Welcome

First of all, welcome to the course. This is unique course and it has evolved over many years. I started teaching it as a seminar course, then as a 6xx level graduate course where I mixed lectures with paper presentations by the students, and then recently I switched it to a 5xx level course so that both undergraduates and graduate students can take it. I am delighted that such a diverse set of students is signed up, including the traditional CS and ECE undergrads and grads, but now also a significant group of Cognitive Science and Neuroscience students. I hope you will get a chance to mix a bit with each other as the semester goes along.

Before we begin, I should tell you a bit about my background and where this course comes from. I did a Bachelor in Science, with a Major in Math and Minor in CS. My interest in vision science began with a course on AI which touched on some basic problems in computer vision and image processing, including one of the most fundamental problems of all: given an image, which is just a 2D array of intensities, can a vision system automatically find edges in images, in particular, 1D contours where the intensities have a high gradient in some property across the contour? This seemed like such a straightforward problem. Its just a matter of writing down a definition of an image edge, and then coming up with some pattern matching technique for finding local regions that satisfied that definition. Right? Alas, no. Wrong. Edge detection turns out to be much more subtle and challenging than that.

As an undergrad, I landed a job in the my last two summers in a neuroscience research lab where I was exposed to a different set of problems in vision science. One that fascinated me the most was how Shannon’s theory of information can be used to quantify the information carried by nerve cells. The lab I was working in was doing such analysis for motor neurons, but the same question could be posed for image coding too, and this idea caught my imagination. Another problem that fascinated me was visually guided reaching. In particular, why do patients with Parkinson’s disease have such trouble reaching for objects? Is it a motor problem? Or a vision problem? Or is it more subtle than that: is it a problem of perceiving the space itself through which the arm needs to move? Can such questions be answered with experiments? Or are they just too high level, and are only helpful for framing the problem?

For my M.Sc. thesis research, I considered the problem of image coding which I mentioned above. However, rather than asking about how efficiently a neurons code images (which frankly I did not have the neuroscience background for), I tried to characterize the amount of information that was inherent in images, based on some simple statistics, namely spatial correlations between pixel intensities. This problem hooked me for several reasons. First, the idea of abstracting away image coding algorithms and only looking at the information inherent in the images themselves appealed to me, from a mathematical point of view. Who cares how neurons do it? That wasn’t the fundamental question? The fundamental question – or so I thought at the time – was, what is the computational problem to be solved and what are the inherent constraints on the problem? Since the problem to be solve is to interpret real images, the question of how much information there is in the image seemed as fundamental as can be.

For my PhD, I decided to change directions and try a different vision problem. I have always been interested in looking at the world around me, rather than exploring my other sense. I owned a digital SLR camera as a teenager and took hundreds of photos and developed them in my darkroom at home. (Thanks to my brother for that. He is the engineering in the family.) I also used to draw a lot when I was a kid. So during my PhD studies I enrolled in several drawing classes and tried to
learn about shading, perspective, composition, and other elements of drawing and to spend nearly every waking hour thinking about what images are, how to make them, and how to analyze them.

One of my delightful discoveries was a lesson in a book *The Natural Way to Draw*. The author explained how to draw drapery surfaces, namely leave the top of the folds white, to make the side of a fold light grey, and make the bottom dark grey. This simple advice struck me as odd: it contradicted the shading model that was used in computer graphics and computer vision for describing the intensity on a curved surface. It turned out that the drawing advice in the book applied to surfaces under diffuse lighting conditions such as on a cloudy day, whereas the computer graphics and vision models applied on a sunny day. This idea led me to study the problem of how to estimate 'Shape from Shading on a Cloudy Day' which turned out to be the main topic of my Ph.D. research.

The training that I received in my MSc and PhD was much more about developing models and algorithms and less about doing experiments. This was fine with me at the time. However, as began to publish my work and attend conferences, I began to appreciate more how experiments are a big part of doing science. To get some training on doing experiments on human vision, in particular, I spent a few years in Germany and continued my shape from shading studies, and tried to ask if the models and algorithms that I had developed in my PhD work had anything non-trivial to say about human vision. Thankfully, the answer was yes, and I will say more about that later.

So that’s my background. What about yours? If you coming from neuroscience or cognitive science then you presumably already have thought about how visual and auditory perception work, and the next little discussion is not for you. It is more for those of you who have only a computer science or electrical engineering background, and who have not ever thought before about what perceptual systems are.

What are you concepts of visual perception? Those of who wear glasses are aware how important optics is for proper vision. Those who are color blind are aware that your visual experience must be different color from those with normal color vision, but the reasons why are subtle. If you have normal color vision then perhaps you think that you see all the colors there are. But that is incorrect: the human vision system is trichromatic only and this is an artifact of evolution, not due to the physics of light or image formation. Those of you who have normal binocular depth perception probably don’t think about it much, unless you go to a 3D movie and then wonder why the 3D looks a bit different than it does in the real 3D world. (Unless you are sitting in the center of the movie theatre, your 3D experience is distorted relative to what the filmmaker intended.) Finally, for those of you who learned how to draw back in high school, you know that there are rules of perspective. that parallel lines in 3D that are projected into an image should meet at a vanishing point, and that if lines are drawn incorrectly, then they will not meet at vanishing points and the drawing will not look right.

What are your concepts of auditory perception? You all know that sound travels as a wave, but perhaps you don’t know what that means from a technical point of view. It means that sound pressure satisfies a particular ‘wave equation’ that describes how air pressure changes locally over space and time. You know that sound is often characterized by frequency and wavelength in particular, musical tones are related to frequency. You also are aware that voices have high and low frequencies – men have lower frequency voices than women typically, and children have high voices than women. You also know that computer speech recognition is a difficult problem, though the reasons why are not obvious – since it seems relatively effortless for us, just like visual object recognition seems so effortless. Finally, your awareness of audition may be heightened if you have
a familiar member or friend who has hearing loss. You may have wondered how hearing aids work, and what is the difference between a hearing aid that is worn on the outside of the head versus a hearing aid that is implanted inside the ear (cochlear implant). But apart from these questions, you may take your hearing for granted. One of my goals in this course is for you to better appreciate how hearing works and problems your auditory system is solving.

**Studying Perception: from illusions to computational models**

One of the hooks for people to study visual perception is to see an example where you 'cannot believe your eyes'. Two of my favorites are below. On the left is *White’s illusion*. The short grey bars on the left appear to be much lighter shade than the short grey bars on the right. But in fact all short grey bars are the same shade of grey. The only differences between the grey bars on the left and right sides is the surrounding contexts. It is somewhat controversial among vision scientists how to explain why this illusion occurs, and the differences in opinion reflect general stances that different scientists take on what a good explanation is.

The example on right is quite different. Here we see two tables and on appears to be long and thin and the other more square. But in fact the two table top shapes are identical in the 2D image except that one is rotated by 90 degrees relative to the other. This means that our visual systems are ignoring the 2D shape and instead are interpreting the shape in terms of a likely 3D scene. One might ask if this 'explanation' accounts for the entire illusion or if there is something else going on. One might also ask if this really is an illusion. It *is* an illusion if we are asked to judge the 2D shapes and ignore the 3D interpretation. But it is perhaps *not* an illusion if we are being ask to judge the shape of a 3D table top, since we would need to know more precisely what is the correct 3D interpretation for such a drawing.

When studying perceptual systems, one often distinguishes aspects that involve the measurement of the environment from aspects that involve interpretation of the environment. This is the distinction of sensation versus perception. Our sensory organs measure a physical stimulus, and our perceptual systems interpret those measurements in terms of a remote event or object that caused the stimulus.

One often speaks of the five senses (vision, hearing, touch, smell, taste). But in the table below I group smell and taste since both involve the same sensory system, called olfaction. Note that there are other senses as well such as knowing and controlling where you limbs and body are in space (propiroception, balance), pain, temperature, nausea, and many other internal feelings that you might have that are related to the hormones. (Frankly I am not an expert at all on that, so I will say no more.)
I would like to discuss briefly what is computational perception, at least, how I think of it. My ideas are heavily influenced by reading David Marr’s book from early 1980’s. I stumbled on this book when I was an undergraduate summer student, hanging out in the Science library and trying to cram as much about vision as I could into those short summer months. I distinctly remember pulling Marr’s hardcover book off the shelf and reading the Introductory chapter. This was an inspiring piece and refreshing piece of writing, compared to the neuroscience and psychology books which were chalk full of facts but didn’t offer a way of thinking. Marr’s book was different, and I was hooked.

I will only sample a few of the ideas here, and I encourage you to find a copy of his book for yourself and at least read the first chapter. One point Marr made is that perception is a process. Although it feels as if perception is instantaneous and although we are unaware of how our percepts of the world come about, the fact is that something is required to turn our sensations into perceptions. Light patterns that arrive on the retina are different from the physical objects out there in the world, and both are different from our brain’s internal representation of the world. Something needs to happen to turn retinal images into internal representations. Part of that something can be called a computation – a manipulation of data from one symbolic representation (e.g. the intensity of light at a pixel) to another symbolic representation (e.g. a surface at some distance). Marr never claimed and I don’t want to claim either that perception is the same thing as computation. But it seems some aspects of perception can be understood best using concepts from computation. There are many subtleties here and we’ll touch on some of them later.

Let’s back off for a minute, and consider that there are many ways to study perception. For example, a behavioral psychologist might look at experiments for measuring human performance at some task, typically in a lab setting. A neuroscientist or neurobiologist would look at the physiology or anatomy of the brain. In particular, a computational neuroscientist might try to pin down neural codes and quantify how information is represented and communicated in the signalling of cells. Throughout this course we will touch on each of these approaches.

A slightly way to slice across perception is to consider different levels. At the highest and most abstract level is the task itself, that is, what one can observe about the system from the outside. What is the input and what is the output? What are the parameters of the task? Marr referred to this as the computational level, and the idea was the pose the perception problem by stating formally what the goal was and to identify how that goal could be achieved by transforming abstract representations of the input into representations of the output. What assumptions validate these transformations? What are the underlying models? For many who have read Marr - in particular, I would speculate for those with a computer science background – this computational level of analysis was the most interesting and novel. In any case, like it or not, this computational level of analysis of perception problems had not been articulated before.

Another level of analysis is that of brain areas and pathways. Within the visual system, there are subsystems that represents color and spatial detail, and motion and binocular depth, and to
some extent these representations are coded in parallel and transmitted along separate pathways. At a lower level are the neural codes. How are properties of the world encoded in the frequency of spikes of neurons and, going backwards, how can one take a spike train of a neuron and 'read' the information about the stimulus that is encoded by that spike train. Neural codes are often described in terms of receptive fields, which in the case of vision are the directions in the visual field that can affect the response of neurons. A related analysis is how a population of cells can encode a stimulus, and how much redundancy there is between cells and why. At any even lower level are the mechanisms of how nerve cells respond to their inputs and communicate with each other.

Why are we talking about these levels? I am not trying to overwhelm you by making you realize how much there is that you don’t know. Rather, I want to point out that there are analogous levels in computer science. At the highest level is the problem specification. For any well defined problem, we can talk about the input and the output. This is analogous to Marr’s computational level.

What about algorithms? Algorithms are considered to be a slightly lower level because, for any problem, there may be multiple algorithms for solving that problem. Similarly, programs in a high level language are considered to be a different level than algorithms because there are many programs (in different languages) that can implement essentially the same algorithm. As we get closer to the hardware level, we get to a different level of description of a computer. Assembly language and machine language are just specific implementations for particular computer architectures. They describe what is happening at a much more specific level, just as describing single neurons or populations of neurons is more specific than discussing general brain areas. Finally, we can go down to the level of gates, circuits and even transistors in the computer. This is analogous to the mechanism of single nerve cells, and how they generate spikes, for example.

This analogy between perceptual systems and computers is meant to be intriguing. But to be meaningful, one probably needs to have a good handle either on perceptual systems or computers. When one first studies computer science, one can easily become overwhelmed with details, and have trouble organizing one’s concepts into different levels. But eventually one does this organization, just as the neuroscientist understands that some aspects of brain function are only meaningful to talk about along with other aspects – since one has to be careful to keep the level in mind. Throughout this course, I will remind you of this issue, and hopefully by the end of the course you will have a better sense of what these levels are.

Why eyes?

Our eyes are very sophisticated optical devices and it took millions of years to evolve\footnote{See interview with Richard Dawkins (outspoken evolutionary biologist at U. Oxford) \url{https://www.youtube.com/watch?v=bwX3fx0Zg5o}} There are many animal eyes which are much less sophisticated than ours. Indeed if we trace back to our ancient ancestors, we would find that their eyes were very poor indeed.

It is not known exactly how eyes evolved, but it is believed that the earliest eyes consisted of a small number of light sensitive cells distributed over a small region on the outer surface of an animal (and hooked up to a primitive nervous system). Let’s suppose the cells are distributed over a concave pit such as in the figure below. Six light sensitive cells are shown. Because the pit is concave, each cell will receive light from a limited set of directions. For the leftmost and rightmost cells, the range of directions of light coming from the scene is shown in the figure.
Suppose it is a sunny day, and the sun is in a direction (to the left) such that it “is visible” to the rightmost cell but not visible to the leftmost cell. By “visible”, I merely mean whether the sun directly illuminates the cell. Thus, the rightmost cell would signal a much brighter (total) light than the leftmost cell.

Now suppose that something to the left of this animal were to move towards the animal and block the sun, i.e cast a shadow, so that the sun would no longer be visible to the rightmost cell. The leftmost cell would still not see the sun, and now the rightmost cell would not see the sun either. These facts together tell the animal something about where the approaching animal is (namely, to the left). A defensive response of the animal therefore might be to move toward the right, i.e. away from the approaching animal. (Or the animal could move to the left, which would be a more aggressive response.)

One way to improve this vision system would be to make the eye more cavelike, as in the figures below on the left, and even more cavelike as on the right. Here we are reducing the aperture by which light can enter the concavity. This reduces the angle of incident light that reaches each photocell of the eye.

Let $A$ be the diameter of the aperture and let $f$ be the distance from the center of the aperture to the sensor surface. Then, $A/f$ is approximately the angle subtended by the aperture, as seen by a photoreceptor. The exact formula depends on the shape of the concavity and the position of the photoreceptor within the concavity: let’s not concern ourselves with this.

The advantage of reducing the aperture is that each cell receives light from a more restricted set of directions which provides the eye with more detailed information about the directional distribution of light arriving at the aperture. The understand this, consider the figures below. The pattern of light in the scene is an alternating dark grey and light grey, and each cell averages over some dark regions and some light regions.
In the figure on the left, the aperture is so big that each cell would see part of a light grey region and part of a dark grey region. If the aperture were slightly larger, so that each cell received light from an equal amount of light and dark regions, then each cell would receive the same total amount of light. In this case, we would say that the light and dark greys had been blurred away completely.

In the figure on the right, the aperture has been decreased, and the pattern of light and dark grey has a higher frequency. This figure is drawn such that the angular width of the aperture is exactly matched to the width of the pattern on the surface. (For the figure on the left, this was not the case.) Notice that the retinal image would still be blurred, since most cells would receive a mix of light from light and dark regions and only a few cells would see only a light gray or only a dark grey. The retinal image would look more like a sinusoid than the piecewise constant intensity (light,dark,light,..).

The disadvantage of using a smaller aperture is that it reduces the amount of light reaching each cell. Eventually, if the aperture becomes a "pinhole", the image will be very dark indeed. This will lead to a decrease in image quality because of internal "noise" in the cell. This should be intuitively obvious: neither your eyes, nor digital cameras work very well at low light levels. We will have more to say about noise later in the course.

**Radians vs. degrees**

We can define an angle either using units of degrees or radians. Recall that $2\pi$ radians is 360 degrees, or $\frac{360 \text{ degrees}}{2\pi \text{ radians}} \approx 57$ degrees per radian. That is, one radian is about 57 degrees. We will use these quantities through the course.