Where's Abe?

Corresponding MATLAB script is lincoln2.m

See: pp. 54-79 of David Marr's monograph, "Vision" (New York: W. H. Freeman and Company; 1982)
See also: Harmon LD & Julesz B (1973) "Masking in visual recognition: Effects of two-dimensional filtered noise." Science, 180:1194-7.

RTB wrote it, 16 May 2003; re-worked for QMBC, June 2011

MATLAB Concepts Covered:
1. Image manipulation: blkproc
2. Edge detection: LoG and thresholding
3. Image analysis in the frequency domain: 'fft2', 'ifft2'
4. Band-rejection filters
5. Critical-band Masking (see accompanying paper: Science 1973)
6. Zero-crossing and the "raw primal sketch"

Discretely sampling and then quantizing an image can render it difficult to recognize. Try this with an image of Abraham Lincoln.

Ex. 1: Load the image of Lincoln ("lincoln.jpg") and display it. Then use 'blkproc' to coarsely sample the image, replacing each block of native pixels with the mean for that block.

The discretized image is very difficult to recognize. Yet, if you squint, you might notice something interesting. What is going on here?

In 1980, David Marr and Ellen Hildreth wrote a very important paper in which they developed the now-famous "vel-squared-G" filter for image processing—we've met this previously as the "LoG" or "Laplacian of Gaussian" filter. Marr went on to argue in his monograph that one important function of early vision was to determine which changes in intensity in the image corresponded to real features in the world, such as edges. He suggested that the way this was accomplished was to compare zero-crossings at multiple spatial scales. Let's see what these actually look like for our pixellated Abe.

Ex. 2: Use the 'log' filter to detect zero-crossings at three different spatial scales, corresponding to sigmas of 4, 7 and 12 pixels.

Marr argued that the nature of the physical world placed constraints on the zero-crossing geometry in the different channels, referring to this as the "spatial coincidence assumption":

> "If a zero-crossing segment is present in a set of independent [LoG] channels over a contiguous range of sizes, and the segment has the same position and orientation in each channel, then the set of such zero-crossing segments indicates the presence of an intensity change in the image that is due to a single physical phenomenon . . . ."

He further argued that if the zero-crossings of the larger channels are "accounted for" by the smaller ones, then the visual system interprets the world as roughly what the smaller channels are "seeing."

Ex. 3: Compare the zero-crossings in the 3 different channels by overlaying them.

So, according to Marr, even though our large channels are seeing something we might recognize as Abe, the fact that our small channels can account for it causes us to see what the small channels are seeing: a bunch of blocks. By squinting, we are removing the small-channel information (essentially, we are applying a bigger Gaussian) and thus revealing the stuff in the larger channels, which look more like Abe.

Ex. 4: Remove high spatial frequency (i.e. small channel) information by blurring the image with a Gaussian filter of sigma = 7.

Not bad. That actually looks more like Abe than the pixellated image we started with. We gained by subtraction. Of course, an engineer would just say that the signal was in the low frequencies and the noise was predominantly in the high, so our Gaussian filter removed noise by getting rid of the high frequencies.

In the preceding section we made a spatial filter and applied it to the image by convolution—that is, sliding our filter across the image and computing the result at each location. But we can do the same thing in the so-called "frequency domain" by first taking the Fourier Transform ('fft2') of the image, multiplying it by our filter, then re-converting to a spatial representation using the Inverse Fourier Transform. This might seem very inefficient, but in many cases it's actually faster than the spatial method. This is because very fast algorithms exist for the Fourier Transform and multiplication is faster than convolution.

Ex. 5: Display the amplitude spectrum of the Lincoln image.

Believe it or not, the same information regarding Abe's image is present in the Fourier-transformed representation. We don't see all of it in this figure because we are only able to render an image of the real part of the transform, which gives us the amplitudes of the various sine waves. The imaginary part gives the phase relationships, which contain critical information.

NOTE: If you'd like a more intuitive notion of what the Fourier Transform does, see fft_demo.m, which simplifies things by using a 1-dimensional signal rather than a 2-dimensional image.

Ex. 6: Low-pass filter the Lincoln image in the frequency domain by simply applying a circular window to pass only the spatial frequencies around the origin. Convert back to a recognizable image using the Inverse Fourier Transform ('ifft2').

Let's return to Marr's way of thinking about this issue. According to him, what would happen if we removed just the middle frequencies? In this case, the zero-crossings in the mid-sized filters would no longer corroborate those in the large or the small channels. The prediction is that the visual system would attribute the two different kinds of information (i.e. that in the small and large channels) to different physical phenomena. To see if this is the case, we need a band-rejection filter. In frequency space, this would consist of an annulus (or donut) of zeros in a sea of ones.

Ex. 7: Use the accompanying help function ('circle') to generate a "band rejection filter," multiply this by the Fourier Transformed "image" and plot the resulting amplitude spectrum.

Interesting. We have the same pretty good rendering of Abe, but he appears as if behind a screen. The high frequency noise is still there (unlike in the low-pass filtered image), but somehow it doesn't interfere with our ability to recognize Abe.

Another framework in which to think about this phenomenon is that of "critical-band masking." The basic idea is that it's not just the presence of noise that matters, but also how close it is (in frequency space, in this case) to the signal. This conceptual framework is laid out nicely in the paper by Harmon and Julesz.