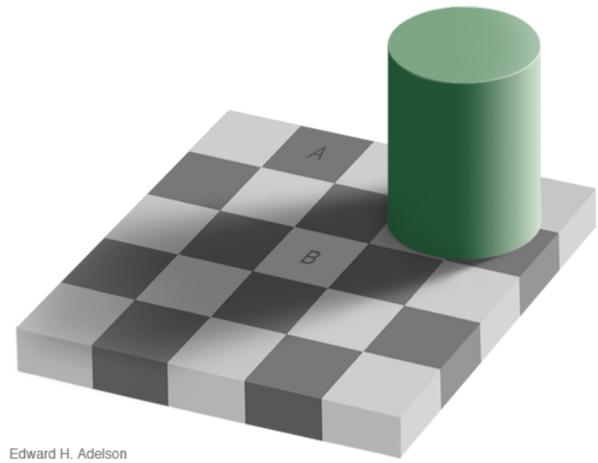
In the next few lectures, we will turn away from discussions of the brain and neural circuits. Instead we will ask about questions at a higher level where we are concerned more with what is the problem being solved.

Surface reflectance versus illumination

We'll start off by considering a familiar situation of a scene in which we have a light source and some surfaces that is being illuminated. The illumination can vary over the surfaces. For example, the some surface points might lie in shadow whereas other parts might be directly illuminated. One nice example is due to Adelson and is called the "checkerboard-shadow" illusion. There is an object (a green cylinder) that sits on a checkerboard and casts a soft shadow on the checkerboard. The illumination comes both from a direct light course (causing a shadow) and there is also ambient illumination that reaches all the surfaces. The illumination is chosen such that a white check in the shadowed region has the same image intensity as a black check in direct illumination.



Think of it this way. The image intensity at a point (x, y) depends on the illumination reaching the surface at that point, multiplied by the reflectance of the surface. That is, the surface only reflects some fraction of the illumination. So reflectance at (x, y) is a number between 0 and 1. The idea of the checkerboard illusion is to choose the illumination of the shadow and non-shadow such that the product is the same for the two points mentioned above (checks A and B in the image):

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

So if $I(x_A, y_A) = I(x_B, y_B)$ and if the shadows suggest that

$$illuminance(x_A, y_A) > illuminance(x_B, y_B)$$

then it follows that

$$reflectance(x_A, y_A) < reflectance(x_B, y_B).$$

I then showed an example photograph illustrating a similar 3D geometry. It showed 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 it has lower intensity. In a second photograph, I replaced the intensities on the illuminated paper with intensities that are equal those of the shadowed paper. Now the illuminated paper appears to be a darker color paper than the shadowed one. This is the same idea as the Adelson illusion. What's basically happening here is that your visual system is taking account of the difference in illumination. The paper on the right *appears* darker now. It is as if the visual system is representing the material color (reflectance), rather than the intensity.

Note that you only *partly* discount the illuminant, in that you are aware of the shadow. The shadowed region looks darker than the unshadowed region (otherwise I would not be able to talk about a shadowed region).

Also note that this example is reminiscent to the "simultaneous contrast" effect that you worked with in Assignment 1, which I'll repeat here on the right.

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.

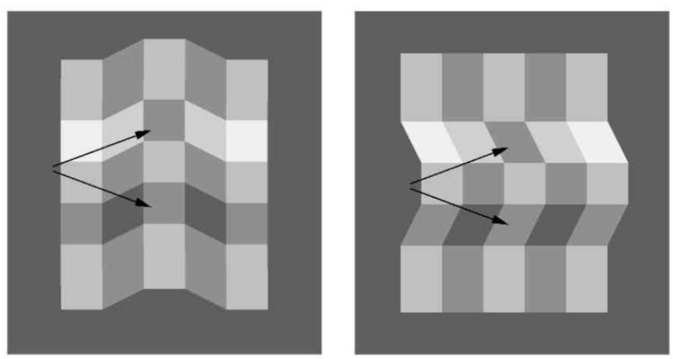
As the relation above says, we can think of an image $I(x, y)$ as consisting of the product of two *intrinsic* images: the $illuminance(x, y)$ which captures the shading and shadows, and the $reflectance(x, y)$ which is the fraction of light arriving at the surface that gets reflected. It is straightforward 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 5x5 random checkerboard pattern on a folded surface. The folding is either along vertical lines or horizontal lines. (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. 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 to 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 has the higher of the two intensities. To explain why the upper tile looks darker, 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. If a tile is the brightest in its 5-group, it is perceived as closer to white, whereas if it is the darkest then it is perceived closer to black. That's it! This idea of comparing within groups takes you a long way – as many other examples show.

In the slides, I then briefly discussed T and X junctions. T-junctions often support occlusion relations, namely one surface in front of another. X-junctions can give rise to transparency effects (or not, depending on the luminance of the image regions in the neighborhood of the junctions). I only sketched out some basic ideas here – enough to give you a flavour.

Color constancy

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.

The spectrum of light $E(x, y, \lambda)$ 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 seen at (x, y). We briefly discussed such spectra in lecture 3 (see slides 10 and 11). 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) more than short wavelength light (< 500 nm), whereas objects that appear blue typically reflect more short wavelength light than long wavelength light. 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.

The perception problem now is similar to what we discussed earlier today. Given the photoreceptor intensities $I_{LMS}L(x, y)$, try to infer the reflectance spectra of the surfaces. This problem is very difficult, since the measured signal $I_{LMS}L(x, y)$ depends on the illuminant, the surface reflectance, and the cone absorptances for L, M, S. The best a vision system can do is to have some approximation. Indeed this is similar to what we saw in the grey level case above. We can't estimate the reflectance exactly, but we often do make estimates that differ from the intensities we are given. The same is true in color.

The ability to judge surface color independent of illumination is called *color constancy*. This ability requires that we discount the illuminant to some extent. There are many examples of why we need to be able to do this. One example is that it helps us in judging the emotional state of other people (whether they are angry or embarrassed), or whether they are sick or healthy or tired. It is also important for judging the food that we eat e.g. judging the ripeness of fruit. Perhaps the biggest reason we need color constancy is that it helps us to identify objects (by their color).

People do make mistakes in judging object color (some of them systematic), but the mistakes are surprising small, given the challenges of the problem.

Let's sketch out a few basic ideas for how color constancy is achieved. First, for conceptual simplicity, suppose that the cone response curves don't overlap.[1] So we can think of three ranges of wavelengths. This lets us treat the three channels as independent. (I'll write RGB from now on.) That is what you did in Assignment 1 when you just deal with the three channels and not with the spectrum variable λ.

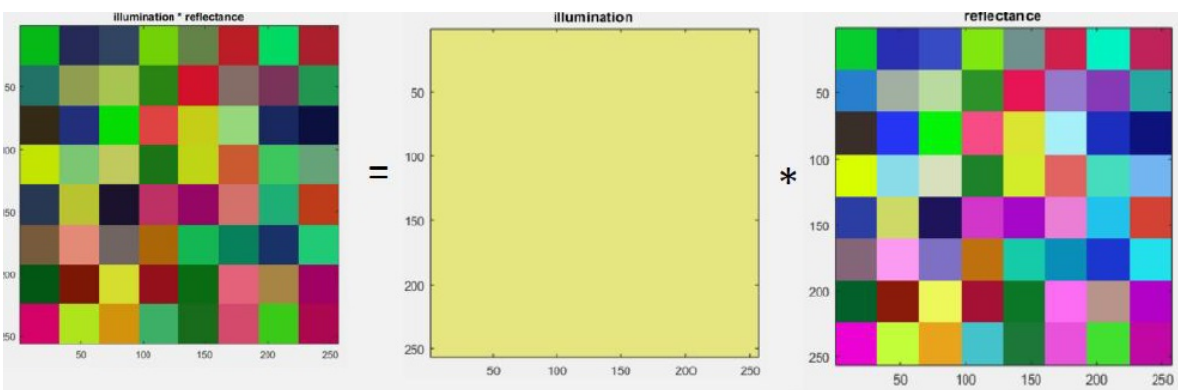
At each point (x, y), we can think of having three intensity values $I_{RGB}(x, y)$ which for simplicity are three illuminance values $illuminance_{RGB}(x, y)$ and three reflectance values $reflectance_{RGB}(x, y)$, which are between 0 and 1. So we can write:

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

So 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

Consider the three images below which are an example of the equation we just saw. The illuminance is constant (yellow) e.g. RGB = (1, 1, .5). On the right is a set of squares with random RGB values which represent the reflectances – take them to be values in 0 to 1. The image on the left is the product of the yellow illuminant and the reflectance. Not surprisingly, the squares on the left all appear more yellowish than the ones on the right. 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, everything in the scene (typically the whole field of view) will be colored by the illuminant.

[1]Computational theories of color constancy don't depend on this, but the details don't concern us.

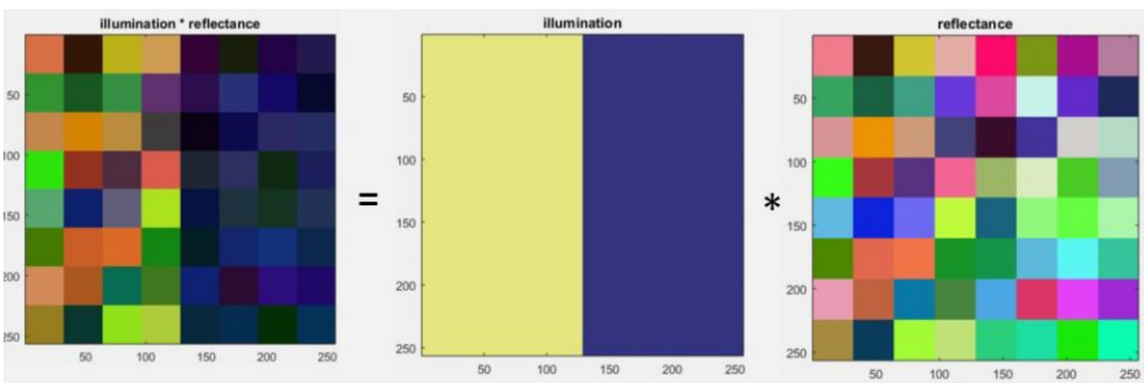
One idea is to normalize each image channel by the maximum value that occurs in that channel. That is, estimate the reflectance to be:

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

This works great if one of the original surfaces was white, i.e. had reflectance (1,1,1). The reason is that the image of this surface would just have the illuminant's RGB value and these RGB values would be the maximum in each of the RGB channels. So normalizing would just cancel the illumination, which is what the visual system would like to do! (Note, however, that if there is no original square that was white, then this method would not produce the correct answer for any of the squares.)

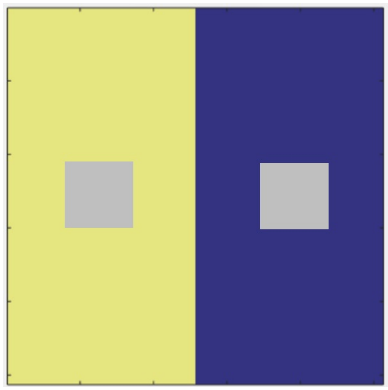
Case 2: the shadow revisited

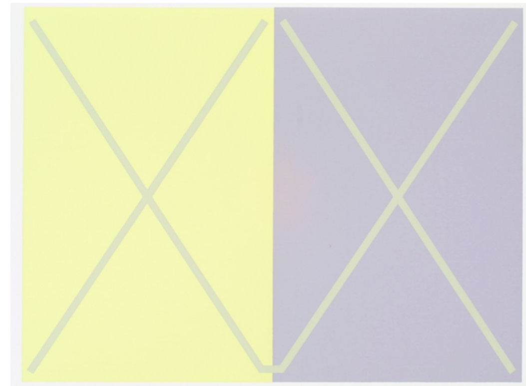
What happens if there is a shadow in the scene ? In natural scenes that are illuminated by sunlight and blue sky, parts of the scene that are not in shadow have yellowish illumination (plus a much weaker blueish illumination from the sky) whereas shadowed regions have just blueish illumination from the sky. This situation is illustrated abstractly in the example below.



How might a vision system discount the illuminant in this case? The idea is similar to what we saw earlier in the lecture. The visual system could treat the left and right halves of the image as two different groups of squares. Of course, when you look at the image on the left, the left half looks yellowish and the right half is dark blueish. But note that *your visual system* is seeing this square image in the context of page of lecture notes where there is a white background. As I mentioned above, in a real scene viewing condition, the whole scene has an illuminant and possibly shadowed regions, and so your visual system treats that situation as very different from the example here.

Let's now go back full circle to the beginning of the lecture where we considered a simple simulteneous contrast display with grey levels only. We can consider a color version of this display too. The two grey squares on the left are identical but, like in the earlier example, the one of the right looks brighter since it is surrounded by a darker surround. There are now color effects too – albeit small ones. When I look at the image below, the grey square on the left looks slightly blueish and the grey square on the right looks slightly yellowish – very very slightly! (Most students in





the class did not see the effect). For the example on the right (due to Joseph Albers), the effect is stronger.

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). Our vision systems often seem to disentangle these factors, allowing us to perceive the surface reflectance somewhat independent of the illumination. How this is done is only partly understood, but the main idea is that we form regions (groups) within the image based on different grouping rules, and we compare surfaces within these regions. For some of the examples we discussed today, the regions were defined by shadow versus non-shadow. But in general, there are other ways of defining groups which have been discovered (details omitted).

[ASIDE: at the end of the lecture, I discussed some other examples.]

