Introduction to Edge Detection

We will next look at the problem of finding an edge in an image. We take a simple model in which the image is the sum of a step function and a noise function,

$$I(x) = au(x) + n(x) \tag{1}$$

where the constant a is sometimes called the "amplitude" of the edge, and $n(x)$ is set of independent, identically distributed noise variables with mean 0 and variance σ_n^2.

To detect an edge, one typically convolves $I(x)$ with some function $f(x)$ such that $I(x) * f(x)$ has a local maximum (or minimum) at $x = 0$. Even for the simplest case of a unit step edge $u(x)$, there are challenging issues that arise.

- There may not be much difference in intensity on the two sides of an edge, that is, the edge amplitude a might be a small number; in this case, the local maximum that is due to the edge might be hidden by other local maxima that are due to the noise;
- Even if we can detect a local maximum in $|I(x) * f(x)|$, the position of this detected local maximum might not be at its correct position $x = 0$, i.e. the noise might cause the local maximum to be shifted. This is important since many algorithms are sensitive to the position of image features. (For example, you saw in Assignment 1 how the vanishing point estimate will vary with the position of the selected image pixels.)

How can you reduce these noise effects? Recall the local averaging filter from last lecture. Local averaging *smooths* out an image. In the case that the image contains a step edge plus noise as in Eq. (1), the smoothing has two effects: it smooths out the noise (and reduces it, which is good), and it also smooths out the edge (which is not good since "the location" of a smoothed edge is less well defined).

Today we will see key ideas about edge detection which were introduced in a classic paper by John Canny[1] 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, namely Eq. 1. (Note this model allows images $I(x)$ to have negative intensities, i.e. noise $n(x)$ can be negative.)

For now, we define $I(x)$, $u(x)$, and $n(x)$ on a discrete variable x. As our argument below goes along, we will find it is sometimes convenient to jump to the continuous case for $I(x)$ and $u(x)$, namely we will sometimes use an integral to approximate a summation. The noise $n(x)$ will always be over a discrete x though.

Detection

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

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

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

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

[1] J. Canny, "A Computational Approach to Edge Detection", IEEE Trans. Pattern Analysis and Mach. Intel. 1986.

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

Canny's first criterion for a good $f(x)$ is that the response to the edge $au(x)$ should be as large as possible, relative to the response to the noise $n(x)$. 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. We would like the (squared) "signal to noise" ratio

$$\frac{a^2 \ (f * u)(0)^2}{(\ (f * n)(x)\ )^{\ 2}}$$

to be as large as possible, *on average*. To capture the "on average" condition, we take the expected value of the denominator, so we would like the following ratio to be as large as possible:

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

Since a and $n(x)$ are fixed, the only way to increase the ratio is to change $f(x)$.

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. Since $f(x)$ has finite support i.e. there are only finitely many x's where $f(x) \neq 0$, we are just computing the variance of a sum of finite number of random variables. But from basic statistics[2], if n_i are independent and identically distributed random variables with mean 0 and variance σ^2, then $\sum c_i n_i$ has variance $\sigma^2 \sum c_i^2$, that is,

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

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 if $f(x)$ is smooth, then we can approximate the summation by an integral,

$$\sum_{-\infty}^{\infty} f(x)^2 \approx \int_{-\infty}^{\infty} f(x)^2 dx,$$

[2] See `http://en.wikipedia.org/wiki/Variance` up and including the part "Sum of uncorrelated random variables."

so

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

We see S:N increases with the edge amplitude a and decreases with variance of the noise. This is no surprise: you should be able to detect a step 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 $f(x)$ 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 using a horizontally scaled version of $f(x)$,

$$f_s(x) \equiv f(sx)$$

for some scale factor $s > 0$. So, for example, if $s > 1$ then $f_s(x)$ would be a horizontally squeezed version of $f(x)$.

Rewriting the $f(x)$ integrals in Eq. 2), we get for any x that:

$$\int_{-\infty}^{0} f_s(x)dx = \int_{-\infty}^{0} f(sx)dx = \frac{1}{s}\int_{-\infty}^{0} f(sx)dsx \ = \ \frac{1}{s}\int_{-\infty}^{0} f(x)dx$$

and

$$\int_{-\infty}^{\infty} f_s(x)^2 dx = \int_{-\infty}^{\infty} f(sx)^2 dx = \frac{1}{s}\int_{-\infty}^{\infty} f(sx)^2 dsx = \frac{1}{s}\int_{-\infty}^{\infty} f(w)^2 dw, \text{ where } w = sx$$

Thus, recalling Eq. 2,

$$\frac{a^2 \ (\int_{-\infty}^{0} f(sx)dx)^2}{\sigma_n^2 \int_{\infty}^{\infty} f(sx)^2 dx} \approx \frac{a^2 \ \frac{1}{s^2}(\int_{-\infty}^{0} f(x)dx \ )^2}{\frac{1}{s} \ \sigma_n^2 \int_{\infty}^{\infty} f(x)^2 dx}$$

So,

$$(S:N)_{f_s(x)} = \frac{1}{s}(S:N)_{f(x)}$$

Thus, the S:N using $f_s(x)$ would be $1/s$ times as great as the S:N using $f(x)$. For example, if $s = 2$, then $f_s(x)$ would be squeezed relative to $f(x)$ and the $S:N$ would be half as great. Based on this consideration, we might conclude that we should use as large i.e. stretched an $f(x)$ as possible to detect an edge. The larger $f(x)$ would average out the noise more.

Localization

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

To find a local maximum of $f(x) * I(x)$, we find the x such that

$$a\frac{d}{dx}(f * u)(x) + \frac{d}{dx}(f * n)(x) = 0. \tag{3}$$

Note that $n(x)$ is defined on discrete x only, and so this equation only makes sense (strictly speaking) if the derivative is a discrete derivative. In this case, $u(x)$ would also need to be defined on discrete x.

Rather than introducing a discrete version of $u(x)$, we will stay with the continuous $u(x)$ when examining the first term. Recall that convolution is commutative and that the derivative operator is a convolution, and so we can rewrite

$$((\frac{d}{dx}f) * u))(x) = (f * \frac{du}{dx})(x) = (f * \delta)(x) = f(x).$$

For the second term, we use the fact that convolution is commutative to get

$$\frac{d}{dx}(f * n)(x) = (\frac{df}{dx} * n)(x).$$

Hence we want to find x such that

$$af(x) + (\frac{df}{dx} * n)(x) = 0. \tag{4}$$

We are 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

$$f(x) \approx f(0) + f'(0)x = f'(0)x.$$

Thus, the response $f(x) * au(x)$ is approximately $af'(0)x$. Letting $\hat{x}$ be an x that satisfies Eq. (4), we get

$$af'(0)\hat{x} = -(\frac{df}{dx} * n)(\hat{x})$$

Since $\hat{x}$ will have mean 0, i.e. $\mathcal{E}(\hat{x}) = 0$, the variance of $\hat{x}$ is

$$\mathcal{E}(\hat{x}^2 - (\mathcal{E}\hat{x})) = \mathcal{E}(\hat{x}^2) = \frac{\mathcal{E}((\frac{df}{dx} * n)(\hat{x}))^2}{a^2 f'(0)^2}.$$

We evaluate the numerator similarly as before,

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

Notice that the right side doesn't depend on $\hat{x}$ since the noise is independent and identically distributed.

Approximating the sum as an integral, we get

$$\mathcal{E}(\hat{x}^2) = \frac{\sigma_n^2}{a^2}\frac{\sum_x f'(x)^2}{f'(0)^2} \approx= \frac{\sigma_n^2}{a^2}\frac{\int_{-\infty}^{\infty} f'(x)^2 dx}{f'(0)^2}. \tag{5}$$

Note the variance of the estimated edge position $\hat{x}$ 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 if we scale $f(x)$? Notice that

$$\frac{df_s(x)}{dx} = \frac{df(sx)}{dx} = s\frac{df(sx)}{d(sx)} = sf'(sx)$$

and, in particular,

$$f_s'(0) = sf'(0).$$

Also,

$$\int_{-\infty}^{\infty} f_s'(x)^2 dx = \int_{-\infty}^{\infty} s^2 f'(sx)^2 dx = s\int_{-\infty}^{\infty} f'(sx)^2 d(sx) = s\int_{-\infty}^{\infty} f'(u)^2 du.$$

Thus,

$$\mathcal{E}(x^2)_{f_s} = \frac{\sigma_n^2 \int f_s'(x)^2 dx}{a^2 \ f_s'(0)^2} = \frac{s\sigma_n^2 \int f'(x)^2 dx}{s^2 \ a^2 f'(0)^2} = \frac{\sigma_n^2 \int f'(x)^2 dx}{s \ a^2 f'(0)^2} = \frac{1}{s}\mathcal{E}(x^2)_f$$

Thus using $f_s(x)$ instead of $f(x)$ would multiply the variance of the estimated edge position $\hat{x}$ by a factor s. Thus, if we use a large $f(x)$ to fight noise (small s), we would pay a price in not being able to localize the edge as accurately.

This tradeoff between detection and localization is a fundamental which arises, not just for edge detection, but for many problem in computer vision. Using a bigger filter $f(x)$ improves your ability to detect some property of the image in the presence of noise, but it reduces your ability to localize that property.