[RUSTLING]

[SQUEAKING]

[CLICKING]

JOSH MCDERMOTT: Let's talk about depth perception. So the problem here that we're going to talk about today is that the input to the visual system consists of two-dimensional images.

The world is three dimensions, and people are pretty good at estimating the three dimensional structure. You can hold the shape in your hand and visually be able to assess the shape. You can tell how far things are away from you, and so on and so forth.

So the extraction of three dimensional structure from the two-dimensional images that the eyes receive is important for lots of things-- for navigating through space, for interacting with objects and so forth. And the most reliable source of information about the three-dimensional structure of the world is visual, at least for humans. Some other animals use sonar and other things like that.

All right, so the keys to the solutions-- one key is the fact that you have two eyes. And we're not going to talk about that in this lecture. That'll be for next time. The other key are assumptions that we make about the world that we have presumably learned or internalized over evolutionary development, coupled with implicit knowledge of the physics and the geometry of light that comes from objects to the eyes.

And so we're going to talk a lot about depth cues, right? We've talked about cues previously in the class. When we talk about a cue, we mean a source of information, so some aspect of the stimulus that provides information, in this case, about a variable that we care about, depth.

So there are lots of visual cues to 3D shape and depth. So stereopsis is the one that is best known, but you can see three dimensions without it.

So if you just close one eye, the world doesn't really look that different. You can still tell that certain things are further away than others, and you can see the three-dimensional shape of objects pretty well.

So this is a partial list of cues to depth. We'll talk about a whole bunch of these-- shape from shading and drop shadows, linear perspective, texture gradients, familiar size, height and field, aerial perspective. Last time we talked about the kinetic depth effect. There's also motion parallax. Those are depth cues for motion-- occlusion, focal blur. And this is not an exclusive list.

So the first thing I want to talk about today is shape from shading. So we've alluded to shape from shading previously in the class, and it refers to the fact that variations in luminance are often interpreted as shape.

And the basis from shape from shading is that many surfaces are approximately Lambertian. So we talked about Lambertian reflectance in the context of lightness perception. So a surface is Lambertian if the amount of reflected light is determined by the angle between the incident light ray and the surface normal vector. But the surface, otherwise, scatters like equally in all directions.

And so this is a picture that kind of depicts this. So the black arrow is the surface normal. So this is a horizontal surface, so the normal is perpendicular to that. The yellow arrow is the direction of the illumination. And then, the little blue arrows are scattered light rays.

And so what you're supposed to take away from this is that irrespective of the direction of the illumination, the light gets scattered equally in all directions. But you can see that the light is-- the most light reflects off the surface when the illumination is perpendicular to the surface. And as the illumination becomes more oblique, there's less light that's scattered. So this is the characteristic of a Lambertian surface.

All right, and so the basis of shape from shading is kind of like the opposite of what is shown in this diagram. So this diagram is showing a situation where the surface orientation is fixed and the illumination direction is varied. And that doesn't happen very often in the world.

So typically, in the world, there will be one light source, and so the direction of illumination will be fixed in some particular direction. But the orientation of the surface will vary because an object is curved.

And so for a fixed lighting direction, the amount of reflected light will then vary with the surface orientation. And so it follows from that that if you knew the direction of illumination, and if you also knew that the intensity changes in the image were only due to shading, then you could infer shape.

And so there are a bunch of classic demonstrations that people do seem to do this under many conditions. So this is a classic stimulus and illusion, if you will, where the physical stimulus is these circles. And inside each circle, there's a luminance gradient. And it just goes from light to dark or from dark to light.

And so when you look at this, most people will see a set of bumps and a set of craters. OK. So does anybody want to hazard a guess as to why some of the circles look like bumps and others like craters? Yeah.

AUDIENCE: Well, because you expect that the light is coming from above, so it looks like it would be illuminating [INAUDIBLE] from above. Then, it makes sense that there'd be lightness on top of that, shading on the bottom.

JOSH MCDERMOTT: Yeah.

AUDIENCE: [INAUDIBLE]

JOSH MCDERMOTT: Yeah, that's exactly right. So the visual system seems to have a prior that favors illumination from above. So we very commonly-- by default, we'll assume that illumination comes from above. And so if the illumination was coming from below, then the inferred shape would be the reverse of what it is, right? So the things that are bumps would be craters and vice versa.

So shape from shading is important in art. So bas-relief is a style of sculpture that relies exclusively on shape from shading. So there's some material here, and somebody carves some 3D structure into it, but then, all of the luminance variation just comes purely from shape. So it's illuminated from a particular direction, and then, you're able to see the 3D structure of the shape from shading.

You can also see evidence of shape from shading in lots of real-world photographs. So this is a photograph of a crater either on the moon or Mars. I forget which. But it doesn't really look like a crater here, right? So the prior favoring light from above is again at work. So this is the crater. So it's just flipped upside down here. So there's lots of examples of photographs of things where if they're taken from an orientation where the illumination is coming from the direction that you don't expect, they look funny.

This is one that I just saw the other day that's kind of striking. So this is a picture of sand dunes. But it really doesn't look like sand dunes, right? It looks like a whole bunch of little craters in the surface. So here's the other orientation, and you can see the correct 3D structure.

This one-- this is a really pretty awesome demonstration of the same thing. So this is a physical thing that was actually engineered. So there's some 3D shapes that are embossed into this thing. And then, it's lit from a particular direction. But it's on this table that can rotate, and so when you look at it now, it kind of looks like this is sticking out, and this is sticking out.

But whoa. Now it flips, right? OK, it's going to get even cooler. So now, they're going to pour some liquid so you can verify which parts are actually the valleys. OK, watch closely.

Yeah, so you know that the places where the liquid is present are kind of indented in the surface, but your priors on the direction of the illumination are causing you to misperceive the shape from shading. So again, a great example of how perception is kind of encapsulated and sometimes cognitively impenetrable.

OK, so that's shape from shading. So again, key principles of shape from shading, it works for Lambertian surfaces. Really, the kind of most straightforward way to infer shape from shading is to know or assume a direction of illumination. Sometimes we assume the wrong direction of illumination, so you get the shape right, shape wrong. And critically, you have to also assume that the intensity variation is due to shape rather than due to, say, paint.

So in these actual-- this is actually something you could print out on a piece of paper. And so the intensity variation here would be due to paint, and so you misperceive things.

Now, in the real world, things are actually much more complicated, because shapes often do have reflectance variation on it, and people are actually pretty good. If you take an object that has reflectance variation, you're pretty good at actually correctly estimating the shape.

So your visual system has this amazing ability to actually kind of separately infer the contributions due to shading and the contribution due to reflectance. And that's kind of a little bit beyond the scope of what we're talking about here and not terribly well understood but remarkable.

Another very powerful cue to depth is shadows. So this is just a display that has a few shapes, and you can see that some are in front of the others. And you can tell that this occludes this, and this occludes this. So there are these occlusion cues present here.

But then, when we add shadows to the display, these are typically called drop shadows, OK? You now actually see the objects as being separated in depth, right?

So shadows are a very powerful cue to depth. The visual system-- so typically, in the real world, they are caused by one object occluding the illumination, so blocking the illumination. And so the region of shadow is typically dark in the image, OK?

And so the visual system seems to have internalized that regularity. Because when you artificially make the shadows lighter rather than darker, and you can do this in graphics programs, then you don't get the sense of depth. And what does this look like to people actually?

AUDIENCE: It looks like [INAUDIBLE].

JOSH MCDERMOTT: What's that?

AUDIENCE: It looks like the inverted image, the casting white shadows.

JOSH MCDERMOTT: Yeah, or spray paint kind of, right? To me, it looks like somebody took a can of spray paint and sprayed it. So again, it really doesn't have the same effect of just a shadow-- doesn't look like one. And if you just invert the whole image then, it also doesn't really produce the sense of depth. So shadows have to be dark.

In computer graphics programs, you can play around with both shape from shading and drop shadows. And these things, again, they're happening all the time in the world, and you typically are not-- you just don't even really notice them. So this is just a picture of eggs.

And in this case, it's pretty obvious that the eggs are sitting on a table, because you get these kind of shadows that suggest that the eggs are sitting on the table and blocking the light. If you remove the shadows, the eggs kind look like they're floating in space in front of the table, right?

But if you get rid of the shading, so you just color all the eggs white, now you lose the 3D shape. So normally, you look at these eggs and they all just look white, right? So it's like your visual system is kind of separating out the contribution of the shading from that of the reflectance.

And remember how when we looked at the Retinex algorithm, when we ran the Retinex algorithm on an actual image, Retinex would make the mistake of actually interpreting some of the shading that's due to shape as being due to illumination. Because Retinex assumes that all of the gradual variation in luminance is due to illumination, which is not a correct model of the world. So there's lots of luminance variation due to shading.

So the visual system is really sensitive to the relationships between the positions of shadows and the depth relationships of objects to the surfaces on which they sit. So the drop shadows kind of have to align correctly for an object to sit on a surface.

So you remember, in the very first class, we showed this demonstration of a ball that kind of rolls across the screen. And depending on the trajectory of the shadow, you see these very different 3D trajectories.

And there's lots of these really cool-- I kind of collect these images, where just by accident, the shadow's kind of in the wrong place, at then it causes things to look like they're levitating. So this is a situation where I guess there were a lot of flagpoles on this particular beach, and there's one that just happens to be casting a shadow. The person's standing on some platform on the sand, and the alignment of the shadow and the surface are such-- are good enough that the person looks like they're kind of floating.

Here's a situation where there's, I guess, just some dirt on the ground that kind of happens to be in approximately the right place and approximately the right color to plausibly be a shadow and so causes the trash can to float. This is a cat with supernatural powers. Again, same kind of thing, just dark patch in the right-- in the right place.

You can also get some pretty funny instances of this involving water, because the shadows will be cast on the bottom of the ocean in this particular case and can cause the boat to seem like it's floating.

This is another one that I came across, where the particular angle of the sun causes the shadow to be in a place where you can see the train is floating. So you start looking for these things, and you see them all over the place.

So shadows are caused by the geometric relationship between an object and a light source. And so when you're using the shadow to infer depth, you're kind of implicitly taking advantage of that relationship. So the visual system has, like, internalized something about the relationship between light sources and objects and shadows.

So you might imagine that what happens is you actually are building or you're inferring, like, an entire scene layout in your head, where there's a light source in a particular place and shadows that are cast relative to objects. And that doesn't actually seem to be what is happening.

So this is kind of an interesting demonstration, where you get pretty good depth from the shadows in the top two circles of circles. Logically, that display doesn't really make sense, because you can tell from the direction of the shadows relative to the objects that in this case, the light is coming from the upper left. In this case, the light must be coming from the bottom left for the shadows to be where they are. And so you couldn't really actually generate an image like this without some really contrived setup with two sources of light that were only cast on part of the scene. But it doesn't seem to bother you.

On the other hand, down here, where the positioning of the shadows has kind of been randomized locally, that really seems to kill the sense of depth for the most part. So there's some degree of local consistency that is really required for this to happen, but it doesn't seem like you're kind of modeling the global scene.

And this is like-- we've seen some other hints of this so far in the class. Remember when we were talking about lightness illusions, we saw evidence that people weren't really modeling a full kind of global version of the scene? There's more local, kind of mid-level stuff going on. Any questions about shadows?

So another regularity that influences depth is the geometry of the geometrical relationships between where objects are in the world and where they project on the retina. And so very often, we're standing on a ground plane, and so things that are further away from us tend to be on the ground plane. And so things that are further away tend to be higher in the visual field. And so you can get these effects like this. And also, as things move are further away, they also tend to project smaller images.

And so you can get these effects. So the one on the left gives like a pretty compelling sense that the surface is receding in depth, because the objects get smaller as they go up in the visual field. That's more or less what would happen in a typical situation in the world. The one on the right has much-- has a much weaker effect, because it violates that geometric regularity that is typical of what you would observe in the world. So again, the common theme here is the visual system has internalized the regularities of the world and uses those to solve this ill-posed problem of depth perception.

So this is an example-- it's kind of an example of a texture gradient, in the sense that you have all these repeated elements. And the texture changes in some systematic way that's kind of consistent with something receding in depth. And those are actually super common.

And this is a case where an artist decided to trick the visual system. So this is a funny pattern that got painted on a perfectly flat floor. But instead of interpreting this as what it is, which is a flat floor that's got a weird, warped texture on it, you infer a situation where the depth is kind of-- where the shape of the floor is kind of curved, and the texture is uniform and just pasted on the curved surface.

So, of course, the inference is ill posed. And so you've got to have priors to constrain the inference. In this case, the prior's that the texture is uniform, because textures in the world tend to be uniform rather than having the weird kind of modeled shape. Here's another pretty cool example of the same kind of thing. All right, so texture gradients are another really powerful cue to depth.

OK, another very important thing that is related to depth perception is what's called Emmert's Law. So Emmert's Law describes the relationship between the size that we perceive things to be, the distance that we perceive them to be away from us, and the visual angle that they subtend.

So the idea is shown here, which is that you can have two objects of different sizes. And if the big one is further away, it will subtend the same visual angle as a small one that's closer. So there's this relationship between visual angle, inferred distance, and inferred size, because in the actual world, there would be a relationship between visual angle, distance, and size.

And so there's lots of these instances where either you know how big something is, because it's a familiar object, and so you use that to infer how far away it is, or you know how far away something is, and you use that to infer the size. And so we're going to do an experiment to verify this.

And so this experiment is going to involve staring at one of the lights in the ceiling. So you're going to do this for, I think, about 30 seconds. So get kind of close to a light and look up at it. And the purpose of doing this is we're going to burn a temporary afterimage into your eyes.

And the reason that we're doing that is that when you generate an afterimage in the eye, right, that's covering some fixed portion of your retina. So it corresponds to a particular kind of visual angle that would normally project onto the retina. So the idea is that after you stare at this thing, you're going to get this afterimage. It's going to have a fixed extent on your retina.

So now what we're going to do-- all right, hopefully, you got a pretty good afterimage. So now, I want you to look at your hand, OK? All right. So that's going to be some particular size. Now, take your hand down and look at the wall, OK? And the thing should look a lot bigger. You might have to keep blinking to get the afterimage image to restore itself.

All right, so what we just did is we generated a stimulus of a fixed extent on the retina. But you can cause it. So now, I'm like looking at it on the back wall, and it's enormous. So you can cause it to look different sizes by looking at things at different distances. And so it seems to be the case that the visual system assumes that that retinal stimulation is due to something that happened at the surface that you're looking at. And so you look at something that's close, and you assume that the thing is small. If you look at something that's further away, you infer that it's large. Yeah.

AUDIENCE: Why does blinking cause it [INAUDIBLE]?

JOSH MCDERMOTT: Oh, it's just-- it's refreshing the adaptation. Yeah, so you're kind of temporarily bleaching the photoreceptors, and I think probably what's happening is that there may be some downstream adaptation that you then kind of reset with the blinking. Yeah. So you're just making the afterimage visible again. I don't know the exact mechanism, actually.

OK, so Emmert's Law-- so the relationship between how big things look and how far away they seem. So this is-- that's one example of this, and it's a pretty good party trick.

And you see Emmert's Law in lots of different settings. So here, this is a situation where the big balls look a lot closer than the small balls. So this is a case where people-- the visual system seems content to assume that the objects are all the same size, and thus the changes in size are attributed to differences in distance.

Here's a situation where you know the relative size of these body parts. You know the hands are of a certain size relative to somebody's face. And so the fact that the hand is bigger in the one on the left leads you to conclude that it's a lot closer. So again, hand looks closer, and this all kind of happens implicitly.

And so again, if you look around, you can find funny instances where of plays some tricks on perception. This is kind of an interesting photo that I came across, where the girl in the middle kind of looks like she's floating. And so I think the reality is that she's a lot taller than her brothers, right, but your visual system seems to assume that these things are all probably the same size, and thus that the girl has to be closer than she actually is, which means she's floating in the air or something like that.

Here, we've got some really big pigeons, OK? So this is a kind of a funny photo. Can anybody tell what's going on here? Yeah.

AUDIENCE: [INAUDIBLE]

JOSH MCDERMOTT: Yeah, but you don't notice the ledge at first, right? So it sort of looks like they're right next to the car. So if you assume that the cars and the birds are at the same distance, then they're about the same size on the right, which means they're about the same size physically. So you've got giant pigeons.

All right, Emmert's Law. Any questions about Emmert's Law? So in some cases, the size is unambiguous, and thus, that kind of determines the apparent distance. In some cases, the distance is less ambiguous, and that will determine the size. Yeah.

AUDIENCE: [INAUDIBLE] there has been no large sign [INAUDIBLE] it's like why do we perceive the pigeions as large [INAUDIBLE] and not small cars?

JOSH MCDERMOTT: Yeah, I think it's a good question. Yeah, I don't know the answer to that, and I'm not sure. I'm not sure that it's completely known. And, I mean, you might imagine, well, most of us have a lot of experience with toy cars, so you might think that it could be the opposite. Yeah. Yeah, no, I think the absolute size of these things is-- it's interesting that you anchor with the car rather than the pigeon. Yeah.

OK, another cue to depth is aerial perspective. So in general, things that are further away are more blurry, because as light travels through the air, it gets kind of-- it gets scattered. And things also tend to look a little bit more bluish.

And so you can see this oftentimes in photographs. This is a photograph of a beautiful landscape. And the stuff that's further away is lower contrast, blurrier, and a little bit bluer. So again, it's like a regularity of the way that optics works that we seem to have learned.

Another really important depth cue is linear perspective. So linear perspective is due to the fact that lines that are parallel in the world will generally converge in the image, unless they lie in a plane that's parallel to the image plane. So if you have lines that are in a plane that's parallel to the image plane, those will remain parallel in the image. But if they're not in the image, they will converge. And so you can see in this particular case, this line and this line are a good example of that, or that and that.

OK, so the question is, like, now how do we know that lines are parallel in the world, right? In principle, these lines might not be parallel in the world, and that's-- does anybody have an idea? Why are lines parallel in the world? What would cause things to be parallel?

AUDIENCE: Humans?

JOSH MCDERMOTT: Humans, yeah. I think that's one possibility. Yeah. Maybe also gravity. But people like to construct rectangles, so there's a lot of things in human-constructed environments that tend to be parallel.

And so this is a classic illusion where depth from both linear perspective and from texture-- so again, we've got linear perspective, so these lines kind of are converging, makes it look like that's a lot further away than this. There's a texture gradient. And what's cool about this is that it looks like there's a really big monster chasing a smaller monster. But, in fact, the two monsters are exactly the same size in the image.

So this is, again, Emmert's Law. So you've got the same visual angle being subtended by this monster and that monster. The depth cues indicate that this one is further away than this one. So Emmert's Law tells you that this has got to be proportionally bigger than this one.

Here's another kind of cool example. So we've got two cigarettes, and they're positioned on a drawing that is making use of linear perspective to cause this to look like it's closer to you than this. The two cigarettes are physically the same size, and thus, they're lying on a table, so they're the same size on the retina. But they look very different, right?

This is a thing that I just saw that's kind of interesting. So check this out. See if you can figure out what's going on. So this is an artist who's drawing this thing. It's going to look pretty 3D.

Now, he's going to jump. Notice that the artist has changed in size, all right? So it looks like this person is changing in size. Anybody want to try to explain why the person looks kind of small here, and bigger here, and very big there? Yeah.

AUDIENCE: He's getting physically close to the camera, but you don't think he is because he's-- like, he's just turning [INAUDIBLE].

JOSH MCDERMOTT: Yeah. Yeah. So in reality, this person is jumping horizontally by a pretty big distance. So the distance to the camera is changing. You are misperceiving this person as actually kind of jumping up and down vertically, right? And if you assume that the distance is remaining constant, then the person's got to be changing in size. So in this case, the depth cues here are so powerful that they overcome the presumably strong prior you have that people stay the same size. Yeah.

OK, all right. So another point I want you to take away from this is that we see in 3D, right? So what you perceive is your inference of the three-dimensional structure of the world. Now, that is derived from these two-dimensional retinal inputs. And it's very difficult for you to actually counteract the visual systems recovery of 3D shape.

So this is a classic demonstration that was constructed by Roger Shepard, who's a cognitive psychologist, who also kind of was an amateur artist. And so he published a book of these illusions that he came up with. And this is probably the most famous.

And so the illusion here is that the parallelograms here and here are physically the same shape-- the same size, same shape in the image. You're shaking your head. You don't believe me, right? Yeah, it doesn't look possible.

OK, so here's one way to think about-- to see this. So this is the image kind of rotated and overlaid. You can see that they're actually the same. Here's another animation. OK. Yeah. But it's really kind of hard to believe.

And so the idea is that these two different shapes in the image look like kind of different shapes in the world, because the tables are kind of oriented differently and angled differently. And you infer these-- so you're inferring these different 3D shapes in the world, and it's really impossible to access the 2D shapes in the image. Yeah.

AUDIENCE: Maybe this is another illusion in that the image, but is it relevant that the parallelogram seems slightly askew? The bottom of the table or of the long table doesn't look parallel to the top of the long table to me.

JOSH MCDERMOTT: Yeah, it doesn't to me either, but this one doesn't quite look exactly rectangular either. Yeah.

AUDIENCE: [INAUDIBLE]

JOSH MCDERMOTT: I don't think that's important. I think this is just-- that's just another version. I think in the original version, they actually-- they look pretty well matched.

Yeah, but the point is that-- so this distance in the image-- this is the same as this distance in the image. But when you look at it, it looks like it's a much greater distance because of the foreshortening effect of projective geometry. So you have this thing that's getting projected onto an image, and you're inferring the three-dimensional shape in the world overcoming that projected process or inverting it correctly.

OK, so that's kind of an important theme. And then, the thing that we'll end with here is this issue of ambiguity and bistability. So one of the amazing things about perception and vision in particular is that we're constantly solving these little posed problems. So three-dimensional interpretation in the world in particular is always ambiguous, especially if you look at the world with one eye. So there's lots of different possible three-dimensional structures that are consistent with the image that you observe.

But usually, we see the one interpretation that is apparently deemed most likely by our visual system. So vision is typically quite stable. But sometimes, especially in these more impoverished displays that we often make use of in perceptual science, there are multiple interpretations that are equally likely, and the percept is bistable.

So this is called the Necker cube. It's like a wireframe cube. So if you just look at this one on the left, you'll be able to see it in two different orientations. So one orientation is that this square is out in front. The other orientation is that this square is out in front. And those are both consistent with the image, and if you look at it, your percept will kind of flip from one to another, maybe every five or ten seconds or something like that.

OK, raise your hands if you've seen it flip. OK, almost everybody. Yeah.

And so there are lots of famous examples like this. So this is kind of combining two famous illusions. In this particular case, the face can be seen as an older woman from the side or a younger woman kind of from the back. And this person is holding their hands up in a silhouette, and that's the duck-rabbit illusion. You can see it as a duck or as a rabbit.

OK, so this is a cartoon that I saw recently. So the bartender says, can I see your ID? Wait, never mind. Wait, yeah, I need to see your ID. Wait, no-- constantly switching between these interpretations of someone who's underage and someone who's over age.

And so this bistability, we had a question about this last time. Bistability is super interesting, the fact that you don't get stuck in one interpretation. And one idea is that this is a situation where the posterior-- so the probability of the world, given the stimulus, is multimodal. So there's multiple interpretations that are both probable, and maybe your visual system is sampling from the posterior so that you sometimes see one thing, sometimes see another.

Maybe it relates to adaptation and what happens is you see one thing, and so some neurons kind of adapt. And then, the other interpretation kind of ends up getting favored in some way-- could be potentially due to implementation constraints. We don't really know, but it's fairly prominent. And then, these are some other examples of bistable things, analogous to the duck-rabbit illusion.

All right, let's end there. So next time, we'll resume talking about monocular depth cues. And then, we'll get into stereopsis, how you derive depth information from having two eyes.