Shading and shape (continued from lecture 11)

Last lecture we examined the problems of perceiving surface shape from texture and shading. The discussion was not at the level of neural coding, but rather it was at the level of what problem was to be solved. What are the 3D scene properties that we mean when we say "shape' (e.g. depth, depth gradient – slant and tilt – and curvature). How are these properties related to image intensities?

We begin today by considering a few variations on the shape from shading problem. We assume that the physical intensity of light reflected from an image depends on the angle between the surface normal and some light source direction **L** which we assume to be constant i.e. the source is far away like the sun:

$$I(X,Y) = \mathbf{N}(X,Y) \cdot \mathbf{L}$$

where

$$\mathbf{N} \equiv \frac{1}{\sqrt{(\frac{\partial Z}{\partial X})^2 + (\frac{\partial Z}{\partial Y})^2 + 1}} (\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}, -1) \ .$$

Note that intensity $I(X,Y)$ is being defined in XY variables of 3D scene coordinates XYZ – where $Z(X,Y)$ is a depth map – rather than as $I(x,y)$. Similarly the depth map is defined on (X,Y) rather than on (x,y). We do so because it is simpler to define scene planes (tangent planes and surface normals) on (X,Y).

The above model holds only when $\mathbf{N} \cdot \mathbf{L} \ \geq 0$, since it is meaningless to have negative intensities. It can happen that the inner product of **N** and **L** is less than zero, and in this case the surface is facing away from the light source and would not be illuminated by the source. The illuminance from the source would be zero (not negative). We could consider this case in the model by writing $I(X,Y) = max(\mathbf{N}(X,Y) \cdot \mathbf{L}, 0)$.

We refer to the situation above in which $\mathbf{N}(X,Y) \cdot \mathbf{L} < 0$ as an *attached shadow*. In this situation, the surface received no direct illumination from the source. This is distinguished from a *cast shadow* where $\mathbf{N}(X,Y) \cdot \mathbf{L} > 0$ but the light source is not visible because it is occluded by some other object.

Linear shape from shading

One variation of the above shading model occurs when the surface is nearly planar (Z is approximately constant) and has low relief bumps and dents on it, and is illuminated from a direction that is oblique to the surface normal. By "low relief", specifically we mean that the partial derivatives of Z with respect to X and Y are small i.e. $\frac{\partial Z}{\partial X} \approx 0$ and $\frac{\partial Z}{\partial Y} \approx 0$ and so

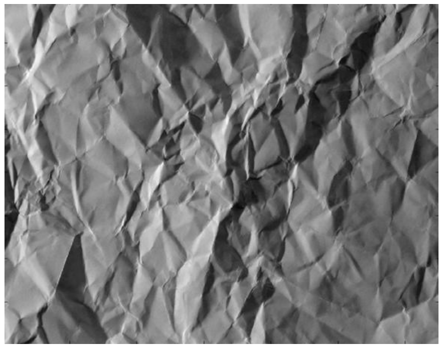
$$\frac{1}{\sqrt{(\frac{\partial Z}{\partial X})^2 + (\frac{\partial Z}{\partial Y})^2 + 1}} \approx 1$$

In this case, we obtain an approximation:

$$I(X,Y) \ \approx \ (\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y}, -1) \cdot (L_X, L_Y, L_Z).$$

An example is shown below of uncrumpled paper illuminated by a light source that is off to the side. The surface has a slant near 0, so that the Z variable in the linear equation above corresponds to the Z axis of the viewer.

The image looks like a real surface, not just some random patterns of grey level intensities. Moreover, you perceive it to be a relatively flat surface. You also may perceive the light direction to come from the left. How sure are you about that? And why do you not think the illumination is coming from the right? We will return to these questions a few lectures from now.



To understand the shading effects better for this model, let's consider a simple example of a surface depth map:

$$Z(X,Y) = Z_0 + a \sin(k_0 X).$$

An example would be hanging drapery (curtains). The frequency is some constant k_0 which is the number of radians per length in X variable. (We could put in a factor of 2π to make the units of k_0 number of cycles per unit distance.)

What is the linear shading model for this example? Computing partial derivatives

$$\frac{\partial Z(X,Y)}{\partial X} = a k_0 \cos(k_0 X), \qquad \frac{\partial Z(X,Y)}{\partial Y} = 0$$

and plugging into the shading model gives:

$$I(X,Y) = a\ k_0\ L_X \cos(k_0 X) - L_Z.$$

Notice that the intensity is 90 degrees out of phase with the depth map, i.e. sine versus cosine. So the maximum and minima of intensity don't occur on top of the depth hills and valleys. Rather, they occur on the sides of the slopes. Also notice that $L_Z < 0$ for this model to make sense, since we need to have positive intensities and the cosine oscillates between positive and negative values. Also notice that $a\ k_0\ L_X$ cannot be too large, otherwise the intensity will become negative when the cosine is negative. This creates an attached shadow effect. For this particular surface, whenever there is an attached shadow there is also a cast shadow, as was observed in class.

Shape from shading on a cloudy day

Another shading model[1] addresses quite a different lighting condition, namely a high relief surface under diffuse lighting such as on a cloudy day. The sunny day model cannot capture this shading because there is not a single light source direction **L**. Rather on a cloudy day there are many light source directions. Indeeed there is a hemisphere (sky) of directions.

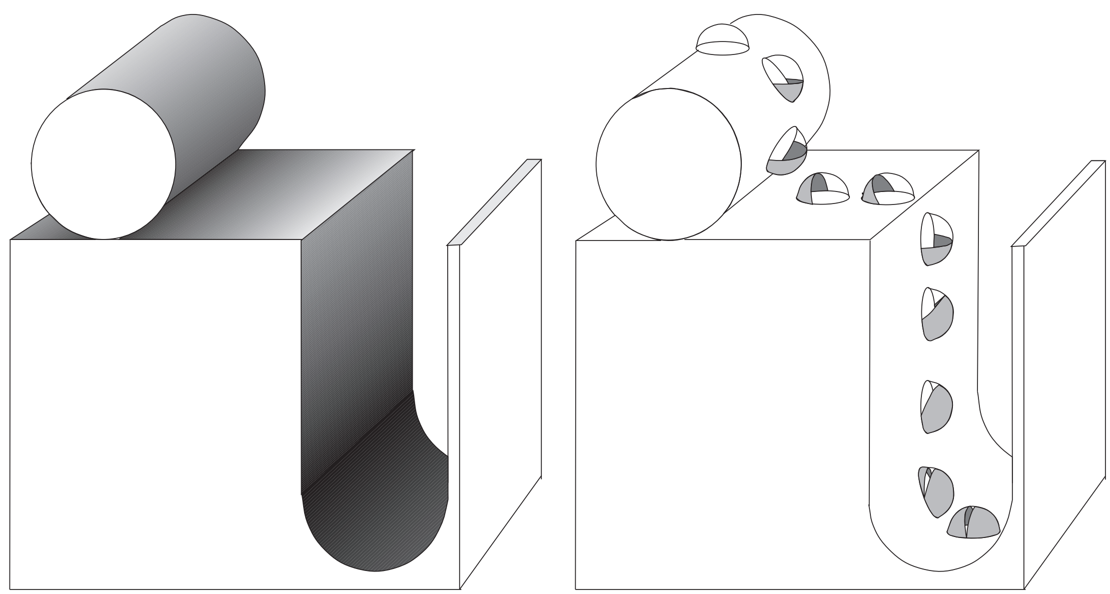
[1] introduced by yours truly in my Ph.D. thesis

Although light comes from all directions on a cloudy day, the surface is not uniformly illuminated. The reason is that not all of the sky is visible from every point on the surface, and this varying sky visibility is a shadowing effect. We can integrate the previous model over directions $\mathcal{V}(\mathcal{X}, \mathcal{Y}, \mathcal{Z})$ in which the sky is visible:

$$I(X,Y) \; = \; \int_{\mathcal{V}(\mathcal{X},\mathcal{Y},\mathcal{Z})} \mathbf{N} \cdot \mathbf{L} \; d\mathbf{L}$$

This implicitly assumes the sky is equal intensity in all directions (which isn't true, but we'll assume it for simplicity).

The above model is much more complicated than the sunny day model because now both the surface normal and the region of the visible sky vary along the surface. The graphic on the right shows the amount of the hemispheric sky that is visible for different points on the surface. At the top of the cylinder, the entire sky is visible and this is the brightest point in the scene.[2] As we go around the cylinder towards the bottom, the amount of sky that is visible decreases. Similarly as we go down into the valley the amount of visible sky decreases. Note that the amount of visible sky can change along the surface because of cast shadow effects.



One can simplify the model by ignoring the $\mathbf{N} \cdot \mathbf{L}$ term and just considering:

$$I(X,Y) \; = \; \int_{\mathcal{V}(\mathcal{X},\mathcal{Y},\mathcal{Z})} d\mathbf{L}$$

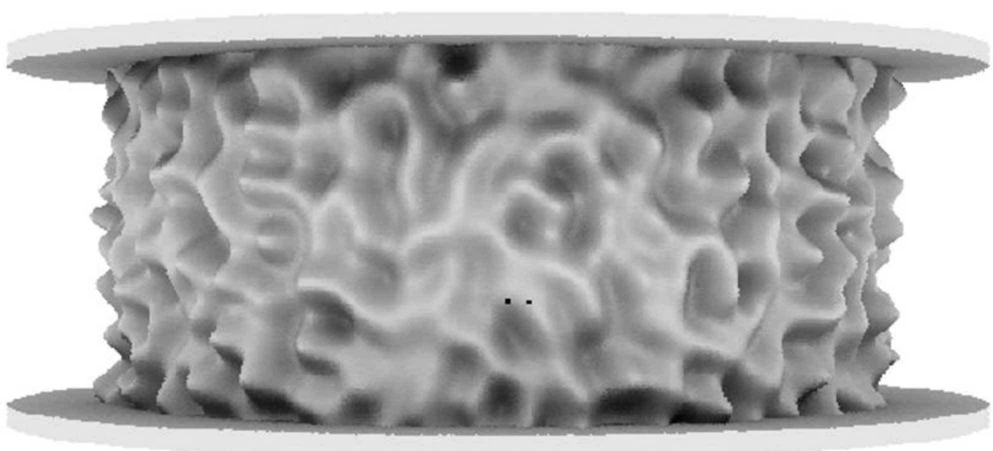
which says that the surface intensity at each point is proportional to the fraction of the sky that is visible. FYI, I was able to come up with a computer vision algorithm for computing a depth map $Z(X,Y)$ that is consistent with this model, given an image $I(X,Y)$.

A few years after my PhD, I carried out perception experiments that studied how people perceive shape from shading under diffuse lighting. These experiments used rendered images such as the one

[2]The shading doesn't indicate this property. Evidently I wasn't careful enough when I made this figure many years ago.

shown below. Notice the little local intensity maxima in the valleys of this rendered surface. These little maxima are due to the surface normal in the valley turning to face directly towards the part of the sky that is vaible (rather than facing a side wall of the valley). For these visible parts of the sky and for points at the bottom of valleys, $\mathbf{N} \cdot \mathbf{L}$ tends to be close to 1. This leads to a local peak in intensities in the bottom of the valleys, which you can easily see in the rendered images (which used fancy computer graphics that took account of sky visiblity and surface normal effects).

In my experiments, I wanted to know if people would be fooled by these local intensity maxima. Their task was to judge the relative depths of pairs of points (such as the little black squares in the image). Sometimes the darker point was deeper, but sometimes the darker point was shallower such as in the case of a point on the side wall versus a point at the bottom of the valley. I found that in many cases people correctly identified that the brighter point was deeper. It was as if people correctly attributed the local intensity maxima to the surface normal effect rather than attributing it to a small hill within a bigger valley.



All the above shading models assume that the surface has a constant reflectance, and that all intensity variations are due to changes in the amount of illumination. But surfaces can have reflectance variations too. The rest of this lecture will examine situations in which the reflectance (and illumination) change.

Lightness versus Brightness

[See the slides to accompany illustrations for the text below.]

I began with an example photograph showing two pieces of white paper laying on a carpet. One paper was in shadow and one paper was not. The paper in shadow naturally receives less illumination from the light source and so its images has lower intensities. Although it is physically darker and appears less bright too (this distinction will be discussed below), it still seems to be a white piece of paper, as if the visual system takes account of the shadowing effect.

A second photograph replaces the intensities on the illuminated paper with intensities that are equal those of the shadowed paper. Remarkably, now the illuminated paper appears to be a darker color paper than the shadowed one. Again, the visual system is taking account of the lighting effect (or discounting the effect of the illumination and shadow). The paper on the right *appears* darker now as if this is the best way for the visual system to explain how the two papers have the same physical intensity. To see this, consider:

$$I(x,y) \;\equiv\; illuminance(x,y) \;\times\; reflectance(x,y)$$

and note that if $I(x_1,y_1) = I(x_2,y_2)$ and if the shadows suggest that

$$illuminance(x_1,y_1) < illuminance(x_2,y_2)$$

then it follows that

$$reflectance(x_1,y_1) > reflectance(x_2,y_2).$$

Is this the correct way of the thinking about what the visual system is doing? Indeed some vision scientists shun these sorts of explainations, and prefer to explain everything in terms of neural coding. But notice that this example is just another version of the simultaneous contrast effect which you saw back in Assignment 1. Perhaps you can explain this effect in terms of neural coding (and lateral inhibition). However, as you saw in Assignment 1 with White's effect, sometimes the simple models also predict the wrong thing.

For today, let's not wring our hands over this issue. Instead let's just try to understand the computational problem that is being solved. The problem is to discount (or at least partially discount) the effects of illumination. The idea is that it isn't as useful for the visual system to estimate the exact magnitude of the intensity at each point in an image. Rather it is more useful to know the reflectance of the surfaces.

As the relation above says, we can think of an image $I(x,y)$ as consisting of the product of two *intrinsic* images: the $illuminance(x,y)$ which captures the shading and shadows, and the $reflectance(x,y)$ which is the fraction of light arriving at the surface that gets reflected. It is straightforward to model the physics of light reflecting off a surface, such as the models above. But how to model the perception of such situations?

First, we need to distinguish between physical and perceptual quantities. The term *luminance* refers to the physical intensity of the light reflected from a surface, whereas the term *brightness* refers to the *perceived* intensity. The two are (hopefully obviously) not the same thing – not just because physical quantities are differnet from perceptual quantities, but also because the light in one image patch might be physically more intense than the light in another patch, and yet the first might be perceived as less intense (less bright).

Second, people also sometimes are capable of judging the reflectance of surfaces. I emphasize that reflectance refers to the fraction of light that gets reflected from a surface, and it is physical quantity. One uses the term *lightness* to describe the *perceived reflectance*. When you look at a surface and judge its colour (grey vs. black vs white, etc), you are making a lightness judgment – not a brightness judgment.

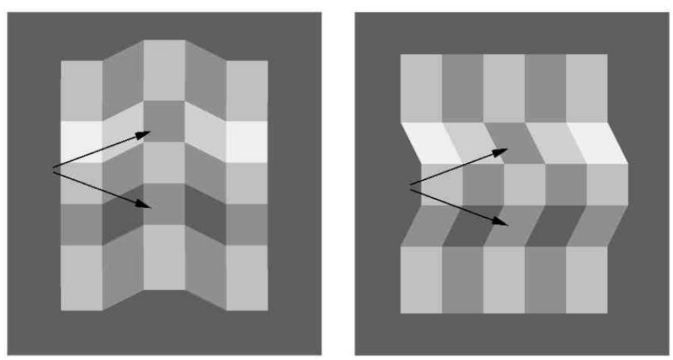
Distinguishing lightness judgments from brightness judgments is difficult. If you run an experiment and you ask people off the street to make different judgments, they typically don't know what you are asking. Even people who are worked in the field sometimes get confused. This is especially a problem when we are looking at pictures, rather than physical objects in a real 3D scene. The following example is susceptible to this problem, but I'll describe it anyhow because it is so nice in other ways.

Adelson's corrugated plain illusion below shows a random checkerboard pattern, which has been folded. The folding is either along vertical lines or horizontal lines, and the two images each naturally contain five groups of five tiles. (Both images are consistent with either a concave or convex folding, but let's not deal with that now.)

Consider the four square tiles that the arrows point to. In fact, they all have the same shade of grey – same physical intensity. In the example on the left, the two tiles that are pointed to appear in the same vertical group of five tiles which all lie in a common plane. The two tiles appear to be the same, whether we are judging brightness or lightness. Indeed it is difficult to say whether our percepts are brightness or lightness since we do not have strong cues about illumination.

In the example on the right, the two tiles that are pointed two now appear to have different shade of grey: the tile on the top appears darker. The most basic explanation for this is that the tile on the top is grouped with the four other tiles in the same row (respecting the 3D interpretation of the folding). The upper tile is the darkest tile in its row. The lower tile belongs to a row that has only two intensities, and the lower tile is in the more intense group.

To explain why the upper tile looks darker (in this example, we don't distinguish brightness from lightness), we suppose that the visual system only compares a tile with others in the same 5-tuple which appear to lie on a common plane and hence have roughly the same illumination (since there is no evidence of illumination changing, like a shadow). If a tile is the brightest in its group, it is perceived as closer to white, whereas if it is the darkest then it is perceived closer to black. That's it, and that idea of comparing within groups takes you a long way – as many other examples show.



Color constancy

When we discussed the basics of color earlier in the course, we used the terms hue and saturation. These properties of color (along with lightness) are often used to recognize objects. In particular, we are very sensitive to the color of skin, when judging the emotional states of others (embarassment, anger, or whether someone has had a bad night). We also judge the ripeness and edibility of food: think meat, bananas, oranges, etc. But our discussion of color earlier in the course did not distinguish between the illumination and the surface reflectance. So let's address this now.

Recall the relationship from lecture 3:

$$I_{LMS} = \int C_{LMS}(\lambda) \; E(\lambda) d\lambda$$

which describes the linear response of a photoreceptor as the sum over all wavelengths of the product of the absorption and the spectrum of the light that arrives at that point on the retina. For a color image, we need to add a pixel position dependence:

$$I_{LMS}(x,y) = \int C_{LMS}(\lambda)\ E(x,y\lambda)d\lambda$$

Notice that the cone absorption C doesn't depend on position. We are assuming that LMS cones have the same properties at all positions.

We also need to say more about spectrum $E(x,y,\lambda)$ which arrives at the retina. In particular we are interested in cases the light is reflected from a surface. Just as in the grey level case of brightness and lightness, we would like to know how a vision system can discount the illuminant color.

The spectrum of light $I(x,y,\lambda)$ arriving at a point in the image depends on the spectrum of the light source (as a function of wavelength) and the percentage of light that is reflected by a surface at each wavelength. Here are few details (not mentioned in class):

- *illuminance*: Each light source has a characteristic spectrum. Sources that emit light because they are hot have a spectrum that depends on their temperature. A fire, tungsten light bulb, the sun all have quite different spectrum. The sun's spectrum is relatively flat over the range of visible light, whereas a tungsten light bulb has much more energy at long wavelengths than short wavelengths. The spectra of sources such as fluorescent light and CRT phosphors are must more spiky as a function of λ than are natural source spectra (which are relatively smooth)

- *surfaces reflectance*: when light reflects off pigmented surfaces – paints and dyes – certain wavelengths are reflected more than others. Foliage is green because it contains a pigment chlorophyll. (In the Fall, because of changes in temperature and other factors, the chlorophyll pigment breaks down. This why leaves change color and lose their green.)

Suppose light is emitted by a source and has a certain amount of energy per wavelength. Call this spectra the $illuminance(\lambda)$. Suppose this source light is then reflected from a surface. For each wavelength, a proportion of the incident light is reflected and this proportion is $reflectance(\lambda)$. For example, objects that appear red reflect long wavelength light ($>$ 600 nm) better than short wavelength light ($<$ 500 nm). The spectrum of reflected light is the wavelength by wavelength product,

$$I(\lambda)\ \equiv\ illuminance(\lambda)\ \times reflectance(\lambda)$$

Since these values can vary along the surfaces and across the image, they depend on image position (x,y), so we write

$$I(x,y,\lambda)\ \equiv\ illuminance(x,y,\lambda)\ \times\ reflectance(x,y,\lambda).$$

This is similar to what we saw above in the black and white domain, but now we have put wavelength into the equation. (Note that the illuminance implicitly includes shading effects too.)

The perceptual problem now is to take the photoreceptor intensities $I_L(x,y), I_M(x,y), I_S(x,y)$ and to infer as much as possible about the reflectance spectra of surfaces in the scene and the illuminant spectrum.

You may think this problem is hopelessly impossible. However, this pessimism ignores the facts of our everyday experience. We are able to judge the colors *of object surfaces* quite well, and discount the illuminant to some extent. This observation holds both informally (our day-to-day experience) and when you go into the lab and do careful experiments with people, asking them to judge color of surfaces as you vary the illumination. *People do make mistakes* (some of them systematic), but the mistakes are surprising small. The ability to see colors such that they appear roughly the same under different illumination is called *color constancy*.

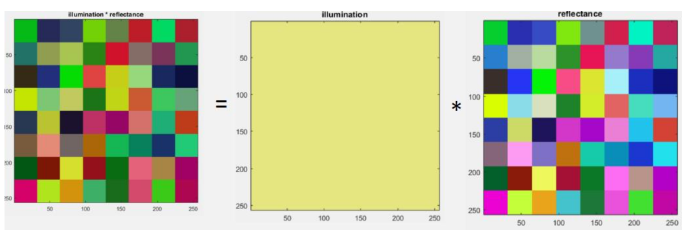
Let's sketch out a few basic ideas for how this can be done. First, if we suppose that the cone response curves don't overlap, then we can think of three ranges of wavelengths. This let's us treat the three channels (LMS or RGB) as independent. (I'll write RGB instead of LMS to be consistent with the slides.) At each point (x, y), we can think of having three intensity values $I_{RGB}(x, y)$ and three illuminance values $illuminance_{RGB}(x, y)$ and three reflectance values $reflectance_{RGB}(x, y)$, so we can write:

$$I_{RGB}(x, y) \equiv illuminance_{RGB}(x, y) \times reflectance_{RGB}(x, y).$$

We are essentially ignoring the details *within* each of the three frequency bands. This is an approximation which let's us cut to heart of the problem, as follows.

Case 1: uniform illuminance (grey world, von Kries)

It often happens that a surface has roughly constant illuminance. As an example, consider the first row below. The checkerboard on the left has colors that are more yellowish than the surface reflectance (on the right) because the illumination is yellow.



The image on the left is literally just the one on the right, multiplied by the one in the center – point by point and channel by channel.

In the real world, the vision system's task is to take the image on the left and to discount the (yellow) illuminant. Obviously we don't do this completely when looking at the little images here; the images on left and right appear different. But this is because we are also comparing these little images to the white page that surrounds them ! In the real world, all the objects that are visible will be colored by the illuminant.

One key idea for discounting the illuminant for the vision system to assume that the surface reflectances in the scene are grey on average.[3]. Then the vision system could take the average RGB value of the scene, and if it not neutral colored (grey) then the vision system could normalize the

[3] This was suggested by a student in the class, and indeed the idea has been tested and holds some water. Its called the *grey world assumption*.

image channel-by-channel by dividing by the average intensity for each color.

$$(\frac{I_R(x,y)}{mean_{x,y} I_R(x,y)}, \frac{I_G(x,y)}{mean_{x,y} I_G(x,y)}, \frac{I_B(x,y)}{mean_{x,y} I_B(x,y)})$$

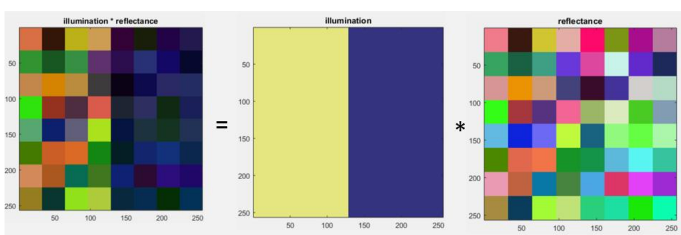
The result is that the new average over the image would be (1,1,1). Of course, you would need to scale those values down if you wanted them to represent reflectances, since the maximum reflectance is 1. For example, you could further divide all channels by a scalar which is the maximum value over all changes, which would ensure that all values are at most 1.

A second approach, which I listed on a slide but ran out of time to discuss is to normalize each image channel by the maximum value that occurs in that channel. That is, compute

$$(\frac{I_R(x,y)}{max_{x,y} I_R(x,y)}, \frac{I_G(x,y)}{max_{x,y} I_G(x,y)}, \frac{I_B(x,y)}{max_{x,y} I_B(x,y)})$$

This is not the same solution as the grey world one above, but the idea is similar: scale down brighter channels to try to discount the illumiant.

Case 2: the shadow revisited



A more challenging problem is the case of a shadow. In natural scenes that are illuminated by sunlight and blue sky, parts of the scene that are not in shadow have yellowish illumination (plus a much weaker blueish illumination from the sky) whereas shadowed regions have just blueish illumination from the sky. This situation is illustrated abstractly in the example here. How might a vision system discount the illuminant in this case?

The take home message from today is that the intensities and colors that we measure with our eyes are the product of a few different factors (literally) and that our vision systems often seem to disentangle these factors, allowing us to perceive the illuminance separately from the surface reflectance. How this is done is only partly understood.

One final side note is worth mentioning: shading and shadowing primarily affect the intensity (and less so the hue and saturation). We often perceive changes in intensity as being due to shading (shape) and shadows (either attached or cast) and we often rely on geometric cues to help us figure out the 3D situation, such as the boundaries of the surfaces in the Adelson example. Once these intensity effects from shading and shadow effects have been accounted for, the visual system can more easily rely on simple normalization processes to discount the color of the illuminant.

