Binocular disparity (review)

We have discussed the geometry of binocular vision a few times in the course, for example, in lectures 1 and 7. In lecture 1, we considered image position x_l and x_r on a projection plane $Z = f$, and we noted that $\frac{x}{f}$ was an angular distance from the viewing direction (optical axis), under the assumption that this angle is sufficiently small. In human vision, we are typically concerned with visual directions rather than positions on a projection plane, and so we will simply let x_l and x_r be a visual angle. Essentially, we are just letting $f = 1$.

If the eyes are parallel, then

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

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

$$\text{disparity (radians)} = (x_l - x_r) - (\theta_l - \theta_r)$$

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). Thus,

$$\text{disparity (radians)} = T_X(\frac{1}{Z} - \frac{1}{Z_{vergence}})$$

Since the brain controls the vergence, the brain in principle has information about the depth on which the eyes are verging. We will return to this observation later.

Disparity space

Point that are closer to the eye than the vergence distance have positive disparity, which is referred to as *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, which is referred to as *uncrossed* disparity since one uncrosses one's eyes to bring the disparity of such points to 0.

One way to think about disparity is to consider the space (x_l, x_r) as a 2D space. If we consider the 2D case of the XZ plane, then for any vergence of the eyes, any other point in (X, Z) will define directions x_l and x_r in the two eyes, and hence will define a point in *disparity space* (x_l, x_r). See slides 8-11 for some basic examples.

Note that, according to the above models, disparity depends on depth and so all the points at a constant depth Z correspond to a constant disparity. Thus, constant disparity lines are of the form $d = x_l - x_r$ which are just lines parallel to $x_l = x_r$. See the same slides as above. We will use this idea of disparity space through the lecture today.

Random dot stereograms

There are many ways to study biological stereo vision, including measuring from cells in the brain as we have seen, modelling them. Another very common way is to create stereo displays and examine what depth people see when they look at such displays. Let's consider a classical and very commonly used display which is known as a stereogram.

 ©Michael Langer

To motivate this, consider one longstanding question in binocular stereovision: How does the eye/brain match corresponding points in the left and right images? Up until the 1950's, 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.

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.

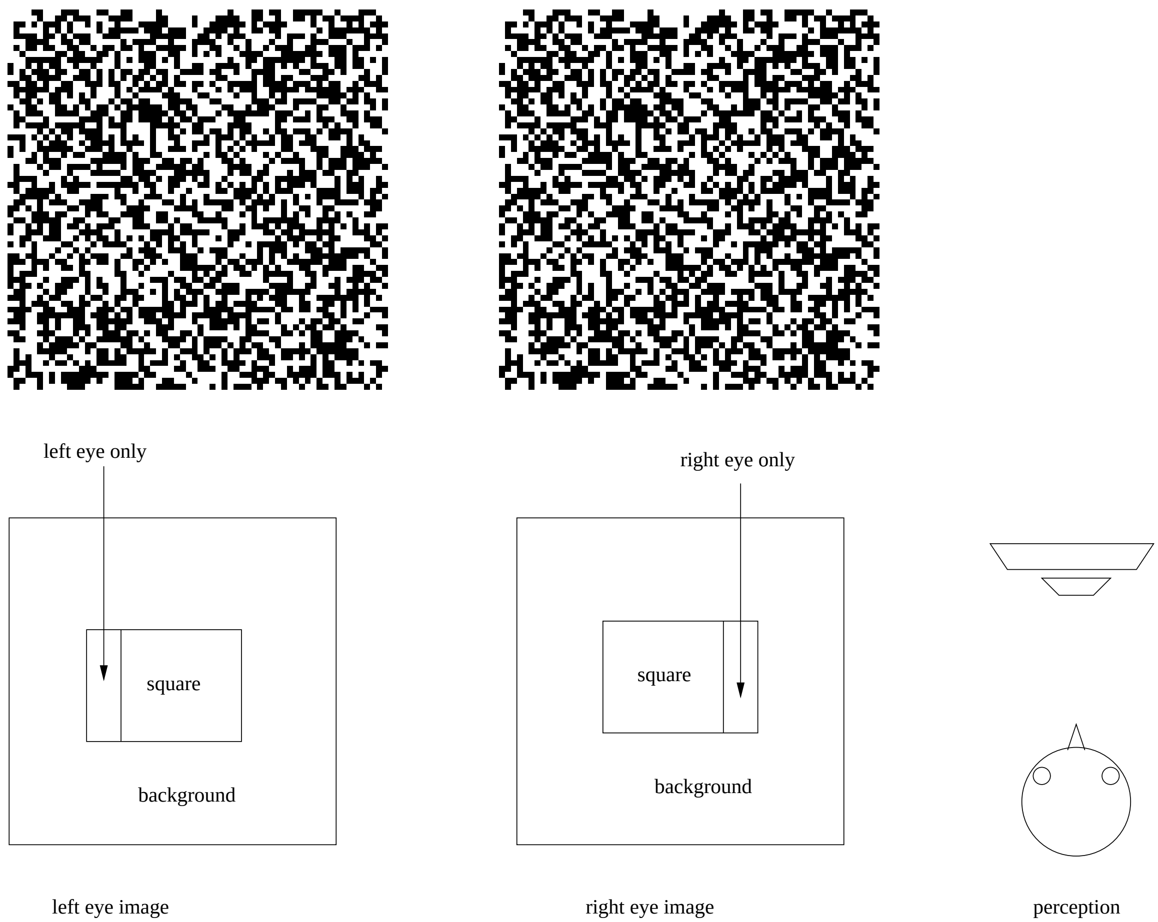
Julesz carried out many experiments with RDSs. These are described in detail in his classic book from 1971 and in a paper[1]. His results are very important in understanding how stereo vision works. They strongly suggest the human visual system (HVS) does not *rely* on matching familiar *monocular* features 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 value 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)
- 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.

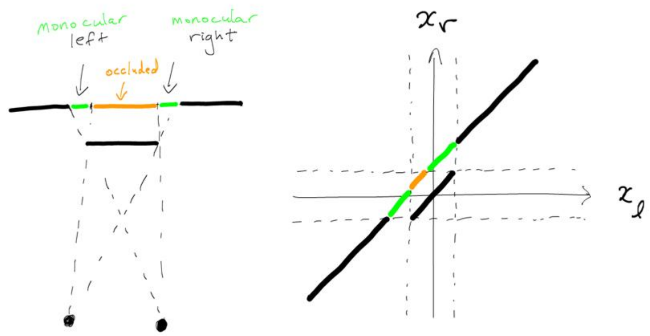
[1] B. Julesz, "Binocular depth perception without familiarity cues", Science, 145:356-362 (1964)



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. Take 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 $y = y_0$, 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). *The slides have essentially the same picture.*



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, you will explore this a bit further.

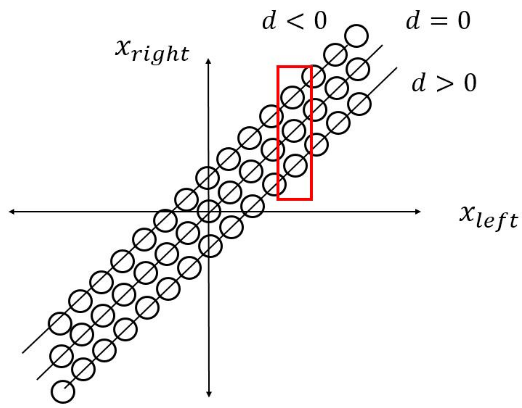
Binocular complex cells and Disparity Space

Now that you have a better idea of disparity space, let's use it to think about binocular complex cells. Let's think about what it means to have a family of cells that are tuned to different disparities.

The figure(s) below considers just a 1D case where the variable is x. Each binocular cell has two monocular receptive fields, centered at x_{left} and x_{right} and so we can indicate the binocular receptive field with a disk.[2] If the monocular receptive field centers are at the same position in the two eyes, $x_{left} = x_{right}$, and then this cell would be tuned to a disparity of 0. This is just the case of the example above. If the monocular receptive field center for the left eye is to the right of the monocular receptive field center for the right eye, then this cell would be tuned to a positive disparity. If the monocular receptive field center for the left eye is to the left of the monocular receptive field center for the right eye, then this cell would be tuned to a negative disparity. See the $d > 0$ and $d < 0$ zones in the plot below.

The idea for the figure is that, for each x_{left} (say) the visual system "considers" the set of binocular complex cells whose left monocular receptive field is centered there. See cells highlighted in red. The best estimate of the disparity would correspond to that of the binocular cell in the (red) set that gave the peak response. (In Assignment 2, you will look at two different ways of making such cells. One is based on differences and the other is based on sum, and you get a peak in the response (min or max, respectively) when the actual disparity is equal to what the cell is tuned for.

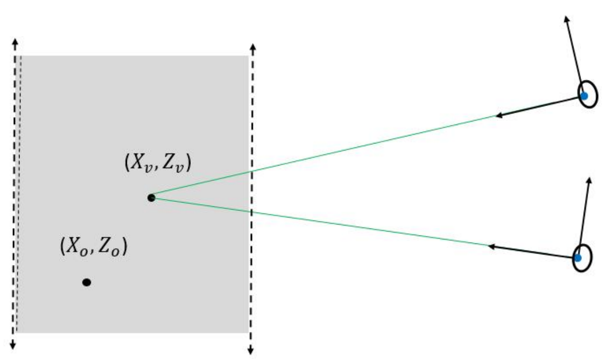
[2] or square, if you prefer – I use disk because I'm thinking of a Gaussian in each dimension and the product of two 1D Gaussians is circularly symmetric.



Binocular fusion

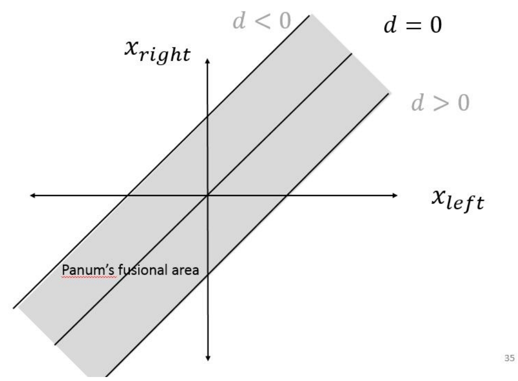
Recall that if you look at some object in the distance and you hold your finger close to your face, then you will see two images of your finger. Your visual system is not able to 'fuse' the left and right images of your finger because the disparity is too big. This creates 'double vision' or *diplopia*.

If the disparity of a 3D point is close enough to 0, then your visual system does fuse the left and right images of the point. This limited range of fusion disparities defines *Panum's fusional area*. Equivalently, Panum's fusional area is a range of depths in front of and beyond the vergence depth – see grey area in figure below below. Note 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.[3] One often refers to the largest disparity that can be fused as *Dmax*.



[3]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.

Panums fusional area can also be illustrated in disparity space, as shown below. The similarity of this figure and the disparity space figure on the previous page now leads to an important idea: *Panum's fusional area is believed to be due to a limited range of disparities of disparity tuned cells,* namely the reason we cannot fuse large disparities is that we don't have V1 cells that are tuned to these disparities.



Binocular disparity and blur

Binocular disparity and blur give very similar information about depth.

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

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

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

So, if the visual system is verging on the same depth as it is accommodating then, for points at depth Z in the scene, we have

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

Indeed one does typically attempt to accommodate at the same depth as one verges – since the scene point one is "looking at" (i.e. $(x, y) = (0, 0)$) should be in focus. The above relationship specifies how two disparity and blur covary for scene points that are *not* at this common vergence/accommodation distance.

This close coupling between the accommodation and vergence systems is a problem for 3D displays such as in 3D cinema. Binocular disparities are used in 3D cinema to cue the visual system that points in the scene lie at different depths. 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. Because vergence and accommodation are coupled, normally your accommodation system follows along and adjusts the lens power so that you are accommodating at the same depth that you are now verging. But for 3D

 ©Michael Langer

cinema this coupling creates a problem, since the screen is at a constant depth. If you verge your eyes to a point that is not at screen depth, then you will accommodate to a point that is not at screen depth, and the retinal image will become blurred. The accommodation system will then try to find a different depth to focus on to bring the image into sharp focus, and it will succeed when it drives towards screen depth. However, again because the systems are coupled, this will drive the vergence back to the screen depth 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.