Edge detection

Today we will see key ideas about edge detection which were introduced in a very well known article [1] by John Canny. The edge detection method he proposed is now called the Canny edge detector. We begin with the 1D case. We assume the image is a sum of a step edge plus noise

$$I(x) = Au(x) + n(x).$$

Here we are ignoring the fact that real images don't have negative intensities as can easily occur with this model.

For now, we are defining $I(x)$ and $u(x)$ on a discrete variable x, since that is how $n(x)$ is defined. As our argument below goes along, we will find it is sometimes convenient to jump to the continuous case, namely we will sometimes use an integral to approximate a summation.

[BEGIN ASIDE: I will be careful not to perform an integration over the noise. Although it is more possible to define noise $n(x)$ over a continous variable x, this is technically difficult to do (and involves concepts of Brownian motion and so forth – we really don't want to go there). Moreover, continuous noise is not relevent for our problem, since by "noise" we really just mean a perturbation of an intensity on a *discrete* pixel. While it would be ok to talk about a continuous image irradiance function (defined on a continuous sensor plane), it just wouldn't make sense to talk about a continuous noise function in our particular problem. Noise is added to (discrete) pixels. END ASIDE]

Detection

To detect the edge, we convolve $I(x)$ with some $f(x)$

$$(f * I)(x) = A\ (f * u)(x) + (f * n)(x).$$

What should our $f(x)$ be? The basic idea is that $f(x)$ should compute a derivative. We could use the D function seen at the beginning of last lecture, but maybe this is not the best function. So for now, let's just assume that $f(x)$ that is anti-symmetric filter,

$$f(x) = -f(-x).$$

In particular, our assumption implies that $f(0) = 0$ and $\int f(x)dx = 0$. We will also assume that $f(x)$ has finite support i.e. it is only non-zero in a finite neighborhood of $x = 0$.

Canny's first criterion for a good $f(x)$ is that the response to the edge $u(x)$ should be as large as possible, relative to the response to the noise. That is, the ratio of "signal" to "noise" should be as large as possible. One typically doesn't distinguish positive from negative responses here, and so we will just work with squared responses. Specifically we would like the (squared) "signal to noise" ratio

$$\frac{(A\ ((f * u)(0))^2}{\mathcal{E}\{\ (\ (f * n)(x)\ )\ ^2\}}$$

to be as large as possible. Note that the $\mathcal{E}$ in the denominator is the "expected value" operator, i.e. $n(x)$ is a random variable. (See below for details.)

[1] J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Mach. Intel. 1986. Cited nearly 9000 times according to Google Scholar.

First we examine the numerator. Applying the definition of $u(x)$ from last lecture (using the continuous $u(x)$) gives:

$$((f * u)(0))^2 = (\int_{-\infty}^{0} f(x)dx)^2$$

For the denominator, we have

$$\mathcal{E}\{ \ (\ (f * n)(x) \ )^2 \} = \mathcal{E}\{ \ (\ \sum_{-\infty}^{\infty} f(x')n(x - x') \ )^2 \}.$$

which we rewrite as follows. We are assuming that noise variables $n(x)$ have mean 0 and variance σ_n^2. and $f(x)$ has only finite support i.e. there are only finitely many x's where $f(x) \neq 0$. Then we are just computing the variance of a sum of random variables. But from basic statistics, if n_i are independent and identically distributed random variables with mean 0 and variance σ^2, then $\sum a_i n_i$ has variance $\sigma^2 \sum a_i^2$, that is,

$$\mathcal{E}\{\sum_i a_i n_i\}^2 = \sigma^2 \sum_i a_i^2.$$

So in the case of noise with mean 0 and variance σ_n^2, we get

$$\mathcal{E}\{(\sum_{-\infty}^{\infty} f(x')n(x - x'))^2 \ \} = \sigma_n^2 \sum_{-\infty}^{\infty} f(x')^2.$$

If the support of $f(x)$ is much greater than a few pixels and $f(x)$ is sufficiently smooth, then we can approximate the summation over $f(x')^2$ by an integral, i.e. $\sum f(x)^2 \approx \int f(x)dx$, so

$$\frac{S}{N} \equiv \frac{A^2 \ (\ \int_{-\infty}^{0} f(x) \ )^2}{\sigma_n^2 \ \ \int_{-\infty}^{\infty} f(x)^2 dx} \quad (1)$$

Note that S:N increases with the edge amplitude A and decreases with variances of the noise. This should correspond to your intuition that you will be able to detect the edge better if the step is large and the noise is small.

What more can we say about $f(x)$? First of all, observe that multiplying $f(x)$ by some number c to get $cf(x)$ would not change the above ratio since we would get c^2 in both the numerator and denominator and these would cancel out.

What if we were to stretch or squeeze $f(x)$ in the x direction? Let's compare the S:N that we get using $f(x)$ to the S:N we would get if we used a $g(x)$ such that

$$f(x) = g(ax)$$

for some scale factor $a > 0$. So, for example, if $a > 1$ then $f(x)$ would be a squeezed version of $g(x)$ or equivalently $g(x)$ would be a (horizontally) stretched out version of $f(x)$.

Rewriting the $f(x)$ integrals in Eq. 1), we get

$$\int_{-\infty}^{0} f(x)dx = \int_{-\infty}^{0} g(ax)dx = \frac{1}{a}\int_{-\infty}^{0} g(ax)dax \ \ = \ \frac{1}{a}\int_{-\infty}^{0} g(x)dx$$

and
$$\int_{-\infty}^{\infty} f(x)^2 dx = \int_{-\infty}^{\infty} g(ax)^2 dx = \frac{1}{a}\int_{-\infty}^{\infty} g(ax)^2 dax = \frac{1}{a}\int_{-\infty}^{\infty} g(x)^2 dx$$
Thus, substituting into Eq. 1 gives
$$\frac{S}{N} = \frac{A^2\ (\int_{-\infty}^{0} f(x))^2}{\sigma_n^2 \int_{\infty}^{\infty} f(x)^2} \approx \frac{A^2\ (\int_{-\infty}^{0} g(x)dx\ )^2}{a\ \sigma_n^2 \int_{\infty}^{\infty} g(x)^2 dx}$$

So,
$$(S:N)_f = \frac{1}{a}(S:N)_g$$
i.e. the S:N using $g(x)$ would be a times as great as the S:N using $f(x)$. For example, if $a = 2$, then $g(x)$ would be stretched out relative to $f(x)$ and the $S:N$ would be twice as great using $g(x)$ rather than $f(x)$. Based on this consideration, we might conclude that we should use as large an $f(x)$ as possible to detect an edge. (Intuitively, a large $f(x)$ would average out the noise more, and so it makes sense that a larger $f(x)$ would help us detect the edge.)

Localization

A second issue is that we also want to accurately locate the edge, namely we want the the local maximum of $f(x) * I(x)$ to be close to the position of the true edge (which the model says is at $x = 0$). Because there is noise within $I(x)$, we are not guarenteed that the maximum $f(x) * I(x)$ will be found exactly at $x = 0$. Let's have a look at where the maximum occurs.

To find a local maximum of $f * I(x)$, by definition we need to find the value of x (call it $x = x_0$) where $\frac{d}{dx} f * I(x) = 0$, i.e. to solve
$$A\frac{d}{dx}(f * u)(x_0) + \frac{d}{dx}(f * n)(x_0) = 0.$$
Note that x_0 is not necessarily 0, because of the noise.

We rewrite the "signal" term by taking advantage of the commutativity of convolution (and the fact that taking a derivative is a convolution),
$$((\frac{d}{dx}f) * u))(x_0) = (f * \frac{du}{dx})(x_0) = (f * \delta)(x_0) = f(x_0).$$
We assuming $f(x)$ is smooth and anti-symmetric (in particular $f(0) = 0$), so if we take a Taylor series approximation of $f(x)$ around $x = 0$ and ignore anything but first order terms, we get
$$f(x_0) \approx f(0) + f'(0)x_0 = f'(0)x_0.$$
Thus, the signal component of the response is approximately $Af'(0)x_0$. Thus,
$$Af'(0)x_0 = -\frac{df(x)}{dx} * n(x))$$
We can then write the variance of x_0 as
$$\mathcal{E}(x_0^2) = \frac{\mathcal{E}(\frac{df(x)}{dx} * n(x))^2}{A^2 f'(0)^2}$$

We evaluate the numerator similarly as before,

$$\mathcal{E}(\frac{df(x)}{dx} * n(x))^2 \ = \ \mathcal{E}\{\sum_x \ f'(x)n(u-x))^2\} = \sigma_n^2 \sum_x \ f'(x)^2$$

and, approximating the sum as an integral as before, we get

$$\mathcal{E}(x^2) = \frac{\sigma_n^2}{A^2} \frac{\sum_x f'(x)^2}{f'(0)^2} \approx= \frac{\sigma_n^2}{A^2} \frac{\int_{-\infty}^{\infty} f'(x)^2 dx}{f'(0)^2} \tag{2}$$

and so we see that the variance of the position x_0 grows with $\frac{\sigma_n^2}{A^2}$. If we wish this variance to be small, again we would like A to be large and σ_n to be small.

What about $f(x)$? If we multiply $f(x)$ by a constant c, then we don't change the location x_0. What if we stretch $f(x)$?

Notice that

$$f'(x) = \frac{df(x)}{dx} = \frac{dg(ax)}{dx} = \frac{dg(ax)}{d(ax)} \frac{d(ax)}{dx} = g'(ax)a \ .$$

and so

$$\int_{-\infty}^{\infty} f'(x)^2 dx = \int_{-\infty}^{\infty} \ g'(ax)^2 a^2 dx = a \int_{-\infty}^{\infty} \ g'(ax)^2 d(ax) = a \int_{-\infty}^{\infty} \ g'(x)^2 dx$$

and $f'(0) = ag'(0)$, and so

$$\mathcal{E}(x_0^2) = \frac{\sigma_n^2 \int f'(x)^2 dx}{A^2 f'(0)^2} = \frac{\sigma_n^2 \ a \ \int g'(x)^2 dx}{A^2 a^2 g'(0)^2} = \frac{\sigma_n^2 \ \int g'(x)^2 dx}{A^2 \ a \ g'(0)^2}$$

Thus using $g(x)$ instead of $f(x)$ would boost $\mathcal{E}(x^2)$ by a factor a. Thus, if we use a large $f(x)$ to fight noise, we now pay a price in not being able to localize the edge as accurately.

This result is quite fundamental (and I expect you'll agree it is not obvious). Using a bigger filter $f(x)$ does improve your ability to detect an edge in the presence of noise, but it means that your estimate of the position of the edge will be less accurate.

[ASIDE: Although I didn't discuss it in class, I should point out that there is more to Canny's argument than what I have given you here. In particular, the noise produces many local maxima of $f * I(x)$. Most of these maxima will be small, but there may be more than one local maxima that is large (and is due to the edge near $x = 0$). What is the likelihood of multiple large local maxima occuring near $x = 0$, and how can you choose $f(x)$ to keep that likelihood small? These are some of the other further questions that Canny's paper addresses. In addition, he addresses the case of edge detection in 2D images, which I will discuss next lecture.]