Edge Detection in Digital Image Processing

Debosmit Ray

Thursday, June 06, 2013.

1. Introduction

In this paper, I discuss the mathematical theorems and algorithms used in image processing. Digital Image Processing is the use of computer algorithms to perform image processing on digital images. Since the use of complex algorithms are allowed, digital image processing can offer both more sophisticated performance at simple tasks, and the implementation of methods which would be impossible by analog means. The uses include feature extraction and pattern recognition, for which to occur, the identification of the edges is very important. Here, I will emphasize on the *Canny Edge Detection* and the *Sobel Edge Detection*. With the fast computers and signal processors available in the 2000's, digital image processing has become the most common form of image processing and is general used because it is not only the most versatile method but also the cheapest. Firstly, images are a measure of parameter over space, while most signals are measures of parameter over time. Secondly, they contain a great deal of information; image processing is any form of information processing for which the input is an image, such as frames of video; the output is not necessarily an image, but can be, for instance, it can be a set of features of the image.

Edge detection is a process of locating an edge of an image. Detection of edges in an image is a very important step towards understanding image features. Edges consist of meaningful features and contain significant information. It significantly reduces the image size and filters out information that may be regarded as less relevant, thus preserving the important structural properties of an image. Most images contain some amount of redundancies that can sometimes be removed when edges are detected and replaced during reconstruction. This is where edge detection comes into play. Also, edge detection is one of the ways of making images not take up too much space in the computer memory. Since edges often occur at image locations representing object

boundaries, edge detection is extensively used in image segmentation when images are divided into areas corresponding to different objects.

2. Canny Edge Detection

The main aims of the Canny Edge Detector are as follows:

(a) Good detection - There should be a low probability of failing to mark real edge points, and low probability of falsely marking non-edge points. Since both these probabilities are monotonically decreasing functions of the output signal-to-noise ratio, this criterion corresponds to maximizing signal-to-noise ratio. So basically, we need to mark as many real edges as possible.

(b) Good localization - The points marked out as edge points by the operator should be as close as possible to the centre of the true edge. In essence, the marked out edges should be as close to the edges in the real edges as possible.

(c) Minimal response - Only one response to a certain edge. This is implicitly captured in the first criterion since when there are two responses to the same edge, one of them must be considered false. So, the idea is that an edge should be marked only once, and image noise should not create false edges.

To satisfy these requirements, John F. Canny used the calculus of variations - a technique which finds the function which optimizes a given functional. The optimal function in Canny's detector is described by the sum of four exponential terms, but it can be approximated by the first derivative of a Gaussian.

A Now, we need to convert the ideas of detection and localization into a mathematical form that is solvable. For the signal-to-noise ratio and localization, we let the impulse response of the filter to be $f(x)$ and the edge itself to be $G(x)$. We go on to assume that the edge is centered at $x = 0$. Then, the response of the filter to this edge at its center H_G is given by the convolution integral

$$H_G = \int_{-W}^{+W} G(-x) f(x) \, dx \tag{1}$$

assuming that the filter has a finite impulse response bounded by $[-W, +W]$. The root-mean-squared response to the noise $n(x)$ only, will be

$$H_n = n_0 \left[\int_{-W}^{+W} f^2(x)\, dx \right]^{\frac{1}{2}} \tag{2}$$

where n_0^2 is the mean-squared noise amplitude per unit length. The sound-to-noise ratio is defined to be

$$SNR = \frac{\left| \int_{-W}^{+W} G(-x) f(x)\, dx \right|}{n_0 \sqrt{\int_{-W}^{+W} f^2(x)\, dx}} \tag{3}$$

For the location, we need to be able to have some measure that increases as the localization improves. So, we use the reciprocal of the root-mean-squared distance of the marked edge from the centre of the true edge. Recalling that we decided to mark the edges at the local maxima of the function $f(x)$, the first derivative of the response will be zero at these points. Also, since the edges are centered at $x = 0$, in the absence of noise, there should be a local maxima in the response at $x = 0$.

Let $H_n(x)$ be the response to the noise, and $H_G(x)$ be the response to the edge. Suppose there is a local maximum in the total response at $x = x_0$. Then we have

$$H_n'(x_0) + H_G'(x_0) = 0 \tag{4}$$

The Taylor expansion of $H_G'(x_0)$ about the origin gives

$$H_G'(x_0) = H_G'(0) + H_G''(0)x_0 + O(x_0^2) \tag{5}$$

By assumption, $H_G'(0)$, i.e., the response of the filter in the absence of noise has a local maximum at the origin, so the first term in the expansion may be ignored. The displacement x_0 of the actual

maximum is assumed to be small, so we will ignore quadratic and higher terms. In fact by a simple argument we can show that if the edge $G(x)$ is either symmetric or antisymmetric, all even terms in x_0 vanish. Suppose $G(x)$ is antisymmetric, and express $f(x)$ as a sum of a symmetric component and an antisymmetric component. The convolution of the symmetric component with $G(x)$ contributes nothing to the numerator of the SNR, but it does contribute to the noise component in the denominator. Therefore,if $f(x)$ has any symmetric component, its SNR will be worse than a purely antisymmetric filter. A dual argument holds for symmetric edges, so that if the edge $G(x)$ is symmetric or antisymmetric, the filter $f(x)$ will follow suit. The net result of this is that the response $H_G(x)$ is always symmetric, and that its derivatives of odd orders [which appear in the coefficients of even order in (5)] are zero at the origin. Equations (4) and (5) give

$$H_G(0)'' x_0 \approx -H_n'(x_0) \tag{6}$$

Now, $H_n'(x_0)$ is a Gaussian random quantity whose variance is the mean-squared value of $H_n'(x_0)$, and is given by

$$E[H_n'(x_0)^2] = n_0^2 \left[\int_{-W}^{+W} f'^2(x) \, dx \right] \tag{7}$$

where E[y] is the expectation value of y. Combining this result with (6) and substituting for $H_G''(0)$ gives

$$E[x_0^2] \approx \frac{n_0^2 \left[\int_{-W}^{+W} f'^2(x) \, dx \right]}{\left[\int_{-W}^{+W} G'(-x) f'(x) \, dx \right]^2} = \delta x_0^2 \tag{8}$$

where δx_0 is an approximation to the standard deviation x_0. The localization is defined as the reciprocal of δx_0.

$$Localization = \frac{\left| \int_{-W}^{+W} G'(-x) f'(x) \, dx \right|}{n_0 \sqrt{\int_{-W}^{+W} f'^2(x) \, dx}} \tag{9}$$

Equations (3) and (9) are mathematical forms for the first two criteria, and the design problem reduces to the maximization of both of these simultaneously. In order to do this, we maximize the product of (3) and (9). We could conceivably have combined (3) and (9) using any function that is monotonic in two arguments, but the use of the product simplifies the analysis for step edges. At the moment, we will try to maximize the product of the criteria for arbitrary edges,

$$\frac{\left| \int_{-W}^{+W} G(-x) f(x) \, dx \right| \; \left| \int_{-W}^{+W} G'(-x) f'(x) \, dx \right|}{n_0 \sqrt{\int_{-W}^{+W} f^2(x) \, dx} \; \; n_0 \sqrt{\int_{-W}^{+W} f'^2(x) \, dx}} \tag{10}$$

B Now, we move on to eliminate the multiple responses. Earlier, during the specification of the edge detection problem, we decided that edges would be marked at local maxima in the response of a linear filter applied to the image. From the last section, we got that the detection criterion measures the effectiveness of the filter in discriminating between signal and noise at the center of an edge. It did not take into account the behavior of the filter *nearby* the edge center. The first two criteria can be trivially maximized as follows. From the Schwarz inequality, we can show that SNR (3) is bounded above by

$$n_0^{-1} \sqrt{\int_{-W}^{+W} G^2(x) \, dx}$$

and localization (9) by

$$n_0^{-1} \sqrt{\int_{-W}^{+W} G'^2(x) \, dx}$$

Both bounds are attained, and the product of SNR and localization is maximized when $f(x) = G(-x)$ in $[-W, +W]$.

These maxima are so close together that it is not possible to select one as the response to the step while identifying the others as noise. We need to add to our criteria the requirement that the function f will not have "too many" responses to a single step edge in the vicinity of the step. We need to limit the number of peaks in the response so that there will be a low probability of declaring more than one edge. Ideally, we would like to make the distance between peaks in the noise response approximate the width of the response of the operator to a single step. This width will be some fraction of the operator width W. Then, after compiling results from S. O. Rice's *"Mathematical Analysis of Random Noise"* that show that the average distance between zero-crossings of a function g to Gaussian noise is

$$x_{ave} = \pi \left(\frac{-R(0)}{R''(0)} \right)^{1/2} \tag{11}$$

where R(τ) is the autocorrelation function of g. In our case, we are looking for the mean zero-crossing spacing for the function f'. Now since,

$$R(0) = \int_{-\infty}^{+\infty} g^2(x)\, dx \tag{12}$$

and

$$R''(0) = - \int_{-\infty}^{+\infty} g'^2(x)\, dx \tag{13}$$

the mean distance between zero-crossings of f' will be

$$x_{zc}(f) = \pi \left(\frac{\int_{-\infty}^{+\infty} f'^2(x)\, dx}{\int_{-\infty}^{+\infty} f''^2(x)\, dx} \right)^{1/2} \tag{14}$$

The distance between adjacent maxima in the noise response of f, denoted by x_{max}, will be twice x_{zc}. We set this distance to be some fraction k of the operator width.

$$x_{max}(f) = 2x_{zc}(f) = kW \tag{15}$$

This is a natural form for the constraint since the response of the filter will be concentrated in a region of width $2W$, and the expected number of noise maxima is this region N_n where

$$N_n = \frac{2W}{x_{max}} = \frac{2}{k} \tag{16}$$

Fixing k fixes the number of noise maxima that could lead to a false response.

We remark here that the inter-maximum spacing (14) scales with the operator width. That is, we first define an operator f_w which is the result of stretching f by a factor of w, $f_w(x) = f(\frac{x}{w})$. Then after substituting into (14) we find that the inter-maximum spacing for f_w is $x_{zc}(f_w) = w x_{zc}(f)$. Therefore, if some function f satisfies the maximum response constraint (15) for fixed k, then the function f_w will also satisfy it, assuming W scales with w. For any fixed k, the multiple response criterion is invariant with respect to spatial scaling of f.

C Conclusion: So, the above discussed steps are how the Canny Edge Detection Algorithm is developed. It is adaptable to various environments, and its parameters allow it to be tailored to recognition of edges of differing characteristics depending on the particular requirements of a given implementation.

3. Sobel Edge Detection

The Sobel operator is used in image processing, particularly within edge detection algorithms. Technically, it is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. At each point in the image, the result of the Sobel operator is either the corresponding gradient vector or the norm of this vector. The Sobel operator is based on convolving the image with a small, separable, and integer valued filter in horizontal and vertical direction and is therefore relatively inexpensive in terms of computations. On the other hand, the gradient approximation that it produces is relatively crude, in particular for high frequency variations in the image.

(a) **Sobel Filter Design** Most edge detection methods work on the assumption that the edge occurs where there is a discontinuity

in the intensity function or a very steep intensity gradient in the image. Using this assumption, if we take the derivative of the intensity value across the image and find points where the derivative is maximum, then the edge could theoretically be located. The gradient is a vector, whose components measure how rapid pixel values are changing with distance in the x and y directions. Thus, the components of the gradient may be found using the following approximation:

$$\frac{\partial f(x, y)}{\partial x} = \Delta x = \frac{f(x + dx, y) - f(x, y)}{dx} \tag{17}$$

$$\frac{\partial f(x, y)}{\partial y} = \Delta x = \frac{f(x, y + dy) - f(x, y)}{dy} \tag{18}$$

where dx and dy measure distance along the x and y directions respectively. In discrete images, one can consider dx and dy in terms of numbers of pixel between two points. $dx = dy = 1$(pixel spacing) is the point at which pixel coordinates are (i, j) thus,

$$\Delta x = f(i + 1, j) - f(i, j) \tag{19}$$

$$\Delta y = f(i, j + 1) - f(i, j) \tag{20}$$

In order to detect the presence of a gradient discontinuity, one could calculate the change in the gradient at (i, j).This can be done by finding the following magnitude measure:

$$M = \sqrt{(\Delta x)^2 + (\Delta y)^2} \tag{21}$$

and the direction θ is given by

$$\theta = arctan\left(\frac{\Delta y}{\Delta x}\right) \tag{22}$$

The Sobel operator is an example of the gradient method of filter design. The Sobel operator is a discrete differentiation operator, computing an approximation of the gradient of the image intensity function. The different operators in eq. (19) and (20) correspond to convolving the image with the following masks:

$$\Delta x = \begin{bmatrix} -1 & 1 \\ 0 & 0 \end{bmatrix}, \Delta y = \begin{bmatrix} -1 & 0 \\ 1 & 0 \end{bmatrix}$$

After this is done:

   i The top left-hand corner of the appropriate mask is super-
imposed over each pixel of the image.

   ii A value is calculated for Δx or Δy by using the mask coef-
ficients in a weighted sum of the value of pixels (i, j) and its
neighbors.

   iii These masks are referred to as convolution masks or some-
times convolution kernels. Instead of finding approximate gra-
dient components along the x and y directions, approximation
of the gradient components could be done along directions at
$45°$ and $135°$ to the axes respectively.

Now, the masks discussed above are somewhat small and can be
very inconvenient to use. Also, an advantage of using a larger
mask size is that the errors due to the effects of noise are re-
duced by local averaging within the neighborhood of the mask.
An advantage of using a mask of odd size is that the operators are
centered and can therefore provide an estimate that is based on a
center pixel (i, j). The Sobel edge operator masks are given as:

$$\Delta x = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \Delta y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}$$

The operator calculates the gradient of the image intensity at each
point, giving the direction of the largest possible increase from
light to dark and the rate of change in that direction. The result
therefore shows how "abruptly" or "smoothly" the image changes
at that point and therefore how likely it is that part of the image
represents an edge, as well as how that the edge is likely to be
oriented. In practice, the magnitude (likelihood of an edge) cal-
culation is more reliable and easier to interpret than the direction
calculation. Mathematically, the gradient of a two-variable func-
tion (the image intensity function) at each image point is a 2D
vector with the components given by the derivatives in the hori-
zontal and vertical directions. At each image point, the gradient
vector points to the direction of largest possible intensity increase,
and the length of the gradient vector corresponds to the rate of

change in that direction. This implies that the result of the Sobel operator at any image point which is in a region of constant image intensity is a zero vector and at a point on an edge is a vector which points across the edge, from darker to brighter values.

4. Conclusion

Edge detection helps in optimizing network bandwidth and it is needed to keep track of data flowing in and out of the network. It helps to extract useful features for pattern recognition. Although the Sobel operator is slower to compute, its larger convolution kernel smoothes the input image to a greater extent and so makes the operator less sensitive to noise. The larger the width of the mask, the lower its sensitivity to noise and the operator also produces considerably higher output values for similar edges. Sobel operator effectively highlights noise found in real world pictures as edges though, the detected edges could be thick. The Canny edge detector and similar algorithm solve these problems by first blurring the image slightly then applying an algorithm that effectively thins the edges to one pixel.

Transferring a 2-D pixel array into a statistically uncorrelated data set enhances the removal of redundant data, which leads to the reduction of the amount of data required to represent a digital image. Considering data communication these days, especially the internet, massive data transfer causes serious problems for interactive network users and techniques such as these go a long way to enable faster data transfer and solve, to a certain extent, the memory consumption problem.

References

[1] John F. Canny, *A Computational Approach to Edge Detection.* IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. PAMI - 8, No. 6, November, 1986.

[2] John F. Canny, *Finding Edges and Lines in Images.* M.I.T. Artificial Intelligence Lab., Cambridge, Massachusetts, Rep. Al-TR-720, 1983.

[3] R. W. Hamming, *Digital Filters.* Englewood Cliffs, NJ. Prentice Hall, 1983.

[4] Leslie Lamport, *LaTeX: A Document Preparation System.* Addison Wesley, Massachusetts, 2nd Edition, 1994.

[5] U. Michael, *Mathematics properties of the JPEG 2000 wavelet filters.* IEEE Transactions on Image Processing, 12(9), 080-1090.

[6] L. Yasri, N. H. Hamid, *Performance Analysis of FPGA Based Sobel Edge Detection Operator.* International Conference on Electronic Design, December 01-03, 2008.