SIAM J. IMAGING SCIENCES
Vol. 5, No. 1, pp. 343–365

# Edge Detection Filter based on Mumford–Shah Green Function[*]

Sasan Mahmoodi[†]

**Abstract.** In this paper, we propose an edge detection algorithm based on the Green function associated with the Mumford–Shah segmentation model. This Green function has a singularity at its center. A regularization method is therefore proposed here to obtain an edge detection filter known here as the *Bessel* filter. This filter is robust in the presence of noise, and its implementation is simple. It is demonstrated here that this filter is scale invariant. A mathematical argument is also provided to prove that the gradient magnitude of the convolved image with this filter has local maxima in discontinuities of the original image. The Bessel filter enjoys better overall performance (the product of the detection performance and localization indices) in Canny-like criteria than the state-of-the-art filters in the literature. Quantitative and qualitative evaluations of the edge detection algorithms investigated in this paper on synthetic and real world benchmark images confirm the theoretical results presented here, indicating the scale invariant property of the Bessel filter. The numerical complexity of the algorithm proposed here is as low as any convolution-based edge detection algorithm.

**Key words.** edge detection filter, Canny-like criteria, Mumford–Shah functional, Green function, modified Bessel function, derivative of Bessel

**AMS subject classifications.** 44A35, 93E11, 60G35, 68U10, 62H35

**DOI.** 10.1137/100811349

## 1. Introduction.

Boundary and edge detection is a well-established topic in image processing. Linear filtering is one of the most popular edge detection methods due to its simplicity in implementation and low numerical complexity. Derivative filters as a type of linear filters have been extensively studied in the literature. Early examples of derivative filters include Roberts, Sobel, and Prewitt filters. Maximum signal to noise ratio (SNR), localization, and least multiplicity of response of a filter are introduced by Canny as criteria in his seminal paper on computational edge detection [1] to design optimal filters for feature detection, in particular edge detection. His calculations lead to an optimal filter which looks like derivative of Gaussian ($DroG$) to detect edges. This observation leads him to choose $DroG$ as an optimal filter for edge detection. An interesting property of $DroG$ is to smooth (regularize) the image before the derivative operation. Canny's work stimulated further research in the field. By approximating Canny's optimal filter, an efficient recursive infinite impulse response (IIR) filter is then proposed by Deriche in [2] to design a filter similar to $DroG$ in frequency domain. Deriche's filter enjoys a constant computational time regardless of the filter size. Canny's $DroG$ is an approximate to his optimal filter, and therefore with the localization

---

[*] Received by the editors October 11, 2010; accepted for publication (in revised form) December 15, 2011; published electronically March 8, 2012. This work was supported in part by the IST program of the European Community under the PASCAL2 Network of Excellence, the IST-2007-216886, and the PinView project under grant 216529.

http://www.siam.org/journals/siims/5-1/81134.html

[†] School of Electronics and Computer Science, Building 1, Southampton University, Southampton, SO17 1BJ, UK (sm3@ecs.soton.ac.uk).

term that he has defined, his calculations do not precisely lead to $DroG$. The localization index is therefore reformulated in [3] by considering the distribution of zero crossings based on the theory of stochastic processes to demonstrate that $DroG$ is precisely an optimal filter when such a localization index is considered. Canny's optimization method to find an optimal filter is based on local maxima detection.

Such a method is adapted to design zero crossing edge detector filters in [4]. An efficient implementation of such filters is also proposed in [4] by realizing an IIR filter with recursive filtering techniques to construct filters requiring constant computation time with respect to filter size. Although this filter enjoys lower computational time with larger filter sizes than Canny's, it is more sensitive to noise due to the zero crossing method used in edge detection. An optimal IIR filter is also designed in [5] by defining an effective width for the infinite length filter. This filter, whose implementation is simple and computationally more efficient than $DroG$, outperforms Canny's filter and enjoys constant computational time regardless of filter size. Petrou and Kittler [6] employ Canny-like criteria to propose an optimal filter for ramp edges by arguing that edges in real world images are not perfect step functions and are closer to ramps. The Petrou–Kittler filter outperforms $DroG$ and Deriche's filters for ramp edges. Separable one-dimensional (1D) optimal filters for ramp edges are further proposed in [7] to replace a two-dimensional (2D) optimal filter developed in [6] to improve the efficiency of the Petrou–Kittler edge detection algorithm, while the quality of performance is not compromised. This separable 1D implementation of the Petrou–Kittler filter is adaptable to the slope of ramp edges and is faster than the Petrou–Kittler 2D optimal filter, while it enjoys the same quality of performance. Shen and Castan [8] also propose an infinite symmetric exponential filter by optimizing only SNR and localization criteria. A recursive realization is also implemented for this filter in [8] to demonstrate that such a filter has a better performance in insensitivity to noise, precision of edge localization, and reduced computational complexity than Canny and Deriche's filters.

A template matching method is developed by optimizing discriminative signal to noise ratio (DSNR) instead of SNR in [9]. This method is robust under noise and occlusion, and its optimization leads to a Wiener filter, which is an optimal step exponential filter in response to a bipolar step (signum) function. This filter shows superiority to $DroG$ in terms of Pratt's figure of merit [10] and detects edges with more details of the image than $DroG$ as the filter size increases. The template matching method is also employed by Wang, Rao, and Ben-Arie in [11] to propose an optimal filter for bipolar ramp edges by using a ramp as an template for edges. This filter demonstrates less distortion and higher figures of merit than the optimal filters for ramp edges proposed in [1, 6, 7]. Demigny also employs Canny's criteria for the optimal design of filters for edge detection in [12]. However, the approach proposed in [12] differs from Canny's in [1] in two ways: (i) Discrete signals rather than continuous ones are considered for the design of an optimal filter. (ii) Pulses instead of step functions are considered to take into account the effects of neighboring edges. Demigny concludes that the optimization of the product of detection performance and localization indices leads to an optimal filter outperforming $DroG$, Deriche, and Shen filters. Canny-like criteria are optimized directly in two dimensions by Jacob and Unser [13] to design optimal steerable filters originally introduced in [14]. In this method, a Gaussian filter is used as a basic window function. A steering filter is then considered to consist of the weighted sum of partial derivatives with various orders of the basic Gaussian filter. By optimizing Canny-like criteria, the weights associated with each portion of the steering filter are calculated. The steering filters proposed by Jacob and Unser [13] demonstrate better results than Canny's at the presence of noise. Their designing technique is also employed to develop filters to better detect discontinuities forming sharp corners.

The technique of scale multiplication is also proposed in [15] to employ multiple scales to enhance the localization criterion in the *DroG* filter for edge detection. On the other hand, Mumford and Shah [16] propose a variational model for segmentation consisting of three terms: (1) fidelity term indicating the desired solution should be as close as possible to the original image, (2) smoothing term responsible for the smoothness of the desired solution, (3) contour length minimization term regularizing (smoothing) the contour(s) representing discontinuities and removing undesired contours. In their influential paper, Mumford and Shah [16] introduce a Green function associated with a linear positive-definite self-adjoint operator obtained by minimizing their functional in the absence of boundaries [16]. This Green function is known to be the modified Bessel function of the second kind and zero degree.

Our contribution in this paper is to propose an edge detection filter based on this Green function and explore its mathematical properties when it is used for edge detection. We demonstrate that such a filter produces scale invariant edges and enjoys better overall performance measured by the product of detection performance and localization indices. The rest of this paper consists of the following sections: in section 2, we briefly discuss the Mumford–Shah (MS) functional and the modified Bessel function of the second kind and zero degree and explain how this Bessel function can be used for edge detection. Implementation issues, detection performance, and localization indices associated with this Bessel function are discussed in section 3; and numerical results are presented in section 4. Finally, the paper concludes in section 5.

## 2. MS Green function.

The original MS model introduced in [16] proposes minimizing the following energy term:

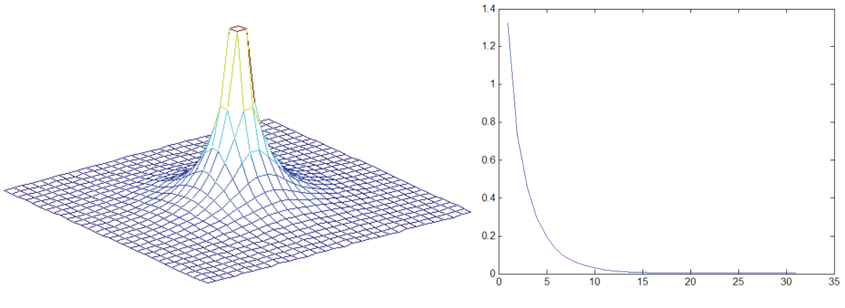
$$E(u,K) = \mu \int_{\Omega\backslash K} |\nabla u|^2 dxdy + \int_{\Omega} |u-g|^2 dxdy + \beta\,|\Gamma|\,, \tag{1}$$

where $\Omega, u \in C^1, g, \Gamma, \mu,$ and $\beta$ are the image domain, the smoothed piecewise-continuous image, the original image, a finite union of smooth arcs, and nonnegative constant coefficients, respectively. In the absence of boundaries, the Green function associated with a linear positive-definite self-adjoint operator obtained by minimizing functional (1) with respect to $u$ for the whole plane is a modified Bessel function of the second kind and zero degree [16]. A closed form of this Green function is written as

$$h(x,y;v,w) = K_0\left(\sqrt{\frac{((x-v)^2+(y-w)^2)}{\mu}}\right), \tag{2}$$

where $K_0(.)$ is the modified Bessel function of the second kind and zero degree and $(x,y),(v,w) \in R^2$. A modified Bessel function of the second kind and zero degree is one of the solutions of the modified Bessel differential equation with zero degree, i.e.,

$$x\frac{d^2K_0}{dx^2} + \frac{dK_0}{dx} - xK_0 = 0, \qquad x > 0.$$



(a) (b)

**Figure 1.** *The MS Green function for* $\mu = 10$ : (a) 3*D view,* (b) *a cross-section.*

An integral representation of this modified Bessel function is written as [17]

$$K_0(x) = \int_0^{+\infty} \cos\left(x \sinh\left(t\right)\right) dt = \int_0^{+\infty} \frac{\cos(xt)}{\sqrt{1+t^2}} dt, \qquad x > 0.$$

This modified Bessel function is singular at $x = 0$, i.e., $\lim_{x \to 0} K_0(x) \to +\infty$. The asymptotic behaviors of $K_0(x)$ for small and large values of $x$ are as follows [16, 17]:

$$K_0(x) \sim \ln \frac{1}{x}, \qquad \text{for small } x > 0,$$
$$K_0(x) \sim \sqrt{\frac{\pi}{2x}} e^{-x}, \qquad \text{for large } x > 0.$$

In the absence of any boundary, the solution $u$ is calculated by using Green function (2) as

$$u(x,y) = \int_{R^2}\int h(x,y;v,w) g(v,w) dv dw. \tag{3}$$

According to (2), $h(x,y;v,w)$ is a linear space invariant function, i.e., $h(x,y;v,w) = h(x-v, y-w)$. Therefore (3) can be written as a convolution:

$$u(x,y) = h^* g, \tag{4}$$

where * represents the 2D convolution. A three-dimensional (3D) view of Green function (2) for $(v,w) = (0,0)$ is shown in Figure 1(a). The cross-section of this Green function is also depicted in Figure 1(b). In this paper, by employing (4), we present two theorems (Theorems 1 and 2) to establish a theoretical framework to develop an edge detection algorithm based on convolution by using the modified Bessel function of the second kind and zero degree. It should be noted that it is not the objective of this work to find the solutions of the nonlinear MS functional using a linear method. However, we aim here to explore the mathematical properties of the convolution of the modified Bessel function shown in Figure 1 with an image containing discontinuities and to propose an edge detection algorithm based on these properties. Hence in our mathematical arguments, the function $u(x,y)$ is treated as the

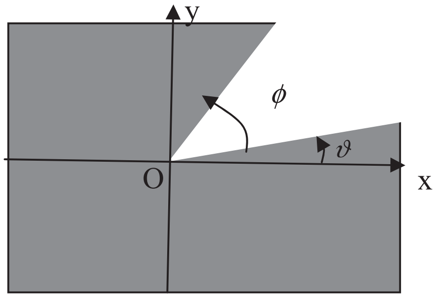


**Figure 2.** *An original image containing discontinuities forming a sharp corner at the center of coordinates.*

image calculated according to (4) by convolving the original image $g$ with the Bessel function $h$ given in (2) and shown in Figure 1, and $u(x, y)$ is therefore known in the rest of this paper as the convolved image. Theorems 1 and 2 presented in this section lay a theoretical foundation to develop a practical edge detection filter for digital images defined on a finite lattice. Throughout this paper, we name step or jump discontinuities in the grey scale values of images as image discontinuities. According to the theorems proved in this section, this Bessel filter is scale invariant and therefore enjoys a high localization index enabling the edge detection method proposed here to detect discontinuities forming sharp corners regardless of the filter size. In Theorem 1, we prove that the gradient magnitude of the convolved image $u$ is unbounded in the discontinuities of the original image $g$ shown in Figure 2.

Theorem 1. *The gradient magnitude of the convolved image $u : R^2 \to R^+$ calculated by (4) is unbounded on the discontinuities of a given piecewise-constant image $g : \Omega \to R^+$, where $\Omega$ and $R^+$ are the image domain and the set of positive real numbers, respectively.*

The proof to Theorem 1 is presented in Appendix A.

In the following theorem, we prove that $|\nabla u|$ has local maxima in the discontinuities (including sharp corners) of the image of Figure 2. This is an important property of the Bessel filter which many of edge detection filters in the literature do not possess. The results obtained from Theorems 1 and 2 are the main reasons for the scale invariance property and the high localization index of the Bessel filter.

Theorem 2. *The gradient magnitude of the convolved image $u : R^2 \to R^+$ calculated in (4) has local maxima on discontinuities of a given piecewise-constant image $g : \Omega \to R^+$.*

The proof of Theorem 2 is given in Appendix B. Theorems 1 and 2 are proved for piecewise-constant images; however, we conjecture that these theorems are also true for piecewise-continuous images. The numerical and experimental results presented in section 4 support such a conjecture.

## 3. Edge detection and its implementation issues.

We use the main results obtained from Theorems 1 and 2 to propose an edge detection algorithm. The convolution of function $h$ in (2) with the original input image $g$ is calculated as the first step for the algorithm proposed here to calculate the convolved image $u$. Then according to Theorem 1, $|\nabla u|$ is examined to find regions for which $|\nabla u|$ is infinity. As shown in Figure 1, Bessel function (2) is singular at the center, which is not numerically tractable for implementation. It is therefore important for numerical purposes to regularize function $h$.

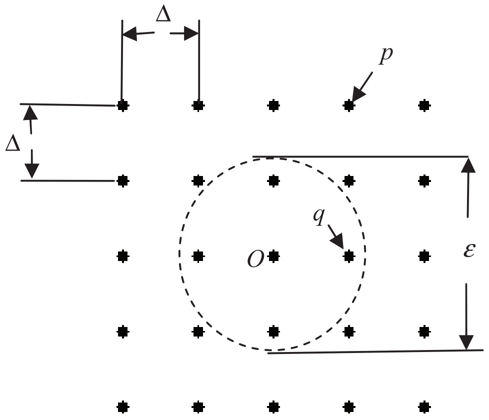


**Figure 3.** *A* $5 \times 5$ *grid used for the construction of a Bessel filter for the case of* $2 < \frac{\varepsilon}{\Delta} < 2\sqrt{2}$.

Here we suggest the following regularized and normalized function known as *Bessel* filter in this paper.

$$h_\varepsilon(x) = \begin{cases} \dfrac{K_0\left(\frac{|x|}{\sqrt{\mu}}\right)}{K_0\left(\frac{\varepsilon}{2\sqrt{\mu}}\right)}, & |x| > \dfrac{\varepsilon}{2}, \\ 1, & |x| \le \dfrac{\varepsilon}{2}. \end{cases} \tag{5}$$

We notice that Theorem 1 does not hold with the regularization method indicated in (5); however, the gradient magnitude of the convolved image $u$ will still have large values and be maximized on the image discontinuities provided that $\varepsilon$ is small. The above statement is true for images defined on both the continuous domain ($R^2$) and the discrete domain (finite subsets of $Z^2$). Again by choosing small values for $\varepsilon$ in this paper, the requirement $\varepsilon \to 0$ indicated in the proof of Theorem 2 is also met, and therefore the conclusion drawn from Theorem 2 still holds for both the continuous and discrete domain images.

The algorithm proposed here is implemented in the MATLAB 7.3 environment. We have exploited the built-in *besselk* function in MATLAB to construct the Bessel filter. Figure 3 helps us understand how the regularized Bessel filter proposed in (5) can be constructed. A $5 \times 5$ window forming the window grids of the filter is shown in this figure. It is noted that $\Delta$ and $\varepsilon$ in the figure are the sampling distance and regularizing parameter, respectively. The dashed line circle in Figure 3 represents the circle with radius $|x| = \frac{\varepsilon}{2}$. According to (5), the values of the filter in all points inside this circle should be unity and the filter values in all points outside of this circle should be $\frac{K_0(\frac{|x|}{\sqrt{\mu}})}{K_0(\frac{\varepsilon}{2\sqrt{\mu}})}$, where $|x|$ is the distance between the central point $O$ and the point outside of the circle. Therefore the distance ($|x|$) between a point in the window and the central point $O$ is calculated. If $|x| > \frac{\varepsilon}{2}$, then the point is outside of the dashed line circle (such as point $p$ shown in the figure) and the value of $\frac{K_0(\frac{|x|}{\sqrt{\mu}})}{K_0(\frac{\varepsilon}{2\sqrt{\mu}})}$ is therefore assigned to the filter at point $p$. For a point inside the circle (i.e., $|x| < \frac{\varepsilon}{2}$ such as for points $q$ and $O$), unity is assigned to the filter at this point. For $\frac{\varepsilon}{\Delta} < 1$, the dashed circle contains

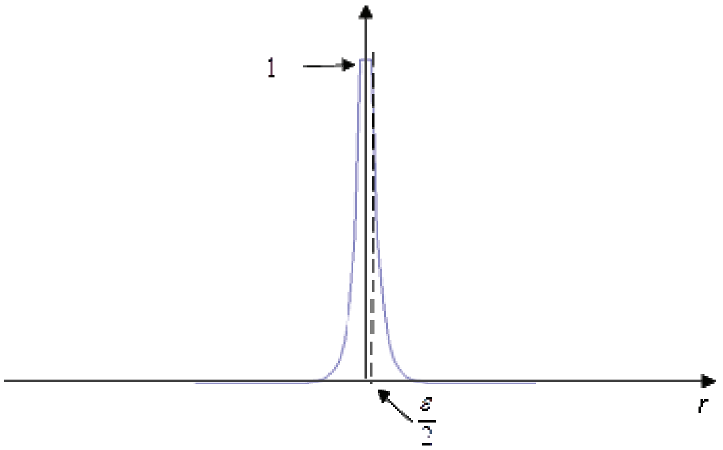


**Figure 4.** *The cross-section of the regularized filter proposed in* (5).

only the central point $O$, and therefore the value of the filter at only the central point $O$ is set to unity. We consider $\Delta$ being unity throughout this paper. Having constructed the filter, the derivatives of the filter with respect to $x$ and $y$ are convolved with the original image to compute $\frac{\partial u}{\partial x}$ and $\frac{\partial u}{\partial y}$, which can then be used to calculate $|\nabla u| = \sqrt{(\frac{\partial u}{\partial x})^2 + (\frac{\partial u}{\partial y})^2}$. Due to Theorem 2, the gradient magnitude $|\nabla u|$ is then examined to find maximum points rather than infinity and nonmaximum suppression is applied to avoid multiple line detection for edges. A cross-section of the constructed regularized Bessel filter is also depicted in Figure 4.

Since this edge detection algorithm is based on the convolution of a Bessel filter with any input image, the numerical complexity of the algorithm could be the same as other convolution-based edge detection algorithms. However, the filter proposed here is implemented by using an FFT algorithm, and its numerical complexity is therefore of the order of $O(N\mathrm{Log}N)$ for an $N \times N$ input image.

Let us now consider the regularized filter (5) to compare its detection performance and localization indices with some existing filters in the recent literature. We therefore consider the 1D version of filter (5) due to the 1D nature of the performance indices used in the literature. The first and second derivatives of filter (5) are required for the calculation of the detection performance and localization indices. These derivatives are calculated in Appendix C. The detection performance $\Sigma_\varepsilon$ as defined in [1] is written as

$$\Sigma_\varepsilon = \frac{\left|\int_{-\infty}^{0} f_\varepsilon(x)dx\right|}{\sqrt{\int_{-\infty}^{+\infty} f_\varepsilon^2(x)dx}}, \tag{6}$$

where $f_\varepsilon(x)$ is calculated in the Appendix C. By using (C-1) (in Appendix C), the detection performance can be written as

$$\Sigma_\varepsilon = \frac{\frac{1}{\sqrt{\mu}} K_0\left(\frac{\varepsilon}{2\sqrt{\mu}}\right)}{\sqrt{2\int_{+\frac{\varepsilon}{2}}^{+\infty}\left(K_0'\left(\frac{x}{\sqrt{\mu}}\right)\right)^2 dx}}, \tag{7}$$

where $K_0(x)$ and $K_0'(x)$ are the modified Bessel function of the second kind and zero degree and its derivative, respectively. Localization index $\Lambda_\varepsilon$ as defined in [1] is written as

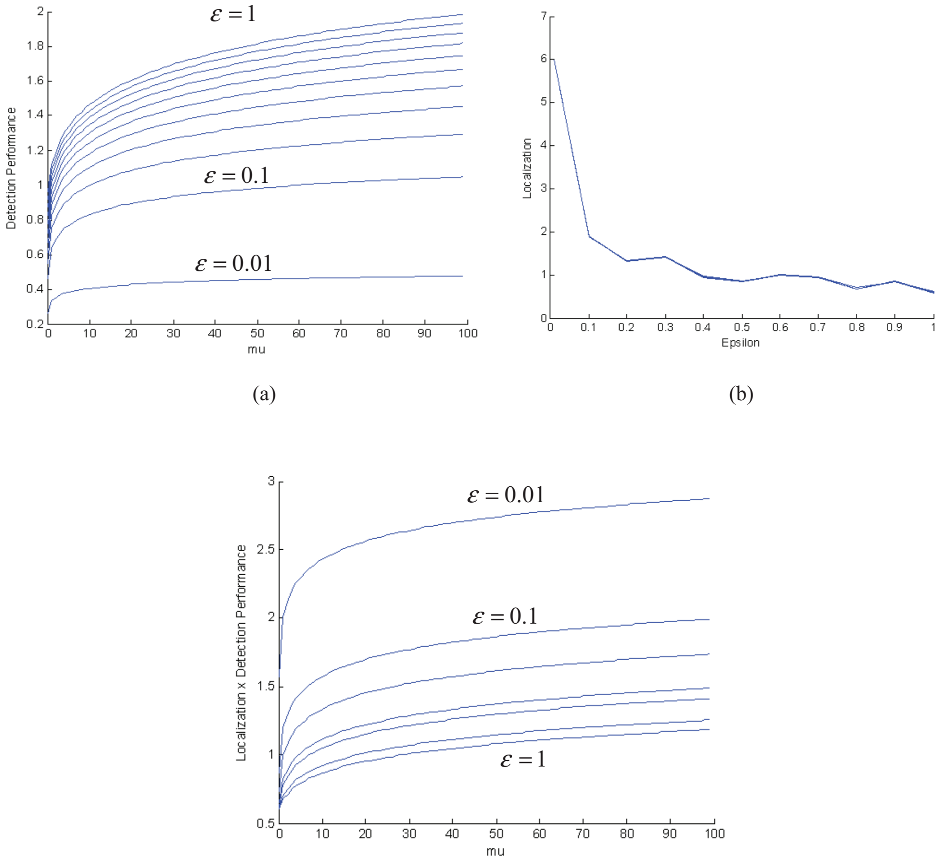
$$\Lambda_\varepsilon = \frac{|f_\varepsilon'(0)|}{\sqrt{\int_{-\infty}^{+\infty} f_\varepsilon'^2(x)dx}}. \tag{8}$$

For large values of $\varepsilon$, according to (C-2), $f_\varepsilon'(0) = 0$, and therefore $\Lambda_\varepsilon = 0$. However, as $\varepsilon \to 0, \Lambda_\varepsilon$ can be written as

$$\Lambda_\varepsilon = \frac{K_0''\left(\frac{\varepsilon}{2\sqrt{\mu}}\right)}{\sqrt{2\int_{+\frac{\varepsilon}{2}}^{+\infty}\left(K_0''\left(\frac{x}{\sqrt{\mu}}\right)\right)^2 dx}}, \tag{9}$$

where $K_0''(x)$ is the second derivative of the Bessel function. In the above calculations for $\Sigma_\varepsilon$ and $\Lambda_\varepsilon$, it is assumed that $\varepsilon > 0, \varepsilon \to 0$ but $\varepsilon \neq 0$. In Figure 5, the detection performance and localization indices for the Bessel filter are numerically calculated with respect to $\mu$ and $\varepsilon$ using (7) and (9). It is clear from Figure 5(a) that as $\varepsilon \to 0$, the detection performance decreases. However, by increasing $\mu$ (for a given value for $\varepsilon$), the detection performance improves. Figure 5(b) shows how the localization changes when $\mu$ and $\varepsilon$ are varied. Changes in $\mu$ have almost no effect on localization (for a given $\varepsilon$, the localization does not change with respect to $\mu$). This independence in localization with respect to parameter $\mu$ is a consequence of Theorem 2. The scale invariance property associated with the Bessel filter is also concluded from Theorem 2 and numerically supported in Figure 5(b). As lower values are chosen for $\varepsilon$, the localization increases. Figure 5(c) shows how the product of the detection performance and localization ($\Sigma_\varepsilon\Lambda_\varepsilon$) change with respect to $\varepsilon$ and $\mu$. As shown in Figure 5(c), the product $\Sigma_\varepsilon\Lambda_\varepsilon$ increases as $\varepsilon$ decreases or $\mu$ increases. In our implementation for discrete images defined on some subset of $Z^2$, we choose $\varepsilon$ to be small enough so that $\frac{\varepsilon}{\Delta} \in [0.01, 1]$, where $\Delta$ is sampling distance, which is considered unity in this paper. For implementation purposes, a Bessel filter with the same size as the input image in frequency domain is constructed. An FFT algorithm is then employed to apply the filter to the input image in the frequency domain. However, we need to determine the window size for a Bessel filter for the following reasons:

(i) We want to ensure that the parameters of the constructed Bessel filter (which has the same size as the input image) are selected correctly to avoid any truncation error.
(ii) We want to compare the performance of the Bessel filter with the other edge detection algorithms employing convolution (rather than an FFT algorithm) for their implementation in the literature where all the filters under investigation have a certain size. Such a treatment allows us to therefore find appropriate parameters of the Bessel filter for a certain size.
(iii) We want to investigate the effects of scaling in edge detection by changing the filter size.



(a)

(b)

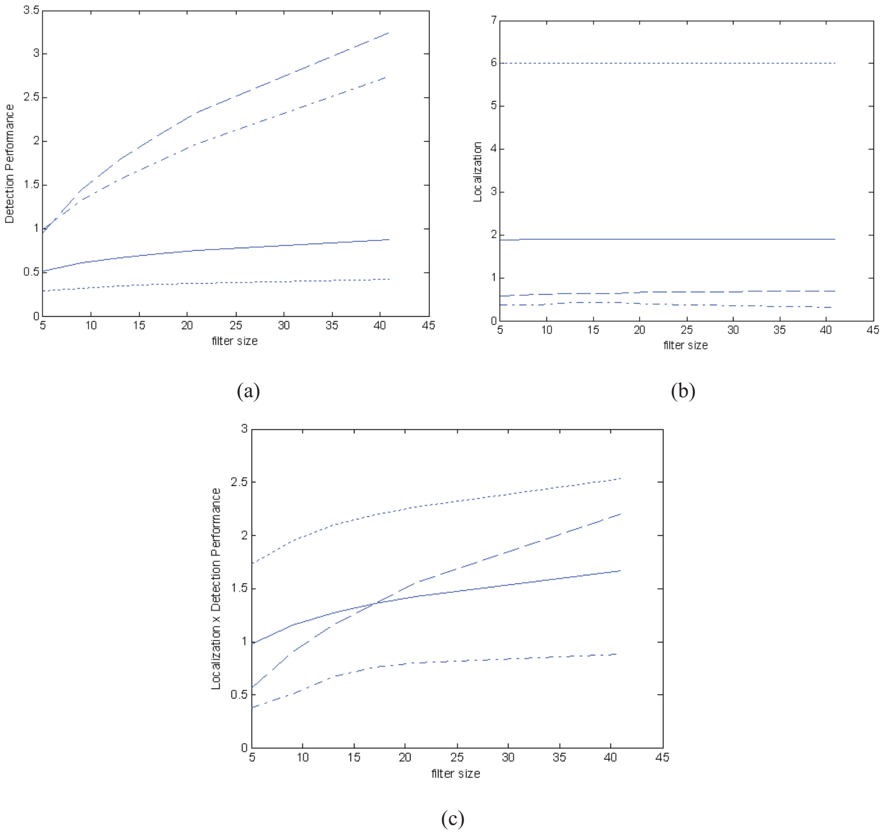
(c)

**Figure 5.** *The Bessel filter indices* (a) *the detection performance and* (b) *localization and* (c) *the product of the detection performance and localization* $\Sigma_\varepsilon \mathrm{A}_\varepsilon$ *are plotted with respect to* $\varepsilon$ *and* $\mu$ *(see the text for more details).*

To this end, we define a truncated Bessel filter to find an appropriate value for $\mu$ for a Bessel filter with a certain size:

$$h_\varepsilon(r,d) = \begin{cases} 1, & 0 \le r < \dfrac{\varepsilon}{2}, \\ \dfrac{K_0\left(\frac{r}{\sqrt{\mu}}\right)}{K_0\left(\frac{\varepsilon}{\sqrt{\mu}}\right)}, & \frac{\varepsilon}{2} < r \le d, \\ 0, & r > d. \end{cases}$$

We use the term $E(d) = \sum_{r=0}^{d} |h_\varepsilon(r) - h_\varepsilon(r,d)|$ to determine a filter size for the Bessel filter with respect to $\mu$. We notice that for $\frac{d}{\sqrt{\mu}} = 5$, $E$ becomes negligible ($E = 0.0045$). Therefore, in this paper an equivalent minimum size for the filter proposed here is defined as $W = (10\sqrt{\mu} + 1) \times (10\sqrt{\mu} + 1)$. It should be noted that in practice the size of the Bessel filter is the same size as the input image. The relationship between the size and the Bessel filter parameter mentioned above helps us only to calculate the filter parameter $\mu$ according to a certain size for comparison purposes. Intuitively this ensures that a Bessel filter with



(a) (b)

(c)

**Figure 6.** *The comparison of indices for Bessel filter with* $\varepsilon = 0.01$ *(dotted line), Bessel filter with* $\varepsilon = 0.1$ *(solid line), DroG (dash dot line), and Demigny (dashed line):* (a) *The detection performance index* $\Sigma$*,* (b) *localization index* $\Lambda$*, and* (c) $\Sigma\Lambda$*.*

parameter $\mu$ and size $(10\sqrt{\mu} + 1)$ termed as "equivalent minimum size" produce the same results as a Bessel filter with parameter $\mu$ and the size of the input image. Hence conclusive comparisons can be made between our filter whose implementation is based on FFT and other filters which are based on convolution in the literature. By employing equivalent minimum size, it is also possible to investigate how the detected edges are affected by varying the scale (and as a result the equivalent minimum size). In the rest of this paper, if in any experiment a size is indicated for a Bessel filter, it therefore denotes the equivalent minimum size. This is also true for a Rao–Ben-Arie filter [9] which is implemented by using an FFT algorithm.

Figure 6 depicts the detection performance, localization and their product with respect to the filter size of *DroG*, Demigny, and Bessel filters with two different values for $\varepsilon$. As shown in the figure, the localization of the filter proposed here increases and its detection performance decreases as lower values of $\varepsilon$ are chosen, which confirms the result shown in Figure 5. Detection performance for the Demigny filter demonstrates the best performance (Figure 6(a)). On the other hand, the localization of the Bessel filter for both values of $\varepsilon$ is superior to the localization of the other two filters (Figure 6(b)). Finally, Figure 6(c) depicts the product of the detection performance and localization. As indicated in [1], the product of the indices is more important than the individual indices. The Bessel filter with $\varepsilon = 0.01$ demonstrates the best overall performance (product of indices) over the whole range of filter sizes. The main reason for this superior performance of the Bessel filter is due to the high localization index which is a consequence of its asymptotic properties near its singularity as demonstrated by Theorems 1 and 2.

**4. Numerical results.** In this section, the filter proposed here is compared with some convolution-based techniques in the literature such as *DroG*, Rao–Ben-Arie, and Jacob–Unser edge detection algorithms [1, 12, 9, 13]. We notice that it is difficult to compare edge detection algorithms with their optimal parameters for a specific image. This is because the optimal performance of a filter depends on its parameters in a specific application. However, for a fair comparison, in this section we compare the performance of the filters when the same size is used for all filters. The three Jacob–Unser filters employed in this paper are a wedge filter with wedge angle $\frac{\pi}{2}$ and two third-order edge detectors with filter smoothing parameters 0.09 and 0.2 (called edge 1 and edge 2 filters, respectively, in this paper; see Table 1 in [13] for more details). It is also noted that the size mentioned in this section for Bessel and Rao–Ben-Arie filters is equivalent minimum size as explained in section 3. However, the real size of the Bessel and Rao–Ben-Arie filters are the size of the input image.

In the first experiment, the six filters (*DroG,* Rao–Ben-Arie, Bessel, and three Jacob–Unser filters) are applied on the synthetic star image shown in Figure 8(a) to compare the accuracy of the detected edges. As depicted in Figure 7, the error term increases with respect to filter size (and therefore the scale) for all filters except for the Bessel filter. The error term for the Bessel filter remains almost unchanged with respect to its size (and its scale). We note that the error term measured in Figure 7 is due to the neighboring edges at sharp corners in the "star" image. In the case of sharp corners, it is well known that such an error term for *DroG* increases by raising the filter size (and its standard deviation). However, this error term for a Bessel filter is the lowest, as expected from the theorems proved in section 2. We

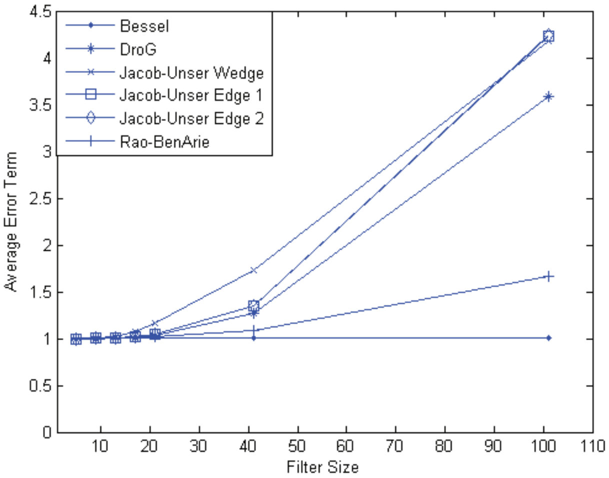
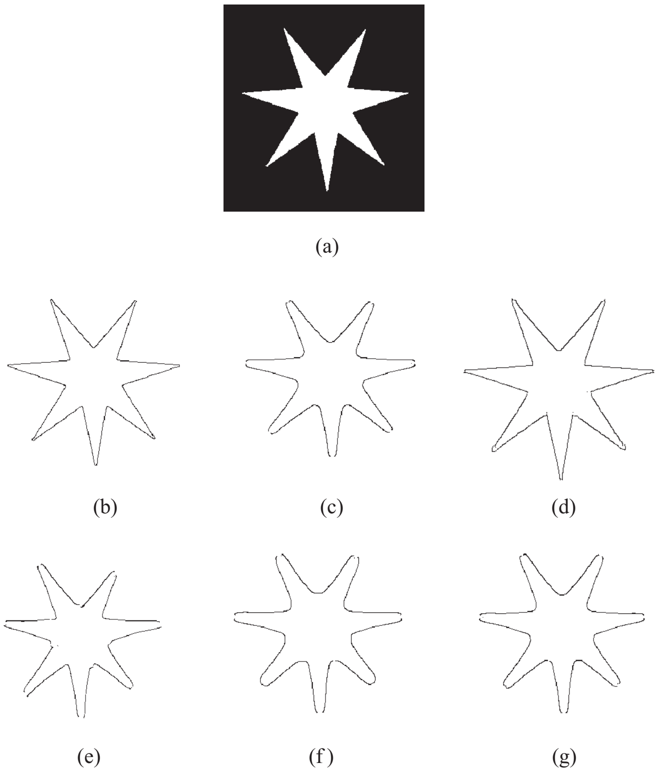


**Figure 7.** *The mean error between the detected edges and true discontinuities in the "Star" image with respect to filter sizes* 5, 9, 13, 17, 21, 41, 101 *for the filters discussed in this paper (see the text for more details).*

notice that the second most accurate edge map (after the Bessel filter) is associated with the Rao–Ben-Arie filter (see Figure 7). The high accuracy of edge detection achieved by the Bessel filter is due to the mathematical properties of the Bessel function. This is demonstrated in great detail in the proofs of Theorems 1 and 2. In fact, the reason that such a high accuracy in edge detection is obtained by using the Bessel filter is its scale invariance property as proved in Theorem 2. There is a family of filters enjoying a high accuracy in edge detection. It is proved in [18] that the condition a filter is required to possess in order to produce highly accurate edge detection is that the filter value at the center asymptotically approach infinity (see Lemma 2.6 in [18]). It is interesting to note that the experimental results indicate that this scale invariance property of the Bessel filter is applicable to the detection of not only edges but also other features such as Blobs, ridges, and corners [19].

Let us now briefly explain how the average error term between true discontinuities in the original image and detected edges are measured in the above experiment. We initially employ a Euclidean distance transform algorithm [20] to create a distance map for the binary images used in this paper. Having detected the edge maps using the edge detection algorithms mentioned earlier, the average error for all edge maps are then calculated by averaging the distances obtained from the distance map for all edge pixels. In this experiment, filter parameters are calculated according to the filter size. For example, for Canny and Jacob–Unser algorithms, standard deviation $= \frac{\text{filter size}}{8}$, and for the Bessel filter $\mu = \frac{(\text{filter size-1})^2}{100}$ is set. The edge maps detected by the six filters for the filter size of 41 are depicted in Figure 8. As shown in this figure, the sharp corners are detected properly by the Bessel and Rao–Ben-Arie filters. Visually the edge maps associated with the Bessel and Rao–Ben-Arie filters look similar. However, as demonstrated in Figure 7, the error associated with the Rao–Ben-Arie filter is more than the Bessel filter. The other filters with the same size have smoothed the sharp corners. In line with Theorems 1 and 2, as will be shown here in the rest of the experiments, the Bessel filter also demonstrates the highest accuracy in edge detection in cases where an object has curved boundaries. For a fair comparison among edge detection algorithms discussed here, we use a narrow strip (referred to as threshold scaling strip (TSS) in this paper) in which brightness increases logarithmically from left to right as shown in Figure 9 to help us determine the equivalent threshold levels for various algorithms while the same size for filters is chosen. Since there are various edges with different strengths in this strip, lower threshold levels result in the detection of more vertical line edges in the strip. For the remaining experiments, this strip is therefore added to the bottom of images. Threshold levels for all algorithms are adjusted so that the same numbers of vertical lines in the strip are detected.

The TSS is added to the bottom of the noisy synthetic image in [13] (see Figure 10(a) in this reference) as shown in Figure 10(a). The Bessel ($\frac{\varepsilon}{\Delta} = 1$), Rao–Ben-Arie, and Jacob–Unser filters with filter sizes 21 and 41 are applied to this noisy image. Visual inspection of Figure 10 shows that edge maps produced by the Bessel and Rao–Ben-Arie filters are more accurate than those of the Jacob–Unser filter. This figure also shows that as the filter size increases, the Jacob–Unser filter produces edge maps with more distortions and deformations than the others. Figure 11 shows the edge maps detected by filters discussed in this paper applied on a noisy image taken from a text book to which some noise is added in the data acquisition process. A close inspection of this figure shows that the shapes of some letters are distorted in the edge maps of the text book image generated by *DroG*, Demigny, Jacob–Unser,



(a)

(b) (c) (d)

(e) (f) (g)

**Figure 8.** *The edge map of* (a) *the "Star" image detected with filter size* 41 *by* (b) *Bessel filter with* ($\varepsilon = 0.01$) (c) *DroG filter,* (d) *Rao–Ben-Arie filter,* (e) *Jacob–Unser wedge filter,* (f) *Jacob–Unser edge* 1 *filter,* (g) *Jacob–Unser edge* 2 *filter.*

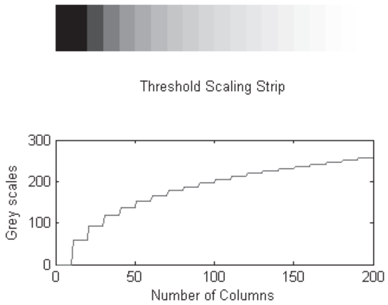
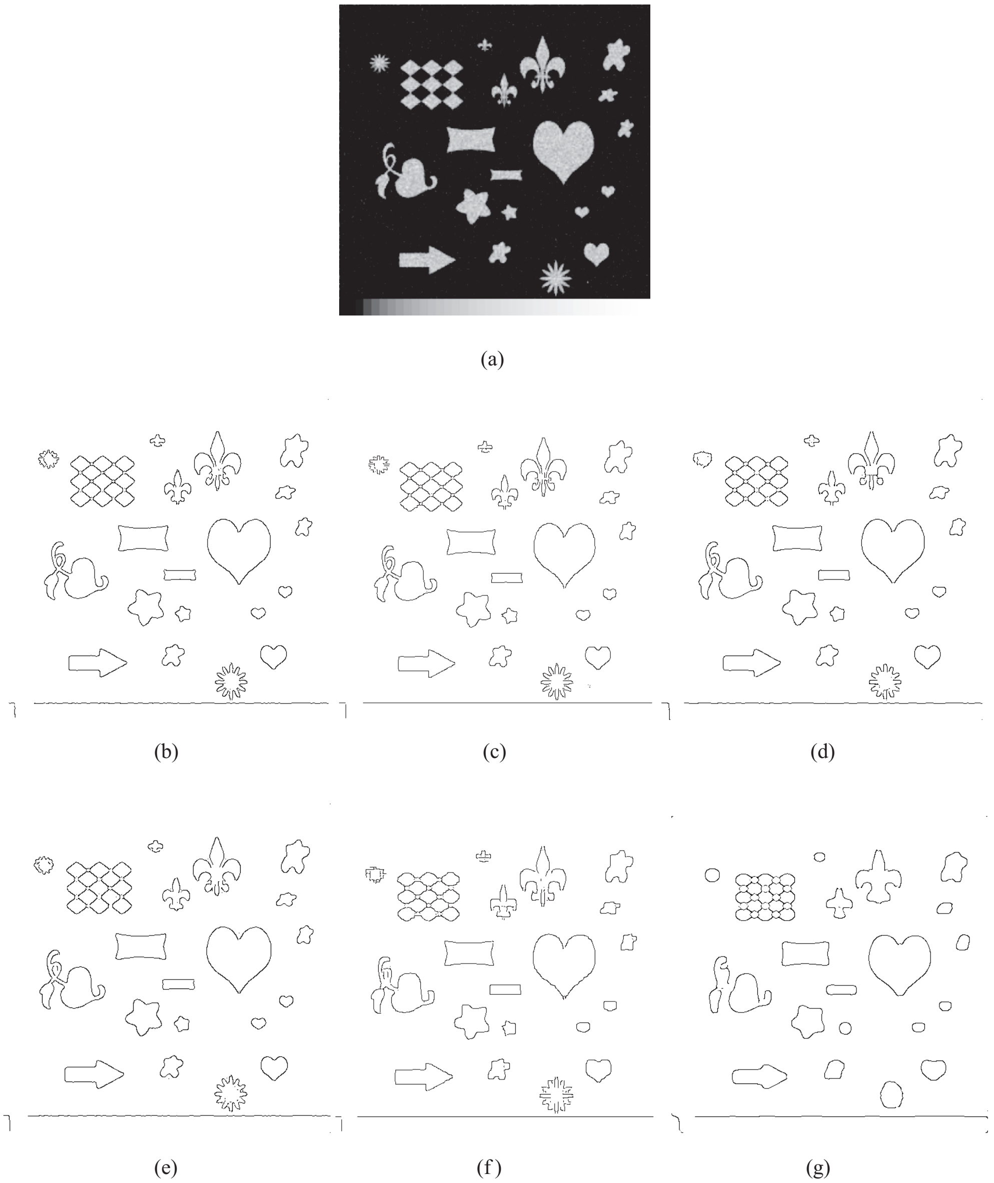


**Figure 9.** *A narrow strip of various logarithmically increasing brightness used in some experiments in this paper: (top) threshold scaling strip, (bottom) a cross-section of the strip.*

and Rao–Ben-Arie filters in the text image. The Bessel filter has, however, produced the most accurate result. The letter "t" throughout the text is a clear example. Other examples include letter "l" in the word "Sobel," "g" in the word "edge," "r" in "contribution," "a" in "values," "m" in "motivation," and so on. This shape distortion in edge detected letters becomes more severe by raising the filter sizes in Demigny, *DroG,* Rao–Ben-Arie, and Jacob–Unser filters as shown in Figure 12. This figure illustrates the edge detection results for the same text



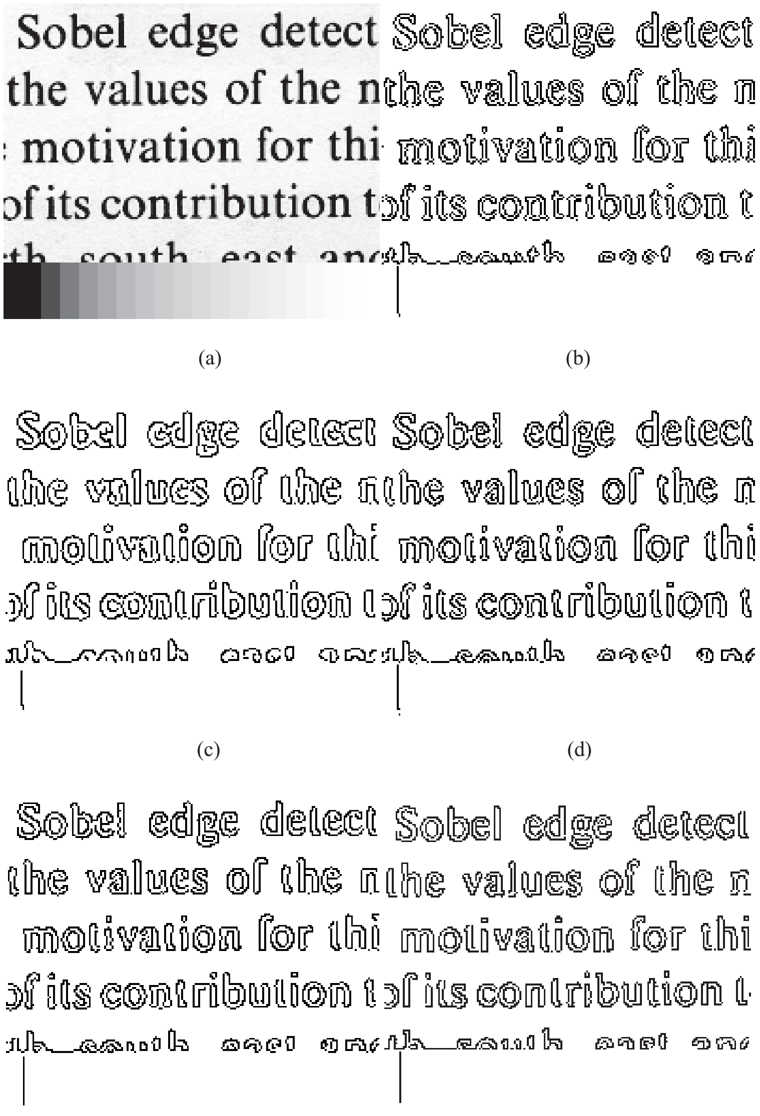
(a)

(b) (c) (d)

(e) (f) (g)

**Figure 10.** *A comparison of edge detection performance for Bessel, Demigny, and Jacob–Unser edge* 1 *filters:* (a) *original noisy image,* (b) *Bessel filter size* $= 21$ *with* $\frac{\varepsilon}{\Delta} = 1$, (c) *Rao–Ben-Arie filter size* $= 21$, (d) *Jacob–Unser edge* 1 *filter size* $= 21$, (e) *Bessel filter size* $= 41$ *with* $\frac{\varepsilon}{\Delta} = 1$, (f) *Rao–Ben-Arie filter size* $= 41$, (g) *Jacob–Unser edge* 1 *filter size* $= 41$ *(the threshold levels of all algorithms are adjusted to detect the first vertical line in the TSS).*

image of Figure 11(a) for Bessel, Rao–Ben-Arie, and Demigny filters with filter size 13. In this figure and with this filter size, the letters are still detected in great detail by the Bessel filter; however, the edges detected by Rao–Ben-Arie and Demigny show more distortion and deformation. The experimental results depicted in Figures 11 and 12 confirm the superiority of the Bessel filter concluded from Figure 6(c), which is due to the best product of indices



(a) (b)

(c) (d)

(e) (f)

**Figure 11.** *Edge maps of a noisy text image produced by edge detection filters with filter size* 7*:* (a) *original noisy text image,* (b) *Bessel filter with* $\frac{\varepsilon}{\Delta} = 0.01$*,* (c) *DroG filter,* (d) *Demigny filter,* (e) *Jacob–Unser edge* 1 *filter, and* (f) *Rao–Ben-Arie filter (the thresholds in all algorithms are adjusted to detect the first vertical line in the TSS).*

associated with the Bessel filter in comparison with the filters investigated in this paper. This high localization index possessed by the Bessel filter is a consequence of its scale invariance property implied by Theorem 2.

Gaussian noise with standard deviation 50 is added to the Cameraman image to produce a noisy image with SNR $= 7.2$ as shown in Figure 13(a). As shown in the figure, in the presence of noise, the Bessel filter produces competitive results in comparison with the Jacob–Unser filters, whilst *DroG* and Demigny filters have detected more false edges. However, as

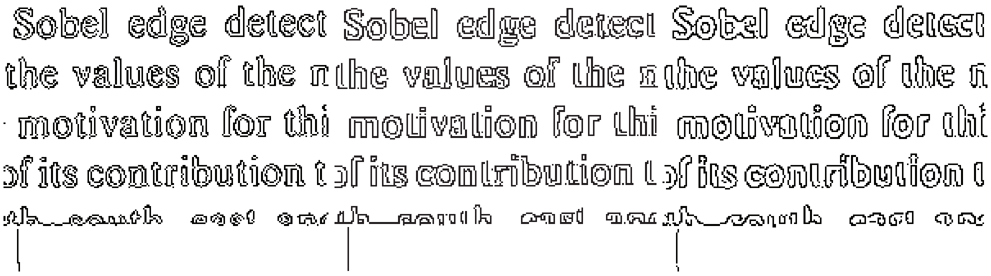
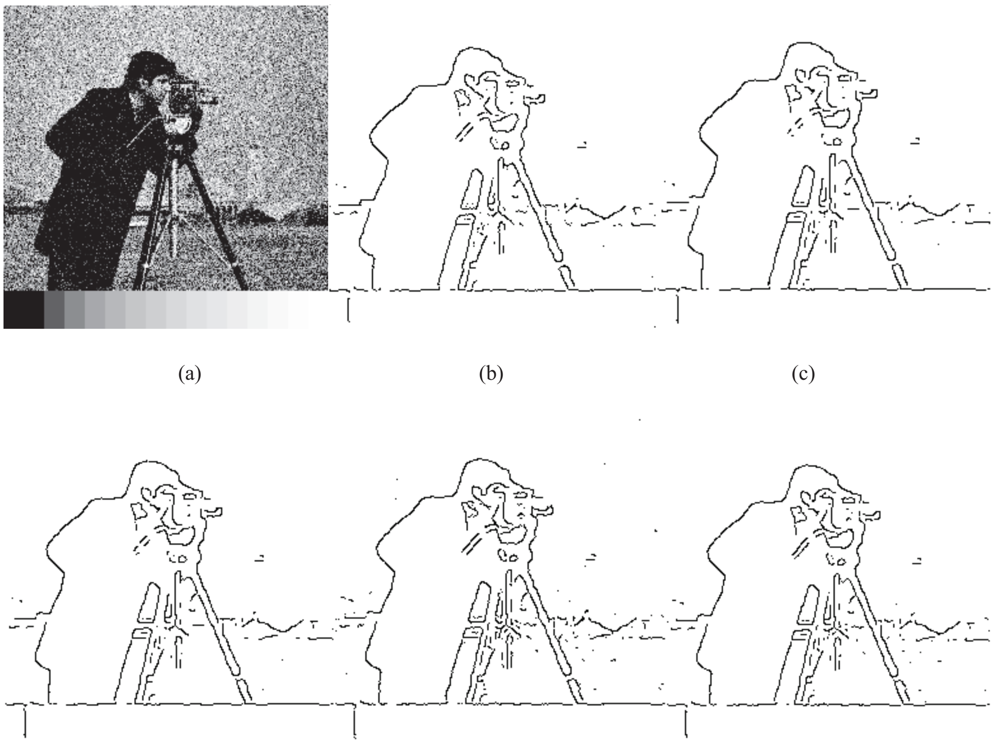


**Figure 12.** *Edge maps generated by the some filters with filter size* 13 *discussed in this paper:* (a) *Bessel filter with* $\frac{\varepsilon}{\Delta} = 0.01$, (b) *Rao–Ben-Arie filter,* (c) *Demigny filter (the thresholds in all algorithms are adjusted to detect the first vertical line in the TSS).*

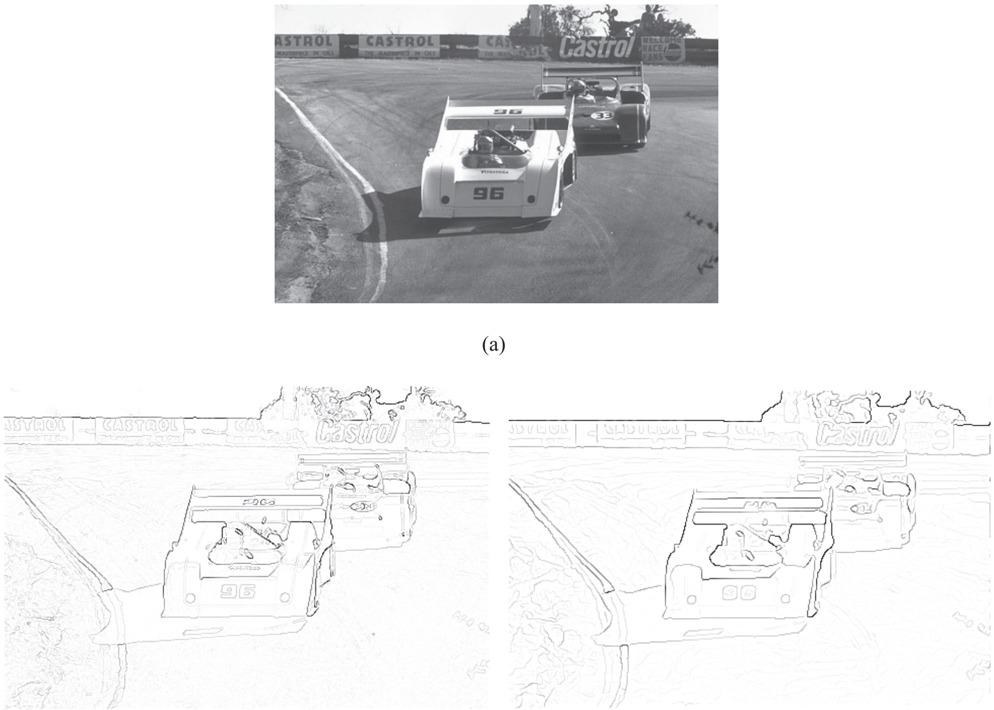


(a) (b) (c)

(d) (e) (f)

**Figure 13.** *The comparison among edge detection filters (with filter size* 13*) investigated in this paper and applied on images with Gaussian noise* (SNR = 7.2)*:* (a) *noisy* 256 × 256 *Cameraman image contaminated with Gaussian noise with standard deviation* 50*; edge maps produced by* (b) *Bessel filter with* $\frac{\varepsilon}{\Delta} = 1.6$, (c) *Jacob–Unser edge* 1 *filter,* (d) *Jacob–Unser edge* 2 *filter,* (e) *DroG filter,* (f) *Demigny filter (the thresholds in all algorithms are adjusted to detect the first vertical line in the TSS).*

demonstrated in Figures 10, 11, and 12, the Bessel filter outperforms popular filters due to its higher localization index (and hence higher overall performance). Soft edges can be produced by placing the absolute values of the image gradient on edges. Such soft edge maps are then



(a)

(b)

(c)

**Figure 14.** *Bessel and Rao–Ben-Arie filters (with filter size* 13*) are employed to produce soft edge maps for a real world image:* (a) *original real world image,* (b) *soft edge map produced by Bessel filter with* $\frac{\varepsilon}{\Delta} = 0.01$*,* (c) *soft edge map produced by Rao–Ben-Arie filter.*

produced by applying Bessel and Rao–Ben-Arie filters (with size 13) on a real world image shown in Figure 14(a). There are many details which are detected with higher accuracy by a Bessel filter. For example, number "96" on the front car and the signs "CASTROL" in the background (left side of the image) are completely readable in Figure 14(b). However, it is very difficult (if not impossible) to read these details in the edge map produced by a Rao–Ben-Arie filter as shown in Figure 14(c). It is noted that no hard thresholding is used in these soft edge maps.

## 5. Conclusion.

The MS Green function is singular at the center; and a regularization method for this function is therefore presented here to propose an edge detection filter known as the Bessel filter in this paper. The mathematical framework required to use this filter for edge detection is established here. The results obtained from the theorems proved here indicate that the Bessel filter enjoys a scale invariant property for edge detection. Such a mathematical framework is indeed applicable to a family of isotropic filters with singularity at the center. The contribution of this paper is to introduce and investigate the properties of the Bessel filter as a member of such a family. The detection performance and localization indices are numerically calculated for this filter and compared with some recent popular edge detector filters in the literature. This numerical calculation demonstrates that the product of the detection performance and localization indices of the Bessel filter is superior. Quantitative measurements of edge detection accuracy performed on synthetic images containing both curved boundaries and sharp corners confirm higher overall performance on Canny-like criteria. The Bessel filter and some recent popular filters are applied to some real world and benchmark images to demonstrate that the Bessel filter produces superior results confirming both higher overall Canny-like performance and higher accuracy in quantitative measurements performed on synthetic images. It is also observed that by increasing the filter size (and therefore the scale), the edges detected by the other filters investigated in this paper are smoothed, deformed, and shifted away from the true edges. However, the Bessel filter produces scale invariant edges. We also notice that while at the presence of mild noise the Bessel filter outperforms the state-of-the-art filters due to its higher localization index, it also produces competitive results in comparison with Jacob–Unser filters and still outperforms Demigny and *DroG* filters when the input image is contaminated with the excessive noise. The results presented here demonstrate promising prospects for the edge detection method proposed here; however, some issues not discussed in this paper are open for further research:

(1) It is interesting to employ the filter proposed here in a class of steerable functions as in [13] and therefore develop a ridge detection algorithm based on the Bessel filter.

(2) The Bessel filter proposed here can be extended for feature extraction in 3D volumetric images.

## 6. Appendices.

**Appendix A.** *Proof of Theorem* 1. Before presenting the proof of this theorem, we briefly explain the structure of the proof. The convolution (in Cartesian coordinates) between the Bessel function shown in Figure 1 and an input image (a piecewise-constant function) depicted in Figure 2 at the center of the coordinate system (point $O$) is evaluated in the polar coordinate system. Next, we calculate the directional derivative of this convolution (at point $O$) along a line passing through point $O$ with any given slope between $\tan(\vartheta)$ and $\tan(\vartheta+\varphi)$ (see Figure 2). By exploiting the asymptotic properties of the Bessel function, we then prove that this directional derivate is infinity (unbounded). We immediately conclude that the gradient magnitude of the convolved function at point $O$ is also unbounded. To explain the proof in details, let us now consider the piecewise-constant image shown in Figure 2 described in polar coordinates as

$$g(r,\theta) = aH(r)\left(H(\theta-\vartheta) - H(\theta-\vartheta-\phi)\right), \tag{A-1}$$

where $H(.)$, $a$, $r$, and $\theta$ are the Heaviside step function, a constant representing luminance, and radial and angular coordinates, respectively, and $\vartheta, \phi \in [0, 2\pi)$ are constants.

In Figure 2, the origin of the coordinate system, point $O$, is placed on the image discontinuity forming a sharp corner. We aim to show that the gradient magnitude of $u$ at $O$ (i.e., $|\nabla u|_O$) is infinite. By using (4), let us evaluate $u_n = \frac{\partial u}{\partial \mathbf{n}}|_{(0,0)}$, where $\mathbf{n} = \cos(\theta_n)\mathbf{i} + \sin(\theta_n)\mathbf{j}$,

$\theta_n \in (\vartheta, \vartheta + \phi)$, and $\mathbf{i}$ and $\mathbf{j}$ are the unit vectors along the $x$ and $y$ axes, respectively:

$$\begin{aligned}\frac{\partial u}{\partial \mathbf{n}} = \frac{\partial h}{\partial \mathbf{n}} * g = \operatorname*{Lim}_{\varepsilon \to 0} \int_{-\infty}^{-\varepsilon} \int_{-\infty}^{-\varepsilon} g(x - x', y - y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy' \\ + \operatorname*{Lim}_{\varepsilon \to 0} \int_{-\infty}^{-\varepsilon} \int_{\varepsilon}^{+\infty} g(x - x', y - y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy' \\ + \operatorname*{Lim}_{\varepsilon \to 0} \int_{\varepsilon}^{+\infty} \int_{-\infty}^{-\varepsilon} g(x - x', y - y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy' \\ + \operatorname*{Lim}_{\varepsilon \to 0} \int_{\varepsilon}^{+\infty} \int_{\varepsilon}^{+\infty} g(x - x', y - y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy'.\end{aligned} \tag{A-2}$$

At point O (0,0), the above integration can therefore be written as

$$\begin{aligned}\frac{\partial u}{\partial \mathbf{n}}(0,0) = \frac{\partial h}{\partial \mathbf{n}} * g = \operatorname*{Lim}_{\varepsilon \to 0} \int_{-\infty}^{-\varepsilon} \int_{-\infty}^{-\varepsilon} g(-x', -y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy' \\ + \operatorname*{Lim}_{\varepsilon \to 0} \int_{-\infty}^{-\varepsilon} \int_{\varepsilon}^{+\infty} g(-x', -y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy' \\ + \operatorname*{Lim}_{\varepsilon \to 0} \int_{\varepsilon}^{+\infty} \int_{-\infty}^{-\varepsilon} g(-x', -y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy' \\ + \operatorname*{Lim}_{\varepsilon \to 0} \int_{\varepsilon}^{+\infty} \int_{\varepsilon}^{+\infty} g(-x', -y') \frac{\partial h(x', y')}{\partial \mathbf{n}} dx' dy'.\end{aligned}$$

It is more convenient to write the above integration in polar coordinate system, i.e.,

$$\frac{\partial u}{\partial \mathbf{n}}(0,0) = \operatorname*{Lim}_{\varepsilon \to 0} \int_{\varepsilon}^{+\infty} \int_{0}^{2\pi} g(-r, -\theta) \frac{\partial h}{\partial \mathbf{n}} r \, dr d\theta \tag{A-3}$$

or

$$\frac{\partial u}{\partial \mathbf{n}}(0,0) = a \operatorname*{Lim}_{\varepsilon \to 0} \int_{\varepsilon}^{+\infty} \int_{\vartheta + \pi}^{\vartheta + \phi + \pi} \frac{\partial h}{\partial \mathbf{n}} r \, dr d\theta. \tag{A-4}$$

On the other hand, $\frac{\partial h}{\partial \mathbf{n}}$ can be calculated as

$$\frac{\partial h}{\partial \mathbf{n}} = \vec{\nabla} h \cdot \mathbf{n} = \frac{\partial h}{\partial x} \cos(\theta_n) + \frac{\partial h}{\partial y} \sin(\theta_n),$$

where

$$\frac{\partial h}{\partial x} = \frac{\partial h}{\partial r} \frac{\partial r}{\partial x} = \cos(\theta) \frac{\partial h}{\partial r}$$

and

$$\frac{\partial h}{\partial y} = \frac{\partial h}{\partial r} \frac{\partial r}{\partial y} = \sin(\theta) \frac{\partial h}{\partial r}.$$

Therefore the term $\frac{\partial h}{\partial \mathbf{n}}$ can be written as

$$\frac{\partial h}{\partial \mathbf{n}} = \frac{\partial h}{\partial r} \cos(\theta - \theta_n). \tag{A-5}$$

Equation (A-4) can then be written as

$$\text{(A-6)}\qquad \frac{\partial u}{\partial \mathbf{n}}(0,0) = \frac{a}{\sqrt{\mu}} \operatorname*{Lim}_{\varepsilon\to 0} \int_{\varepsilon}^{+\infty} \int_{\vartheta+\pi}^{\vartheta+\phi+\pi} r \frac{\partial K_0\left(\frac{r}{\sqrt{\mu}}\right)}{\partial r} \cos{(\theta-\theta_n)} dr d\theta.$$

Therefore

$$\text{(A-7)}\qquad \frac{\partial u}{\partial \mathbf{n}}(0,0) = \frac{a}{\sqrt{\mu}} \left(\sin(\vartheta-\theta_n) - \sin(\vartheta+\phi-\theta_n)\right) \operatorname*{Lim}_{\varepsilon\to 0} \int_{\varepsilon}^{+\infty} r \frac{\partial K_0\left(\frac{r}{\sqrt{\mu}}\right)}{\partial r} dr.$$

Equation (A-7) can be integrated by using integration by parts:
(A-8)

$$\frac{\partial u}{\partial \mathbf{n}}(0,0) = \frac{a}{\sqrt{\mu}} \left(\sin(\vartheta-\theta_n) - \sin(\vartheta+\phi-\theta_n)\right) \operatorname*{Lim}_{\varepsilon\to 0} \left[ \left( rK_0\left(\frac{r}{\sqrt{\mu}}\right)\right)\Big|_{\varepsilon}^{+\infty} - \int_{\varepsilon}^{+\infty} K_0\left(\frac{r}{\sqrt{\mu}}\right) dr \right].$$

Using asymptotic behaviors of $K_0(r)$ discussed in section 2, as $r = \varepsilon \to 0$ and $r \to +\infty$, and l'Hôpital's rule for indeterminate terms, we can determine the asymptotic behavior of term $rK_0(\frac{r}{\sqrt{\mu}})$ in (A-8) when $r = \varepsilon \to 0$ and $r \to +\infty$, i.e.,

$$\operatorname*{Lim}_{r\to 0} \left[ rK_0\left(\tfrac{r}{\sqrt{\mu}}\right)\right] = \operatorname*{Lim}_{r\to 0} \left[ -r\text{Log}\left(\tfrac{r}{\sqrt{\mu}}\right)\right] = 0$$

and

$$\operatorname*{Lim}_{r\to +\infty} \left[ rK_0\left(\tfrac{r}{\sqrt{\mu}}\right)\right] = \operatorname*{Lim}_{r\to +\infty} \left[ r\sqrt{\frac{\pi}{2r}} \exp\left(\tfrac{-r}{\sqrt{\mu}}\right)\right] = 0.$$

Equation (A-8) can therefore be rewritten as

$$\text{(A-9)}\qquad \frac{\partial u}{\partial \mathbf{n}}(0,0) = \frac{a}{\sqrt{\mu}} \left(\sin(\vartheta+\phi-\theta_n) - \sin(\vartheta-\theta_n)\right) \operatorname*{Lim}_{\varepsilon\to 0} \left( \int_{\varepsilon}^{+\infty} K_0\left(\tfrac{r}{\sqrt{\mu}}\right) dr \right);$$

as $\varepsilon \to 0$, the integration term in (A-9) approaches infinity.

Since $\frac{\partial u}{\partial \mathbf{n}}(0,0) = \vec{\nabla} u|_O \cdot \mathbf{n}$, then

$$|\nabla u|_O = +\infty.$$

We notice that for $\phi = \pi$ the wedge shown in Figure 2 changes to an edge along a straight line on which point $O$ can be regarded as an arbitrary point. In such a case, (A-9) is reduced to

$$\text{(A-10)}\qquad \frac{\partial u}{\partial \mathbf{n}}(0,0) = \frac{-2a}{\sqrt{\mu}} \left(\sin(\vartheta-\theta_n)\right) \operatorname*{Lim}_{\varepsilon\to 0} \left( \int_{\varepsilon}^{+\infty} K_0\left(\tfrac{r}{\sqrt{\mu}}\right) dr \right).$$

The absolute value of the right-hand side of (A-10) also approaches infinity as $\varepsilon \to 0$, hence $|\nabla u|_O = +\infty$. We also note that by considering a regular curve approximated by an infinite number of infinitesimal straight line segments and by assuming point $O$ to be on one of these infinitesimal straight line segments, we exploit the above argument obtained from (A-10) to conclude that $|\nabla u|_O$ approaches infinity as $\varepsilon \to 0$ for this case as well. It is also noted that for boundaries forming straight lines and/or regular curves the unit vector **n** is conveniently considered to be the normal unit vector to the boundary. ■

**Appendix B.** *Proof of Theorem* 2. To begin, let us briefly explain the structure of the proof. The directional derivative of the convolution between the isotropic Bessel function (Figure 1) and the input image shown in Figure 2 along a straight line passing through point $O$ with a slope between $\tan(\vartheta)$ and $\tan(\vartheta+\varphi)$ (see Figure 2) is considered. We initially assume that the local maximum associated with point $O$ is located at point $O_0$ with polar coordinates $(r_0, \theta_0)$. By using the asymptotic properties of the Bessel function, we finally prove that point $O_0$ is located at the center of coordinates; i.e., point $O_0$ is the same as point $O$. To explain the proof in details, let us now consider again $g$ in Figure 2. We now aim to prove that $|\nabla u|$ has a maximum at point $O$ as $\varepsilon \to 0$. By considering $U(0,0) = \frac{\partial u}{\partial \mathbf{n}}(0,0)$, we can calculate $\frac{\partial U}{\partial \mathbf{n}}|_{(0,0)}$ as

$$\left.\frac{\partial U}{\partial \mathbf{n}}\right|_{(0,0)} = \int_0^{+\infty}\int_0^{2\pi}\left(r\frac{\partial^2 h}{\partial r^2}\cos^2(\theta-\theta_n) + \frac{\partial h}{\partial r}\sin^2(\theta-\theta_n)\right) g\,dr\,d\theta,$$

where $h$ and $g$ are given in (2) and (A-1). By integrating the above equation with respect to $\theta$ and integrating by parts with respect to $r$ for the first term in the above equation, we can write

$$\begin{aligned}\left.\frac{\partial U}{\partial \mathbf{n}}\right|_{(0,0)} &= -a \operatorname*{Lim}_{\varepsilon\to 0}\left(\sin\left(2\left(\vartheta+\phi-\theta_n\right)\right) - \sin\left(2\left(\vartheta-\theta_n\right)\right)\right)\int_{\varepsilon}^{+\infty}\left(\frac{dh_\varepsilon}{dr}\right)dr\\ &= a\left(\sin\left(2\left(\vartheta+\phi-\theta_n\right)\right) - \sin\left(2\left(\vartheta-\theta_n\right)\right)\right)\operatorname*{Lim}_{\varepsilon\to 0} h\left(\varepsilon\right).\end{aligned} \tag{B-1}$$

Let us assume that $U$ has a local maximum associated with point $O$ in $(r_0, \theta_0)$, where $r_0$, representing the distance between the local maximum and the central point $O$, is a small displacement, $\theta_0 \neq \vartheta$, and $\theta_0 \neq \vartheta+\varphi$. We notice that the local maxima with $r_0 \neq 0$ and angle $\theta_0 = \vartheta$ or $\theta_0 = \vartheta+\varphi$ are associated with points on lower and upper discontinuity edges (along straight lines) of the wedge and therefore not associated with point $O$. A Maclaurin series can be written in the neighborhood of maximum point, i.e.,

$$\left.\frac{\partial U}{\partial \mathbf{n}}\right|_{(r_0,\theta_0)} = \left.\frac{\partial U}{\partial \mathbf{n}}\right|_{(0,0)} + r_0\left.\frac{\partial^2 U}{\partial \mathbf{n}^2}\right|_{(0,0)} + r_0\theta_0\left.\frac{\partial^2 U}{\partial \mathbf{n}_\perp \partial \mathbf{n}}\right|_{(0,0)} + O(r_0^2),$$

where $\mathbf{n} = \cos(\theta-\theta_n)\mathbf{e}_r - \sin(\theta-\theta_n)\mathbf{e}_\theta$, $\mathbf{n}_\perp = \sin(\theta-\theta_n)\mathbf{e}_r + \cos(\theta-\theta_n)\mathbf{e}_\theta$ ($\mathbf{n}_\perp$ is perpendicular to $\mathbf{n}$), and $\mathbf{e}_r$ and $\mathbf{e}_\theta$ are unit vectors along the radial and angular coordinates, respectively. Since $r_0$ is a small displacement, we ignore terms with higher degrees. On the other hand, $\frac{\partial U}{\partial \mathbf{n}}|_{(r_0,\theta_0)}$ vanishes, because $U$ has a local maximum at $(r_0, \theta_0)$, i.e.,

$$\left.\frac{\partial U}{\partial \mathbf{n}}\right|_{(r_0,\theta_0)} = 0 = \left.\frac{\partial U}{\partial \mathbf{n}}\right|_{(0,0)} + r_0\left.\frac{\partial^2 U}{\partial \mathbf{n}^2}\right|_{(0,0)} + r_0\theta_0\left.\frac{\partial^2 U}{\partial \mathbf{n}_\perp \partial \mathbf{n}}\right|_{(0,0)}.$$

We can then calculate $r_0$ as

$$r_0 = \frac{-\left.\frac{\partial U}{\partial \mathbf{n}}\right|_{(0,0)}}{\left.\frac{\partial^2 U}{\partial \mathbf{n}^2}\right|_{(0,0)} + \theta_0\left.\frac{\partial^2 U}{\partial \mathbf{n}_\perp \partial \mathbf{n}}\right|_{(0,0)}}. \tag{B-2}$$

The second directional derivatives of $U$ along the unit vectors $\mathbf{n}$ and $\mathbf{n}_\perp$ can also be calculated as

$$\text{(B-3)} \qquad \left.\frac{\partial^2 U}{\partial \mathbf{n}^2}\right|_{(0,0)} = a\left(\sin^3(\vartheta-\theta_n)-\sin^3(\vartheta+\phi-\theta_n)\right)\operatorname*{Lim}_{\varepsilon\to 0}\left(\int_\varepsilon^{+\infty}\frac{1}{r}\frac{dh}{dr}dr\right),$$

$$\text{(B-4)} \qquad \left.\frac{\partial^2 U}{\partial \mathbf{n}_\perp \partial \mathbf{n}}\right|_{(0,0)} = a\Big(\sin^2(\vartheta+\phi-\theta_n)\cos\left(\vartheta+\phi-\theta_n\right) - \sin^2(\vartheta-\theta_n)\cos\left(\vartheta-\theta_n\right)\Big)\operatorname*{Lim}_{\varepsilon\to 0}\left(\int_\varepsilon^{+\infty}\frac{1}{r}\frac{dh}{dr}dr\right).$$

By using (B-1), (B-3), and (B-4), equation (B-2) can be written as

$$r_0 = \operatorname*{Lim}_{\varepsilon\to 0}\frac{-a\ \left(\sin\left(2\left(\vartheta+\phi-\theta_n\right)\right)-\sin\left(2\left(\vartheta-\theta_n\right)\right)\right)h\left(\varepsilon\right)}{\left(\sin^3(\vartheta-\theta_n)-\sin^3(\vartheta+\phi-\theta_n)+\theta_0\left(\sin^2(\vartheta+\phi-\theta_n)\cos\left(\vartheta+\phi-\theta_n\right)-\sin^2(\vartheta-\theta_n)\cos\left(\vartheta-\theta_n\right)\right)\right)\int_\varepsilon^{+\infty}\frac{1}{r}\frac{dh}{dr}dr}.$$

The above term is indeterminate as $\varepsilon\to 0$. By using l'Hôpital's rule for indeterminate terms, we can therefore write $r_0$ as

$$r_0 = \operatorname*{Lim}_{\varepsilon\to 0}\frac{a\ \left(\sin\left(2\left(\vartheta+\phi-\theta_n\right)\right)-\sin\left(2\left(\vartheta-\theta_n\right)\right)\right)\frac{dh}{d\varepsilon}}{\left(\sin^3(\vartheta-\theta_n)-\sin^3(\vartheta+\phi-\theta_n)+\theta_0\left(\sin^2(\vartheta+\phi-\theta_n)\cos\left(\vartheta+\phi-\theta_n\right)-\sin^2(\vartheta-\theta_n)\cos\left(\vartheta-\theta_n\right)\right)\right)\frac{1}{\varepsilon}\frac{dh}{d\varepsilon}} = 0.$$

Therefore as $\varepsilon\to 0$, the local maximum point for $U=\frac{\partial u}{\partial \mathbf{n}}$ approaches point $O$. Therefore $|\nabla u|$ at point $O$ has a local maximum. For $\phi=\pi$ the boundary passing through point $O$ is a straight line, and $r_0$ therefore vanishes. Boundaries forming regular curves passing through point $O$ are considered to consist of infinite numbers of infinitesimal straight line segments. For any point on such infinitesimal line segments, the above argument is applicable. The unit vector $\mathbf{n}$ is considered to be normal to boundaries forming a straight line and/or a regular curve. ■

**Appendix C.** In this appendix, we calculate the first and second derivatives of regularized filter (5). These derivatives are required in the evaluation of the detection performance and localization indices in (6) and (8). The first derivative of regularized filter (5) is calculated as

$$\text{(C-1)} \qquad f_\varepsilon(x) = \begin{cases} \dfrac{K_0'\left(\frac{x}{\sqrt{\mu}}\right)}{\sqrt{\mu}K_0\left(\frac{\varepsilon}{2\sqrt{\mu}}\right)}, & x > \dfrac{\varepsilon}{2}, \\ 0, & |x| < \dfrac{\varepsilon}{2}, \\ \dfrac{-K_0'\left(\frac{-x}{\sqrt{\mu}}\right)}{\sqrt{\mu}K_0\left(\frac{\varepsilon}{2\sqrt{\mu}}\right)}, & x < \dfrac{-\varepsilon}{2}, \end{cases}$$

where $K_0'(x)$ is the derivative of the modified Bessel function of the second kind and zero degree with respect to $x$. The second derivative of $h_\varepsilon(x)$ is then written as

$$\text{(C-2)} \qquad f_\varepsilon'(x) = \begin{cases} \dfrac{K_0''\left(\frac{|x|}{\sqrt{\mu}}\right)}{\mu K_0\left(\frac{\varepsilon}{2\sqrt{\mu}}\right)}, & |x| > \dfrac{\varepsilon}{2}, \\ 0, & |x| < \dfrac{\varepsilon}{2}. \end{cases}$$

**Acknowledgment.** We would like to thank Professor M. Hammerton for his proofreading.

## REFERENCES

[1] J. F. Canny, *A computational approach to edge detection*, IEEE Trans. Pattern Anal. Mach. Intell., 8 (1986), pp. 679–698.

[2] R. Deriche, *Using Canny's criteria to derive a recursively implemented optimal edge detector*, Int. J. Comput. Vis., 1 (1987), pp. 728–737.

[3] H. D. Tagare and R. J. P. deFigueiredo, *On localization performance measure and optimal edge detection*, IEEE Trans. Pattern Anal. Mach. Intell., 12 (1990), pp. 1186–1190.

[4] S. Sarkar and K. L. Boyer, *Optimal infinite impulse response zero crossing based edge detectors*, CVGIP: Image Understanding, 54 (1991), pp. 224–243.

[5] S. Sarkar and K. L. Boyer, *On optimal infinite impulse response detection filters*, IEEE Trans. Pattern Anal. Mach. Intell., 13 (1991), pp. 1154–1171.

[6] M. Petrou and J. Kittler, *Optimal edge detectors for ramp edges*, IEEE Trans. Pattern Anal. Mach. Intell., 13 (1991), pp. 483–491.

[7] M. Petrou, *Separable 2D filters for the detection of ramp edges*, IEE Proc. Vis. Image Signal Process., 142 (1995), pp. 228–231.

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

[9] K. R. Rao and J. Ben-Arie, *Optimal edge detection using expansion matching and restoration*, IEEE Trans. Pattern Anal. Mach. Intell., 16 (1994), pp. 1169–1177.

[10] W. K. Pratt, *Digital Image Processing*, 4th ed., John Wiley & Sons, New York, 2007.

[11] Z. Wang, K. R. Rao, and J. Ben-Arie, *Optimal ramp edge detection using expansion matching*, IEEE Trans. Pattern Anal. Mach. Intell., 18 (1996), pp. 1092–1097.

[12] D. Demigny, *On optimal linear filtering for edge detection*, IEEE Trans. Image Process., 11 (2002), pp. 728–737.

[13] M. Jacob and M. Unser, *Design of steerable filters for feature detection using Canny-like critera*, IEEE Trans. Pattern Anal. Mach. Intell., 26 (2004), pp. 1007–1019.

[14] W. T. Freeman and H. Adelson, *The design and use of steerable filters*, IEEE Trans. Pattern Anal. Mach. Intell., 13 (1991), pp. 891–906.

[15] P. Bao and X. Wu, *Canny edge detection enhancement by scale multiplication*, IEEE Trans. Pattern Anal. Mach. Intell., 27 (2005), pp. 1485–1490.

[16] D. Mumford and J. Shah, *Optimal approximations by piecewise smooth functions and associated variational problems*, Comm. Pure Appl. Math., 42 (1989), pp. 577–688.

[17] G. B. Arfken and H. J. Weber, *Mathematical Methods for Physicists*, 6th ed., Harcourt Academic Press, San Diego, CA, 2005.

[18] S. Mahmoodi, *Scale-invariant filtering design and analysis for edge detection*, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci., 467 (2011), pp. 1719–1738.

[19] S. Mahmoodi and S. Gunn, *Scale space smoothing, image feature extraction and Bessel filters*, in Lecture Notes in Computer Science-17th Scandinavian Conference in Image Analysis, Ystad, Sweden, 2011.

[20] B. Heinz, J. Gil, D. Kirkpatrick, and M. Werman, *Linear time Euclidean distance transform algorithms*, IEEE Trans. Pattern Anal. Mach. Intell., 17 (1995), pp. 529–533.