Chapter 6

Lecture 6: Color and color constancy

Tuesday, Feb. 22, 2010 MIT EECS course 6.869, Bill Freeman and Antonio Torralba
Slide numbers refer to the file "06color2011"

slides 2, 3

Why does a visual system need to use color? Suppose you were building a vision system for a robot, or trying to understand the visual system of humans–what benefit does processing color give?

Color offers a higher dimensional description of surface properties than simply the scalar attenuation of an overall light intensity. This richer information about surface reflectivity allows us to infer material properties from local measurements. The benefit to an organism is to infer properties of food (is the fruit ripe?) or other people (is my child healthy?). We use color to help us group regions of similar material, to find skin.

slides 4-7

In this lecture, we first cover color physics, then color perception.

Electromagnetic waves surround us, at energies (and corresponding wavelengths) ranging from x-rays to microwave radiation to ultraviolet, visible, infrared, and thermal radiation.

In experiments summarized by his drawing, shown here, Isaac Newton revealed many intrinsic properties of light. A pinhole of sunlight comes in through the window shade, and a lens focuses the light onto a prism. The prism then divides the white light up into many different colors. These colors seem to be elemental: if you take one of the component colors and pass it through a second prism, it doesn't split into further components; it just bends.

Such experiments led to our understanding of light and color. Visible light is electromagnetic radiation of wavelengths between roughly 400 and 700 nm. Sunlight arriving through the atmosphere has a broad distribution of light of those wavelengths. At an air/glass interface, light bends in a wavelength-dependent manner, so a prism disperses the different wavelength components of sunlight into different angles, which we see as different colors.

slides 8-10 Light changes when it bounces off surfaces in the world and this is the reason vision works at all. We can summarize the changes to light with a "bi-directional reflectance distribution function", or BRDF. The BRDF gives the fractional change in spectral power of the reflected light as a function of the angles of incidence and viewing relative to the surface, and as a function of the wavelength of light.

To simplify our study of surface appearance, we will assume surfaces reflect diffusely and that the observed power spectrum of the reflected light is simply a wavelength-by-wavelength product of the illumination power spectrum and the surface reflectance spectrum. If we assume diffuse reflection and only concern ourselves with relative power spectra, rather than absolute intensities, then we can ignore incident and outgoing light directions in this way.

This wavelength-by-wavelength multiplication is also a good model for spectral changes to light caused by viewing an illuminant through an attenuating filter. The illumination power spectrum is multiplied at each wavelength by the transmittance spectrum of the attenuating filter.

slides 11, 12 Here are some example light spectra. The spectrum of blue sky is on the left, and the spectrum of a tungsten light bulb (which will look orangish) is on the right. Some reflectance spectra are on the right. A white flower reflects spectral power almost equally over all visible wavelengths. A yellow flower reflects in the green and red.

slide 13 It's helpful to develop the skill of being able to look at a light power spectrum and to know roughly what color that spectrum would correspond to. Here is a rough description of what wavelengths correspond to what perceived colors, with a reference spectrum showing roughly what each individual wavelength, viewed by itself, looks like. (I think of it as a cartoon color model, a hard-edged approximation to a very mushy reality). The visible spectrum lies roughly in the range between 400 and 700 nm. We can divide into three one-hundred nm bands, which, from short to long wavelengths, corresponds to blue, green, and red (again, speaking in broad strokes). These are often called the additive primary colors, which we'll write more about.

White light is a mixture of all spectral colors. There are three more possible combinations of the three one-hundred nm bands of wavelengths, and each can be associated with a color name: Cyan is a mixture of blue and green, or roughly spectral power between 400 and 600 nm. In printing applications, this is sometimes called "minus red", since it is the full spectrum, with the red band subtracted. Blue and red, or light in the 400-500nm band, and in the 600-700nm band, is called magenta, or minus green. Red and green, with spectral power from 500-700 nm, make yellow, or minus blue.

slides 14,15, 16 Given individual colors, we often want to mix them together. There are two main ways that spectra combine when we mix colors together. While the precise way two spectra combine may depend on the details of the corresponding physical process, these two methods are a good model.

The first way is called additive color mixing. This is the way spectra combine when you project two lights simultaneously, so they are summed in our eye. CRT color televisions, DLP projectors, and colors viewed very closely in space or time all exhibit additive color mixing. The spectrum of the mixed color is a weighted sum of the spectra of the individual components. For example, in the additive color mixing model, red and green combine to give yellow.

The second way colors combine is called subtractive color mixing, but might make more sense to be called multipliciative color mixing. Under this mixing model, the spectrum of the combined color is proportional to the product of the subtractively mixed components. This color mixing occurs when light passes through a sequence of attenuating spectral filters, such as with photographic film, paint, optical filters, and crayons. Under the subtractive color mixing model, cyan and yellow combine to give green, since the cyan filter attenuates the red components of white light, and a subsequent yellow filter would remove the remaining blue components, leaving only the green spectral region of the original white light.

The overhead projector demonstration of these effects brings our cartoon color models back to the real world, where they mostly hold, but not perfectly.

slides 17-21 Before we turn to color perception, let's introduce a mathematical model for light spectra that makes them much easier to work with. In general, when modeling the world, we want to keep everything as simple as possible, and that usually means working with as few degrees of freedom as possible. Color spectra seem like relatively high-dimensional objects, since we can pick any combination of numbers, from 400 to 700 nm, as we'd like. Even sampling at every 10 nm of wavelength, that gives us 31 numbers for each spectrum.

It is common to use low-dimensional linear models to approximate real-world reflectance and illumination spectra. Any given spectrum, say $a(\lambda)$, is approximated as some linear combination of "basis spectra". For example, a 3-dimensional linear model of $a(\lambda)$ would be

$$\begin{pmatrix} \vdots \\ a(\lambda) \\ \vdots \end{pmatrix} \approx \begin{pmatrix} \vdots & \vdots & \vdots \\ a_1(\lambda) & a_2(\lambda) & a_3(\lambda) \\ \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \omega_1 \\ \omega_2 \\ \omega_3 \end{pmatrix} \tag{6.1}$$

The basis spectra can be found from a collection of training spectra. If we write the training spectra as columns of a matrix, D, then performing a singular value decomposition on D yields

$$D = U * S * V' \tag{6.2}$$

where U is a set of orthonormal spectral basis vectors, S is a diagonal matrix of singular values, and V' is a set of coefficients. The first n columns of U are the basis spectra that can best approximate the spectra in the training set, in a least squares sense.

Here's a simple demonstration that, at least for a subset of all surface reflectance spectra, this works quite well. The "Macbeth Color Checker", a tool of color scientists and engineers, is a standard set of 24 color tiles, made the same way year after year. (So iconic that this woman, a dedicated color scientist, I presume, as tatooed a Macbeth color checker on her arm! Alas, I'm sure the tatoo colors are only an approximation to the real Macbeth colors).

The reflectance spectra of each Macbeth color chip has been measured. All those spectra are pretty well approximated by a 3-dimensional linear model, as you can see from these plots.

slide 22-24 That's as far as we'll describe the physical properties of colors. Now, we turn to the human perception of color. This section, we divide into two sub-sections, according to whether we do or don't make a particular assumption, namely: that the perceived color of a light depends only on the spectral power distribution of that light entering the eye. First, we'll describe the fine and powerful color measurement and management systems that are based on that assumption being true. Then, we'll show how that assumption really isn't true, and discuss those ramifications.

slides 25-28 Why do we need to quantify color perception, anyway? Why can't we just see the colors and let that be it? It turns out to have all sorts of useful applications to be able to assign coordinates to a color percept, and to be able to adjust machinery to create colors that match some desired input.

We have color standards for foods, for example. Shown is a USDA kit for evaluating color of foods, and a chart showing french fry color standards. Companies can trademark colors, so we need to be able to specify what is being trademarked.

slides 29-30 So what is the machinery of human vision? Here is a drawing of the rod and cones of the eye, and some of the nerve cells connecting them. The tall ones are the rods and the short ones are the cones, both at the top layer. By the way, where does the light come in, in this drawing? Differently than how you or I might design things, the light comes in at the bottom, passes through the nerve fibers and blood vessels, then reaches the photosensitive detectors at the top of the image. I think there may be benefits to having the photodetectors on the inside, where they can more easily receive nourishment from blood vessels.

slide 31 Slicing the retina the other way gives these two views. On the left is how the cones appear at the fovea, in the center of the field of view. The image on the right shows their relative size in the periphery. In between them, on the right, are the black-and-white, and movement sensitive photodetectors, the rods.

slide 32 The cones come in 3 different types, denoted L, M, and S, for whether they are sensitive to the long, middle, or short wavelengths of the visible spectrum. The spectral sensitivity curves are shown here.

In some sense, this figure shows the whole story of spectrum-based color perception. Three detectors, with the spectral sensitivities shown here, signal their response to an incoming light. The fact that there are three different detectors means that we'll 3 numbers to describe a perceived color in the world (or 2 numbers, if we normalize for intensity). The shapes of those curves tell us which real-world spectra will look the same to us (because they'll give the same trio of photoreceptor responses) and thus will tell us how make one color look like another one.

But that's jumping ahead. To help understand how color is measured, and the experiments that taught us what we know about color vision, let's examine the psychophysical experiments that were done to learn what we know about color perception.

slide 33 Color measurement is mostly about color matching. Because we know the eye has three classes of photoreceptors, and through experimentation, we know that we'll need three different colored light sources in order to try to match a given test light. We shine a controllable combination of the primary lights on one half of a bipartite white screen, and the test light on the other half. I mentioned that in this part of the class, we're assuming that the color appearance is entirely determined by the spectrum of the light. To ensure that this is the case, care must be taken to view the color comparisons under repeatable, controlled conditions. A grey surround field is placed around the viewing aperture, giving a view to the subject that looks something like that of (b).

Here will be a way we can assign a number to a color: Pick a set of primaries, and see what combination of the primaries is required to match any given color.

slides 34-41 Let's see that in action. Here's a mock-up of the color matching experiment. We shine the test light on the left, and originally all the primary lights are turned off, so it's black.

Now we light up some combination of the primaries. No, it doesn't match. We keep adjusting their amounts until we get a color match. So (**slide 37**) , we now have a (reproducible) representation for the color at the left: if you take these amounts of each of the selected primaries, you'll match the input color.

What if our three selected primiaries don't let us reach the test color? Here's an example of that. But **slide 41** it turns out we can always match any input test color if we "add negative light", which means to add positive light to the other side of the test comparison.

slide 42 Human color matching has elegant properties that help us describe colors using linear algebra. Most every desirable linear property is satisfied with such color matching experiments. Here's one of them: if color A_1 matches color B_1, and color A_2 matches color B_2, then the sum of colors A_1 and A_2 will match the sum of colors B_1 and B_2.

slide 43 That tells us that if we represent a color by the amount of the 3 primaries needed to make a match, or any number proportional to that, then we'll be able to use a nice vector space representation for color, where the observed linear combination laws will be obeyed.

That's the psychophysics. We also have in the back of our heads the mechanistic view for how colors generate signals in our brain: the light power spectrum gets projected onto the 3 photoreceptor classes spectral sensitivity curves, generating three numbers, the L, M, and S cone responses, which are the signal for that color (bottom row of slide). If we can adjust the primary color amounts, a_1, a_2, and a_3, so that their sum generates the same set of photoreceptor signals when projected onto the photoreceptor spectral response curves (far right in figure), we have matched the color.

slide 44

This very naturally leads to a linear algebraic (and geometric) interpretation. of color perception. To be concrete, let the space of all possible spectral signals be a 31 dimensional. We'll allow one dimension to describe the spectral power, sampled at every 10 nanometers. In this figure, we depict that as a 3-dimensional space. The generation of cone response for a given spectral sensitivity curve can be thought of as projecting a 31-dimensional signal onto a basis function and recording the resulting projection length. If we record the response of the three different spectral sensitivity curves, we are measuring the projection of our 31-d vector onto each of three linearly independent basis vectors. Thus, a triplet of cone responses maps onto some coordinate in a 3-dimensional subspace (depicted as a 2-d plane here) of the original 31-dimensional space.

Viewed that way, then the task of color measurement is simply the task of finding the projection of any of the possible 31-d spectra into the special 3-d subspace defined by the cone spectral response curves. Any basis for that 3-d subspace will serve that task. Equivalently, we seek to predict the cone responses to any spectral signal, and projection of the spectral signal onto any 3 independent linear combinations of the cone response curves will let us do that.

So we can define a color system by simply specifying its 3-d subspace basis vectors. And we can translate between any two such color representations by simply applying a general 3x3 matrix transformation to change basis vectors. Note, the basis vectors do not need to be orthogonal, and most color system basis vectors are not.

slide 46 Long before scientists had measured the L, M, and S spectral sensitivity curves of the human eye, others had measured equivalent bases through psychophysical experiments. It is interesting to observe how such curves could be measured psychophysically.

We start with a set of *any* three linearly independent primary lights, ie, none of the three spectra can be written as a linear combination of the other two. The idea is this: if we find spectral curves which, when taking the projection of an input spectrum, give us the controls for each primary to match the input color, then we have found a basis for the 3-dimensional cone response space. This is because 3-d projection vectors that always lead to matched colors must be basis vectors for that same 3-d space.

Here's what we can do to find such basis vectors, called "color matching functions", for any give set of primary lights. We exploit the linearity of color matching and find the primary light values contributing to a color match, one wavelength at a time. So for every pure spectral color as a test light, we measure the combination of these three primaries required to color match light of that wavelength. For some wavelengths and choices of primaries, the matching will involve negative light values, and remember that just means those primary lights must be added to the test light to achieve a color match.

slide 46 is an example of such a measured color matching function, for a particular choice of primaries, monochromatic laser lights of wavelengths 645.2, 525.3, and 444.4 nm. We can see these matches are behaving as we would expect: when the spectral test light wavelength reaches that of one of the primary lights, then the color matching function is 1 for that primary light, and 0 for the two others.

Because of the linearity properties of color matching, it's easy to derive how to control the primary lights in order to match *any* input spectral distribution, $t(\lambda)$. Let the three measured color matching functions be $c_i(\lambda)$, for $i = 1, 2, 3$. Let the matric bfC be the color matching functions arranged in rows,

$$C = \begin{pmatrix} c_1(\lambda_1) & \dots & c_1(\lambda_N) \\ c_2(\lambda_1) & \dots & c_2(\lambda_N) \\ c_3(\lambda_1) & \dots & c_3(\lambda_N) \end{pmatrix} \tag{6.3}$$

Then, by linearity, the primary controls to yield a color match for any input spectrum $\vec{t} = \begin{pmatrix} t(\lambda_1) \\ \vdots \\ t(\lambda_N) \end{pmatrix}$ will be $C\vec{t}$.

slide 49 We can check this line of reasoning. If the color matching functions are indeed basis vectors for the same 3-d space that the cone response functions project signals into, then we should be able to write one set of basis functions as a linear combination of the other. This figure shows that you can indeed to that: a best-fitting 3x3 matrix linear combination of the cone response functions (data points on the figure) are shown overlaid on a set of color matching functions.

slides 50, 51 So there is an infinite space of color matching basis functions to pick, so it's natural to ask whether any one choice of bases is better than another. One natural choice might be the cone spectral responses themselves, but those were only measured relatively recently, and many other systems were tried, and standardized on, earlier.

One standard you should know about, because it's so common, is the CIE XYZ color space. Again, a color space is simply a table of 3 color matching functions, which must be a linear combination of all the other color matching functions, because they all span the same 3-d subspace of all possible spectra. The CIE color matching functions were designed to be all-positive at every wavelength. They're shown in **slide 51**. What might be the benefit of having an all-positive set of color matching functions? I believe they were selected so that it would be simple to build a machine that used color filters of those spectral responses to directly measure the color coordinates of a signal. Why not use the cone responses themselves, another all-positive set of color matching functions? I don't know; maybe those weren't known then.

A bug with the CIE color matching functions is that there is no all positive set of color primaries associated with those color matching functions. But if the goal is to simply specify a color from an input spectrum, then any basis can work, regardless of whether there is a physically realizable set of primaries associated with the color matching functions.

To find the CIE color coordinates, one projects the input spectrum onto the 3 color matching functions, to find coordinates, called tristimulus values, labeled X, Y, and Z. Often, these values are normalized to remove overall intensity variations, and one calculates $x = \frac{X}{X+Y+Z}$ and $y = \frac{Y}{X+Y+Z}$.

slides 52, 53

One final topic for the model where power spectral density determines color is metamerism, when two different spectra necessarily look the same to our eye. There is a huge space of metamers: any two vectors describing light power spectra which give the same projection onto a set of color matching functions will look the same to our eyes.

There's a sense that our eyes are missing much of the possible visual action. There's a 31-dimensional space of colors out there, and we're only viewing projections onto a 3-d subspace of that.

But in practise, the projections we observe do a pretty good job of capturing much of the interesting action in images. Given how much information is not captured by our eyes, hyperspectral images (recorded at many different wavelengths of analysis) add some, but not a lot, to the pictures formed by our eyes.

slides 54

Let us summarize our discussion of color so far. Under certain viewing conditions, the perceived color depends just on the spectral composition of light arriving at the eye (we move to more general viewing conditions next). Under such conditions, there is a simple way to describe the perceived color: project its power spectrum onto a set of 3 vectors called color matching functions. These projections are the color coordinates. We standardize on particular sets of color coordinates. One such set is the CIE XYZ system.

By the way, how would you translate from one set of color coordinates to another, say, for notation, from the color coordinates in a unprimed system to those in a primed system? Place the spectra of a set of primary lights into the columns of a matrix $\mathbf{P}$. If we take the color coordinates, $\vec{x}$, as a 3x1 column vector and multiply them by the matrix $\mathbf{P}$, we get a spectrum which is metameric with the input spectrum whose color coordinates were $\vec{x}$. So to convert $\vec{x}$ to its representation in a primed coordinate system, we just have to multiply this metameric spectrum by the color matching functions for the primed color system:

$$\vec{x}' = \mathbf{CP}'\vec{x} \tag{6.4}$$

The color translation matrix $\mathbf{CP}'$ is a 3x3 matrix.

slides 55-58 Now, on to the third part of this lecture, where we address the assumption that a given spectral power distribution always leads to the same color percept.

In the color constancy demo, we'll show an example where the identical spectral distribution arriving at your eye leads to a very different color percept. What's going on? The visual system needs to perceive the color of surfaces, but the data it gets is the wavelength-by-wavelength product of the surface color and the illuminant color. So our visual system needs to "discount the illuminant" and present a percept of the underlying colors of the surfaces being viewed, rather than simply summarizing the spectrum arriving at the eye.

The ability to perceive or estimate the surface colors of the objects being viewed, and to not be fooled by the illumination color, is called "color constancy"–you perceive a constant color, regardless of the illumination. People have some degree of color constancy, although not perfect color constancy.

For the case where there is just one illumination color in the image, if we know either the illuminant or all the surface colors, we can estimate the other from the data. So, from a computational point of view, you can also think of the color constancy task as that of estimating the illuminant spectrum from an image.

This week's problem set is to take part in an illumination spectrum estimation contest sponsored by the Optical Society of America (OSA). There is a $1000 first prize, by the way. This is the contest: there are 10 images (shown here in a screen-grab from the contest web page) for which we're given the L, M, and S cone responses (simulated). From those responses, the challenge is to estimate the illumination spectrum that was used to create each image. The spectrum is constant over any image, and different for each one.

The OSA contest is open until July, although they'll show the current rankings of all submissions between now and then, updated about every two weeks. To give the class the closure and excitement of a result for competition before the summer, we'll offer a token prize, an Ipod Shuffle, to the class member with the best score on the first contest list posted after the problem sets are all in.

slide 59

Now that we're motivated, let's examine the computation required to achieve color constancy. Here's the rendering equation, showing, in our model, how the L, M, and S cone responses for the jth patch are generated. The equation itself is shown, as well as a graphical diagram showing the vector and matrix sizes that I hope makes things a little clearer. We have some unknown illuminant, described by, say, a 3-dimensional vector of coefficients for the illumination spectrum basis functions. For this jth color patch, we have a set of surface reflectance spectrum basis coefficients, let's say also 3-dimensional. The term-by-term product of the resulting spectra (the quantity in parenthesis in the top equation) is our model of the spectrum of the light reaching our eye. That spectrum then gets projected onto spectral responsivity curves of each of the three cone classes in the eye, resulting in the L, M, and S response for this jth color patch. (An equation for the RGB pixel color values would be the same, with just a different matrix $\mathbf{E}$). If we make N distinct color measurements of the image, then we'll have N different versions of this equation, with a different vector $\vec{x}^s_j$ and different observations $\begin{pmatrix} L_j \\ M_j \\ S_j \end{pmatrix}$ for each equation.

Like various other problems in vision, this is a bilinear problem. If we knew one of the two sets of variables, we could find the other trivially by solving a linear equation (using either a least squares or an exact solution). It's a very natural generalization of the a b = 1 problem that Antonio talked about last week.

Let's notice the degrees of freedom. We get 3 numbers for every new color patch we look at, but we also add 3 unknowns we have to estimate (the spectrum coefficients $\vec{x}^s_j$), as well as the additional three unknowns for the whole image, the illumination spectrum coefficients $\vec{x}^i$. If only surface color spectra had only two degrees of freedom, we'd catch up and potentially have an over-determined problem if we just looked at enough colors in the scene. Unfortunately, 2-dimensional surface reflectance models just don't work well in practice.

slide 60

So how will we solve this? Let's look at two well-known simple algorithms, and then we'll look at a Bayesian approach.

Bright equals white If we knew the true color of even a single color patch, we'd have the information we needed to estimate the 3-d illumination spectrum. One simple algorithm for estimating or balancing the illuminant is to assume that the color of the brightest patch of an image is white. (If you're working with a photograph, you'll always have to worry about clipped intensity values, in addition to all the non-linearities of the camera's processing chain). If that is the kth patch, and $\vec{x}^W$ are the known spectral basis coefficients for white, then we have

$$\vec{y}_k = \begin{pmatrix} L_k \\ M_k \\ S_k \end{pmatrix} = \mathbf{E}^T(\mathbf{A}\vec{x}^W \;.*\; \mathbf{B}\vec{x}^i) \tag{6.5}$$

which we can solve for the unknown illuminant, $\vec{x}^i$.

slide 61

How well does it work? It works sometimes, but not always. On the left is a picture for which the bright equals white algorithm would probably work (although I haven't checked it. On the right is one where I don't think it would work.

slide 62 The bright equals white algorithm estimates the illuminant based on the color of a single patch, and we might expect to get a more robust illuminant estimate if we use many color patches in the estimate. A second heuristic that's often used is called the **grey world assumption**: the average value of every color in the image is assumed to be grey.

Taking the sum over all samples j on both sides of the rendering equation, and letting $\vec{x}^G$ be the spectral basis coefficients for grey, gives

$$\frac{1}{N}\sum_j \begin{pmatrix} L_j \\ M_j \\ S_j \end{pmatrix} = \mathbf{E}^T(\mathbf{A}\frac{1}{N}\sum_j \vec{x}_j^s \; .* \; \mathbf{B}\vec{x}^i) \tag{6.6}$$

$$= \mathbf{E}^T(\mathbf{A}\vec{x}^G \; .* \; \mathbf{B}\vec{x}^i) \tag{6.7}$$

$$\tag{6.8}$$

Then, again, we just have a linear equation to solve for $\vec{x}^i$.

slide 63 This assumption can work quite well, although, of course, we can find images for which it would completely mess up, such as this forest scene here.

slide 64

Using just part of the data (the brightest color, or the average color) gives sub-optimal results. Why not use all the data, make a richer set of assumptions about the illuminants and surfaces in the world, and treat this as a **Bayesian estimation problem**? That's what we'll do now, and what you'll continue in your homework assignment.

To remind you, in a Bayesian approach, we seek to find the posterior probability of the state we want to estimate, given the observations we see. We use Bayes rule to write that probability as a (normalized) product of two terms we know how to deal with: the likelihood term and the prior term. Letting $\vec{x}$ be the quantities to estimate, and $\vec{y}$ be the observations, we have

$$P(\vec{x}|\vec{y}) = kP(\vec{y}|\vec{x})P(\vec{x}) \tag{6.9}$$

where k is a normalization factor that forces that the integral of $P(\vec{x}|\vec{y})$ over all $\vec{x}$ is one.

The likelihood term tells us, given the model, how probable the observations are. If we assume additive, mean zero Gaussian noise, the probability that the jth color observation differs from the rendered parameters follows a mean zero Gaussian distribution. Remembering that the observations $\vec{y}_j$ are the the L, M, and S cone responses,

$$\vec{y}_j = \begin{pmatrix} L_j \\ M_j \\ S_j \end{pmatrix} \tag{6.10}$$

we have

$$P(\vec{y}_j|\vec{x}^i, \vec{x}_j^s) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp \frac{-|\vec{y}_j - \vec{f}(\vec{x}^i, \vec{x}_j^s)|^2}{2\sigma^2}, \tag{6.11}$$

For an entire collection of N surfaces, we have

$$P(\vec{x}|\vec{y}) = P(\vec{x}^i)\prod_j P(\vec{y}_j|\vec{x}^i, \vec{x}_j^s)P(\vec{x}_j^s) \tag{6.12}$$

reminder: Here's what's inside the rendering function, $\vec{f}(\vec{x}^i, \vec{x}_j^s)$. We assume diffuse reflection from each colored surface. Given basis function coefficients for the illuminant, $\vec{x}^i$, and a matrix $\mathbf{B}$ with the illumination basis functions as its columns, then the spectral illumination as a function of wavelength is the column vector $\mathbf{B}\vec{x}^i$. We also need to compute jth surface's diffuse reflectance spectral attenuation function, the product of its basis coefficients times the surface spectral basis functions: $\mathbf{A}\vec{x}^s_j$ In our diffuse rendering model, the reflected power is the term-by-term product (we borrow Matlab notation for that, .*) of those two. The observation of the jth color is the projection of that spectral power onto the eye's photoreceptor response curves. If those photoreceptor responses are in the columns of the matrix, $\mathbf{E}$, then the forward model for the three photoreceptor responses at the jth color is:

$$\vec{f}(\vec{x}^i, \vec{x}^s_j) = \mathbf{E}^T(\mathbf{A}\vec{x}^s_j \;.*\; \mathbf{B}\vec{x}^i). \tag{6.13}$$

Eq. (6.13), in component form

We can find the linear solution of Eq. (6.13), for a given illuminant vector, $\vec{x}^i$, and assuming no noise in the observations. Let's write everything out in component form, in order to do the calculation carefully. Let's assume we're only fitting the x^s_{kj} to the jth color patch observation, and therefore omit all subscripts j, for simplicity. So x^s_k will mean the kth reflectance basis component (of the jth patch). w indexes wavelength values.

$$\begin{aligned} y_n &= \sum_w E_{nw} \sum_k A_{wk} x^s_k \sum_m B_{wm} x^i_m && (6.14)\\ &= \sum_k x^s_k \sum_w E_{nw} A_{wk} \sum_m B_{wm} x^i_m && (6.15)\\ & && (6.16) \end{aligned}$$

If we define the n, k components of a matrix D to be

$$D_{nk} = \sum_w E_{nw} A_{wk} \sum_m B_{wm} x^i_m, \tag{6.17}$$

then we have

$$\vec{y} = \mathbf{D}\vec{x}^s \tag{6.18}$$

If $\mathbf{D}$ is invertible, then we have

$$\vec{x}^s = \mathbf{D^{-1}}\vec{y} \tag{6.19}$$

slide 65

I want to remind us about the similarities between the 1 = ab problem, and our color constancy problem. While we can't draw out the high-dimensional likelihood function for the color constancy problem, conceptually it's very similar to that of this 1=ab problem (and various other vision problems share these same characteristics). Many different settings of the parameters can explain our data, giving what we call a "likelihood ridge". For the 1=ab problem, that gives a 1-d ridge of parameter settings that explain the same data. For the color constancy problem, with 3-d data, surface parameterizations and illuminant parameterization, the likelihood ridge is 3-dimensional.

How do we pick from the many feasible solutions on the ridge, all with the same likelihood value? The priors will let us distinguish different values on the likelihood ridge.

slide 66 In the problem set, you'll fit Gaussians to model the observed prior distribution of surface and illuminant basis function coefficients.

slide 67 Another difference for different positions along the likelihood ridge is the "width" of the ridge. At some positions, only a very precise specification of all the parameter values will explain the observations. At other positions, we have more slop in the parameters, and many different nearby parameter settings also explain the data. In a Bayesian framework, this is most naturally quantified with the *loss function*, which specifies the penalty for guessing wrong. Let $\hat{\vec{x}}$ be your estimate of the parameters, $\vec{x}$. Then $L(\hat{\vec{x}}, \vec{x})$ is the loss incurred by guessing $\hat{\vec{x}}$ when the true value was $\vec{x}$. With the posterior probability, we can calculate the expected loss, $\bar{L}(\hat{\vec{x}}, \vec{x})$

$$\bar{L}(\hat{\vec{x}}, \vec{x}) = \int_{\vec{x}} L(\hat{\vec{x}}, \vec{x}) P(\vec{x}|\vec{y}) \tag{6.20}$$

We often use a loss function which is only a function of $\hat{\vec{x}} - \vec{x}$.

slides 68-70

Bayes rule lets us find a posterior probability for $\vec{x}$, given the observations $\vec{y}$. The final stage of Bayesian estimation is to go from that function, $P(\vec{x}|\vec{y})$, to a single best guess value, $\hat{\vec{x}}$, a point estimate.

The two most commonly used point estimates are called the MAP and MMSE estimates, and in your homework for this material, you'll use either one of those for this color constancy problem.

The MAP estimate stands for "maximum a posteriori", Latin for "just take the max of the posterior distribution". While quite a natural thing to do, it can suffer from various problems, since it only depends on the single maximum valued point of the posterior probability. The loss function implied by the MAP estimate assigns a constant penalty for all guesses, except for a precisely correct guess, which receives high reward. This penalty structure doesn't make sense for perceptual tasks, for which nearly the right answer is often just as good as precisely the right answer.

Another very common estimate is MMSE, "minimum mean squared error". This is the mean of the posterior probability, which, as the name implies results in an estimate which minimizes the expected squared error in the estimated parameter. The homework assignment, excerpted on these slides, gives details of how to find each of those estimates for this problem. Again, for perceptual problems, this loss function often doesn't make sense, although in practice an MMSE estimate is often very good. (Although for the 1 = ab problem, limited to $0 < a, b < 4$, the MMSE estimate is relatively far away from any feasible solution to the problem).