CS 280 Computer Vision
Spring 2018 Efros, Malik, Yu Lecture 8

1 Lecture

Aliasing causes different signals to appear as *aliases* of each other when sampled. For example, we might undersample our signal and see it as something totally different. In order to defeat aliasing – to *anti-alias* – we get rid of high frequencies (*filter*) and then sample below the Shannon limit. This ensures that we don't lose information over the course of the sampling.

To get rid of high frequencies, we employ low-pass filtering (e.g. blurring with a Gaussian).

Filtering downgrades our visual content such that the sampling is enough.

When doing filtering, we can convolve with a small Gaussian, then subsample to cut each dimension in half, then convolve with another Gaussian, then subsample again, then convolve with another Gaussian... and so on. The end result will be a Gaussian pyramid, which is useful for coarse-to-fine searches over translations or searches over scale (e.g. for template matching).

To sharpen, we can add the high frequencies back into the original image:

$$f + \alpha(f - f * g) = (1 + \alpha)f - \alpha f * g = f * ((1 + \alpha)e - \alpha g)$$

f is the image, $f * g$ is the blurred image, and e is the identity. Hence $f - f * g$ represents the high frequencies (whatever was taken away by the low-pass filter). Note: the right formulation is useful because the filter doesn't depend on f and can be precomputed.

Frequencies

The human eye is more sensitive to the middle frequencies than to the high and low frequencies. When we're closer to something, we see the high frequencies as middle frequencies; when we're farther from something, we see the low frequencies as middle frequencies. We can therefore create *hybrid images*, which are perceived differently according to the viewing distance, by adding a low frequency image to a high frequency one.

A Fourier transform is just a change of basis. It's a way to represent information as a linear combination of differently-oriented sine curves. We're representing an image (losslessly!) by coefficients for these basis images. In other words, a Fourier transform separates information based on the low frequencies and the high frequencies. It's merely a different way to analyze a signal (e.g. in terms of frequency content).

Band-pass filtering: a way to perform a Fourier transform in image space. We start with a Gaussian pyramid. *Recall that a Gaussian pyramid is a way to peel away low-frequency information, a little at a time.* We want each image in the pyramid to be a separate band of frequencies (the lowest image would be the highest frequencies, and then it would have bands of lower and lower frequencies).

To obtain these images, we can simply subtract successive low-pass images from each other (i.e. those in the Gaussian pyramid). This gives us a *Laplacian pyramid*. Each image will be its own separate sub-band.

Human Vision

We don't see the entire image as one thing. Our eyes move and we get little samples of the visual world. These movements are called **saccadic eye movements**, and they happen several times each second. We focus on a certain part of the image and take a snapshot, which will be sharp around the center (the fovea) and blurry at the periphery. And then we move somewhere else. Basically, we see the image in patches, in contrast to a computer which sees it all as one thing.

There are two types of light-sensitive receptors: *cones* and *rods*. **Cones** are for color vision (there are three types) and need a lot of light to operate. **Rods** are gray-scale sensors which are highly sensitive and don't need a lot of light.

The distribution of rods and cones is non-uniform on the retina. The fovea (at the center) has a high density of cones and allows us to see color and high-frequencies well. Away from the fovea we mostly have a bunch of rods which, being distributed more sparsely, only allow for blurrier and grayer vision.

For this reason we mostly see color and high frequencies at the center, and gray and low frequencies on the periphery. The cones at the fovea are also much less sensitive, meaning we need good lighting to read.

Note: our eye contains more green cones than red or blue cones. Thus we are more sensitive to green.

Color

Color is not really a physical property. It's a psychological thing. The physical thing is electromagnetic radiation. We have electromagnetic waves with different frequencies, going from very low gamma rays to very high radio waves. And out of this spectrum of waves, there is a tiny, tiny sliver between 400 nm and 700 nm that our eyes are sensitive to.

We see these waves as somewhat different things. They are more bluish at the lower end of the frequency spectrum, and reddish at the higher end. Also, it turns out that the human sensitivity curve (as a function of wavelength) is bell-shaped and peaks at the middle of the spectrum. This signifies that we're more sensitive to green than purple or red.

We see light in the range from 400 nm to 700 nm due to an evolutionary adaptation: our sun, i.e. our main creator of photons, is emitting photons in this range. Therefore, we have adapted to see this range of wavelengths.

Each photon can be thought of as having a certain wavelength, from 400 to 700 nm. And then every patch of length can be completely described physically by its spectrum – the number of photons per time unit at each wavelength ("how many blue photons did I get? How many red photons did I get?...").

Not everything is an emitter. Most objects are simply reflectors, and have their own reflectance spectra. It turns out that tomatoes absorb everything except red – they're red-haters! They take in pretty much everything but send back red.