that the object will be in focus. Do the accommodation and vergence systems "talk to" each other. It turns out that they do.

Fourth, is defocus blur a depth cue ? When I change my focus to make the image of some object sharp, am I getting depth information about that object? Or am I just making it more sharp? We'll return to these questions later in the course.

