The Digital Edge

Robert M Haralick

Virginia Polytechnic Institute and State University
Departments of Electrical Engineering and Computer Science
Blacksburg Virginia 24061

I. Introduction

What is an edge in a digital image? The first intuitive notion is that a digital edge occurs on the boundary between two pixels when the respective brightness values of the two pixels are significantly different. Significantly different may depend upon the distribution of brightness values around each of the pixels.

We often point to a region on an image and say this region is brighter than its surrounding area, meaning that the mean of the brightness values of pixels inside the region is brighter than the mean of the brightness values outside the region. Having noticed this we would then say that an edge exists between each pair of neighboring pixels where one pixel is inside the region and the other is outside the region. Such edges are referred to as step edges.

Step edges are not the only kind of edge. If we scan through a region in a left right manner observing the brightness values steadily increasing and then after a certain point observe that the brightness values are steadily decreasing we are likely to say that there is an edge at the point of change from increasing to decreasing brightness values. Such edges are called roof edges.

It is, therefore, clear from our use of the word edge that edges refer to places in the image where there appears to be a jump in brightness value or a jump in brightness value derivative.

In some sense this summary statement about edges is quite revealing since in a discrete array of brightness values there are jumps in the literal sense, between neighboring brightness values if the brightness values are different, even if only slightly different. Perhaps more to the heart of the matter, there exists no definition of derivative for a discrete array of brightness values. The only way to interpret jumps in value and jumps in derivatives when referring to a discrete array of values is to assume that the discrete array of values comes about as some kind of sampling of a real-valued function of defined on a bounded and connected subset of the real plane R^2. The jumps in value and jumps in derivative really must refer to points of discontinuity of f and to points of discontinuity in the partial derivatives of f.

Edge finders should then regard the digital picture function as a sampling of the underlying function f, where some kind of random noise has been added to the true function values. To do this, the edge finder must assume some kind of parametric form for the underlying function f, use the sampled brightness values of the digital picture function to estimate the parameters, and finally make decisions regarding the locations of discontinuities and the locations of discontinuities of partial derivatives based on the estimated values of the parameters.

Of course, it is impossible to determine the true locations of discontinuities in value or derivatives based upon a sampling of the functions. The locations are estimated by function approximation. Sharp discontinuities will reveal themselves in high values for estimates of first partial derivatives. Sharp discontinuities in derivative will reveal themselves in high values for estimates of the second partial derivatives. This means that the best we can do is to assume that the first and second partial derivatives of any possible underlying image function has known bounds. Therefore, any estimated first or second order partials which exceed these known bounds must be due to discontinuities in value or in derivative of the underlying function.

In this paper, we assume that in each neighborhood not having discontinuities in value or derivatives the underlying function f takes the parametric form of a polynomial in the row and column coordinates and that the sampling producing the digital picture function is a regular equal interval grid sampling of the square plane which is the domain of f.

CH1595-8/81/0000/0285$00.75 © 1981 IEEE

These underlying functions are easy to represent as linear combinations of the polynomials in any polynomial basis set. That polynomial basis set which permits the independent estimation of each coefficient would be the easiest to use. Such a polynomial basis set is the discrete orthogonal polynomial basis set.

In section II.1 we discuss the one dimensional family of discrete orthogonal polynomials. In section II.2 we discuss how arbitrary two dimensional polynomials can be computed as linear combinations of the tensor product of one dimensional discrete orthogonal polynomials. In section II.3, we discuss how the discretely sampled data values are used to estimate the coefficients of the linear combinations: coefficient estimates for exactly fitting or estimates for least square fitting are calculated as linear combinations of the sampled data values.

Having used the pixel values in a neighborhood to estimate the underlying polynomial function we can now determine the value of the partial derivatives at any location in the neighborhood and use those values in edge finding. Having to deal with partials in both the row and column directions makes using these derivatives a little more complicated than using the simple derivatives of one dimensional functions. Section III discusses how a direction isotropic magnitude of the first and second derivative can be defined in terms of the respective first and second order partials. The assumption that the underlying function f has bounded derivatives can now be precisely stated: the direction isotropic magnitudes of the first and second order derivatives at locations inside any regions are smaller than some specified number. Any neighborhood, therefore, which produces an estimated function whose direction isotropic magnitude of first and second order derivatives exceed these bounds has an edge in the neighborhood at those locations where the bounds are exceeded.

Thus, the question of finding an edge reduces to the question of whether at any location the direction isotropic magnitudes of the first and second order derivatives of the estimated underlying function exceed certain bounds. This question is not as easy to answer as it might seem on the surface. The pixel values are random variables because of noise. This randomness propagates itself to the calculated estimates of the coefficients of the linear combinations of the polynominals taken from the discrete orthogonal polynomial set and finally to the direction isotropic derivative magnitudes which are based on these coefficients. A future paper will discuss the statistical discrimination problem.

II Discrete Orthogonal Polynomials

II.1 Discrete Orthogonal Polynomial Construction Technique

Let the index set R be symmetric in sense that $r \varepsilon R$ implies $-r \varepsilon R$. Let $P_n(r)$ be the n^{th} order polynomial. We define the construction technique for discrete orthogonal polynomials iteratively.

Define $P_0(r) = 1$.

Suppose $P_0(r), \ldots, P_{n-1}(r)$ have been defined. In general, $P_n(r) = r^n + a_{n-1}r^{n-1} + \ldots + a_1 r + a_0$. $P_n(r)$ must be orthogonal to each polynomial $P_0(r), \ldots, P_{n-1}(r)$. Hence, we must have the n equations

$$\sum_{r \varepsilon R} P_k(r)(r^n + a_{n-1}r^{n-1} + \ldots + a_1 r + a_0) = 0, \quad k=0,\ldots,n-1$$

These equations can be solved for the n coefficients $a_0, \ldots, a_{n-1}$.

The first five polynomial functions formulas are

$$P_0(r) = 1$$

$$P_1(r) = r$$

$$P_2(r) = r^2 - \mu_2/\mu_0$$

$$P_3(r) = r^3 - (\mu_4/\mu_2)r$$

$$P_4(r) = \frac{r^4 + (\mu_2\mu_4-\mu_6)r^2 + (\mu_2\mu_6 - \mu_4^2)}{\mu_0\mu_4 - \mu_2}$$

where

$$\mu_k = \sum_{s \varepsilon R} s^k$$

II.2 Two Dimensional Discrete Orthogonal Polynomials

Two dimensional discrete orthogonal polynomials can be created from two sets of one dimensional discrete orthogonal polynomials by taking tensor products. Let R and C be index sets satisfying the symmetry condition $r \varepsilon R$ implies $-r \varepsilon R$ and $c \varepsilon C$ implies $-c \varepsilon C$. Let $\{P_0(r),\ldots,P_N(r)\}$ be a set of discrete polynomials on R. Let $\{Q_0(c),\ldots,Q_M(c)\}$ be a set of discrete polynomials on C. Then the set $\{P_0(r)Q_0(c),\ldots,P_n(r)Q_m(c),\ldots,P_N(r)Q_M(c)\}$ is a set of discrete polynomials on RXC.

The proof of this fact is easy. Consider whether $P_i(r)Q_j(c)$ is orthogonal to $P_n(r)Q_m(c)$. when $n \neq i$ or $m \neq j$. Then

$$\sum_{r \varepsilon R} \sum_{c \varepsilon C} P_i(r)Q_j(c)P_n(r)Q_m(c)$$

$$= \sum_{r \varepsilon R} P_i(r)P_n(r) \sum_{c \varepsilon C} Q_j(c)Q_m(c).$$

Since $n \neq i$ or $m \neq j$ one or other of the sums must be zero.
Examples

Index Set	Discrete Orthogonal Polynomial Set
$\{-1/2,\ 1/2\}$	$\{1,\ r\}$
$\{-1,\ 0,\ 1\}$	$\{1,\ r,\ r^2 - 2/3\}$
$\{-2/3,\ -1/2,\ 1/2,\ 3/2\}$	$\{1,\ r,\ r^2 - 5/4,\ r^3 - 41/20r\}$
$\{-2,\ -1,\ 0,\ 1,\ 2\}$	$\{1,\ r,\ r^2 - 2,\ r^3 - 17/5,\ r^4 + 3r^2 + 72/35\}$
$\{-1,0,1\} \times \{-1,0,1\}$	$\{1,\ r,\ c,\ r^2 - 2/3,\ rc,\ c^2 - 2/3,\ r(c^2 - 2/3),\ c(r^2 - 2/3),\ (r^2 - 2/3)(c^2 - 2/3)\}$

II.3 Fitting Data With Discrete Orthogonal Polynomials

Let an index set R with the symmetry property $r \varepsilon R$ implies $-r \varepsilon R$ be given. Let the number of elements in R be N. Using the construction technique, we may construct the set $\{P_0(r),\ldots,P_{N-1}(r)\}$ of discrete orthogonal polynomials over R.

For each $r \varepsilon R$, let a data value $d(r)$ be observed. The exact fitting problem is to determine coefficients $a_0,\ldots,a_{N-1}$ such that

$$d(r) = \sum_{n=0}^{N-1} a_n P_n(r)$$

The orthogonality property makes the determination of the coefficients particularly easy. To find the value of some coefficient, say a_m, multiply both sides of the equation by $P_m(r)$ and then the sum over all $r \varepsilon R$.

$$\sum_{r \varepsilon R} P_m (r)d(r) = \sum_{n=0}^{N-1} a_n \sum_{r \varepsilon R} P_n$$

Hence,

$$a_m = \sum_{r \varepsilon R} P_m(r)d(r) / \sum_{r \varepsilon R} P_m^2(r)$$

The approximate fitting problem is to determine coefficients $a_0,\ldots,a_K$, $K \leq N-1$ such that

$$e^2 = \sum_{r \varepsilon R} [d(r) - \sum_{n=0}^{K} a_n P_n(r)]^2$$

is minimized. To find the value of some coefficient, say a_m, take the partial derivative of both sides of the equation for e^2 with respect to a_m. Set it to zero and use the orthogonality property to find that again

$$a_m = \sum_{r \varepsilon R} P_m(r)d(r) / \sum_{r \varepsilon R} P_m^2(r)$$

The exact fitting coefficients and the least squares coefficients are identical for $m = 0,\ldots,K$.

Fitting the data values $\{d(r) | r \varepsilon R\}$ to the polynomial

$$Q(r) = \sum_{n=0}^{K} a_n P_n(r)$$

now permits us to interpret $Q(r)$ as a well behaved real-valued function defined on the real line. To determine

$$\frac{dQ}{dr}(r_0)$$

we need only to evaluate

$$\sum_{n=0}^{N} a_n \frac{dP_n}{dr}(r_0)$$

In this manner, any derivative at any point may be obtained. Simarily for any definite integrals.

III. Direction Isotropic Derivatives Magnitudes

In this section we seek to determine those linear combinations of squared partial derivatives of two dimensional functions which are invariant under rotation of the domain of the two dimensional function. To illustrate what we mean, let us first consider the simple bilinear function $f(r',c') = k_1 + k_2 r' + k_3 c'$. If we rotate the coordinate system by θ, and call the resulting function g we have in the row (r',c') coordinates

$$r' = r \cos\theta + c \sin\theta$$

$$c' = -r \sin\theta + c \cos\theta$$

and $g(r,c) = k_1 + k_2(r \cos\theta + c \sin\theta) + k_3(-r \sin\theta + c \cos\theta)$

$= k_1 + (k_2 \cos\theta - k_3 \sin\theta)r + (k_2 \sin\theta + k_3 \cos\theta)c$

Now, $[\frac{\partial f}{\partial r'}(r',c')]^2 + [\frac{\partial f}{\partial c'}(r',c')]^2 = k_2^2 + k_3^2$

and $[\frac{\partial g}{\partial r}(r',c')]^2 + [\frac{\partial g}{\partial c}(r',c')]^2 = (k_2\cos\theta - k_3\sin\theta)^2 + (k_2\sin\theta + k_3\cos\theta)^2$

$= k_2^2 + k_3^2$

Hence the sum of the squares of the first partials is the same constant $k_2^2 + k_3^2$ for the original function or for the rotated function. This direction isotropic derivative magnitude is immediately recognized as the squared gradient. In the remainder of the section we explicitly develop the direction isotropic derivative magnitude for the first, and the second derivatives of an arbitrary function. Then we state the theorem giving the formula for the isotropic derivative magnitude of any order.

Proceeding with the first order case, upon writing the rotation equation with r' and c' as the independent variables we have

$$r = r'\cos\theta - c'\sin\theta$$
$$c = + r'\sin\theta + c'\cos\theta$$

Let the rotated functions be g. Then $g(r,c) = f(r',c')$.

Now expressing $\frac{\partial f}{\partial r'}(r',c')$ and $\frac{\partial f}{\partial c'}(r',c')$ in terms of

$\frac{\partial f}{\partial r}(r',c')$ and $\frac{\partial f}{\partial c}(r',c')$ we have

$$\frac{\partial f}{\partial r'} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial r'} + \frac{\partial f}{\partial c}\frac{\partial c}{\partial r'} = \frac{\partial f}{\partial r}\cos\theta + \frac{\partial f}{\partial c}\sin\theta$$

$$\frac{\partial f}{\partial c'} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial c'} + \frac{\partial f}{\partial c}\frac{\partial c}{\partial c'} = \frac{\partial f}{\partial r}\sin\theta + \frac{\partial f}{\partial c}\cos\theta$$

Then,

$$\left(\frac{\partial f}{\partial r'}(r',c')\right)^2 + \left(\frac{\partial f}{\partial c'}(r',c')\right)^2$$

$$= \left(\frac{\partial' f}{\partial r}(r',c')\cos\theta + \frac{\partial f}{\partial c}(r',c')\sin\theta\right)^2 + \left(\frac{-\partial f}{\partial r}(r',c')\sin\theta + \frac{\partial f}{\partial c}(r',c')\cos\theta\right)^2$$

$$= \left(\frac{\partial f}{\partial r}(r',c')\right)^2 (\cos^2\theta + \sin^2\theta) + \left(\frac{\partial f}{\partial r}(r',c')\frac{\partial f}{\partial c}(r',c')(\cos\theta\sin\theta - \cos\theta\sin\theta)\right)$$
$$+ \left(\frac{\partial f}{\partial c}(r',c')\right)^2 (\sin^2\theta + \cos\theta^2\theta)$$
$$= \left(\frac{\partial f}{\partial r}(r',c')\right)^2 + \left(\frac{\partial f}{\partial c}(r',c')\right)^2 = \left(\frac{\partial g}{\partial r}(r',c')\right)^2 + \left(\frac{\partial g}{\partial c}(r',c')\right)^2$$

Thus for each point in the unrotated coordinate system

$$\left(\frac{\partial f}{\partial r'}(r',c')\right)^2 + \left(\frac{\partial f}{\partial c'}(r',c')\right)^2$$

produces the same value as

$$\left(\frac{\partial f}{\partial r}(r',c')\right)^2 + \left(\frac{\partial g}{\partial c}(r',c')\right)^2$$

in the rotated coordinate systems, where $g(r,c) = f(r',c')$.

Proceeding in a similar manner for the second order partials we have

$$\frac{\partial^2 f}{\partial r'^2} = \frac{\partial^2 f}{\partial r^2}\cos^2\theta + 2\frac{\partial^2 f}{\partial r \partial c}\cos\theta\sin\theta + \frac{\partial^2 f}{\partial' c^2}\sin^2\theta$$

$$\frac{\partial^2 f}{\partial r'\partial c'} = \frac{-\partial^2 f}{\partial r'}\cos\theta\sin\theta + \frac{\partial^2 f}{\partial r \partial c}(\cos^2\theta\ \sin^2\theta) + \frac{\partial^2 f}{\partial' c^2}\cos\theta\sin\theta$$

$$\frac{\partial^2 f}{\partial c'^2} = \frac{\partial^2 f}{\partial r}\sin^2\theta - 2\frac{\partial^2 f}{\partial r \partial c}\cos\theta\sin\theta + \frac{\partial^2 f}{\partial c^2}\cos^2\theta$$

Looking for some constant λ which makes

$$\left(\frac{\partial^2 f}{\partial r'^2}\right)^2 + \lambda\left(\frac{\partial^2 f}{\partial r'\partial c'}\right)^2 + \left(\frac{\partial'^2}{\partial r'^2}\right)^2$$

$$= \left(\frac{\partial^2 f}{\partial r^2}\right)^2 + \lambda\left(\frac{\partial^2 f}{\partial r d c}\right)^2 + \left(\frac{\partial^2 f}{\partial c^2}\right)^2$$

we discover that there does exist exactly one and its value is $\lambda = 2$. Thus for each point in the unrotated coordinate system,

$$\left(\frac{\partial^2 f}{\partial r'^2}(r;c)\right)^2 + 2\left(\frac{\partial^2 f}{\partial r'\partial c'}(r;c')\right)^2 + \left(\frac{\partial f^2 f}{\partial c'^2}(r;c')\right)^2$$

produces the same value as

$$\left(\frac{\partial^2 f}{\partial r^2}(r;c')\right)^2 + 2\left(\frac{\partial^2 f}{\partial r\partial c}(r;c')\right)^2 + \left(\frac{\partial'^2 f}{\partial c^2}(r;c')\right)^2$$

The direction isotropic second derivative magnitude is, therefore,

$$\left(\frac{\partial^2 f}{\partial r^2}\right)^2 + 2\left(\frac{\partial^2 f}{\partial r\partial c}\right)^2 + \left(\frac{\partial^2 f}{\partial c^2}\right)^2$$

Higher order direction isotropic derivative magnitudes can be constructed in a similar manner. The coefficients of the squared partials continue in a binomial coefficient pattern for the dimensional case which we have been discussing. To see this consider the following theorem which states that the sum of the squares of all the partials of a given total order is equal, regardless of which orthogonal coordinate system it is taken in. Specializing this to two-dimensions and third order partials, the direction

$$\left(\frac{\partial^3 f}{\partial c\partial r\partial r}\right)^2 + \left(\frac{\partial^3 f}{\partial c\partial r\partial c}\right)^2 + \left(\frac{\partial^3}{\partial c\partial c\partial r}\right)^2 + \left(\frac{\partial^3}{\partial r\partial c\partial c}\right)^2$$

$$\left(\frac{\partial^3 f}{\partial c\partial r\partial r}\right)^2 + \left(\frac{\partial^3 f}{\partial c\partial r\partial c}\right)^2 + \left(\frac{\partial^3 f}{\partial c\partial c\partial r}\right)^2 + \left(\frac{\partial^3 f}{\partial r\partial c\partial c}\right)^2$$

$$= \left(\frac{\partial^3 f}{\partial r^3}\right)^2 + 3\left(\frac{\partial^3 f}{\partial r^2\partial c}\right)^2 + 3\left(\frac{\partial^3 f}{\partial r\partial c}\right)^2 + \left(\frac{\partial^3 f}{\partial c^3}\right)^2$$

It should be clear from this example that the binomial coefficients arise because of the commutivity of the partial differential operators.