Illumination and reflectance spectra

Digital color images have three intensity values per pixel (RGB) which define a color at each pixel. We would like to understand better what these intensities mean. A good place to start our examination of color is with Isaac Newton and his prism experiment (late 17th century). Newton observed that when a beam of sunlight is passed through a prism, the beam is spread out into a fan of different color lights – like a rainbow. The theory that explains Newton's experiments is now well known. Light is composed of electromagnetic waves, with wavelengths ranging from 400-700 nanometers[1] and these waves are refracted by different amounts when they pass through the prism. (The same rainbow effect occurs in a lens too, but typically the spread is very small and anyhow lens designers correct for it – so its not a problem.)

Up to now we have defined radiance as a function of spatial position and direction. But radiance also depends on wavelength. That is, the radiance of light at a point $\mathbf{X}$ and in direction $\mathbf{L}$ and at a wavelength λ is $L(\mathbf{X}, \mathbf{l}, \lambda)$. Often we use the term *spectral power distribution* (SPD) to talk about something that is a function of wavelength. We will see other examples later.

The radiance of light coming from a real source such the sun, a light bulb, or a blue sky each have a different spectra $L(\lambda)$. Some light sources have more power at longer wavelengths, whereas others have more power at shorter wavelengths. A candleflame has relatively more power at long wavelengths than does direct sunlight. A blue sky has relatively more power at short wavelengths than does direct sunlight. Sources from many artificial lights (fluorescents) have most of their power concentrated in a very small range of wavelengths.

When a surface reflects light, the spectrum changes: the outgoing spectrum is different from the incoming spectrum, namely

$$L_{out}(\mathbf{x}, \mathbf{l}_{out}, \lambda) = \int f(\mathbf{x}, \mathbf{l}_{in}, \mathbf{l}_{out}, \lambda) \; L_{in}(\mathbf{x}, \mathbf{l}_{in}, \lambda) \; \mathbf{n}(\mathbf{x}) \cdot \mathbf{l}_{in} \; d\Omega_{in}.$$

If the surface is Lambertian, then this equation reduces to

$$L_{out}(\mathbf{x}, \lambda) = \int f(\mathbf{x}, \lambda) \; L_{in}(\mathbf{x}, \mathbf{l}_{in}, \lambda) \; \mathbf{n}(\mathbf{x}) \cdot \mathbf{l}_{in} \; d\Omega_{in}.$$

Notice that there is no interaction between different wavelengths.[2] The power is reweighted wavelength by wavelength. The function $f(\mathbf{x}, \lambda)$ is the *spectral reflectance* distribution at $\mathbf{x}$. Notice that it can (and typically does) vary over points on a surface $\mathbf{x}$. This just corresponds to our common experience that surface materials are non-uniform – they have reflectance (color) variations across them.

What can we say about $f(\mathbf{x}, \lambda)$ for the case of a Lambertian surface? If the values of $f(\mathbf{x}, \lambda)$ are very small, than the radiance leaving the surface will be small. In this case, the surface appears *dark* colored – it reflects relatively little light over all wavelengths. On the other hand, if the values of $f(\mathbf{x}, \lambda)$ are large, then the surface will be light colored – the lightest surface being white. Notice that there are limits on how large $f()$ can be, since a surface cannot reflect more light than arrives at it!

[1] A nanometer is 10^{-9} meters. e.g. for a wavelength of 500 nm, you need about 2000 "waves" to cover one millimeter.

[2] Surfaces that are fluorescent do have these interactions, but they are rare.

What if $f(\mathbf{x}, \lambda)$ has smaller values at short wavelengths but relatively larger values at long wavelengths? In this case, there will be relatively more long wavelength light reflected, and the reflected light will be more "reddish" (long wavelength) than the incoming light. [ASIDE: This issue of how color perception depends on spectral wavelength distributions is very interesting, and there is much to say. But not in this course, unfortunately – other fish to fry!]

Image irradiance

Let's say a bit more about the image irradiance $E(\mathbf{x}, \lambda)$. Let's drop the dependence on λ, because this is not important for the point I wish to make next. (You can add a λ parameter to each function below if you wish.)

The sensor plane is at a distance Z_s from the lens. Take a small patch of area a_s on the sensor plane, e.g. the area of a sensor element, and let $\mathbf{x}$ be a point in this area. Now reverse project the points on this patch through the lens and out to the world to a set of points of area a_o at the conjugate depth Z_s^*. These two sets of points at depths Z_s and Z_s^* have normals that are parallel to the optical axis. Let $\mathbf{n} \cdot \mathbf{l} = \cos\alpha$, where α is the angle between the center of the area a_s and the optical axis.

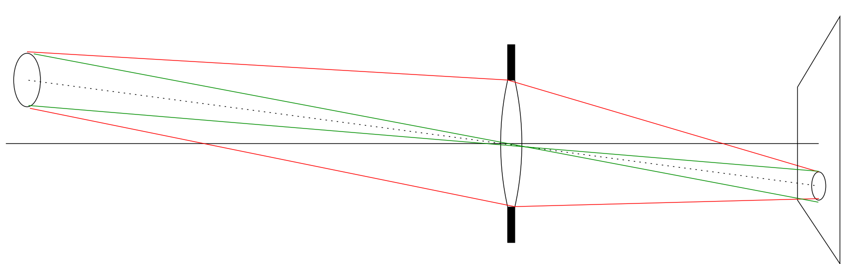
The solid angles Ω_s and Ω_s^* (in green below) subtended by the two areas when seen from the center of the lens are the same, namely

$$\Omega_s = \frac{a_s \cos\alpha}{(Z_s/\cos\alpha)^2} = \Omega_s^* = \frac{a_s^* \cos\alpha}{(Z_s^*/\cos\alpha)^2}.$$

Why? The numerator is the area of a patch, foreshortened by $\cos\alpha$. The reason for foreshortening is that, to define a solid angle, we project the patch onto a unit sphere – in this case, a unit sphere centered at the lens center. So it is only the area of the patch projected in the direction of the dotted line (in the figure below) that matters. The denominator is $(Z/\cos\alpha)^2$ which is just the distance squared from the lens center to the center of the patch. We divide the area by this distance in order to know the (projected) area on the *unit* sphere.

The above equations give:

$$\frac{a_s^*}{a_s} = (\frac{Z_s^*}{Z_s})^2.$$



Consider the rays from a_s^* that arrive at the lens. These rays pass through the lens and arrive at a_s. These rays are represented by the red bundle in the figure.

Assuming all of these rays have the same radiance $L(\mathbf{x})$, the radiant power ϕ (Watts) of light rays coming from the area a_s^* and reaching the lens is

$$\phi = L(\mathbf{x})\ a_{lens} \cos\alpha\ \Omega_s^*,$$

where $a_{lens} = \frac{\pi A^2}{4}$, and A is the diameter of the lens (recall lecture 4). The $\cos\alpha$ term is there because radiance has units Watts per steradian per unit area, where the area is perpendicular to the direction. But the lens is not perpendicular to the direction of the solid angle, so we need to foreshorten the lens area in that direction (by multiplying by $\cos\alpha$).

The above light power that reaches the lens then gets focussed on the sensor area a_s. Thus, the power per unit area reaching the sensor is

$$E(\mathbf{x}) = \frac{\phi}{a_s}.$$

Substituting from above gives:

$$E(\mathbf{x}) = (\frac{\pi A^2}{4})\frac{L(\mathbf{x})\cos^4\alpha}{Z_s^2}$$

The $\cos^4\alpha$ term has the effect of darkening the edges of the image, relative to the center of the image. It is called *vignetting*. It can be a significant effect for wide angle lenses, since α is large and so $\cos\alpha$ can be significantly different from 1.

If $Z_s \approx f$ (which is often the case), then the image irradiance is inversely proportional to the square of the f-number.

$$E(\mathbf{x}) = \frac{\pi A^2 L(\mathbf{x})\cos^4\alpha}{4f^2}$$

As mentioned in lecture 4, typical cameras allow users to vary the $f - numbers(\#)$ in powers of $\sqrt{2}$, namely: $1, 1.4, 2, 2.8, 4, 5.6, 8, 11, 16, 22, 32, 45, 64, 90, 128$. This allows users to step the irradiance up or down by factors of 2.

Trichromatic images (RGB)

Digital cameras make images with red, green, and blue components (RGB). This is done by dividing each pixel into four sub-pixels, and putting a partially transparent filter in front of each sub-pixel. The filters preferentially remove the energy at different wavelengths. The R filter allows the longer wavelengths through but absorbs the short and medium. The G filter allows the medium wavelengths through, but absorbs the short and long. The B filter allows the short through, but absorbs the medium and long. There are typically two G subpixels and only one R and B. (See the Bayer filter `http://en.wikipedia.org/wiki/Bayer_filter`.)

How can we related this filtering idea to the terminology we introduced earlier. Let $E(\mathbf{x}, \lambda)$ be the (spectral) irradiance of the light reaching a sub-pixel. When we place an R,G, or B filter in front of that sub-pixel, some fraction $C_R(\lambda), C_G(\lambda)$, or $C_B(\lambda)$ of the light is allowed through. The spectral irradiance reaching the pixel at wavelength λ would be $C_{R,G,B}(\lambda)\ E(\mathbf{x}, \lambda)$. Adding up the irradiance over all λ gives an RGB triplet at each pixel:

$$E_{R,G,B}(\mathbf{x}) = \int C_{R,G,B}(\lambda)\ E(\mathbf{x}, \lambda) d\lambda$$

Note that the (sub)pixel itself cannot distinguish different wavelengths and has no "memory" of what the distribution $E(\mathbf{x}, \lambda)$ was before it was filtered.

Shutter speed

Another important factor in determining the amount of light reaching each sensor is the time duration t over which the sensor is exposed to the light. Recall that we have been talking about the power (energy per unit time). So, to calculate the total light energy, we need to multiply by time that the sensor is exposed to the light.

In photography, the inverse of the time t that the sensor is exposed to the light is called the *shutter speed*. A shutter speed of 60, for example, means that the sensor is exposed for is open for 1/60 second. i.e. The "shutter" is open for this long. It is standard in photography to allow the user to vary the shutter speed ($\frac{1}{t}$) roughly in powers of 2, namely, $1, 2, 4, 8, 15, 30, 60, 125, 250, 500, 1000$. In fact, the actual values are powers of 2, e.g. $1, 2, 4, 8, 16, 32, 64, 128, 256, 512, 1024$.

Recall that typical cameras allow users to vary the f-numbers in powers of $\sqrt{2}$, which steps the irradiance by factors of 2. This suggests that if a user increases the shutter speed by one step (a factor of 2) and decreases the f-number by one step (a factor of $\sqrt{2}$), the image irradiance will not change. One needs to be careful in making such a claim, however. To change the f-number, you either change the focal length or change the aperture size. If you change the focal length, then you will change the field of view angle and hence change your $\mathbf{x}$. The alternative is to keep the focal length fixed and change the aperture (which is usually what happens when the user is allowed to change f-number). However, notice that if $\mathbf{x}$ is not in focus, then you will change the amount of blur at $\mathbf{x}$ (recall lecture 4), which might not be what you want.

Exposure, camera response function, and dynamic range

When you expose a (sub-)pixel to a certain irradiance for a certain time t, the pixel receives a certain power. This is called the *exposure* ξ:

$$\textit{Exposure} \qquad \xi = E_{R,G,B}\ t\ .$$

This exposure determines the 8-bit RGB intensity values (in a range 0 to 255) that you get at a pixel, as we explain next.

In many scenes that you wish to photograph, the range of image irradiances is too great to be captured in 8 bits. Some regions of the image will either be overexposed (meaning that these regions are mapped to the maximum intensity, say 255) or underexposed (meaning that all intensities are mapped to 0). One of the big challenges of shooting a good photograph is avoiding areas that are under or over exposed. This requires choosing the camera position and orientation in 3D, and also setting the camera parameters: the focal length, the aperture, and the shutter speed.

The RGB values $I_{RGB}(\mathbf{x})$ of the image produced by the camera depend on the exposure ξ via some non-linear response function ρ. Typically this function ρ is approximately the same for R, G, and B. Then,

$$I_{RGB} = \rho\ (\xi_{RGB}\ )$$

where

$$\rho : [0, \infty) \to \{0, 1, 2, \ldots, 255\}.$$

The function $\rho()$ typically has an S shaped curve, as the sketch on the left on the next page illustrates. The function takes the value 0 for low exposures (underexposed) then ramps up over some range of exposures, and then maxes out at 255 beyond some maximum exposure (overexposed).

The *dynamic range* of a camera is the range of exposures values that can be distinguished by the camera (up to quantization effects that are due to finite precision i.e. 8 bits per RGB). This is roughly the range of exposure values that do not under or overexpose the pixels. Typically *one defines dynamic range as a ratio of max:min exposures that the camera can measure without over- or under-exposing.* Similarly, we refer to the dynamic range of a scene to be the max:min ratio of exposures that this scene produces. If there are light sources visible in the scene (or specular reflections, such as off the black helmet in Assignment 1 Question 2), then the dynamic range will be quite high.

High dynamic range imaging (HDR) – see Assignment 1 Question 2

Photographers are often faced with scenes have a dynamic range that is greater than the dynamic range of the camera. There is no way to shoot such a scene without over or underexposing some pixels. Fortunately, researchers have developed a method to expand the dynamic range of images which requires taking multiple photos with different shutter speeds.[3] This method is called *high dynamic range imaging.*

Note that to properly expose the darker regions of the scene, you would need a slower shutter speed (large t, or small $\frac{1}{t}$) whereas to properly exposue a very bright regions of the image you would need a very fast shutter speed (small t, or large $\frac{1}{t}$). If you take a sequence of images with various shutter speeds, then pixel will be properly exposed (not near 0 or 255) in one of the images. Since any pixel will be properly exposed for *some* shutter speed $1/t$, one can in principle use that image to infer the exposure ξ_{RGB} values via:

$$\rho^{-1}(I_{RGB}) = \xi_{RGB}.$$

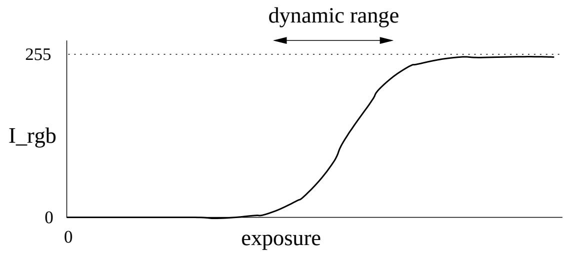
i.e. for any pixel, there should be some photograph in which the exposure at that pixels falls into the operating range of the sensor.

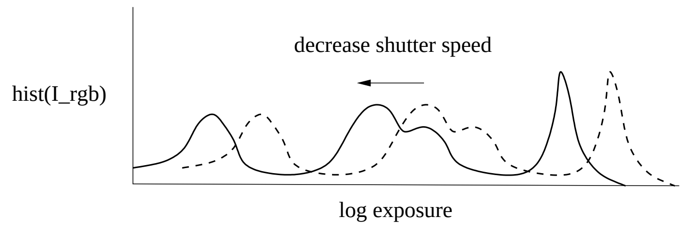
To understand better what is happening here, consider the log (base 2) of the exposure,

$$\log \xi_{RGB} \;=\; \log E_{R,G,B}(\mathbf{x}) + \log(t)$$

and note that changing the shutter speed just shifts the log of the exposure. If the shutter speeds are consecutive factors of 2 (which is typical with cameras), then doubling the shutter speed $1/t$ decrements the log exposure (at each pixel) by 1. This is illustrated in the figure below right, which shows the *histogram* (in this case, the number of pixels at each quantized log exposure value.). The histogram gets shifted to the left when you decrease the shutter speed. In order for pixels with some irradiance value E to be properly exposed in at least one of the images, the shutter speed must chosen so that the exposure falls into the region in which the camera is reliable.

[3]"Recovering High Dynamic Range Radiance Maps from Photographs" by Paul Debevec and Jitendra Malik, SIGGRAPH 1996. `http://www.debevec.org/Research/HDR/`





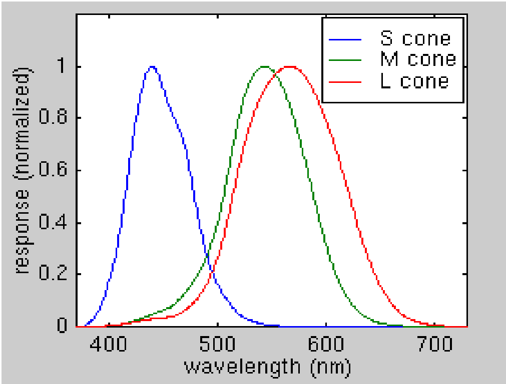
ASIDE (not covered in class)

Why do we represent colors using only three dimensions (RGB) ? The reason is that human vision itself is *trichromatic*. At the back of your eye is an array of *photoreceptor cells* that respond to light. These are similar to pixels, except that there is no notion of shutter speed. They respond continuously.

There are two classes of photoreceptor cells in the human eye. One class, called *cones*, responds to high light levels such as during the day. These are the ones that are used in color vision.[4]

The cone cells come in three types called L, M, and S, where L is for long wavelength (red), M is for medium wavelength (green) and S is for short wavelength (blue). You can think of them as RGB cones rather than LMS, if you like. Each cone type is defined by a pigment (a protein) that absorbs certain wavelengths of light better than it absorbs others. e.g. the pigment in the L cones absorbs longer wavelength light better than medium or shorter wavelength light. That is, an L cone is more sensitive to long wavelength light.

The three curves below (figure is in color) show the weights that determine the responses of each of the three cone types to the range of wavelengths of visible light. Each of the curves has been normalized so that its maximum is 1. These curves are called *spectral sensitivity functions*. You can think of these functions as playing the same role as the filters in front of the pixels in a digital camera.



[4]The second class of photoreceptor cells, called *rods*, responds at low light levels such as at night. These do not play any role in color vision. At low light levels, we do not see in color – we only see black and white.