This lecture carries forward some of the topics from early in the course, namely defocus blur and binocular disparity. The main emphasis here will be on the information these cues carry about depth, rather than on how blur and binocular disparity information is coded in the visual system.

Depth from Blur

Lecture 2 examined how blur depends on depth. You learned about accommodation, namely by changing the focus distance of the eye, you can bring surfaces at certain depths into sharp focus and cause surfaces at other depths to become more blurred. Accommodation can be used to judge the 3D depth of points – at least in principle, since the eye controls accommodation. This does not seem to be how people judge the depths of all points in a scene, however. We do not sequentially "scan" through *all* focus settings, and the reason we don't is presumably since estimating depth is just one of many things the visual systems needs to do. That said, some depth information is available from focus, and so we would like to better understand what that depth information is and how the visual system might use it.

One idea is that if the visual system can estimate the depth at which it is currently focusing (by controlling the shape of the lens to bring a desired point into focus), and if it can also estimate the current aperture (which it can, since the eye controls the pupil size), and if it can estimate the blur width at various points – for example, the width of a blurred edge – then the visual sysem can compute the distance in diopters between any blurred point and the focal plane. Recall the relation derived in Exercise 2 Question 4:

$$\text{blur width in radians} = A \mid \frac{1}{Z_{object}} - \frac{1}{Z_{focalplane}} \mid$$

Note the absolute value on the right side of this equation, which is due to the fact that blur occurs for points farther than the focal plane and also points closer than the focal plane. From blur alone, we have a two -fold depth ambiguity.

Interestingly, the eye does not hold the focal distance constant. Rather the eye's focus distance is continuously oscillating. The amount of oscillation is small: the amplitude is roughly 1/3 of a diopter. But this may be enough to resolve the ambiguity. For example, if an object is closer (or further) than the focal plane, then moving the focal plane closer to the eye will decrease (or increase) the blur; the opposite holds when the focal plane is moved further from the eye. In particular, the two-fold ambiguity mentioned in the previous paragraph is easily resolved – at least in principle.

Blur on a Slanted Plane

It is very common to have objects in a scene that are large planes, or at least can be approximated as such over a large region. Examples are the ground we walk on, walls, and ceilings. Let's consider the blur that arises on a slanted plane.

Let the scene depth map be a slanted plane,

$$Z = Z_0 + mY$$

where Z_0 is the depth of the plane at the point that intersects the Z axis. Assume that we are focussed on that depth. Note that this scene has a floor or ceiling slope only. A more general plane would have a slope component in the X direction also.

Recalling $\frac{y}{f} = \frac{Y}{Z}$, we divide by Z to get

$$1 = \frac{Z_0}{Z} + m\frac{y}{f}$$

or

$$Z_0(\frac{1}{Z_0} - \frac{1}{Z}) = m\frac{y}{f}.$$

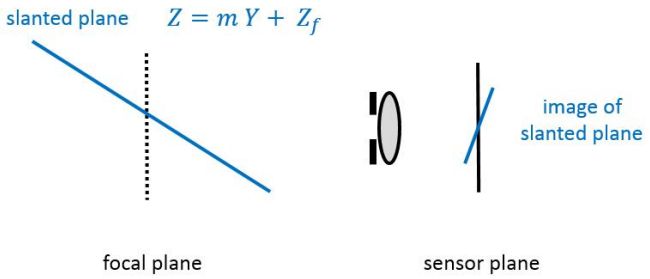
From page 1, the blur width w in radians is

$$w = A|\frac{1}{Z_0} - \frac{1}{Z}|$$

and so

$$w = \frac{mA}{Z_0} \mid \frac{y}{f} \mid.$$

Thus, the blur width on a slanted plane increases linearly with the image coordinate y. This linear dependence scales with A, with the focus distance Z_0 in diopters, and with the slope m of the plane.



Long ago photographers tried to take advantage of this linear dependence. One idea was to build a camera whose sensor plane was slanted slightly in the direction of the depth gradient.[1] When the slant is chosen appropriately, the sensor plane becomes aligned with the focussed image of the points on the slanted plane and one obtains a perfectly focussed image – something that otherwise is not possible to do, especially not with a wide aperture.

Another idea is to tilt the lens in in the opposite direction, so as to *increase* the gradient of blur in the y direction. Examples are shown below. The perceptual effect is that one misinterprets the overall scale of the scenes: the scenes appear to be photos of small toy worlds, rather than photos of large scale environments. While there is some controversy on what is causing his perceptual effect, the general idea is that the large blur gradient needs to be 'explained' by one of the variables in the above equation. Having an extremely large ground plane slant m is not possible, since the perspective cues suggest a particular slant m which is not extreme – I'll discuss perspective cues next lecture. Having a large aperture A is also not possible, since the aperture needed to get such blur gradient in a large scene would be much larger than the aperture of our eyes – we simply don't experience large scale scenes with such a blur gradient. The most likely culprit seems to be the variable Z_0 which is the distance to the point on the optical axis – indeed, making Z_0 small by scaling the whole scene down would explain the large blur gradient, while holding perspective cues constant.

[1] The configuration was a called a *tilt-shift* lens. Details omitted since I just want to give the basic idea.

Binocular Stereopsis (and its relation to blur)

We have discussed the geometry of stereo a few times in the course, for example, in lectures 1 and 6. If the eyes are parallel, then

$$\text{disparity (radians)} = \frac{x_l}{f} - \frac{x_r}{f} = \frac{T_X}{Z}$$

and if the left eye and right eye are rotated by angles θ_l and θ_r relative to the Z axis, then :

$$\text{disparity (radians)} = (\frac{x_l}{f} - \frac{x_r}{f}) - (\theta_l - \theta_r) = T_X(\frac{1}{Z} - \frac{1}{Z_{vergence}})$$

It is easy to show that $\theta_l - \theta_r$ is the *vergence angle*, namely the angle defined by the three points (left eye, scene point where eyes are verging, right eye).

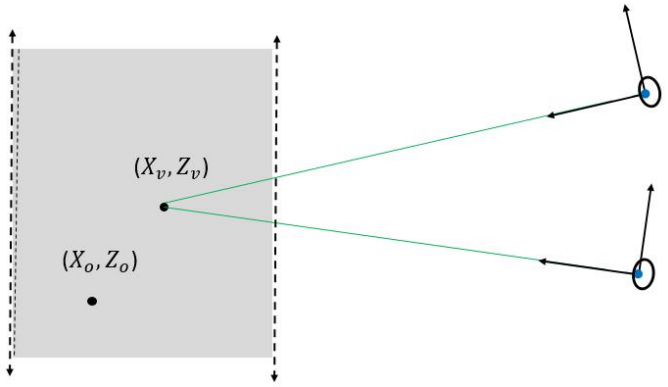
Since the brain controls the vergence, the brain in principle deterimines the depth on which the eyes are verging, and so this depth information is available. This is similar to accommodation as we'll see below, namely the brain controls the shape of the lens and so the brain 'knows' the depth of points that are in focus. Indeed the mechanisms of binocular vergence and accommodation are coupled: when the vergences angle is changed, so does the power of the lenses – at least to the extent possible. (Recall that as you get older, the range of accommodation decreases.) I will discuss blur again a bit later in the lecture.

Crossed and uncrossed disparities, binocular fusion

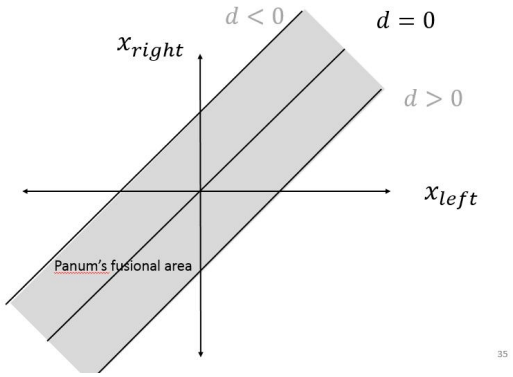
Point that are closer to the eye than the vergence distance have positive disparity, or *crossed* disparity since one needs to cross one's eyes to bring the disparity of such points to 0. Points that are further than the vergence distance have negative disparity, or *uncrossed* disparity since one uncrosses one's eyes to bring the disparity of such points to 0.

If the magnitude of the disparity of a 3D is small enough, then one can perceptually fuse the left and right images of the point, rather than seeing two images of these points – i.e. 'double vision' or *diplopia*. This limited range of fusion disparities defines *Panum's fusional area* which is equivalently a range of depths in front of and beyond the vergence depth – see grey area in figure below below.

That is, for any vergence distance, Panum's fusional "area" is really a 3D volume such that visible points in this volume are fused by the visual system.[2] One often refers to the largest disparity that can be fused as *Dmax*. Experiments have shown that Dmax depends on several scene factors, including the visual angle of the object being fused, the eccentricity, and the pattern on the object.



Panums fusional area can also be illustrated in disparity space, as shown below.



Binocular disparity and blur

Binocular disparity and blur give very similar information about depth.

$$\text{disparity in radians} = T_X \mid \frac{1}{Z_{object}} - \frac{1}{Z_{vergence}} \mid$$

where T_X is often called the 'interocular distance' or IOD.

$$\text{blur width in radians} = A \mid \frac{1}{Z_{object}} - \frac{1}{Z_{focalplane}} \mid$$

[2] In fact the iso-disparity surfaces in the scene are not depth planes, since the retina is not a planar receptor array. But let's not concern ourselves with this detail.

So, if the visual system is verging on the same depth as it is accommodating then

$$\frac{\text{disparity}}{\text{blurwidth}} = \frac{T_X}{A}.$$

With $T_X = 6cm$ and $A = 6mm$, this would give a ratio of 10:1. Indeed one does typically attempt to accommodate at the same depth as one verges – since the scene point we are looking at should be in focus. The above relationship specifies how two cues covary for points that are not on the vergence/accommodation distance. This covariance is presumably important for controlling accommodation and vergence. Indeed the neural control systems that control vergence and accommodate are closely coupled. (Details omitted.)

This close coupling between the accommodation and vergence systems is a problem for 3D cinema. Binocular disparities are used in 3D cinema to make the scenes appear 3D, and yet images are all presented at the display plane – the movie screen or your TV or laptop screen. When you look at an object that is rendered in 3D, you make a vergence eye movement to bring that object to zero disparity. However, normally your accommodation system follows along and adjusts the lens power so that you are accommodating at the same depth that you are verging. But for 3D cinema that is incorrect, since the screen is always at the same depth. If you verge your eyes to a point that is rendered with a non-zero disparity on the screen, then you will verge to a 3D point with depth difference than the screen. In that case, your accommodation system will get conflicting information if it follows the vergence system, namely the image on the screen will become blurred. The system will try to find a different depth to focus on to bring the image into sharp focus. However, this will drive the vergence back to the screen and away from the object that you are trying to verge on. There is no way to resolve this conflict, unless you can decouple the two systems. Most people cannot do this, which is why 3D displays give many people headaches and general viewing discomfort.

The other problem with 3D cinema is that the disparities are designed for a particular viewing distance and position, namely in the middle of the cinema audience. Anyhow who has sat in the front row at a 3d movie or way off to the side is familiar with this problem.

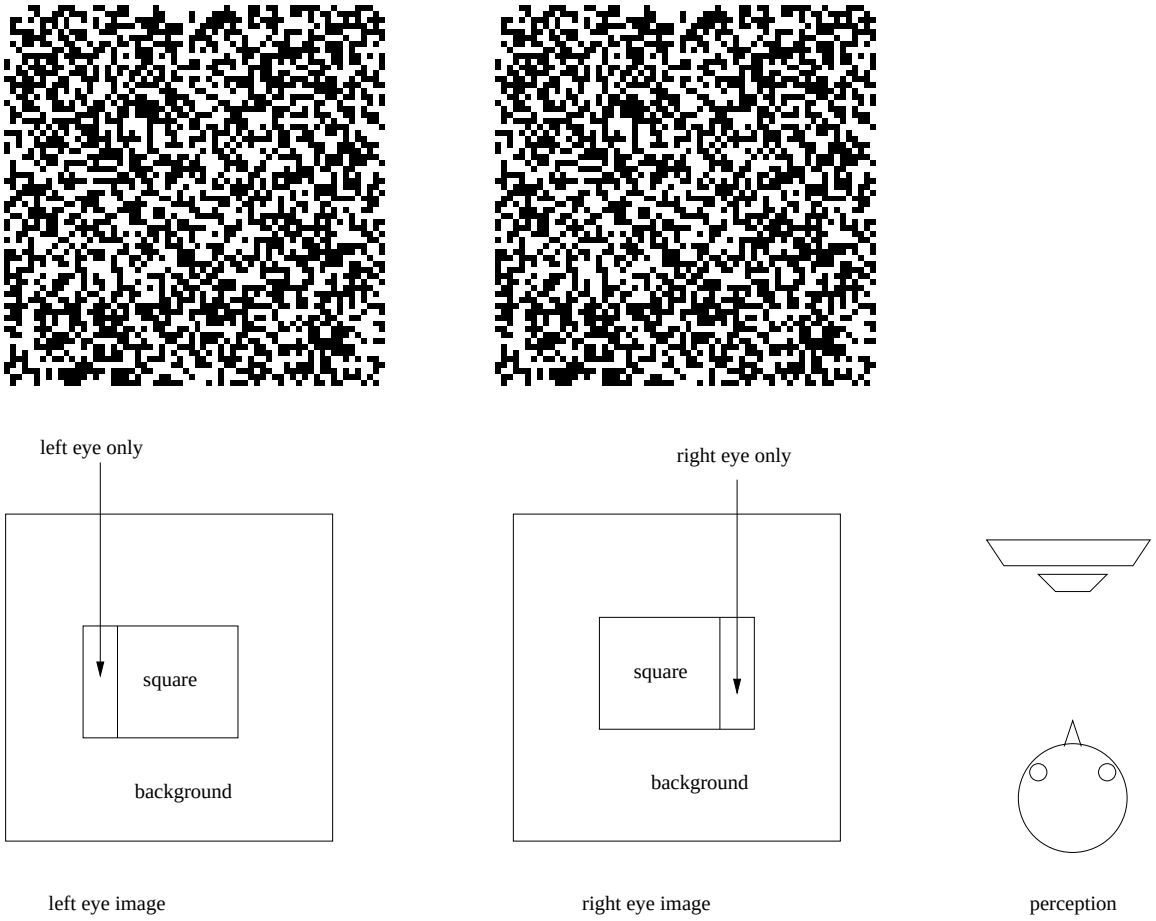
Random dot stereograms

One longstanding question in binocular stereovision is: How does the eye/brain match corresponding points in the left and right images? Up until the middle of the 20th century, it was believed that the brain solved this correspondence problem by finding some familiar pattern such as a line or edge or corner in the left image and matched it to the same familiar pattern in the right image, and vice-versa. This makes sense intuitively, since it was known that the brain follows certain rules for organizing local image regions into small groups of patterns. (We will discuss "perceptual organization and grouping" later in the course.)

In the 1960's, engineers and psychologists became interested in the process of binocular correspondence and fusion, and started using computers to address the problem – they did perception experiments using computer generated images. Computer scientists also began experimenting with writing computer vision programs using digital image pairs. One important type of image that was used was the *random dot stereogram* (RDS). RDS's were invented by Bela Julesz at Bell Labs. An RDS is a pair of images (a "stereo pair"), each of which is a random collection of white and black (and sometimes gray) dots. As such, each image contains no familiar features. Although each image on its own is a set of random dots, there is a relation between the random dots in the two images.

The random dots in the left eye's image are related to the random dots in the right eye's image by shifting a patch of the left eye's image relative to the right eye's image. There is a bit more to it than that though as we'll see below.

Julesz carried out many experiments with RDSs. These are described in detail in his classic book from 1971 and in a paper[3]. His results are very important in understanding how stereo vision works. They strongly suggest the human visual system (HVS) does not *rely* on matching familiar *monocular* features to solve the correspondence problem. Each image of a random dot stereogram is random. There are no familiar patterns in there, except with extremely small probability.



The construction of the random dot stereograms is illustrated in the figure below. First, one image (say the left) is created by setting each pixel ("picture element") randomly to either black or white. Then, a copy of this image is made. Call this copy the right image. The right image is then altered by taking a square patch and shifting that patch horizontally by d pixels to the left, writing over any pixels values. The pixels vacated by shifting the patch are filled in with random values. This procedure yields four types of regions in the two images.

- the shifted pixels (visible in both left and right images)

[3] B. Julesz, "Binocular depth perception without familiarity cues", Science, 145:356-362 (1964)

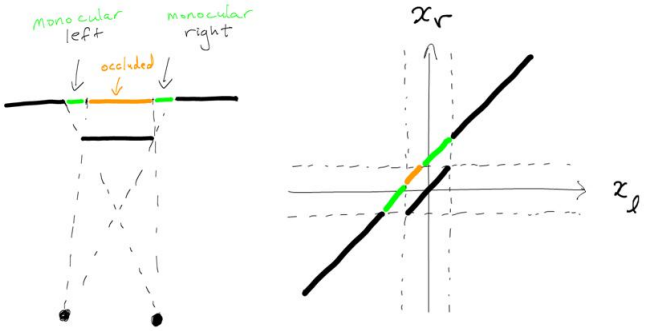
- the pixels in the left image that were erased from the right image, because of the shift and write; (left only)
- the pixels in the right image that were vacated by the shift (right only)
- any other pixels in the two images (both left and right)

To view a stereogram such as shown above, your left eye should look at the left image and your right eye should look at the right image. (This is difficult to do without training.) If you do it correctly, then you will see a square floating in front of a background.

Disparity space

Let's relate the above example to a 3D scene geometry that could give rise to it. The scene contains two depths: the depth of the square and the depth of the background. Suppose the eyes are verging on the square. We approximate the disparity as 0 on the whole square, and then the background has negative disparity.

Let's consider a 'disparity space' representation of the scene. Specifically consider a single horizontal line $y = y_0$ in the image which cuts across the displaced square. We wish to understand the disparities along this line. The figure below represents this line in the two images using the disparity space coordinate system (x_l, x_r). For each 3D scene point that projects to this line $y = y0$, there is a unique x_l and x_r coordinate, regardless of whether the point is visible in the image. (It may be hidden behind another surface.) Moreover, each depth value Z corresponds to a unique disparity value, since $d = x_l - x_r = T_x/Z$.



Notice that the set of lines that arrive at the left eye are vertical lines in the figure on the right, and the set of lines that arrive at the right eye are horizontal lines in the figure on the right. Similarly, each horizontal line in the figure on the left represents a line of constant depth (constant disparity). Each diagonal line in the figure on the right represents a line of constant disparity (constant depth).

In the sketch, we have assumed that the eyes are verging at a point on the foreground square. The background square has $x_l < x_r$ and so disparity d is negative.

Because of the geometry of the projection, certain points on the background surface are visible to one eye only; others are visible to both eyes; still others are visible to neither eye. Points that are visible to one eye only are called *monocular* points. In the exercises and assignment, you will explore this a bit further.

