A review on edge detection based on filtering and differentiation

Armando J. Pinho and Luís B. Almeida

Abstract – **In this paper we address the issue of edge detection in digital images using linear filtering and differentiation. Our main objective was to join, and in some sense unify, a significant part of the techniques that fall into this paradigm. First, we focus on the question of two-dimensional differentiation and the problems associated with its application in the discrete spatial domain. Afterwards, we review some of the most commonly used filtering techniques for edge detection.**

Resumo – **Ao longo deste artigo abordamos a questão da detecção de contornos em imagens digitais, utilizando técnicas baseadas em filtragem e diferenciação. O nosso principal objectivo foi reunir, e de certa forma unificar, uma parte que julgamos ser significativa dos métodos de detecção de contornos que se enquadram neste paradigma. Em primeiro lugar, começamos por abordar algumas questões relacionadas com a diferenciação a duas dimensões e também com os problemas que advêm da sua aplicação no domínio discreto. Em seguida, proporcionamos uma panorâmica de algumas das técnicas baseadas em filtragem mais usadas na detecção de contornos em imagens.**

Keywords – **Edge detection; Image filtering; Image differentiation.**

I. INTRODUCTION

An important part of the edge detectors proposed during the last three decades can be viewed as being composed, essentially, of two operations: low-pass filtering and differentiation. Right from the beginning of the research on edge detection that differentiation played a fundamental role. In fact, most often, edges are associated with fast variations of the spatial distribution of light intensity and, therefore, are related to the derivatives of the light intensity function. Although the main concept is theoretically simple and intuitive, its application in practice poses several problems that have been motivating a large number of publications focusing on their solutions.

The edge detectors based on differentiation can be viewed as a system built from a first low-pass filtering block, followed by a differentiation block and, finally, by a detector. Frequently, the filtering and differentiation blocks are seen as a single operation, for example implemented using an odd filter. Moreover, some of the oldest edge detectors do not use the filtering stage at all.

We address, mainly, two of the problems associated with this approach. First, we focus on the question of two-

A. Pinho is with INESC / Dep. de Electrónica e Telecomunicações, Universidade de Aveiro, 3810 Aveiro, Portugal, Fax +351–34–370545, Email `ap@inesca.pt`.

L. Almeida is with the Inst. Superior Técnico / INESC, R. Alves Redol, 9, 1000 Lisboa, Portugal, Fax +351–1–525843, Email `lba@inesc.pt`.

dimensional differentiation and the problems related with its application in the discrete spatial domain. Next we review some of the most commonly used filtering techniques for edge detection.

The aim of this paper is, basically, two-fold. First, it describes most of the edge detectors based on a filtering and differentiation paradigm. In this sense, it is a review paper. Second, it aims to provide a way of comparison among some of the most recent techniques, giving, simultaneously, a unified view of them.

II. DIFFERENTIATION OF IMAGES

The edge detection methods based on differentiation use, mainly, first and second order derivatives of the light intensity function. For example, to detect (isolated) step edges we can look for maxima of the absolute value of the first derivative or zero crossings of the second derivative. Since images are two-dimensional functions, we start by presenting some basic results on differentiation in two dimensions.

Let ϱ be an analog image, i.e., ϱ is a function defined in $\mathbb{R}^2 \xrightarrow{\varrho} \mathbb{R}_0^+$. The first order derivative of ϱ can be calculated along some direction, $\mathbf{r}$, using the partial derivatives of ϱ along the main axes,

$$\varrho_x \stackrel{\text{def}}{=} \frac{\partial \varrho}{\partial x}, \quad \varrho_y \stackrel{\text{def}}{=} \frac{\partial \varrho}{\partial y} \tag{1}$$

in the following way:

$$\frac{\partial \varrho}{\partial r} = \frac{\partial \varrho}{\partial x}\frac{\partial x}{\partial r} + \frac{\partial \varrho}{\partial y}\frac{\partial y}{\partial r} = \varrho_x \cos\phi + \varrho_y \sin\phi \tag{2}$$

where ϕ is the angle formed between $\mathbf{r}$ and the x axis, and r is a parameter along the direction defined by vector $\mathbf{r}$. The gradient of ϱ, ∇g, is, by definition, a vector with the same direction as the maximum directional derivative, i.e., for which

$$\frac{\partial}{\partial \phi}\left(\frac{\partial \varrho}{\partial r}\right) = 0 \tag{3}$$

resulting that

$$\phi_\nabla = \arctan\left(\frac{\varrho_y}{\varrho_x}\right) \tag{4}$$

where ϕ_∇ is the angle formed between the gradient vector and the x axis. Therefore, the gradient is given by

$$\nabla \varrho = \varrho_x \vec{x} + \varrho_y \vec{y} = |\nabla \varrho|\, \vec{n} \tag{5}$$

where $\vec{x}$ and $\vec{y}$ are unitary vectors along the respective axes, $\vec{n} = \vec{x}\cos(\phi_\nabla) + \vec{y}\sin(\phi_\nabla)$ is a unitary vector along the gradient direction, and $|\nabla \varrho|$ is the magnitude of the gradient (value of the derivative along that direction) given by

$$|\nabla \varrho| = \sqrt{\varrho_x^2 + \varrho_y^2} \tag{6}$$

According to these definitions, we can say that edge points may be located by the maxima of the module of the gradient, and that the direction of the contour is orthogonal to the direction of the gradient. It is not difficult to see that operators based on the gradient are directional, since they give their maximum response when they are aligned with the orthogonal direction of the contour.

Edge detection based on second order derivatives is frequently performed using one of two operators: the second derivative along the direction of the gradient or the Laplacian. Let us assume that n is a parameter along the direction of $\mathbf{n}$ (direction of the gradient vector). The second derivative of ϱ along the gradient direction, $\partial^2/\partial n^2$, is related with the derivatives along the axes x and y in the following way [1]:

$$\frac{\partial^2 \varrho}{\partial n^2} = \frac{\varrho_x^2 \varrho_{xx} + 2\varrho_x \varrho_y \varrho_{xy} + \varrho_y^2 \varrho_{yy}}{\varrho_x^2 + \varrho_y^2} \qquad (7)$$

where

$$\varrho_{xx} \stackrel{\text{def}}{=} \frac{\partial^2 \varrho}{\partial x^2}, \quad \varrho_{yy} \stackrel{\text{def}}{=} \frac{\partial^2 \varrho}{\partial y^2}, \quad \varrho_{xy} \stackrel{\text{def}}{=} \frac{\partial^2 \varrho}{\partial x \partial y} \qquad (8)$$

The Laplacian of ϱ is given by

$$\nabla^2 \varrho \stackrel{\text{def}}{=} \varrho_{xx} + \varrho_{yy} \qquad (9)$$

and is frequently considered a good approximation to the second derivative along the gradient direction. In fact, these two operators are related by [2]

$$\nabla^2 \varrho = \frac{\partial^2 \varrho}{\partial n^2} + |\nabla \varrho| \kappa \qquad (10)$$

where κ is the curvature of the line of constant intensity that crosses the point under consideration. Using this relation, it is easy to see that, in fact, the Laplacian is a good approximation to the second derivative along the gradient direction, providing that the curvature of the line of constant intensity that crosses the point under consideration is small. Moreover, we can immediately state that the Laplacian is useless in the detection of corners (zones of high curvature).

We can point out, at least, three major advantages of using the Laplacian in relation to the second derivative along the gradient direction. First, it is simple to use, since it only requires the computation of two second order derivatives. Second, it is a linear operator, in opposite to $\partial^2/\partial n^2$, which is non-linear. Finally, but not less important, the Laplacian is a non-directional operator. This characteristic avoids the necessity to determine the most appropriated direction to apply the operator (note that this is required by the $\partial^2/\partial n^2$ operator).

At least two problems arise when we try to apply the differentiation concepts mentioned above to digital images. One, is related to discretization. Since digital images are represented by sets of quantified samples, we need to determine discrete approximations of the differential operators. The other problem is the known amplification of high frequency noise generated by the differentiation operation. In other words, differentiation is an ill-conditioned operation in the sense of Hadamard [3].

During the last three decades several solutions have been proposed to these and other related problems that affect edge detection. Most of these solutions attack simultaneously both problems and some of them will be presented and discussed in this paper.

A. Discrete approximations of differential operators

We define a digital image, g, as a mapping $\mathcal{C} \times \mathcal{R} \stackrel{g}{\longrightarrow} \mathcal{I}$, with $\mathcal{C} \stackrel{\text{def}}{=} \{0, 1, \ldots, N_{\mathcal{C}} - 1\}$, $\mathcal{R} \stackrel{\text{def}}{=} \{0, 1, \ldots, N_{\mathcal{R}} - 1\}$, $\mathcal{I} \stackrel{\text{def}}{=} \{0, 1, \ldots, N_{\mathcal{I}} - 1\}$, and $N_{\mathcal{C}}, N_{\mathcal{R}}, N_{\mathcal{I}} \in \mathbb{N}$. The $(c, r) \in \mathcal{C} \times \mathcal{R}$ are coordinate pairs in a Cartesian system, where c denotes the column and r denotes the row.

Let us consider, therefore, that g is a digital image, obtained from sampling and quantization of an analog image, ϱ. One of simplest way to approximate the first order derivatives ϱ_x and ϱ_y is through the calculation of the first differences along the main axes, i.e.,

$$\begin{aligned} g_c(c, r) &\stackrel{\text{def}}{=} g(c, r) - g(c + 1, r) \\ g_r(c, r) &\stackrel{\text{def}}{=} g(c, r) - g(c, r + 1) \end{aligned} \qquad (11)$$

where $g_c(c, r)$ and $g_r(c, r)$ denote, respectively, the approximations to ϱ_x and ϱ_y around point (c, r). Often, these operators are represented as masks, such as

$$\mathbf{H}_c = \begin{bmatrix} \mathbf{1} & -1 \end{bmatrix} \quad \mathbf{H}_r = \begin{bmatrix} \mathbf{1} \\ -1 \end{bmatrix} \qquad (12)$$

where the bold value indicates the origin of the mask.

This operator has the disadvantage of not being symmetric in relation to the point of interest, which originates a bias in position. One of the ways to avoid this problem consists in using an odd number of mask elements as, for example,

$$\mathbf{H}_c = \begin{bmatrix} -1 & \mathbf{0} & 1 \end{bmatrix} \quad \mathbf{H}_r = \begin{bmatrix} 1 \\ \mathbf{0} \\ -1 \end{bmatrix} \qquad (13)$$

Several other first order derivative approximations along two perpendicular axes have been proposed [4]-[7]. Probably the most known are those proposed by Roberts (which is calculated using a set of axes rotated 45 degrees is relation to the usual orientation),

$$\mathbf{H}_1 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \quad \mathbf{H}_2 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix} \qquad (14)$$

Prewitt,

$$\mathbf{H}_c = \frac{1}{3} \begin{bmatrix} -1 & 0 & 1 \\ -1 & \mathbf{0} & 1 \\ -1 & 0 & 1 \end{bmatrix} \quad \mathbf{H}_r = \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ 0 & \mathbf{0} & 0 \\ -1 & -1 & -1 \end{bmatrix} \qquad (15)$$

Sobel,

$$\mathbf{H}_c = \frac{1}{4} \begin{bmatrix} -1 & 0 & 1 \\ -2 & \mathbf{0} & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad \mathbf{H}_r = \frac{1}{4} \begin{bmatrix} 1 & 2 & 1 \\ 0 & \mathbf{0} & 0 \\ -1 & -2 & -1 \end{bmatrix} \qquad (16)$$

and Frei-Chen (isotropic)

$$\mathbf{H}_c = \frac{1}{2+\sqrt{2}} \begin{bmatrix} -1 & 0 & 1 \\ -\sqrt{2} & \mathbf{0} & \sqrt{2} \\ -1 & 0 & 1 \end{bmatrix}$$

$$\mathbf{H}_r = \frac{1}{2+\sqrt{2}} \begin{bmatrix} 1 & \sqrt{2} & 1 \\ 0 & \mathbf{0} & 0 \\ -1 & -\sqrt{2} & -1 \end{bmatrix} \qquad (17)$$

To use these operators we perform an internal product between the respective mask and the image, as follows

$$g_\alpha(c,r) = \sum_i \sum_j g(c+i, r+j)(\mathbf{H}_\alpha)_{j,i} \qquad (18)$$

where α denotes the direction (1 or 2 for the masks of Roberts and c or r for the others), and $(\mathbf{H}_\alpha)_{j,i}$ denotes the element of row j and column i of mask $\mathbf{H}_\alpha$. Moreover, we consider that point $(i = 0, j = 0)$ corresponds to the origin of the mask (represented in bold).

Most often, the approximations that we presented above have the final objective of calculating the gradient using (4) and (6). Despite the fact that it is enough to compute two directional derivatives in order to calculate the gradient, it was several times suggested that, for noise suppression reasons, it would be better to use more than two directional derivatives. In this case, the gradient would be approximated by the directional derivative with the highest amplitude.

Based on this idea, several sets of directional masks were proposed. One of the most known is probably the one proposed by Kirsch [7], which is formed by the following masks:

$$\mathbf{H}_E = \frac{1}{15} \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} \rightarrow \mathbf{H}_{NE} = \frac{1}{15} \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} \nearrow$$

$$\mathbf{H}_N = \frac{1}{15} \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \uparrow \mathbf{H}_{NW} = \frac{1}{15} \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \nwarrow \qquad (19)$$

where the arrows show the directions of the derivatives approximated by the masks. As can be seen easily, these masks are generated by rotations of 45 degrees of the elements around the central element. Other sets of directional masks can be obtained using similar rotations of the orthogonal masks of Prewitt and Sobel [7].

The angular resolution allowed by these operators of support 3×3 is, at most, of 45 degrees. This means that we are only able to distinguish four different directions. For larger angular resolutions we have to use masks with a larger spatial support. Some of those operators were proposed, and are described, for example, in [4], [7].

In a similar way as first differences for the approximation of the first order derivative, second differences are the simplest approximation to the second order derivative. We define the second differences along the main axes as

$$g_{cc}(c,r) \stackrel{\text{def}}{=} g_c(c-1,r) - g_c(c,r)$$

$$g_{rr}(c,r) \stackrel{\text{def}}{=} g_r(c,r-1) - g_r(c,r) \qquad (20)$$

where $g_{cc}(c,r)$ and $g_{rr}(c,r)$ denote the approximations to ϱ_{xx} and ϱ_{yy}, respectively, around point (c,r). If we substitute (11) into (20) we obtain

$$g_{cc}(c,r) = g(c-1,r) - 2g(c,r) + g(c+1,r)$$

$$g_{rr}(c,r) = g(c,r-1) - 2g(c,r) + g(c,r+1) \qquad (21)$$

that can be represented by the following masks:

$$\mathbf{H}_{cc} = \begin{bmatrix} 0 & 0 & 0 \\ 1 & -\mathbf{2} & 1 \\ 0 & 0 & 0 \end{bmatrix} \quad \mathbf{H}_{rr} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & -\mathbf{2} & 0 \\ 0 & 1 & 0 \end{bmatrix} \qquad (22)$$

Using the definition of Laplacian given by (9) in addition to (21) we obtain a discrete approximation to the Laplacian given by

$$\mathbf{H}_{cc+rr} = \mathbf{H}_{cc} + \mathbf{H}_{rr} = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -\mathbf{4} & 1 \\ 0 & 1 & 0 \end{bmatrix} \qquad (23)$$

Frequently, these masks are represented using the symmetrical of the elements[1] of (23), multiplied by a normalization factor [6], [7]. Some other approximations were proposed to the Laplacian (see, for example, [6], [7]). However, due to the reduced direct use of this operator, we will not discuss it further.

B. Noise reduction and regularization

Most of the early edge detectors were based on simple discrete approximations of differential operators, such as those that we presented above. Among others, became popular the detectors of Roberts, Prewitt, Sobel and Frei-Chen, normally used as estimators of the modulus of the gradient, $|\widehat{\nabla} g|$, through

$$|\widehat{\nabla} g(c,r)| = \sqrt{\sum_\alpha g_\alpha^2(c,r)} \qquad (24)$$

or

$$|\widehat{\nabla} g(c,r)| = \sum_\alpha |g_\alpha(c,r)| \qquad (25)$$

where $g_\alpha(c,r)$ is given by (18). Generally, (25) is used instead of (24), which is more correct, if there is a need to reduce the computational load.

One of the greatest problems related with this kind of operators is their low capacity to reject high frequency noise. A simple and intuitive approach to attenuate this drawback is through the calculation of averages over the samples of the image. This means, in our case, that calculating the estimations of the derivatives using a relatively large set of neighboring samples is more robust to noise than using only two samples. Therefore, we can generalize the definition of first difference given in (11), using the symmetry mentioned in (13), in the following way:

$$g_c(c,r) = \widehat{g}(c+1,r) - \widehat{g}(c-1,r)$$

$$g_r(c,r) = \widehat{g}(c,r-1) - \widehat{g}(c,r+1) \qquad (26)$$

---

[1]Multiplied by -1.

where $\widehat{g}(c, r)$ is an estimate of $g(c, r)$, calculated using a weighted average of the image samples over a neighborhood of $g(c, r)$ and itself, i.e.,

$$\widehat{g}(c, r) = \sum_{(i,j) \in \mathcal{V}} w(i, j)\, g(c + i, r + j) \qquad (27)$$

where $\mathcal{V}$ denotes the neighborhood and $w(i, j)$ denotes the weight of the sample located in point (i, j), such that

$$\sum_{(i,j) \in \mathcal{V}} w(i, j) = 1 \qquad (28)$$

It is not difficult now to verify that the operators of Prewitt, Sobel and Frei-Chen can be explained under this generalized definition of differences, and also understand why they offer a better tolerance to noise when compared with operators having a smaller support, such as the detector of Roberts.

Following this idea of using averages to attenuate the effect of high frequency noise, several operators have been proposed, some of them being (support) enlarged versions of the 3×3 detectors mentioned above. That is the case, for example, of the 7×7 Prewitt based operator, of which we present only the mask used to estimate the horizontal first order derivative [7]

$$H_c = \frac{1}{21} \begin{bmatrix} -1 & -1 & -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & -1 & \mathbf{0} & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 & 1 & 1 & 1 \\ -1 & -1 & -1 & 0 & 1 & 1 & 1 \end{bmatrix} \qquad (29)$$

Right from the early stages of edge detection (and, more generally, image analysis) that was recognized the need to use operators of several dimensions. Rosenfeld *et al.* [8]-[10] proposed an algorithm to detect edges, commonly known as "difference of boxes", that relies on the use of pairs of neighborhoods (one neighborhood on each side of the point under analysis) of several dimensions and orientations. By convenience, they suggested that the neighborhoods should have a square shape and have sizes related to the powers of two. The output value of this operator is just the difference of the mean intensity values, calculated over the pair of neighborhoods. Also, they indicated that one of the possible ways to find the "ideal" operator size, is to look for the largest one that does not originate a significant decrease on the output value, when compared with the output value of the immediately smaller operator.

It is interesting to note that this work of Rosenfeld *et al.*, undertaken during the first steps of edge detection, addresses a question that recently has attracted the attention of many researchers: scale. The technique suggested by Rosenfeld incorporates, maybe for the first time, this important notion in the area of image analysis.

In the sequence of a comment on Rosenfeld's operator ([8]), Argyle proposed an operator based on a Gaussian function "broken" around the central point [11], i.e.,

$$f(x) = \begin{cases} G_\sigma(x) & \text{if} \quad x > 0 \\ -G_\sigma(x) & \text{if} \quad x < 0 \end{cases} \qquad (30)$$

where $G_\sigma(\cdot)$ is a Gaussian function with variance σ^2, given by

$$G_\sigma(x) \overset{\text{def}}{=} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad (31)$$

He justified his proposal arguing that the discontinuities present in both ends of the differences of boxes operator are adverse to a good noise reduction capability. Note that Argyle's operator (Eq. (30)) vanishes at both ends, which, in his opinion, improves noise tolerance. On the other hand, he considered that the discontinuity in the center of the operator is required.

In the follow up of this discussion, Macleod [12] claimed that the central discontinuity is not needed and, moreover, poses some problems, and suggested another operator,

$$f(x) = G_{\sigma_1}(x)\Big(G_{\sigma_2}(x - \sigma_2) - G_{\sigma_2}(x + \sigma_2)\Big) \qquad (32)$$

which is the difference of two Gaussian functions separated by $2\sigma_2$ and modulated by other Gaussian function. Figure 1 shows examples of Macleod's and Argyle's operators.
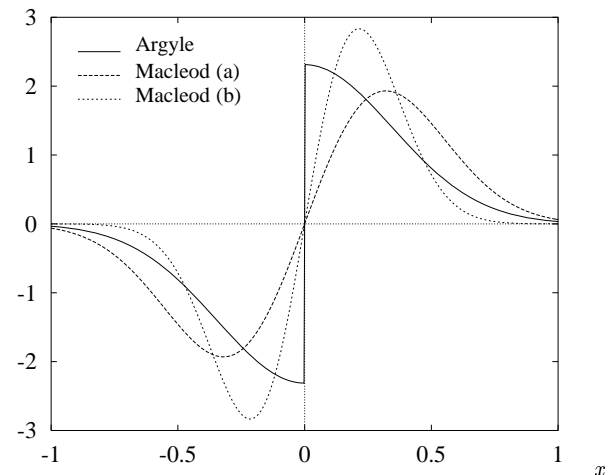

Figure 1 - Graphics of Argyle's operator for $\sigma = 0.345$, and of Macleod's operator (a) $\sigma_1 = 0.5$ and $\sigma_2 = 0.345$, and (b) $\sigma_1 = 0.25$ and $\sigma_2 = 0.345$.

As we pointed out, the main reason for using operators with large support is to improve tolerance to high frequency noise and, therefore, to reduce the number of false responses. More recently, Torre and Poggio [1] addressed this question in a theoretical way. They formulated the differentiation operation as an ill-posed problem in the sense of Hadamard, and they used regularization techniques to transform it into a well-behaved problem.

Let us see the following example [3]. Consider that $\varrho(x)$ is a function affected by a small amplitude noise given by $\epsilon \sin(\omega x)$. The difference between $\varrho(x)$ and $\varrho(x) + \epsilon \sin(\omega x)$ can be made arbitrarily small, if ϵ is made sufficiently small. However, the difference of their derivatives can be quite large if ω is made large.

The above case exemplifies the usual idea that differentiation amplifies high frequency noise. In the sense of Hadamard, this means that the differentiation operation is an ill-posed problem [3], since it violates one of the principles of well-posed problems, in this case that small variations in the data should not originate abrupt changes in the

solution of the problem. The other two principles are the existence of a solution and its uniqueness.

One of the most natural ways of regularizing the operation of numerical differentiation is through the approximation or interpolation of the data using analytic functions. If approximation functions are used (which is advisable for noisy data), then the regularization techniques lead, for example, to the search of a function $z(x)$ such than the functional

$$\sum_{i=1}^{n} \Big(g(x_i) - z(x_i) \Big)^2 + \lambda \int |z''(x)|^2 \, dx \qquad (33)$$

has to be minimized. Eq. (33) $z(x)$ denotes the approximation function of the n points $g(x_i)$.

The first term of (33) measures the fidelity of the approximation, i.e., the distance between the data and the approximation function. The second term of (33) controls the regularity of the solution through a stabilizing functional that, in this case, is the second order derivative. Parameter λ controls the balance of these two constraints.

It was proved that the solution of (33) can be obtained through the convolution of the data with a cubic spline quite similar to a Gaussian function [1], providing that the noisy data are regularly placed along the x axis and that some limit conditions are met. It was also shown that, for a class of inverse problems described through a convolution operation (which includes the differentiation), the regularization can be performed through the convolution of the data with a filter $\tilde{f}_\alpha(x)$ (having Fourier transform $\tilde{\mathfrak{F}}_\alpha(\omega)$) that obeys to the following conditions (Tikhonov conditions):

1. $\tilde{\mathfrak{F}}_\alpha(\omega)$ is limited for $\alpha \geq 0$.

2. $\tilde{\mathfrak{F}}_\alpha(\omega)$ is an even function in relation to ω, i.e. $\tilde{\mathfrak{F}}_\alpha(\omega) = \tilde{\mathfrak{F}}_\alpha(-\omega)$, and belongs to $L^2(-\infty, \infty)$.

3. $j\omega\tilde{\mathfrak{F}}_\alpha(\omega)$ belongs to $L^2(-\infty, \infty)$.

4. $\lim_{|\omega| \to \infty} \tilde{\mathfrak{F}}_\alpha(\omega) = 0, \ \forall \alpha > 0$.

5. $\lim_{\alpha \to 0} \tilde{\mathfrak{F}}_\alpha(\omega) = 1$ and $\tilde{\mathfrak{F}}_0(\omega) = 1$.

Parameter α controls the spatial support of the filter, i.e., its scale, and $j \overset{\text{def}}{=} \sqrt{-1}$ is the imaginary unit.

This set of results that we presented above allows a more rigorous justification to the frequent and intuitive use of averages or low-pass filtering in edge detection. Nevertheless, this does not mean that all (smoothing) filters used for edge detection obey to the Tikhonov conditions. For example, the rectangular filter (simple average) defined by

$$\tilde{f}_\alpha(x) = \begin{cases} \dfrac{1}{\sqrt{2\alpha}} & |x| \leq \alpha \\ 0 & |x| > \alpha \end{cases} \qquad (34)$$

and having Fourier transform

$$\tilde{\mathfrak{F}}_\alpha(\omega) = \sqrt{\frac{2}{\alpha}} \frac{\sin(\omega\alpha)}{\omega} \qquad (35)$$

does not respect the third Tikhonov condition. In the next section we describe some of the most recent edge detectors that explicitly use linear filtering. Some of them were proposed before the work of Torre and Poggio. Some others were proposed afterwards, but without any reference to that theory. Anyway, we will see that all of them obey to the conditions of Tikhonov.

III. LINEAR FILTERS FOR EDGE DETECTION

Some of the most recent edge detectors are explicitly formulated using a combination of linear low-pass filtering followed by differentiation, although frequently these two operations are implemented by a single filter. Scale is typically associated with the dimension of the spatial support of the filter. As we saw, small support filters may not provide sufficient tolerance to high frequency noise. The operators proposed by Argyle (30) and Macleod (32), described above, reveal a concern related to the adaptation of their scale to the particular noise conditions of the images. However, these operators, and also all the others that we described until now, were developed based fundamentally on heuristics.

In the following text we make reference to two kinds of filters: detection filters and smoothing filters. Some of the formulations follow a line that takes to the development of filters that are used to attenuate noise (i.e., regularize the data). The detection is made afterwards through differentiation. Some other formulations derive directly the detection filter. To avoid confusion, we use the following notation: $f(x)$ denotes detection filters, and $\tilde{f}(x)$ denotes smoothing filters.

A. Dickey and Shanmugam

The use of objective criteria towards the development of edge detection operators was initiated, as far as we know, by Dickey and Shanmugam [13], [14]. They derived a filter in the spatial frequency domain which generates a maximum energy output inside a pre-defined interval around the edge point under analysis. This corresponds, for the one-dimensional case, to the maximization of

$$\gamma = \frac{\displaystyle\int_{-I/2}^{I/2} |o(x)|^2 \, dx}{\displaystyle\int_{-\infty}^{\infty} |o(x)|^2 \, dx} \qquad (36)$$

where $o(x) = f(x) * \varrho(x)$ [2] is the response of the filter $f(x)$ to the input signal $\varrho(x)$, characterized by containing an edge point centered in $x = 0$, and I is the resolution of the system, i.e., the interval over which the response is analyzed. To derive the filter they used also the following restrictions: (1) the filter should be linear and space invariant, i.e., $\mathfrak{O}(\omega) = \mathfrak{F}(\omega)\mathfrak{G}(\omega)$; [3] (2) the filter should be frequency limited, where ω_c is the cut-off frequency, i.e., $\mathfrak{F}(\omega) = 0$ for $|\omega| > \omega_c$; (3) the response of the filter to constant signals or of slow variation should be negligible, i.e., $\mathfrak{F}(0) = 0$; (4) the filter should be an even function, to allow easy extension to two dimensions, i.e., $\mathfrak{F}(-\omega) = \mathfrak{F}(\omega)$.

---

[2]Operator $*$ denotes convolution.

[3]$\mathfrak{O}(\omega)$, $\mathfrak{F}(\omega)$ and $\mathfrak{G}(\omega)$ represent, respectively, the Fourier transforms of $o(x)$, $f(x)$ and $\varrho(x)$.

The filter proposed by Dickey and Shanmugam for the detection of step edges is described by

$$\mathfrak{F}(\omega) = \begin{cases} K_1\,\omega\,\psi_1\left(c, \dfrac{\omega I}{2\omega_c}\right) & \text{if } |\omega| < \omega_c \\ 0 & \text{otherwise} \end{cases} \quad (37)$$

where $c = \omega_c I/2$, K_1 is a real constant, and ψ_1 is a prolate spheroidal function of zero order and degree one. Dickey and Shanmugam also suggested an approximation for (37) that avoids the use of prolate functions [14]. This approximation, posteriorly corrected by Lunscher, is given by [15]

$$\mathfrak{F}(\omega) \approx \begin{cases} K_2\,\omega^2\,\exp\left(-\dfrac{c\omega^2}{2\omega_c^2}\right) & \text{if } |\omega| < \omega_c \\ 0 & \text{otherwise} \end{cases} \quad (38)$$

and is valid for $|\omega| < \omega_c c^{-1/4}$.

Dickey and Shanmugam showed that the signal to noise ratio inside the interval of resolution improves when c increases. However, to keep the noise under an acceptable level the bandwidth of the filter should not be too large and, therefore, the resolution is forced to be reduced (remember that c is proportional to the bandwidth, ω_c, and resolution interval, I).

Lunscher [15] pointed out that (38) has a shape similar to the Laplacian of the Gaussian function, filter posteriorly proposed by Marr and Hildreth [16]. Nevertheless, there are several differences between them. For example, while the filter of Dickey and Shanmugam is explicitly frequency limited for $|\omega| = \omega_c$, the filter of Marr and Hildreth is not. Another difference, that produces visible results in the edge maps, is related with the fact that while Dickey and Shanmugam mark the presence of an edge whenever the output of the filter reaches some pre-defined threshold, Marr and Hildreth seek, instead, zero crossings in the output of the filter. Observing that when $c \to 0$ the filter of Dickey and Shanmugam behaves like a Laplacian, then it is easy to see that this implies that edges are marked twice if they are detected by threshold.

B. Marr and Hildreth

From all the filters that we present in this section, the edge detection filter proposed by Marr and Hildreth [16] is the only one that did not result from the optimization of some mathematical criteria. Nevertheless, it is frequently referred to as an "optimal" filter. It was developed based on a set of psychophysical and physiological observations, and also on some properties of the Gaussian function.

Marr and Hildreth not only proposed an edge detector but also a theory about the detection of edges. They pointed out, for example, that the variation of light intensity occurs at several scales, which should imply, also, their detection using detectors tuned for several different scales. Since a single filter cannot be optimal for all possible scales, then they suggested that the images should be first smoothed using several resolutions, and only afterwards scanned for edges.

Their choice of the "optimal" smoothing filter obeyed to two conditions: (1) the filter should be smooth in the frequency domain and approximately frequency limited, to be able to reduce the scale range where intensity changes occur; (2) the filter should be localized in the spatial domain, since the influence of a given point only makes sense in a relatively small neighborhood. Therefore, Marr and Hildreth looked for a linear filter that was able to jointly minimize the (product of the) spread in space and frequency, and found that this requirement is met by the the Gaussian function.

After smoothing, the detection can be done by looking for the maxima of the absolute value of the first derivative or for the zero crossings of the second derivative. Marr and Hildreth proposed the second order derivative and, therefore, formulated the edge detection operation as the problem of finding the zero crossings of

$$o(x,y) = D^2[G_\sigma(x,y) * \varrho(x,y)] \quad (39)$$

where D denotes the differential operator, $\varrho(x,y)$ the image under analysis, and $G_\sigma(x,y)$ a two-dimensional Gaussian function defined as

$$G_\sigma(x,y) \stackrel{\text{def}}{=} \frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (40)$$

Using the rule of the derivative of a convolution we have

$$D^2[G_\sigma(x,y) * \varrho(x,y)] = D^2 G_\sigma(x,y) * \varrho(x,y) \quad (41)$$

which means that the smoothing and differentiation operations can be implemented by a single operation consisting on the convolution of the image with the second order derivative of the Gaussian function.

One of the problems that arises when using second order derivatives is the determination of the most appropriate direction to align the operator. Since we want to find the points where the absolute value of the gradient is a local maximum, using the zero crossings of the second derivative, then the most appropriate direction is precisely the gradient direction in each point. However, that would imply the previous knowledge of the gradient and, therefore, an increase in computational effort.

As previously mentioned in Section II, although the Laplacian does not always coincides with the second order directional derivative along the gradient direction, it reveals several advantages, among them the property of being a rotationally invariant operator. This was the main reason that motivated Marr and Hildreth to the choice of the Laplacian, instead of the more correct, but also more problematic, operator based on the directional second order derivative. The final form of the operator of Marr and Hildreth is, therefore, [16]

$$\nabla^2 G_\sigma(x,y) = -\frac{1}{\pi\sigma^4}\left(1 - \frac{x^2+y^2}{2\sigma^2}\right)\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad (42)$$

and is frequently known as the Laplacian of the Gaussian.

C. Canny

The use of mathematical criteria in order to develop "optimized" edge detection filters was definitively popularized by Canny [17], [18]. The Canny filter is, most probably, the

most known and cited among the several "optimal" filters that have been proposed since then.

Canny started by defining the operation of edge detection in one dimension as the determination of the local maxima resulting from the convolution of a filter with the signal affected by additive white noise, $n(x)$, for which[4]

$$E[n(x)] = 0, \quad n_0^2 \stackrel{\text{def}}{=} E[n^2(x)] \qquad (43)$$

and defined three criteria for the development of the filter. One of them, represents the need of a good detection, i.e., the existence of a low probability that it generates false detections or fails to detect some edges. In other words, it was intended that the filter should possess a high signal to noise ratio. This criterion is defined as

$$C_{\text{SNR}} = \frac{\left| \int_{-\infty}^{\infty} \varrho(-x) f(x) \, dx \right|}{n_0 \sqrt{\int_{-\infty}^{\infty} f^2(x) \, dx}} \qquad (44)$$

where $f(x)$ is a linear filter and $\varrho(x)$ is the profile of an edge localized in $x = 0$. Note that the numerator of (44) corresponds to the response of the filter (in $x = 0$) to the ideal signal (the edge profile), while the denominator is the response of the filter only to the noise.

The second criterion imposes good localization of the detections, i.e., the edges must be marked as closest as possible of their correct positions. Mathematically, this criterion is defined by

$$C_{\text{LOC}} = \frac{\left| \int_{-\infty}^{\infty} \varrho'(-x) f'(x) \, dx \right|}{n_0 \sqrt{\int_{-\infty}^{\infty} f'^2(x) \, dx}} \approx \frac{1}{\sqrt{E[x_0^2]}} \qquad (45)$$

where x_0 denotes the detection position, i.e., (45) is an approximation of the inverse of the standard deviation of the effective detection position [18]. Tagare *et al.* [19] pointed out that (45) was derived using a wrong expression for the variance of the position of the detected edge. However, despite the discussion generated (see [20], [21]), it seems to be only significant for very high noise levels [22].

Finally, the third criterion is intended to limit the number of false responses in the vicinity of the correct detection. Note that (44) controls only the behavior of the filter in the point where the edge occurs. As pointed out by Canny [18], the maximization of only the first two criteria leads to a filter defined by $f(x) = \varrho(-x)$. For a step edge this leads to a filter that is a truncated step, i.e., to the "difference of boxes", which is characterized by good detection and localization capabilities, but also by producing multiple responses in the vicinity of the edge. The idea of this third criterion is, therefore, to maximize the mean distance between the maxima of the responses due to noise, where that

distance is given by [18]

$$x_{\text{max}} = 2\pi \sqrt{\frac{\int_{-\infty}^{\infty} f'^2(x) \, dx}{\int_{-\infty}^{\infty} f''^2(x) \, dx}} \qquad (46)$$

Since (46) changes according to the scale of the filter, i.e., $x_{\text{max}}(f_W) = W x_{\text{max}}(f)$ [18] [5] then, to have a scale invariant criterion, we redefine it to be

$$C_{\text{MUL}} = \frac{x_{\text{max}}}{W} \qquad (47)$$

where W denotes the support (scale) of the filter.

The optimization process proposed by Canny is the following: maximize the product of the C_{SNR} and C_{LOC} criteria, while keeping the multiple response criterion, C_{MUL}, constant and equal to a pre-defined value.

Although this process can be used for developing filters dedicated to a specific and arbitrary edge profile, by far the most studied has been the step edge profile. Using the proposed criteria, Canny derived a family of filters tuned to this profile, i.e. for $\varrho(x) = A u_{-1}(x)$ where $u_n(x)$ denotes the n*th*-order derivative of the delta function and A the amplitude of the step. After some manipulations (44) and (45) present the following form:

$$C_{\text{SNR}} = \frac{A}{n_0} \frac{\left| \int_{-\infty}^{0} f(x) \, dx \right|}{\sqrt{\int_{-\infty}^{\infty} f^2(x) \, dx}} = \frac{A}{n_0} \Sigma(f) \qquad (48)$$

and

$$C_{\text{LOC}} = \frac{A}{n_0} \frac{|f'(0)|}{\sqrt{\int_{-\infty}^{\infty} f'^2(x) \, dx}} = \frac{A}{n_0} \Lambda(f') \qquad (49)$$

Note that the functionals $\Sigma(f)$ and $\Lambda(f')$ do not depend on the characteristics of the image, i.e., they reflect only the properties of filter $f(x)$. It is also important to observe their behavior when a change of scale is made. In that case we have

$$\Sigma(f_W) = \sqrt{W} \, \Sigma(f) \quad \text{and} \quad \Lambda(f'_W) = \frac{1}{\sqrt{W}} \Lambda(f') \quad (50)$$

which means that $\Sigma(f)\Lambda(f')$ is scale invariant.

The maximization of $\Sigma(f)\Lambda(f')$, constrained to the condition of a constant C_{MUL}, leads to a filter described by the following equation [18]

$$f(x) = [a_1 \sin(\omega x) + a_2 \cos(\omega x)]e^{\alpha x} + \\ [a_3 \sin(\omega x) + a_4 \cos(\omega x)]e^{-\alpha x} + c \qquad (51)$$

valid for $x \in [0, W]$, [6] and constrained to the following set of boundary conditions:

$$f(0) = 0, \quad f(W) = 0, \quad f'(0) = s, \quad f'(W) = 0 \quad (52)$$

---

[4]Notation $E[\cdot]$ represents the statistical expectation.

[5]Where $f_W(x) = f(x/W)$.

[6]The expression of the filter for the interval $[-W, 0]$ is obtained using its anti-symmetric characteristic, i.e., $f(x) = -f(-x)$.

(where s is a constant equal to the derivative of $f(x)$ in the origin) that allow the determination of $a_1, \ldots, a_4$, in function of the parameters α, ω, s and c, obtained by numerical optimization.

For an easy comparison among several of the filters represented by (51), we introduce a normalization factor, μ, which is given by

$$\mu \stackrel{\text{def}}{=} \frac{1}{\left| \displaystyle\int_0^W f(x)\,dx \right|} \qquad (53)$$

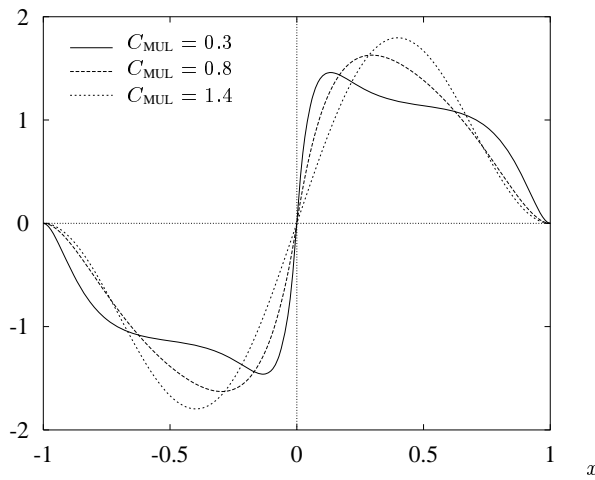while the normalized filters are defined by $f_\mu(x) \stackrel{\text{def}}{=} \mu f(x)$.

Figure 2 - Filters $f_\mu(x)$ obtained for $C_{\mathrm{MUL}} \in \{0.3, 0.8, 1.4\}$.

Figure 2 shows some plots of $f_\mu(x)$ for several values of C_{MUL}. As can be seen, the shape of $f_\mu(x)$ approximates the "difference of boxes" when C_{MUL} decreases.

From the family of filters derived by Canny, he chose one with a probability of error due to multiple responses (p_m) as close as possible to the probability of error of detection (p_f). The relation between these two probabilities of error is given by [17]

$$p_m = r p_f \Rightarrow \frac{|f'(0)|}{\sqrt{\displaystyle\int_{-\infty}^{\infty} f''^2(x)\,dx}} = r\Sigma(f) \qquad (54)$$

Having $c = 1$, the filter that maximizes r is defined by the following parameters:

$$\begin{aligned} \alpha &= 2.05220, \quad \omega = 1.56939, \quad \mu = 3.74235 \\ a_1 &= -0.14868, \quad a_2 = -0.20876, \\ a_3 &= 1.24465, \quad a_4 = -0.79125 \end{aligned} \qquad (55)$$

which attains the following criteria values:

$$\Sigma(f) \approx 0.61\sqrt{W}, \quad \Lambda(f') \approx 1.84/\sqrt{W},$$
$$\Sigma(f)\Lambda(f') \approx 1.12, \quad r \approx 0.58 \ \text{e} \ x_{\max} \approx 1.2W$$

Based on computational complexity arguments, Canny proposed the first derivative of the Gaussian function as a

good and efficient approximation to (55). In order to allow a reasonable comparison between these two filters, we made a numerical adjustment between (55) and

$$f_G(x) = -k\sigma\sqrt{2\pi}G'_\sigma(x) = k\frac{x}{\sigma^2}\exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad (56)$$

where $G'_\sigma(x)$ denotes the first derivative of the Gaussian function defined in (31). The smallest quadratic error calculated in the interval $x \in [-1, 1]$ is obtained for $k \approx 1.022$ and $\sigma \approx 0.345$, for which results a filter with (both functions are plotted in Fig. 3):

$$\Sigma(f) \approx 0.62\sqrt{W}, \quad \Lambda(f') \approx 1.48/\sqrt{W},$$
$$\Sigma(f)\Lambda(f') \approx 0.91, \quad r \approx 0.53, \quad x_{\max} \approx 1.37W$$

Note that these criteria values differ slightly from those provided by Canny [17], [18], because while Canny considered the infinite support of the Gaussian function, we performed the calculations assuming that it is truncated in $x = \sigma/0.345$.
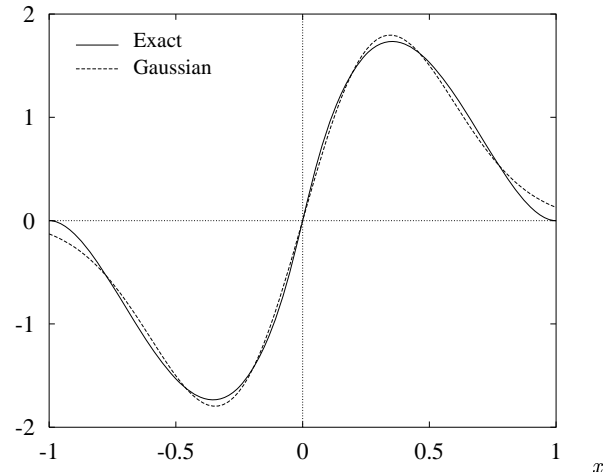

Figure 3 - Filter (55) and its approximation using the first derivative of the Gaussian function.

Comparing the values of the criteria obtained for the Canny filter and also for its approximations, we can observe, essentially, that the filter based on the first derivative of the Gaussian reveals a worse localization capability and an better suppression of multiple responses, when compared to the Canny filter. The signal to noise ratio is similar in both.

Although (51) was derived as a whole, i.e. there is not an explicitly separation between smoothing and differentiation filters, it is straightforward to see it like that, if we consider that the smoothing filter is the integral of (51). Therefore, the smoothing filter corresponding to the detection filter based on the derivative of the Gaussian is the Gaussian function itself.

Up to now, we have been discussing one-dimensional filters. However, our intention is to detect edges in images, which means that we have to extend them to two dimensions. Often, these extensions are made without careful justifications, or purely based on computational efficiency reasons. This was also the case for the extension to two

dimensions of the Canny filter, for which he proposed the use of a Gaussian function perpendicular to the direction of the edge (projection function), followed by the detection function (i.e., the derivative of a Gaussian).

Essentially, Canny pointed out two motivations for using the Gaussian as the projection function. One of them, is the good behavior of the Gaussian as a windowing function, avoiding abrupt transitions at the boundaries. The other, is its easy integration with the detection function, resulting in a two-dimensional Gaussian as the smoothing filter, followed by differentiation. In other words, the edge detection can be performed through the calculation of the derivative, along two directions, of the image filtered by a two-dimensional Gaussian. Yet another important advantage, is the separability of the two-dimensional Gaussian defined in (40), which allows its decomposition into two one-dimensional filters. This means that the filtering can be applied first to columns (rows) and then to rows (columns), reducing the computational burden. We have, therefore,

$$G_\sigma(x, y) * \varrho(x, y) =$$
$$\iint G_\sigma(x - \alpha, y - \beta)\, \varrho(\alpha, \beta)\, d\alpha\, d\beta =$$
$$\int G_\sigma(y - \beta) \left[\int G_\sigma(x - \alpha)\, \varrho(\alpha, \beta)\, d\alpha \right] d\beta = \quad (57)$$
$$G_\sigma(y) * \Big(G_\sigma(x) * \varrho(x, y) \Big) =$$
$$G_\sigma(x) * \Big(G_\sigma(y) * \varrho(x, y) \Big)$$

and also, exemplifying for the case of the operator along the x direction,

$$\frac{\partial}{\partial x}\Big(G_\sigma(x, y) * \varrho(x, y) \Big) = \frac{\partial G_\sigma(x)}{\partial x} * \Big(G_\sigma(y) * \varrho(x, y) \Big) \tag{58}$$

D. Spacek

Spacek was one of the first researchers that followed up the work of Canny. He derived an edge detector with the aim of developing a theory that could integrate edge detection, curve measuring, and motion detection, during the early stages of visual processing [23]. He formulated the optimization criteria in a way quite similar to Canny, although, right from the beginning, Spacek decided to develop an edge detector tuned for step profiles.

However, the combination of the three criteria was performed using a different approach. As we mentioned, Canny derived the filters by making the multiple responses criterion constant and then maximizing the product of the other two. Instead, Spacek used a performance measure, independent of scale, which is basically the product of the three criteria:

$$P(f) = \frac{\left(f'(0) \int_{-1}^{0} f(x)\, dx \right)^2}{\left(\int_{-1}^{0} f^2(x)\, dx \right) \left(\int_{-1}^{0} f''^2(x)\, dx \right)} \tag{59}$$

where $P(f)$ is a functional in $f(x)$. Maximizing (59) we obtain a filter described by the following expression [23]:

$$f(x) = [C_1 \sin(x) + C_2 \cos(x)]e^x +$$
$$[C_3 \sin(x) + C_4 \cos(x)]e^{-x} + 1 \tag{60}$$

with

C_1	C_2	C_3	C_4
-13.3816	2.7953	0.0542	-3.7953

where (60) is defined for $x \in [-1, 0]$. The positive half of the filter is obtained using its anti-symmetric characteristic.

As pointed out by Spacek, the filter defined by (60) can be approximated quite well by the cubic spline

$$c(x) = x^3 + 2x^2 + x \tag{61}$$

defined for $x \in [-1, 0]$. Following the rule stated above, we denote the normalized filters by $c_\mu(x)$ and $f_\mu(x)$, where $\mu = 12$ and $\mu \approx 1.7765$, respectively. In fact, these functions present a great degree of similarity, as can be seen in Fig. 4, where we also show the Canny filter defined by (55).
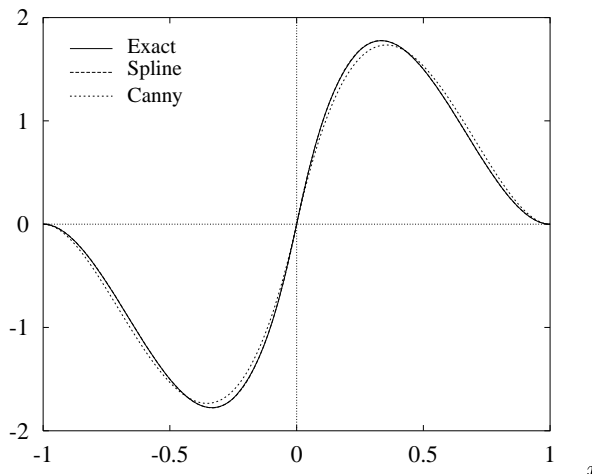

Figure 4 - Filter (60), its approximation using the cubic spline defined in (61), and also the Canny filter described by (55).

One of the greatest problems related to directional filters, as the case of (60), is their extension to two dimensions, since it cannot be performed by rotation. Spacek overcame this using a simple, but ingenious, approach, based on the following property of the differentiation of the convolution:

$$f(x) * \varrho(x) = [\tilde{f}(x) * \varrho(x)]' \tag{62}$$

where $\tilde{f}(x)$ is the integral of $f(x)$. This means that we can first convolve with the integral of the filter, and differentiate afterwards. The advantage is that being $f(x)$ anti-symmetric, then $\tilde{f}(x)$ is symmetric and, therefore, can be easily extended to two dimensions by rotation. Also, it is now easy to interpret this result under the regularization theory: $\tilde{f}(x)$ is, effectively, a regularizing filter, applied just before differentiation.

The work of Spacek was posteriorly extended by Petrou and Kittler, that reduced from six to four the number of arbitrary constants in the derivation of (60), searching, instead, for their optimal value [24]. However, they concluded that the improvements are only marginal and that the cubic spline defined in (61) is, in fact, a very good approximation to the "optimum" filter.

E. Deriche

The filters developed by Canny, Spacek and Petrou *et al.* all have finite support, due to the boundary constraints defined in (52) imposed during the design. Deriche showed [25], [26] that using the same general equation as the one defined by Canny in (51), but assuming that the filter is of infinite support (i.e., $f(\infty) = 0$ and $f'(\infty) = 0$), it is possible to improve the values given by the criteria. That infinite support filter is

$$f(x) = -c_1 e^{-\alpha|x|} \sin(\omega x) \qquad (63)$$

having the following criteria values:

$$\Sigma(f)\,\Lambda(f') = \frac{2\alpha}{\sqrt{\omega^2 + \alpha^2}}, \quad r = \sqrt{\frac{\alpha^2 + \omega^2}{5\alpha^2 + \omega^2}} \qquad (64)$$

Deriche pointed out that the case where $\alpha \gg \omega$ as the most favorable, since the product $\Sigma(f)\,\Lambda(f')$ is maximum for $\omega = 0$. In that case, the filter can be approximated by

$$f(x) = -c_2 x e^{-\alpha|x|} \qquad (65)$$

since $\sin(\omega x) \approx \omega x$ for $\omega x \approx 0$. This approximation provides $\Sigma(f)\,\Lambda(f') \approx 2$ and $r \approx 0.44$.

Deriche proposed recursive implementations for these filters in order to avoid truncation errors and also to improve computational efficiency. However, we will not cover implementation issues in this paper.

F. Poggio, Voorhees and Yuille

Poggio *et al.* [27] defined the edge detection problem as a process that measures, detects, and locates changes in the light intensity of an image, which, according to their opinion, implies the calculation of several derivatives. However, as we already discussed, numerical differentiation is ill-conditioned. To address this problem they proposed the direct use of regularization techniques.

Basically, the idea is to find $z(x)$, starting from a set of data $g(x_i)$, in such a way that $Az = g$. Let us consider that A is an operator that samples z, i.e., $Az|_{x=x_i} = z(x_i)$. In our case, z should be a function sufficiently well behaved, on one hand, to be differentiable, and, on the other hand, to approximate the data $g(x_i)$ placed at points x_i. According to Tikhonov's regularization, we have to look for a function z that minimizes the functional

$$\|Az - g\|^2 + \lambda \|Pz\|^2 \qquad (66)$$

where λ is a regularization parameter and P a stabilizing operator.

Picking the second derivative for P and norm L^2, we arrive at the following expression:

$$\sum_{i=1}^{n} \Big(g(x_i) - z(x_i)\Big)^2 + \lambda \int |z''(x)|^2 \, dx \qquad (67)$$

that has to be minimized through the choice of an appropriate function $z(x)$. Poggio *et al.* showed that for data placed on a regular lattice and assuming that the signal goes to zero

at the infinite or is periodic, then $z(x)$ can be obtained convolving the data with the function

$$\tilde{f}(x) = \frac{1}{2\lambda^{1/4}} \exp\left(\frac{-|x|}{\sqrt{2}\lambda^{1/4}}\right) \cos\left(\frac{-|x|}{\sqrt{2}\lambda^{1/4}} - \frac{\pi}{4}\right) \quad (68)$$

Moreover, they also showed that this function can be approximated by a Gaussian, using $\sigma = \lambda^{1/4}$. Note that $\tilde{f}(x)$ is a smoothing filter. The correspondent detection filter is give by

$$f(x) = \frac{d\tilde{f}(x)}{dx} \qquad (69)$$

G. Sarkar and Boyer

Sarkar and Boyer [28], [29] followed a line of work similar to the one of Deriche for the development of infinite support filters based, mainly, on Canny's criteria. They used identical criteria in what respects the signal to noise ratio and localization ((48) and (49)), and introduced some changes in the criterion of multiple responses. Moreover, they raised the question of the estimation of the filter support. While for the filter proposed by Canny the support is well defined (it is of limited support), this is not the case for Deriche's filter[7].

To attenuate this problem, Sarkar and Boyer proposed the substitution of W in (47) by a more appropriate measure of the width of a filter of infinite response[8]:

$$W_{\mathrm{N}} = \sqrt{\frac{\displaystyle\int_{-\infty}^{\infty} x^2 f^2(x)\,dx}{\displaystyle\int_{-\infty}^{\infty} f^2(x)\,dx}} \qquad (70)$$

which represents the square root of the normalized mean squared deviation of $f^2(x)$, measured in respect to the origin.

Probably the most important conclusion that we can draw from the work of Sarkar and Boyer is that the ("optimum") filter proposed by them (which maximizes the product of the three criteria) is very similar, in shape, to the derivative of a Gaussian. Nevertheless, they pointed out that their filter presents some advantages in terms of computational efficiency.

H. Shen and Castan

Shen and Castan addressed the edge detection problem in quite different manner [30]. Their definition of edge detector is based, explicitly, on a set of low-pass filtering and differentiation operations. They used an optimization process that relies on three factors: (1) the maximization of the detector's response to an edge, (2) the minimization of the smoothing filter's response to noise, and (3) also the

---

[7]Although not discussed by Canny, the derivative of the Gaussian is also of infinite support and, therefore, reveals similar problems. In other words, the support depends on the maximum truncation error that we allow when we approximate the Gaussian by a finite support filter.

[8]This same measure of the spatial width of the filter was also proposed by Tagare *et al.* [19], also in the context of edge detection.

minimization of the detector's response to noise. The normalized global criterion to minimize is given by [30]:

$$C_{\mathrm{N}} = \frac{\sqrt{\int\int_{-\infty}^{\infty} \tilde{f}^2(x)\, dx \; \int_{-\infty}^{\infty} \tilde{f}'^2(x)\, dx}}{\tilde{f}^2(0)} \qquad (71)$$

from which results, after minimization, the smoothing filter

$$\tilde{f}(x) = a b^{|x|} \qquad (72)$$

where $a = -\ln(b)/2$ and $0 < b < 1$. Then, using relation (69), we have

$$f(x) = \begin{cases} a\ln(b)\, b^x & x > 0 \\ -a\ln(b)\, b^{-x} & x < 0 \end{cases} \qquad (73)$$

This filter contains a particularity not present in the other filters, which is the discontinuity for $x = 0$. If we regard the estimation of a derivative in a given point (x_0) as the difference between two estimates of the function calculated on opposite positions relative to x_0, then it seems natural to give more importance (weight) to points near x_0. For example, applying this idea to the derivative of the Gaussian, we see that the largest weights are given to points around $x = \sigma$. In fact, the discontinuity of $f(x)$ in $x = 0$ has a decisive influence in the localization capacity of the filter. Unfortunately, as we will see, it also implies some disadvantages.

Shen and Castan proposed also a measure to assess the localization error of the filters. However, before we go into further details on that measure, we will first use the localization measure proposed by Canny (49) to evaluate the localization capacity of (73) (we remember that this criterion presents high values for filters exhibiting good localization). As we can verify, (49) is proportional to the value of the first derivative of $f(x)$, for $x = 0$. The filter described by (73) has a discontinuity in $x = 0$ which means that, for $x = 0$, the first derivative is infinite. This characteristic points towards a good localization ability of filter (73). In fact, using the localization measure proposed by Shen and Castan

$$L_e = \left| \lim_{\epsilon \to 0} \frac{\epsilon}{\int_{-\epsilon/2}^{\epsilon/2} |\tilde{f}''(x)|\, dx} \right| \int_{-\infty}^{\infty} |\tilde{f}''(x)|\, dx \qquad (74)$$

we obtain $L_e = 0$ for the filter (note that (74) varies inversely to (49)).

If we apply this measure to the Gaussian function we obtain $L_e \approx 2.43\sigma$, which shows that the value presented by Shen and Castan [30] ($L_e \approx 16.53\sigma$) is incorrect, appearing also in other publications such as, for example, [31]-[33]. This results from the incorrect calculation of

$$\int_{-\infty}^{\infty} |G_\sigma''(x)|\, dx \qquad (75)$$

which value is $4/(\sqrt{2\pi e}\,\sigma^2)$ and not $4\sqrt{e}/\sigma^2$ as they presented ($G_\sigma(x)$ is the Gaussian function defined in (31)).

We calculated also the value of L_e for the Canny filter, i.e. for (51) with parameters (55), for which we obtained $L_e \approx 0.636W$. This value is also different from the one presented by Shen and Castan ($L_e \approx 0.81W$). For the comparison of the values of L_e relative to the derivative of the Gaussian and the Canny filter we use the relation $\sigma = 0.345W$, already discussed. In that way, we get $L_e \approx 0.838W$ for the derivative of the Gaussian, which means a worse localization characteristic than the Canny filter (this agrees with the results obtained using Canny's localization criterion).

Let us now go a bit further and compare the localization measure proposed by Shen and Castan (74) with Canny's localization criterion (49). Analyzing the first term of (74) (the one with the limit) we can see that its purpose is the calculation of $|\tilde{f}''(0)|$, i.e. $|f'(0)|$, term also present in (49). The second term also has an equivalent in (49), representing norms of $f'(x)$ (or $\tilde{f}''(x)$): absolute norm in (74) and quadratic norm in (49).

While Canny calculated $\sqrt{E[x_0^2]}$ [17], [18], Shen and Castan opted by calculating $E[|x_0|]$, where x_0 denotes the effective position of the edge point and $E[\cdot]$ denotes the statistical expectation. Using (49) and (74), and knowing that [18], [30]

$$\Lambda = \frac{1}{\sqrt{E[x_0^2]}} \frac{n_0}{A} \qquad (76)$$

and

$$L_e = E[|x_0|] \frac{A}{E[|n(x)|]} = E[|x_0|] \frac{A}{n_0} \sqrt{\frac{\pi}{2}} \qquad (77)$$

we have, for the Gaussian function,

$$\sqrt{E[x_0^2]} \approx 1.15\sqrt{\sigma}\, \frac{\sqrt{E[n^2(x)]}}{A} = 1.15\sqrt{\sigma}\, \frac{n_0}{A} \qquad (78)$$

and

$$E[|x_0|] \approx 2.43\sigma\, \frac{E[|n(x)|]}{A} = 2.43\sigma \sqrt{\frac{2}{\pi}}\, \frac{n_0}{A} \approx 1.94\sigma\, \frac{n_0}{A} \qquad (79)$$

where $n(x)$ is additive white noise characterized by (43), and A is the amplitude of the step edge.

The other performance measure proposed by Shen and Castan (the noise to signal ratio) is also used as an optimization criterion (71), and relates the energy of the response of the filter in $x = 0$ with the energy of the noise at the output of the filter and at the output of the derivative of the filter. Filter (73) has, using this measure, $C_{\mathrm{N}} = 1$, while the derivative of the Gaussian and the Canny filter are characterized, respectively, by $C_{\mathrm{N}} \approx 1.25(1.26)$ [9] and $C_{\mathrm{N}} \approx 1.26$. Using Canny's signal to noise criterion, (48), we obtain $\Sigma = (-\ln b)^{-1/2}$. It is easy to verify that this value goes to zero when the filter narrows, i.e. when $b \to 0$, and goes to infinity when the filter widens, i.e. when $b \to 1$ (note that, in this case, it goes to the infinite support "difference of boxes" filter).

For a better comparison of the values of Σ obtained for the derivative of the Gaussian and for the Shen and Castan

[9]The value in parenthesis corresponds to the filter with support truncated at $x = \sigma/0.345$.

filter, we determine now an equivalent width common to both filters. With that purpose we will use three different measures: (1) the relation between the value of the function in the truncature point and its maximum value

$$A_0 \overset{\text{def}}{=} \left| \frac{f(1)}{\max_{x \in [0,1]} f(x)} \right| \tag{80}$$

(2) the relation between the absolute norm of the function in $[0,1]$ and in $[1,\infty)$

$$A_1 \overset{\text{def}}{=} \frac{\int_1^\infty |f(x)|\,dx}{\int_0^1 |f(x)|\,dx} \tag{81}$$

and (3) the relation between the quadratic norm of the function in $[0,1]$ and in $[1,\infty)$

$$A_2 \overset{\text{def}}{=} \sqrt{\frac{\int_1^\infty f^2(x)\,dx}{\int_0^1 f^2(x)\,dx}} \tag{82}$$

For the derivative of the Gaussian we have (using $\sigma = 0.345W$):

$$A_0 \approx 0.072, \quad A_1 \approx 0.015 \quad \text{and} \quad A_2 \approx 0.028$$

For the Shen and Castan filter: $A_0 = b$, and

$$A_1 = \frac{b}{(1-b)} \approx b, \quad A_2 = \sqrt{\frac{b^2}{(1-b^2)}} \approx b, \text{ for } b \ll 1$$

For the Shen and Castan filter this lead us to the following values of Σ: 0.62 (using A_0), 0.49 (using A_1) and 0.53 (using A_2). In Fig. 5 we can see the comparison between the Shen and Castan filter (for the three values of b calculated above) and the derivative of the Gaussian. Calculating the normalization factor for the filter of Shen and Castan, using (53), and making $W = 1$, we obtain

$$f_\mu(x) = \begin{cases} \dfrac{\ln b}{b-1} b^x & x > 0 \\[2mm] -\dfrac{\ln b}{b-1} b^{-x} & x < 0 \end{cases} \tag{83}$$

Although the filter of Shen and Castan may, at first sight, appear superior than the other filters that we presented, mainly due to its good localization capacity, it suffers from the problem of a deficient rejection of multiple responses near the edge point. As we saw above, the filters based on the three criteria of Canny present a finite first derivative for $x = 0$. This is imposed by the multiple responses criterion which goes to zero if $f(x)$ has discontinuities (see (46)). Therefore, the value of $x_{\max}$ for the filter of Shen and Castan is zero. The main problem that arises from this situation is the difficulty in identifying the maximum (of the response) that corresponds to the edge point, since it is generally surrounded by multiple maxima due to noise.
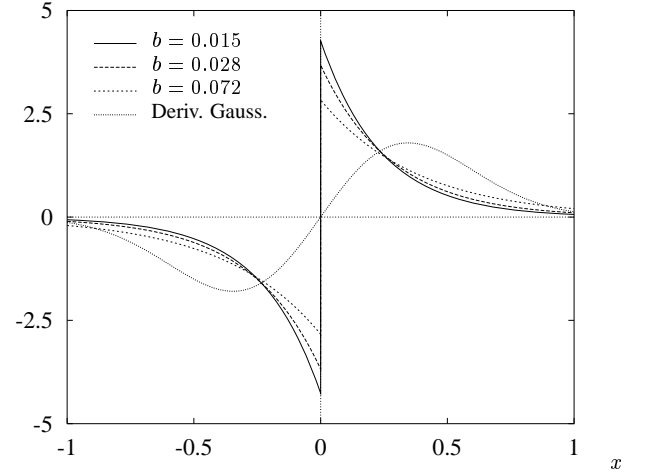

Figure 5 - Comparison of the curves of the derivative of a Gaussian filter and three instances of the Shen and Castan detection filter (for $b = 0.015$, $b = 0.028$ and $b = 0.072$).

IV. Discussion and conclusions

In this paper we presented an overview of several edge detectors based on the filtering and differentiation paradigm. In particular, those based on linear filters are, nowadays, among the most used. Therefore, the following discussion will focus on them.

Throughout the last section we presented and analyzed several linear filters developed according to well defined mathematical criteria (only with the exception of the filter of Marr and Hildreth). Given this list (although, for sure, not complete) we can raise the following question: which filter is the best? It is obvious that, according to each one's optimization criteria each filter shows better performance values than the others. But that is not a fair way to compare them. Around this question there are some considerations that we think they should be brought into discussion.

We start to recall that these filters were developed for one-dimensional signals. The extension to two dimensions were made *a posteriori*, frequently following procedures not well justified. On the other hand, the edge model used considers, at most, one edge point into the support region of the filter[10]. It is also evident that the step profile is, by far, the most used edge model, although some studies have been carried on using other models (see, for example, [34] for ramp edges). Note that the step edge model imposes two conditions difficult to be met in real images: on one hand, the transition is abrupt and, on the other hand, the intensity is kept constant on both sides of the edge. Finally, even the optimization criteria, and specially how to combine them, are not immune to criticisms, given the fact that there is not a clear definition of a real edge.

Even under these and other criticisms, linear filtering has a role of extreme importance in edge detection. Moreover, from what we saw in this paper, it seems that almost all

---

[10]In [30] Shen and Castan performed an additional analysis on the behavior of their filter for multiple edge points, arguing that this is necessary due to the infinite support of the filter. However, it is our opinion that this analysis is also necessary for the finite support filters. In fact, although infinite, it is also true that the response of the filter has a fast decay which, in that aspect, does not differ too much from the finite support filters.

filters exhibit some similarity with the Gaussian function and its derivatives. This observation lead us to believe that the "optimum" linear filter for the detection of step edges should not differ too much from the derivative of the Gaussian and, therefore, the smoothing filter should be based on the Gaussian function. This is not surprising since the Gaussian has been emerging as a very important function in several areas of image analysis and processing and, specially, in multi-resolution analysis.

We can find, in the literature, some unification proposals of some of these filters. For example, Farag showed that the filters of Dickey and Shanmugam, Marr and Hildreth, and Canny are filters of maximum energy for step edges [35]. Shen and Castan also proposed a unification of several band-limited differential operators, based on their filter (see (72)) [33] and on a property stating that applying filter (72) repeatedly is equivalent to filter with a Gaussian. Therefore, all the filters similar to the Gaussian function can also be related. Moreover, they showed that Deriche's (smoothing) filter corresponds exactly to two iterations of their filter.

The regularization theory also contributes to the unification of all these filters. It can be verified that all of them obey the Tikhonov conditions, which means that the (smoothing) filters provide regularization of the data. It is curious to note that even the filter of Shen and Castan is within those conditions (remember the discontinuity of the first derivative). In fact, Poggio et al. [27] noted that if we use the simplest Tikhonov stabilizing functional [11] (i.e., the first derivative operator), then the regularizing filter has a discontinuity in the origin of its first derivative. Moreover, they stated that those filters are not sufficiently smooth to ensure that the second derivatives are well defined and, therefore, they don't advise its use.

Most often, we can separate the smoothing part from the differentiation part of the edge detectors. As we saw, this is the idea behind the formulation of the edge detection problem as a problem of numerical differentiation. This suggests us that, although the filters that we presented had been developed in a more or less tight association with a specific order of derivative (first or second) they might be used with differentiators of different orders [12].

This question lead us to one of the most persistent discussions in the field of edge detection, respecting to the order of the differentiator. Step edges can be marked using the maxima of the first derivative or the zero crossings of the second derivative. However, there are some basic differences between these two approaches although, obviously, they are strongly related.

One is related to the importance of having closed contour lines. Due to some mathematical properties, the edges that are marked using the zero crossings of the second derivative form closed lines or end at the image boundaries. However, this nice property has a price, implying the introduction of spurious edges due to inflection points in the image function (see, for example, [2]). Also, a known drawback of using

second order derivatives is related to high-frequency noise amplification (differentiation amplifies the high-frequency components of the signal).

We agree with De Micheli et al. [36] regarding the idea that most of the linear filters available for edge detection provide results quite similar, making the task of identifying the "best" difficult or even impossible. Nevertheless, due to its simplicity and also to its relation to most of the proposed filters, it seems that the Gaussian function is indeed a well balanced choice.

REFERENCES

[1] V. Torre and T. Poggio. On edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1986, **8**, pp. 147–163.

[2] J. J. Clark. Authenticating edges produced by zero-crossing algorithms. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1989, **11**, pp. 43–57.

[3] M. Bertero, T. A. Poggio, and V. Torre. Ill-posed problems in early vision. *Proceedings of the IEEE*, 1988, **76**, pp. 869–889.

[4] D. H. Ballard and C. M. Brown. *Computer vision*. Prentice-Hall, Inc., 1982.

[5] R. C. Gonzalez and P. Wintz. *Digital image processing*. Addison Wesley, 2nd edition, 1987.

[6] A. K. Jain. *Fundamentals of digital image processing*. Prentice-Hall, Inc., 1989.

[7] W. K. Pratt. *Digital image processing*. Wiley-Interscience, 2nd edition, 1991.

[8] A. Rosenfeld. A nonlinear edge detection technique. *Proceedings of the IEEE*, 1970, **58**, pp. 814–816.

[9] A. Rosenfeld and M. Thurston. Edge and curve detection for visual scene analysis. *IEEE Trans. on Computers*, 1971, **20**, pp. 562–569.

[10] A. Rosenfeld and A. C. Kak. *Digital picture processing*. Volume 2, Academic Press, 1982.

[11] E. Argyle. Techniques for edge detection. *Proceedings of the IEEE*, 1971, **59**, pp. 285–287.

[12] I. D. G. Macleod. Comments on "Techniques for edge detection". *Proceedings of the IEEE*, 1972, **60**, p. 344.

[13] F. M. Dickey and K. S. Shanmugam. Optimum edge detection filter. *Applied Optics*, 1977, **16**, pp. 145–148.

[14] K. S. Shanmugam, F. M. Dickey, and J. A. Green. An optimal frequency domain filter for edge detection in digital pictures. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1979, **1**, pp. 37–49.

[15] W. H. H. J. Lunscher. The asymptotic optimal frequency domain filter for edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1983, **5**, pp. 678–680.

[16] D. Marr and E. Hildreth. Theory of edge detection. *Proc. Royal Society of London, B*, 1980, **207**, pp. 187–217. Reprinted in [37].

[17] J. F. Canny. *Finding edges and lines in images*. Technical Report AI–TR–720, M.I.T. Artificial Intell. Lab., Cambridge, MA, 1983.

[18] J. F. Canny. A computational approach to edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1986, **8**, pp. 679–698.

[19] H. D. Tagare and R. J. P. de Figueiredo. On the localization performance measure and optimal edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1990, **12**, pp. 1186–1190.

---

[11]P in Eq. (66).

[12]Note that this may not be possible for the Shen and Castan filter, due to the reasons that we pointed above.

[20] K. L. Boyer and S. Sarkar. Comments on "On the localization performance measure and optimal edge detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1994, **16**, pp. 106–108.

[21] H. D. Tagare and R. J. P. de Figueiredo. Reply to "On the localization performance measure and optimal edge detection". *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1994, **16**, pp. 108–110.

[22] J. Koplowitz and V. Greco. On the edge localization error for local maximum and zero-crossing edge detectors. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1994, **16**, pp. 1207–1212.

[23] L. A. Spacek. Edge detection and motion detection. *Image and Vision Computing*, 1986, **4**, pp. 43–56.

[24] M. Petrou and J. Kittler. On the optimal edge detector. In *Proc. of the Alvey Vision Conf.*, Manchester, UK, 1988, pp. 191–196.

[25] R. Deriche. Optimal edge detection using recursive filtering. In *Proc. of the 1st Int. Conf. on Computer Vision, ICCV'87*, London, England, 1987, pp. 501–505.

[26] R. Deriche. Using Canny's criteria to derive a recursively implemented optimal edge detector. *Int. J. of Computer Vision*, 1987, **1**, pp. 167–187.

[27] T. Poggio, H. Voorhees, and A. Yuille. A regularized solution to edge detection. *J. of Complexity*, 1988, **4**, pp. 106–123.

[28] S. Sarkar and K. L. Boyer. Optimal, efficient, recursive edge detection filters. In *Proc. of the 10th Int. Conf. on Pattern Recognition, ICPR'90*, Atlantic City, USA, 1990, pp. 931–936.

[29] S. Sarkar and K. L. Boyer. On optimal infinite impulse response edge detection filters. *IEEE Trans. on Pattern Analysis an Machine Intelligence*, 1991, **13**, pp. 1154–1171.

[30] J. Shen and S. Castan. An optimal linear operator for step edge detection. *CVGIP: Graphical Models and Image Processing*, 1992, **54**, pp. 112–133.

[31] S. Castan, J. Zhao, and J. Shen. New edge detection methods based on exponential filter. In *Proc. of the 10th Int. Conf. on Pattern Recognition, ICPR'90*, Atlantic City, USA, 1990, pp. 709–711.

[32] S. Castan, J. Zhao, and J. Shen. Optimal filter for edge detection methods and results. In *Computer Vision - ECCV'90, 1st European Conf. on Computer Vision*, Antibes, France, 1990, pp. 13–17.

[33] J. Shen and S. Castan. Towards the unification of band-limited derivative operators for edge detection. *Signal Processing*, 1993, **31**, pp. 103–119.

[34] M. Petrou and J. Kittler. Optimal edge detectors for ramp edges. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1991, **13**, pp. 483–491.

[35] A. A. Farag. Edge-based image segmentation. *Remote Sensing Reviews*, 1992, **6**, pp. 95–122.

[36] E. De Micheli, B. Caprile, P. Ottonello, and V. Torre. Localization and noise in edge detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 1989, **11**, pp. 1106–1117.

[37] R. Kasturi and R. C. Jain, eds. *Computer vision: principles*. IEEE Computer Society Press, Los Alamitos, CA, 1991.