| | | |
|---|---|---|
| CS 280 | Computer Vision | |
| Spring 2018 | Efros, Malik, Yu | Lecture 9 |

# 1 Lecture

## Physiology of Color Vision

We have three types of cones on the retina: short wave (S), medium wave (M), and long wave (L). As an example, S cones respond most strongly to wavelengths around 440 nm.

Note: the "red" (M) and "green" (L) cones are very close to each other, which may be an evolutionary adaptation from our ancestors needing to pick berries off of trees (and thereby localize red/green berries against the blue sky).

Rods and cones act as filters on the wavelength spectrum. To get the output of a filter, we multiply its response curve by the spectrum and integrate over all wavelengths. This gives us a single reading for each of the red, green, and blue cones (*trichromacy*).

How can we represent an entire spectrum of lighting with three numbers? We actually can't. Much information is lost, and two different spectra might appear indistinguishable (*metamers*).

If it's not too light and not too dark (e.g. at sunset), all four types of our receptors – cones $\times 3$ and rods – will be working. Therefore we can experience more shades of colors.

## Color Spaces

The default color space is RGB. This is great for devices (like cameras) but doesn't correspond to our perceptual experience ("what is brilliant pink?"). It's also not clear where the grays live, or what the hue and saturation are.

We can link the physical space and the perceptual space by considering only physical spectra with normal distributions (since there is no simple functional description for perceived color under all viewing conditions). Gaussians are characterized by their mean, their variance, and their enclosed area. How do these three parameters affect our perception of color?

In this simplified scenario, there are very clean correspondences.

- The mean wavelength of the Gaussian corresponds to the *hue* (green, yellow, etc.) we perceive.
- The variance of the Gaussian corresponds to *saturation*, which measures the peak color with respect to the amount of white light. If we add more and more white light, the peak color will feel more diluted (not as pure). In other words, if we spread out the Gaussian, the color won't appear as pure and the saturation will be lower.
- The area under the Gaussian corresponds to brightness – the brilliance of light we experience.

Say we're given a curve plotting number of photons against wavelength. In the general case, it's difficult to predict perceived color. But if we have a Gaussian, we can produce an accurate qualitative description.

HSV (hue, saturation, intensity) is the perceptual color space that arises from such analysis. In the HSV space, the representation at a certain intensity is a color wheel where the middle point is white (by contrast, on the RGB cube white is at a corner). Red is at angle 0, and going counterclockwise we also have yellow, green, cyan, blue, and magenta. Hue indicates the angle, while saturation indicates the "radius." At the extreme, we have a saturation of 0 or 1 (s.t. at 0 there are no colors). *Saturation measures color purity.*

One caveat with HSV is that it is not *perceptually uniform*: if we pick two points in the color space, the distance is not indicative of our perceived difference between the colors. L*a*b* is another color space that does try to be perceptually uniform. "L" is luminance, while "a" and "b" are chroma (like the cosine and sine of some angles).

**Color constancy:** even if the illumination changes, our visual system can factor out its impact and get to the physical quantity. Formally, it is "the ability to perceive the invariant color of a surface despite ecological variations in the conditions of observation." Note: we do not have color constancy over all global color transformations. If we completely invert the colors, we will be able to see the change in illumination.

**Camera white balancing:** let's do some post-processing to get back to what we experienced in the original scene. With manual white balancing, we choose a color-neutral object in the photos and normalize to that "calibration object" (e.g. we know that a sheet of paper should be white). With automatic white balancing (AWB), we can force the average color of the scene to be gray, or the brightest object to be white.

## Edges, Templates, Textures

An **edge** is a place where the intensity changes rapidly. To characterize rapid change, we can use derivatives (and look for peaks).

### Taking the Derivative with a Convolutional Filter

Recall that an image is simply a function over $x$ and $y$, i.e. a 2D function $f(x, y)$. The partial derivative along the $x$-axis is

$$\frac{\partial f(x,y)}{\partial x} = \lim_{\epsilon \to 0} \frac{f(x+\epsilon, y) - f(x,y)}{\epsilon}$$

For discrete data, this can be simplified as

$$\frac{\partial f(x,y)}{\partial x} \approx f(x+1, y) - f(x,y)$$

To implement this as convolution, the filter should be $[-1, 1]$. We can also use one of the fancy finite difference filters, e.g. that of Prewitt, Sobel, or Roberts.

The **gradient** of an image, $\nabla f = \left[\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right]$, points in the *direction of most rapid increase in intensity* at each point $(x, y)$. The gradient direction is given by

$$\theta = \tan^{-1}\left(\frac{\partial f}{\partial y} / \frac{\partial f}{\partial x}\right)$$

and the edge strength is given by the gradient magnitude

$$\|\nabla f\| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}$$

To handle noise, we should smooth our signal (convolve with a Gaussian) before taking the derivative. Also, since convolution is associative, we can actually do

$$\frac{d}{dx}(f * g) = f * \frac{d}{dx} g$$

which saves one operation.