GRAPHICAL MODELS AND IMAGE PROCESSING **60,** 385–401 (1998)
ARTICLE NO. IP980478

Optimizing Edge Detectors for Robust Automatic Threshold Selection: Coping with Edge Curvature and Noise

Michael H. F. Wilkinson

Centre for High Performance Computing, University of Groningen, Groningen, The Netherlands
E-mail: michael@rc.service.rug.nl

Received October 22, 1997; accepted June 5, 1998

The Robust Automatic Threshold Selection algorithm was introduced as a threshold selection based on a simple image statistic. The statistic is an average of the grey levels of the pixels in an image weighted by the response at each pixel of a specific edge detector. Other authors have suggested that many edge detectors may be used within the context of this method instead. A simple proof of this is given, including an extension to any number of image dimensions, and it is shown that in noiseless images with straight line edges these statistics all yield an optimum threshold. Biases caused by curvature of edges and by noise (uniform Gaussian and Poisson) are explored theoretically and on synthetic 2-D images. It is shown that curvature bias may be avoided by proper selection of the edge detector, and a comparison of two noise bias reduction schemes is given. Criteria for optimizing edge detectors are given and the performances of eight edge detectors are investigated in detail. The best results were obtained using two edge detectors which compute an approximation of the square of the gradient. It is shown that this conclusion can be extended to 3-D. Least sensitivity to noise was obtained when using 3×3 Sobel filter kernels to approximate partial derivatives in x and y.

1. INTRODUCTION

Kittler *et al.* [1, 2] introduced an automatic threshold selection algorithm for segmentation of object-background images dubbed Robust Automatic Threshold Selection (RATS). The RATS algorithm computes thresholds either locally or globally using a weighted average of the grey levels within arbitrary areas of an image. The weight assigned to each pixel is the response of a simple edge detector at that pixel. It has been suggested that many edge detectors other than the original one for which Kittler *et al.* demonstrated the correctness of their approach may be used instead [3, 4]. Criteria for optimization of the edge detector are

1077-3169/98 $25.00

developed from a simple proof of this assumption. It is shown that curvature of the edges may introduce a bias which may increase the size of compact objects systematically for certain types of edge detector. Bias caused by either uniform Gaussian or Poison noise is discussed. An approach to eliminate such a bias discussed by Kittler *et al.* [2] is extended to the wider class of generalized RATS threshold selectors and compared to the use of higher order filters as suggested by Kammoun and Astruc [4]. Finally, a number of edge detectors are compared in curvature bias, noise performance, and sensitivity, both in theory and on computer-generated images. A large part of the discussion focusses on blurred rather than ideal step edges, since any real life image acquired at the ideal (Nyquist) sampling frequency will show blurring of edges. In the images of interest in our laboratory (microscopic images of bacteria) this blurring is of particular importance, since the objects of interest are close in size to the optical limit of resolution.

2. THEORY

2.1. *The Generalized RATS Principle*

Consider an ideal step $S_A(x)$ at $x = 0$ of amplitude A on a constant background B. This function can be split into an even and an odd part,

$$S_A(x) = B + \frac{A}{2} + \frac{A}{2}\,\mathrm{sgn}(x), \tag{1}$$

with $\mathrm{sgn}(x)$ the sign of x. Let $S_A(x)$ be convolved with an even point-spread-function (PSF) $\mathrm{psf}(x)$ with a unity integral over $(-\infty, \infty)$. The result $S_{A,\mathrm{psf}}(x)$ can be written as an even and odd part,

$$S_{A,\mathrm{psf}}(x) = B + \frac{A}{2} + \frac{A}{2}(\mathrm{psf} * \mathrm{sgn})(x), \tag{2}$$

in which $(\mathrm{psf} * \mathrm{sgn})(x)$ denotes the convolution of $\mathrm{psf}(x)$ and $\mathrm{sgn}(x)$. Let a weight function $w(x)$ be an even, integrable function, which is zero outside $[-a, a]$ for some finite a, and which has a nonzero integral over $[-a, a]$; it then follows that

$$\begin{aligned} T &\equiv \frac{\int_{-a}^{a} w(x) S_{A,\mathrm{psf}}(x)\,dx}{\int_{-a}^{a} w(x)\,dx} = \frac{\frac{A}{2}\int_{-a}^{a} w(x)(\mathrm{psf} * \mathrm{sgn})(x)\,dx + \left(B + \frac{A}{2}\right)\int_{-a}^{a} w(x)\,dx}{\int_{-a}^{a} w(x)\,dx} \\ &= B + \frac{A}{2} \end{aligned} \tag{3}$$

since the first term of the numerator is the integral of an odd function over an even interval (and therefore zero). In the case of an even PSF, where the maximum of the first derivative is halfway between object and background intensity, the statistic T is therefore equal to the ideal threshold. Equation (3) is a generalization of the result obtained by Kittler *et al.* [1]. According to (3), which can readily be generalized to more than one dimension, any reasonable function which is symmetric around the step could be used instead of the original edge detector. In the 2-D case, the ideal threshold for any straight edge through the origin can be computed from (3) using any $w(x, y)$ that is even in x and y. Note that (as in the original implementation of RATS) full rotational symmetry of $w(x, y)$ is not required. A simple translation of the coordinates allows application of (3) to any other straight line edge.

2.2. *Curvature Bias*

In practice, more constraints on $w(x, y)$ are needed. Consider an operator which yields a weight function:

$$w_r(x, y) = \begin{cases} 1 & \text{if } (x, y) \text{ is within } r \text{ of an edge} \\ 0 & \text{otherwise.} \end{cases} \tag{4}$$

Apart from the problem of evaluating $w_r(x, y)$ in any practical image, it does comply to all the constraints for (3). However, let the object of interest be a circle of grey level A and radius R. T then becomes

$$T = B + \frac{A(\pi R^2 - \pi(R-r)^2)}{\pi(R+r)^2 - \pi(R-r)^2} = B + \frac{A}{2}\left(1 - \frac{r}{2R}\right). \tag{5}$$

A similar reasoning may be used in 3-D images of spheres, where the threshold computed with $w_r(x, y, z)$ is

$$T = B + \frac{A(R^3 - (R-r)^3)}{(R+r)^3 - (R-r)^3} = B + \frac{A}{2}\left(1 - \frac{3Rr}{6R^2 + 2r^2}\right) \approx B + \frac{A}{2}\left(1 - \frac{r}{2R}\right). \tag{5a}$$

The latter approximation is true if $r \ll R$. In our application we are interested in bacteria, which (to a good approximation) are either spheres or cylinders with hemispherical ends. The 2-D image of such objects would be a rectangle of length aR and width $2R$ with two semicircles of radius R attached to the ends. For such shapes the bias would be

$$T = B + \frac{A(\pi R^2 - \pi(R-r)^2) + 2aRr}{\pi(R+r)^2 - \pi(R-r)^2 + 4aRr} = B + \frac{A}{2}\left(1 - \frac{r}{2(1 + a/\pi)R}\right). \tag{6}$$

In either case T is only ideal if $r \ll R$ or $a \gg \pi$. In other words, we need a function that is nonzero only in a very narrow region around the edge. By segmenting images of ideal circles of different sizes, or rods of different a, and performing a linear fit of the computed threshold as a function of $1/R$ or $1/(1 + a/\pi)$, we may compute the effective radius r_{eff}, as a measure of how sharply any given weight function peaks around the actual step. Similar techniques can also be applied in the 3-D case.

2.3. *Bias Caused by Uniform Gaussian Noise*

Kittler *et al.* [2] have shown that uniform Gaussian noise may cause a bias in T. Let the mean response to edges caused by Gaussian noise with standard deviation σ of $w(x, y)$ be $\eta\sigma$, and the fraction of pixels in the image belonging to the foreground (object) be q. Taking only the noise into account, T will be

$$T_{\text{noise}} = \frac{q(B + A)\eta\sigma + (1 - q)B\eta\sigma}{\eta\sigma} = B + \frac{A}{2} + \left(q - \frac{1}{2}\right)A \tag{7}$$

irrespective of the noise. If the number of object-related edge pixels is small, the signal-to-noise ratio (S/N) is low, or if the response of $w(x, y)$ to object-related edges is small, (6) is an excellent approximation of T. However, we must bear in mind that thresholding is not practical when S/N is low. If we require that fewer than n% of all pixels be misassigned

by threshold at $A/2$, this means that $A/2$ must be larger than the $(100-n)$th percentile of the noise distribution (for symmetric distributions). For Gaussian noise, this means that for 1% misassignment A must be larger than 4.65σ, and for 0.1% error A must be larger than 6.18σ. Hence, expressions for S/N > 5 are probably more meaningful than (7). When S/N $\gg 1$ the approximate expression of T is

$$T = \frac{W_{\text{edge}}\left(B+\frac{A}{2}\right)+W_{\text{noise}}(B+qA)}{W_{\text{edge}}+W_{\text{noise}}} = B+\frac{A}{2}+\frac{W_{\text{noise}}(q-0.5)}{W_{\text{edge}}+W_{\text{noise}}}A \tag{8}$$

in which W_{edge} is the integral over the weight function in the absence of noise and W_{noise} is the integral over the weight function in the absence of true signal. Let X and Y be the image dimensions, ε be the fraction of edge pixels, and $g|A|$ the response of the weight function to a step function of amplitude A. These weights then become

$$W_{\text{edge}} = \varepsilon g|A|XY \quad \text{and} \quad W_{\text{noise}} = \eta\sigma XY.$$

Substitution into (8) yields

$$T = B+\frac{A}{2}+\frac{\eta\sigma(q-0.5)}{\varepsilon g|A|+\eta\sigma}A = B+\frac{A}{2}+\frac{q-0.5}{\varepsilon g|A|/\eta\sigma+1}A. \tag{9}$$

The last term in (8) and (9) is the bias (β) caused by the noise. The bias does depend on the S/N ratio ($=|A|/\sigma$). All parameters in (9) depend solely on the image, except for g and η which depend on the weight function. The bias will be low for a weight function with low response to noise but a high response to edges, i.e., high g/η. Note that if the filters response is not restricted to pixels immediately adjacent to the ideal step, g may be defined in at least two ways: (i) the maximum response to an ideal step over the step amplitude, and (ii) the sum of responses at one side of a step, along a straight line at right angles to the edge. This latter "integrated" form is the appropriate one for the above formulas.

2.4. *Bias Caused by Poisson Noise*

Many practical images, especially when recorded in low-light situations, are dominated by photon noise which has a Poisson distribution. Poisson noise can often be approximated by Gaussian noise fairly well, the key difference in this context being the fact that it is not uniform: σ is proportional to the square root of the signal. If the background B and the step A are photon counts, T becomes

$$T = \frac{B(1-q)\eta\sqrt{B}+(B+A)q\eta\sqrt{B+A}}{(1-q)\eta\sqrt{B}+q\eta\sqrt{B+A}} = B+\frac{q}{q+(1-q)\sqrt{B/\{B+A\}}}A. \tag{10}$$

In this case the bias depends on both q and the inverse image contrast $B/\{B+A\}$. If the objects are far brighter than the background, T will approach $B+A$, rather than $B+A/2$. Hence, unlike the Gaussian case, the bias may be either positive or negative for small q. Equation (10) is an approximation for the low S/N case. When S/N is high, we need a form analogous to (9) for the Poisson case,

$$T = B+\frac{A}{2}+\frac{\left(\frac{q-(1-q)\sqrt{B/\{B+A\}}}{2q+2(1-q)\sqrt{B/\{B+A\}}}\right)}{\frac{\varepsilon g|A|}{\eta\sigma}+1}A, \tag{11}$$

in which

$$\bar{\sigma} = (1-q)\sqrt{B} + q\sqrt{B+A}$$

is the mean noise in the image. Though slightly more complicated than (9), (12) has the same general form. This suggests that similar bias reduction strategies may be used for both uniform Gaussian and Poisson noise.

2.5. *A Technique to Reduce Noise Bias*

Kittler *et al.* [2] use a simple method to reduce noise related bias. If σ is known, they compute, e.g., the 99th percentile of the noise distribution of their weight function. Only pixels where the weight function is larger than this threshold are used. Let this threshold be $\lambda_p\sigma$, and p be the probability that a noise response is higher than $\lambda_p\sigma$. The parameter λ_p depends on the probability density distribution of the noise response of the weight function, and on p. If the mean value of noise pixels higher than $\lambda_p\sigma$ is $\eta_p\sigma$, the mean noise response can be written as

$$\eta\sigma = p\eta_p\sigma.$$

The bias term in (9) then becomes

$$\beta = \frac{p\eta_p\sigma(q-0.5)}{\varepsilon g|A| + p\eta_p\sigma}A \leq \frac{p\eta_p\sigma(q-0.5)}{\varepsilon\lambda_p\sigma + p\eta_p\sigma}A = \frac{p\eta_p(q-0.5)}{\varepsilon\lambda_p + p\eta_p}A \tag{12}$$

since the minimum value of $|A|$ detectable by this strategy is $\lambda_p\sigma/g$. As p goes to zero, $p\eta_p$ also goes to zero, yet λ_p goes to infinity, so we should be able to reduce the bias arbitrarily, but at the expense of a reduced sensitivity to object related gradients. The maximum bias is independent of σ.

This has certain implications for the choice of weight function. The weight function used by Kittler *et al.* has a response to diagonal, blurred edges which is $1/\sqrt{2}$ times the response to horizontal or vertical blurred edges. Therefore, a rectangle tilted by 45° must have a contrast 40% larger than a similar rectangle with edges parallel to the image axes to be detected. Only a weight function with response $g|A|$ equal for any orientation of an edge should be used in this case.

To summarize, the weight function must (i) peak sharply around an edge, (ii) have as low a response to Gaussian noise as possible, (iii) preferably combined with a high response to step functions, and (iv) be invariant to edge orientation. Criteria (i) on the one hand and (ii) and (iii) on the other are slightly at odds with each other. A low response to noise combined with a high response to step functions usually requires smoothing of some form, which in general decreases edge localization [5, 6]. This in turn will usually result in an increase of the radius r of the region around edges where the weight function is nonzero.

2.6. *Higher Order Filters*

The above discussion assumes that the weight function is a linear function of the absolute magnitude of the step size A ($g|A|$). It is for this reason that the mean noise response $\eta\sigma$, the noise threshold $\lambda_p\sigma$, and the mean value of noise pixels above the threshold $\eta_p\sigma$ all

scale linearly with σ. If we have a weight function $w(x, y)$ which has all required properties, it automatically follows that $w^n(x, y)$ is a suitable weight function too, for all positive n. However, if $w(x, y)$ has a sharp global maximum at the step, $w^n(x, y)$ should peak even more sharply around the step, and may therefore have a lower effective radius r_{eff}. This has been pointed out by Kammoun and Astruc [4], who explored the one-dimensional case with asymmetric blurring of the edge. By peaking more strongly around the inflection point of the luminance, T shifted away from $B + A/2$, and nearer to the grey level of the inflection point.

When using $w^n(x, y)$, the mean noise response, the noise threshold, and the mean value of noise pixels above the threshold all scale linearly with σ^n. In fact, the noise threshold becomes $\lambda_p^n \sigma^n$, which does not harm the sensitivity in any way since the response to a step function will be increased equally to $g^n|A|^n$. The maximum bias is also independent of σ.

Kammoun and Astruc explored the effect of n on the particular kind bias discussed above. In a more pragmatic approach, the case of $n = 2$ will be discussed in detail, since $w^2(x, y)$ may be computed more cheaply for some important filters than $w(x, y)$ itself. To use the magnitude of gradient of an image as $w(x, y)$, without directional bias, we compute

$$w(x,y) = \sqrt{\left(\frac{dp(x,y)}{dx}\right)^2 + \left(\frac{dp(x,y)}{dy}\right)^2} \Leftrightarrow w^2(x,y) = \left(\frac{dp(x,y)}{dx}\right)^2 + \left(\frac{dp(x,y)}{dy}\right)^2.$$

Thus, computing $w^2(x, y)$ does not require the use of a (rather costly) square root. If the image is digitized to N levels, then a lookup table of N entries may be used to obtain the squares of the partial derivatives, and these may simply be added. This may be done in integer math without any loss of precision, which can be an advantage on certain computer architectures. A lookup table for the square root may of course be used, but this would need to be on the order of N^2, which becomes impractical for all but the smallest values of N.

Kammoun and Astruc [4] also use higher order filters to reduce the noise bias in the segmentation method. However, higher order filters may also show more sensitivity to random noise, since outliers in the noise distribution are multiplied in a similar way to object related gradients. Both Kittler's bias reduction technique and the use of higher order filters effectively reduce the number of pixels contributing to T, which must increase the variance of T. The familiar trade-off between bias and variance (or localization and sensitivity to noise) is bound to exist.

2.7. *A Selection of First and Second Derivative Based Edge Detectors*

The most obvious choices of operators to compute a suitable weight function $w(x, y)$ from an arbitrary image are gradient operators. Let us introduce

$$\Delta_x(x, y) = p(x - 1, y) - p(x + 1, y) \quad \text{and} \quad \Delta_y(x, y) = p(x, y - 1) - p(x, y + 1).$$

Then the gradient operator used by Kittler *et al.* has the form

$$e(x, y) = \max\{|\Delta_x(x, y)|, |\Delta_y(x, y)|\}. \tag{13}$$

This operator is even in x and y, but does not have full rotational symmetry. To improve

rotational symmetry we have

$$g(x, y) = \sqrt{\Delta_x^2(x, y) + \Delta_y^2(x, y)} \tag{14}$$

and

$$g^2(x, y) = \Delta_x^2(x, y) + \Delta_y^2(x, y). \tag{15}$$

A reduction of the sensitivity to noise may be achieved by using two kernel operators used in the Sobel edge detector:

$$\Delta_{x,\text{Sobel}}(x, y) = \frac{\Delta_x(x, y-1) + 2\Delta_x(x, y) + \Delta_x(x, y+1)}{4} \tag{16a}$$

and

$$\Delta_{y,\text{Sobel}}(x, y) = \frac{\Delta_y(x-1, y) + 2\Delta_y(x, y) + \Delta_y(x+1, y)}{4}. \tag{16b}$$

These forms of the x and y derivatives should have a Gaussian response to noise with $\sigma_{\text{Sobel}} = \sqrt{3/4}\sigma$ instead of $\sigma_\Delta = \sqrt{2}\sigma$. If the edge detectors discussed so far were to be implemented using the Sobel forms of the partial derivatives, the noise would be reduced by $\sqrt{3/8}$ for the $e(x, y)$ and $g(x, y)$ filters, and by $3/8$ in the $g^2(x, y)$ filter. Table 1 lists a number of important parameters for each of the six edge detectors.

Other filters, such as the Laplacian and Marr-Hildreth filters and edge detectors of Weszka and Rosenfeld [7], have been proposed for use in computing T [3], but these all have a number of disadvantages. The Laplacian and Marr-Hildreth filters compute unsmoothed and smoothed approximations of the second derivative, and are therefore odd, rather than even, functions around a (blurred) step edge. Taking the absolute value remedies this, but a zero is found at the inflection point of the luminance, rather than a maximum. It is expected that the resulting effective width will be larger. The main problem with second-derivative-based filters lies in their sensitivity to noise ($g/\eta = 0.224$ for the 3×3 Laplacian kernel rather than 0.564 for $g(x, y)$; see Tables 1a–1c). The large kernel edge detectors of Weszka and Rosenfeld [7], have a good g/η, but are not even around a step edge. In the following discussion of the aforementioned six first derivative edge detectors, the absolute value of the Laplacian (∇^2) computed with the 3×3 kernel,

$$\begin{matrix} 0 & -1 & 0 \\ -1 & 4 & -1 \\ 0 & -1 & 0, \end{matrix}$$

and the absolute value of the Marr-Hilldreth filter ($\nabla^2 G_\sigma$) computed from a 3×3 Gaussian kernel,

$$\begin{matrix} 0.075 & 0.125 & 0.075 \\ 0.125 & 0.2 & 0.125 \\ 0.075 & 0.125 & 0.075, \end{matrix}$$

followed by the Laplacian will be considered in more detail. Higher order filters of the form $g^m(x, y)$ will be considered with $g(x, y)$ as in (14), and $m = 1, 2, \ldots, 10$.

TABLE 1a
Important Parameters of Δ-Based Edge Detectors

Max. response to	$e(x, y)$	$g(x, y)$	$g^2(x, y)$
Gaussian noise	1.596σ	1.772σ	$4\sigma^2$
Ideal edge step A			
Horizontal (=vertical)	A	A	A^2
Diagonal[a]	A	$1.41A$	$2A^2$
Blurred[b] edge step A			
Horizontal (=vertical)	$0.410A$	$0.410A$	$0.168A^2$
Diagonal	$0.297A$	$0.420A$	$0.176A^2$

[a] Ideal diagonal edge has all pixels with $x < y$ zero, all others A, no antialiazing used.

[b] Ideal step edge blurred with 9×9 Gaussian kernel of $\sigma = 1$ [10].

TABLE 1b
Important Parameters of Δ_{Sobel}-Based Edge Detectors

Max. response to	$e_s(x, y)$	$g_s(x, y)$	$g_s^2(x, y)$
Gaussian noise	0.977σ	1.085σ	$1.5\sigma^2$
Ideal edge step A			
Horizontal (=vertical)	A	A	A^2
Diagonal[a]	A	$1.41A$	$2A^2$
Blurred[b] edge step A			
Horizontal (=vertical)	$0.410A$	$0.410A$	$0.168A^2$
Diagonal	$0.290A$	$0.410A$	$0.168A^2$

[a] Ideal diagonal edge has all pixels with $x < y$ zero, all others A, no antialiazing used.

[b] Ideal step edge blurred with 9×9 Gaussian kernel of $\sigma = 1$ [10].

TABLE 1c
Important Parameters of Second Derivative Edge Detectors

Max. response to	$\lvert\nabla^2(x, y)\rvert$	$\lvert\nabla^2 G_{\sigma=0.6}(x, y)\rvert$
Gaussian noise	3.568σ	0.435σ
Ideal edge step A		
Horizontal (=vertical)	A	$0.275A$ ($0.45A$)[c]
Diagonal[a]	$2A$	$0.350A$ ($0.6A$)[c]
Blurred[b] edge step A		
Horizontal (=vertical)	$0.071A$	$0.072A$
Diagonal	$0.086A$	$0.064A$

[a] Ideal diagonal edge has all pixels with $x < y$ zero, all others A, no antialiasing used.

[b] Ideal step edge blurred with 9×9 Gaussian kernel of $\sigma = 1$ [10].

[c] The integrated response over more than one pixel from the edge is given in brackets.

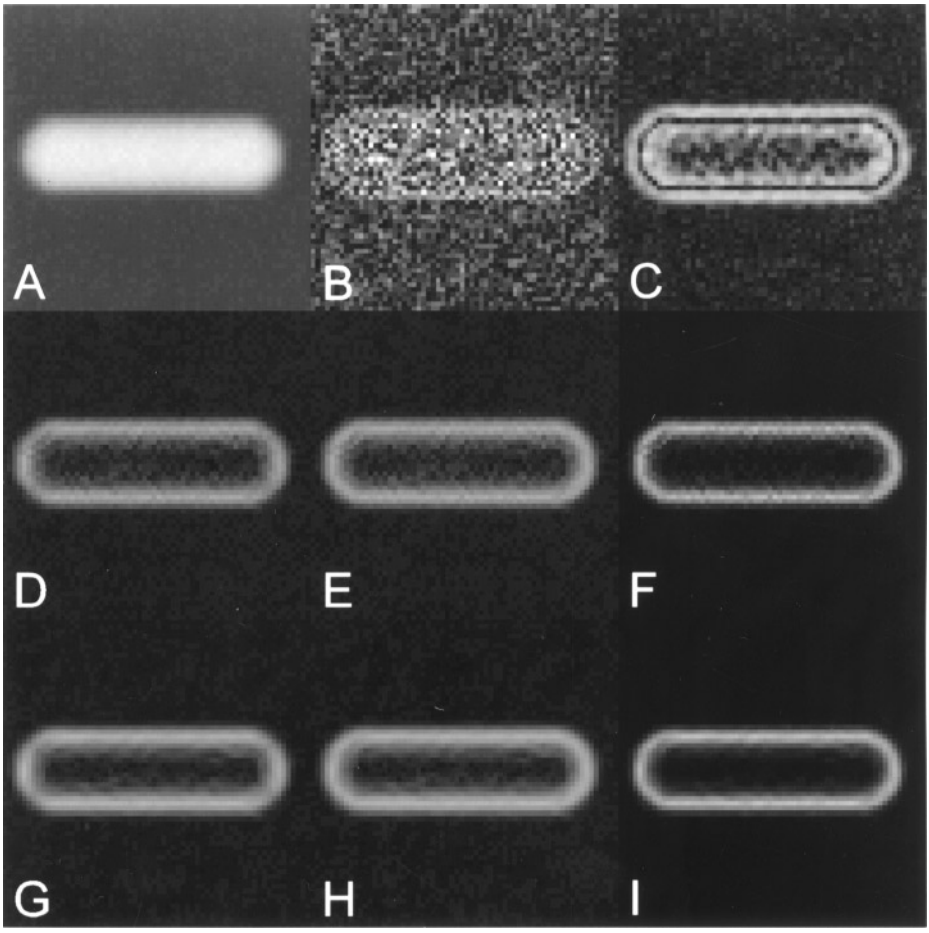

FIG. 1. The effect of 8 different gradient on a simulated shape of a bacterium (A) $e(x, y)$; (B) $g(x, y)$; (C) $g^2(x, y)$; (D), (E), and (F) same as (A), (B), and (C), respectively, but using Sobel filters instead of $\Delta(x, y)$.

Compass gradients have been rejected on two grounds. First, any two orthogonal kernels out of the 8 possible orientations contain all gradient information: both magnitude and direction [8]. Besides, the standard forms of many compass gradient operators (e.g., Prewitt, Sobel) contain directional bias for blurred step edges, overestimating gradients at diagonal edges [8]. This objection could of course be remedied, but the former remains in place.

Figure 1 shows the effect of the eight candidate weight functions operating on a simulated image of a bacterium. It is immediately clear that the Sobel filter versions show less noise than the others and that both $g^2(x, y)$ filters show the edge more sharply than the others.

3. RESULTS

3.1. *Curvature Bias*

The bias caused by nonzero curvature can be determined from the response to images of ideal circles of various diameters. Since (5) holds for a continuous image, only circles with a radius of at least of 4 pixels were used, to reduce quantization effects. A second test

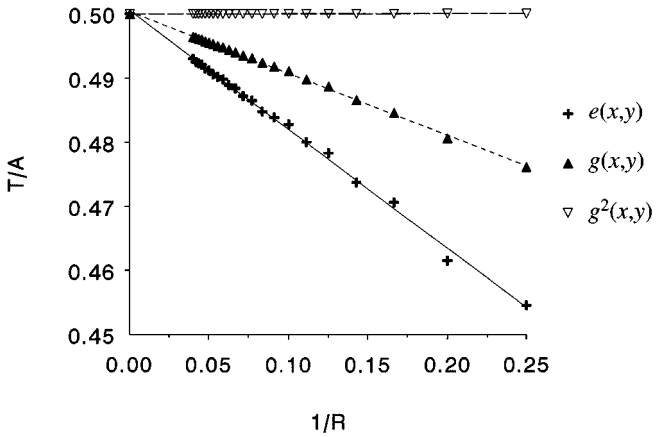

FIG. 2. Curvature bias as a function of object size: T/A as a function of $1/R$ for circles with radii ranging from 4 to 25 pixels, for the three Δ-based gradient filters. The effective radius r_{eff} can be derived from the slope of the least squares linear fit.

used images of rods of a 4 pixel radius, and a ranging from 0 to 12 in increments of 0.5. The computed thresholds were plotted as functions of $1/R$ and $1/(1 + a/\pi)$, respectively. The result of the first simulation for the three filters based on the simple first derivatives are plotted in Fig. 2. Similar linear fits of T/A vs $1/(1 + a/\pi)$ were obtained (data not shown). From both data sets r_{eff} values were computed from a least squares fit slope. The results are listed in Table 2. Rather surprisingly, both the squared first derivative filters and the second derivative filters have an effective diameter of zero for unblurred edges. The same tests were done on the same image sets, but blurred with a 5×5 pixel Gaussian kernel of $\sigma = 0.625$ pixels [9]. This was done to simulate the blurring present in any real image, especially when sampled at or near the Nyquist frequency of the optics. The results of this test are also shown in Table 2. Clearly, the effective radii increase when the images are blurred. However, both in the blurred and unblurred cases, the squared first derivative filters have the smaller r_{eff}.

The effective radius was also computed as a function of order m for filters of the form $g^m(x, y)$ with $m = 1, 2, \ldots, 10$. Figure 3 shows the results of these experiments for blurred and unblurred edges. For small $m = 1$ the effective radii are positive, and for large m the

TABLE 2
Effective Radii (r_{eff}) of Edge Detectors in Pixels

	From R dependence		From a dependence	
Filter	Ideal edges	Blurred edges	Ideal edges	Blurred edges
$e(x, y)$	0.74	1.46	0.74	1.46
$g(x, y)$	0.39	1.40	0.37	1.38
$g^2(x, y)$	0.00	0.77	0.00	0.76
$e_s(x, y)$	0.37	1.44	0.37	1.44
$g_s(x, y)$	0.51	1.45	0.50	1.43
$g_s^2(x, y)$	0.00	0.78	0.00	0.77
$\lvert\nabla^2(x, y)\rvert$	0.000	0.910	0.000	0.933
$\lvert\nabla^2 G_{\sigma=0.6}(x, y)\rvert$	0.000	0.959	0.024	1.000

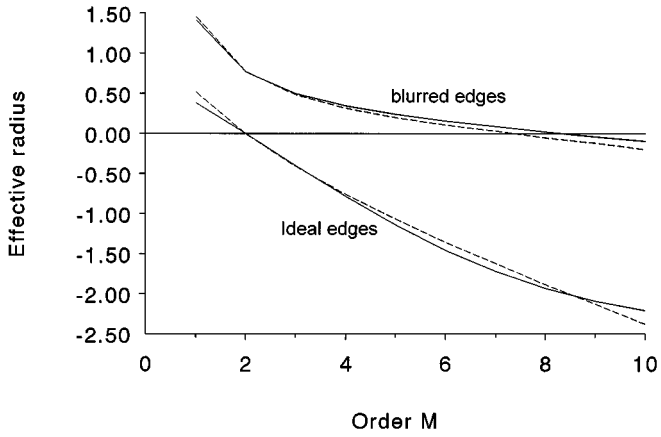

FIG. 3. Effective filter radius r_{eff} as a function of the order m of the filter for Sobel and simple first derivatives: ideal edges and blurred edges.

effective widths becomes negative. The zero-crossing is at $m = 2$ for unblurred edges and at $m \approx 7$ for blurred edges. In none of the cases was the bias sufficiently large to cause errors in the measurement of the surface areas of the objects.

3.2. *Noise Bias*

3.2.1. Gaussian noise. The effect of q on the noise bias has been demonstrated by Kittler *et al.* [2]. To determine the effect of ε on the noise bias, five images consisting of series of alternating horizontal strips of background (B) and foreground ($B + A$) grey level, with $q = 0.25$ and ε ranging from $1/16$ to $1/2$, where made. Gaussian noise with variable σ was added to each pixel. The ratio T/A was computed for a S/N ratio of 10 with and $B = 0$ and $A = 1000$. For the horizontal edges present, each of the filters had a $g = 1$ except for the Marr-Hildreth operator (integrated $g = 0.45$). The mean responses to noise for each filter (η) from Table 1 were used to compute the theoretical values of this ratio as a function of ε, using the high S/N approximation (9). The results are given in Fig. 4. Though there is a slight systematic difference between the theoretical and experimental value, the overall agreement is good. The higher order filters show a distinct advantage over the filters with a response linear in $|A|$. The second derivative filters show a poor performance. In the case of the 5-pixel wide Marr-Hildreth filter, the last point with $\varepsilon = 1/2$ is above the theoretical curve, which can be explained by the fact that the foreground strips were only 2 pixels wide in these images, leading to a higher response of this filter. The same experiment was done with $\varepsilon = 1/4$ and S/N ranging from 5 to 30. Again, experimental and theoretical results were computed. The results are shown in Fig. 5.

3.2.1. Poisson noise. Using one image from the set used above, with $q = 0.25$, $\varepsilon = 1/16$, $A = 100$, and B ranging from 0 to 200, the effect of the contrast on the Poisson noise bias was computed. For each data point the average result of 10 images was used. Figure 6 shows the experimental and theoretical curves for high S/N. Again, there is a slight systematic difference between the experimental data and the predictions from the high S/N approximation. It is also clear that the quadratic filters show the best performance.

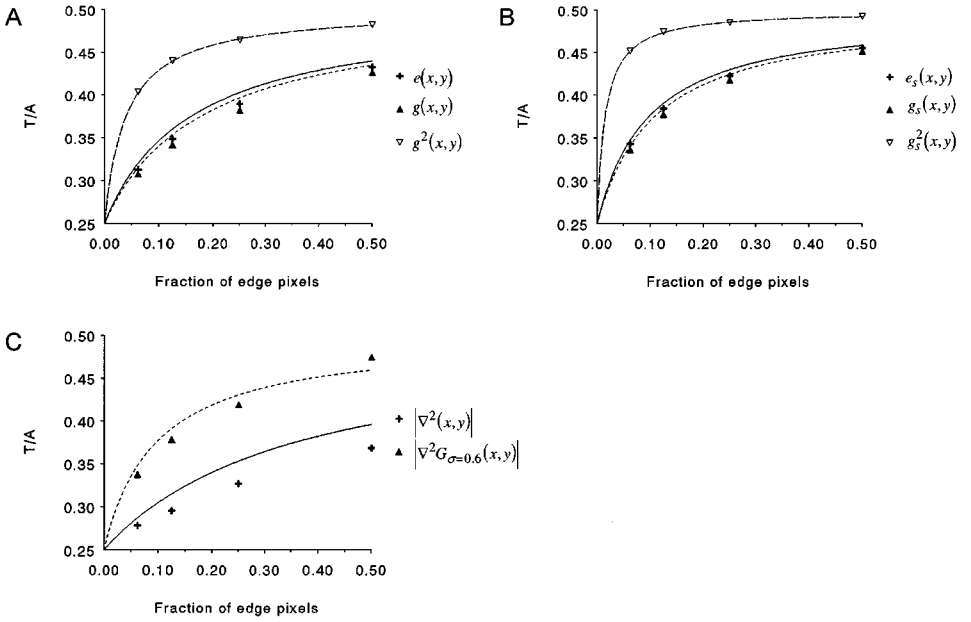

FIG. 4. The effect of the fraction of edge pixels (ε) on the noise bias: T/A vs ε, for uniform Gaussian noise with $q = 0.25$ and S/N = 10. The symbols mark experimental values, the curves were computed from the high S/N approximation. Part (A) shows the results for the Δ-based filters; (B) shows the results for the Δ_{Sobel}-based filters. The solid lines represent $e(x, y)$, the dotted $g(x, y)$ and the dashed $g^2(x, y)$.

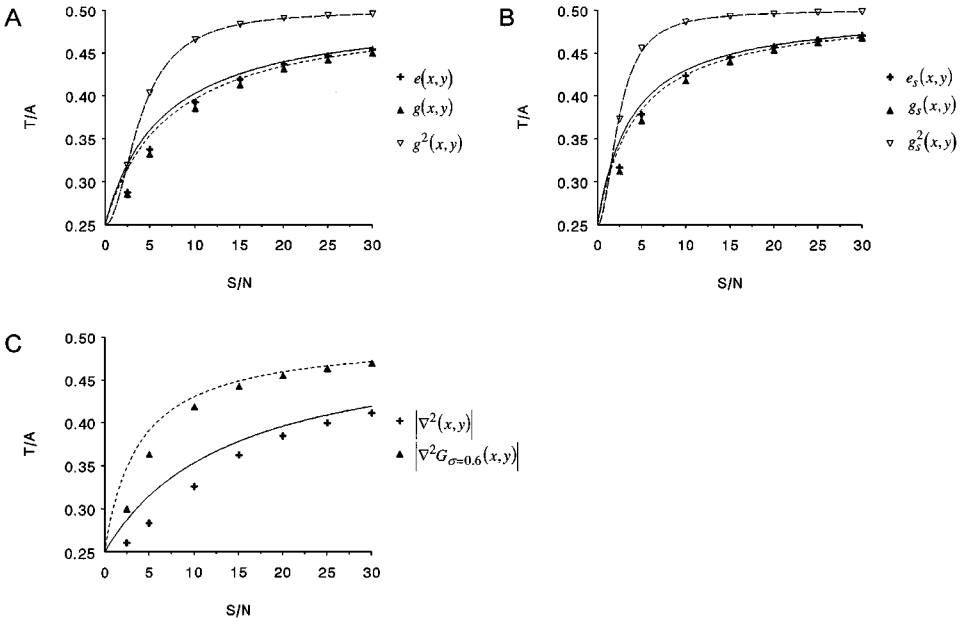

FIG. 5. The effect of S/N on the noise bias: T/A vs S/N for $q = 0.25$ and $\varepsilon = 1/16$. The symbols mark experimental values, the curves were computed from the high S/N approximation. Part (A) shows the results for the Δ-based filters; (B) shows the results for the Δ_{Sobel}-based filters. The solid lines represent $e(x, y)$, the dotted $g(x, y)$ and the dashed $g^2(x, y)$.

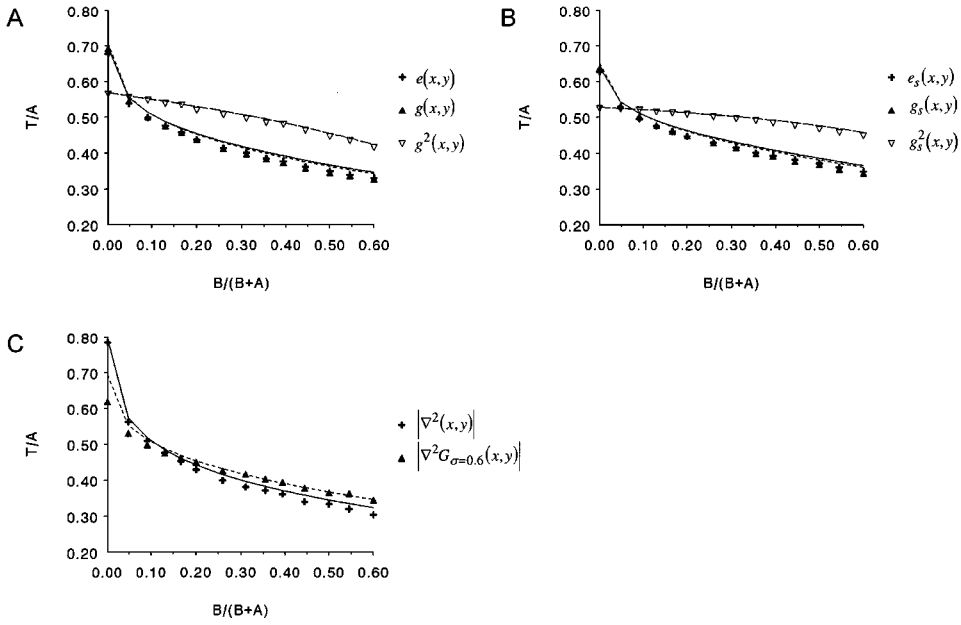

FIG. 6. The effect of Poisson noise on the noise bias for different image contrasts: T/A vs the inverse of the image contrast $B/(B+A)$, for $A=100$, $q=0.25$, and $\varepsilon=1/16$. The symbols mark experimental values, and the curves were computed from the high S/N approximation. Part (A) shows the results for the Δ-based filters; (B) shows the results for the Δ_{Sobel}-based filters. Note that the bias is positive for high image contrast, even though $q<0.5$. The solid lines represent $e(x, y)$, the dotted $g(x, y)$ and the dashed $g^2(x, y)$.

3.3. *The Noise Bias Reduction Techniques*

To demonstrate the noise reduction technique, T/A was computed for 10 images with $q=0.25$, $\varepsilon=1/16$, $A=1000$, and $B=0$, and Gaussian noise with $\sigma=100$ (S/N = 10), for values of λ ranging from 0 (no correction) to 10. To compare results for similar values of λ, the response of the Marr-Hildreth filter was scaled linearly to have a $g=1$. The mean values and standard errors of T/A are shown in Fig. 7. The results for all filters based on the simple first derivative converge on the ideal value of 0.5 for $\lambda\approx 5$. For the Sobel versions, convergence takes place at $\lambda\approx 3$, consistent with a response to noise $\sqrt{8/3}=1.633$ times lower than the "non-Sobel" case. When λ approaches 10 and the noise threshold approaches A, the standard error in T/A increases for all filters.

When applying the method to second derivative filters, the results were less satisfactory. The poor noise performance of these filters is a clear disadvantage when used in this context. Only in the case of the Marr-Hildreth filter could the bias be reduced to approximately zero, but at $\lambda\approx 7$, which sacrifices a great deal of sensitivity. The standard errors are also higher than when using first derivatives.

The same experiment was repeated for the positive bias for Poisson noise, for $q=0.25$, $\varepsilon=1/16$, $A=100$, and $B=0$. A similar convergence at the same values of λ as in the Gaussian case was observed (data not shown).

The statistic T was computed for image sets with Gaussian noise, but using filters of the form $g^m(x, y)$, with $m=1, 2, 3, \ldots, 10$, and $g(x, y)$ as in (14). T/A and standard errors were computed. The results are shown in Fig. 8. Evidently the use of higher orders does reduce the bias in a similar manner to Kittler's method, but the standard error increases very rapidly.

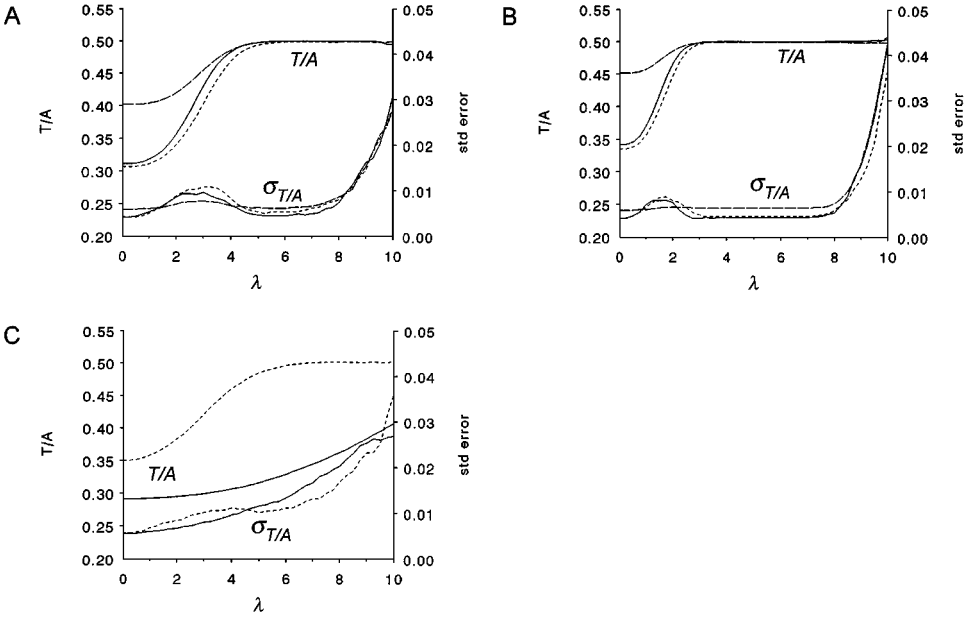

FIG. 7. The noise bias reduction technique for uniform Gaussian noise: T/A and its standard error $\sigma_{T/A}$ vs λ, for $q = 0.25$, S/N = 10, and $\varepsilon = 1/16$. Part (A) shows the results for the Δ-based filters; (B) shows the results for the Δ_{Sobel}-based filters. Note the increase in $\sigma_{T/A}$ as λ approaches S/N (10). The solid lines represent $e(x, y)$, the dotted $g(x, y)$, and the dashed $g^2(x, y)$.

4. DISCUSSION

It has also been shown that, apart from the noise bias described by Kittler *et al.* [2], another bias may be caused by curvature of edges. The discussion of noise bias has been extended to include Poisson noise, which shows different behavior at high image contrasts, yielding positive rather than negative biases for images containing predominantly background pixels.

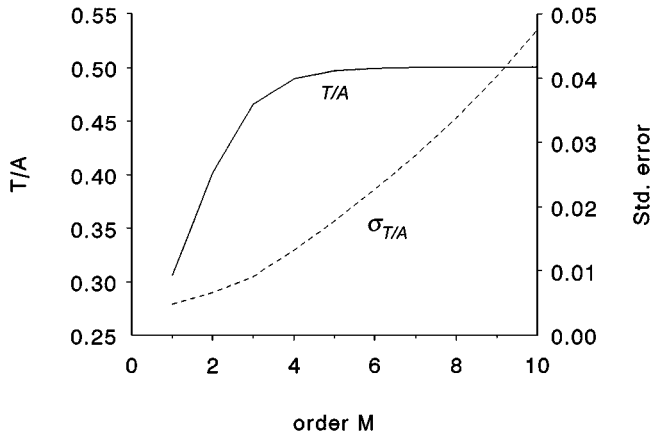

FIG. 8. T/A and its standard error $\sigma_{T/A}$ as a function of the order m of edge detectors of the form $g^m(x, y)$ for $q = 0.25$, S/N = 10, and $\varepsilon = 1/16$. The solid line denotes T/A, the dotted $\sigma_{T/A}$. As m increases, the bias is reduced at the expense of a higher standard error.

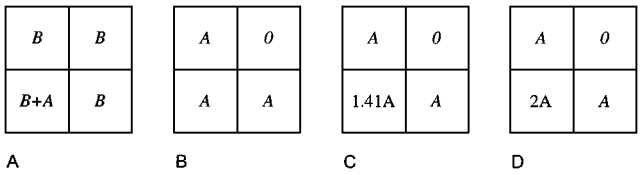

FIG. 9. Discrete curvature bias of a 90° corner for the three Δ-based filters: (A) shows the grey levels at the pixels at the corner point; (B) shows the values of $e(x, y)$, yielding a local threshold of $T = B + A/3$; (C) shows $g(x, y)$, yielding $T = B + A/2.41$; (D) shows $g^2(x, y)$, yielding $T = B + A/2$.

In all, 16 edge detectors have been compared, regarding both forms of bias. It has been shown that proper selection of the edge detector can reduce the curvature bias, especially by the use of higher order filters. The fact that the curvature bias for ideal edges is zero for the quadratic first derivative filters and for second derivative filters may be explained by looking at the response of each filter to a corner of an object (Fig. 9). At such a corner the $e(x, y)$ filter assigns the same weights to all edge pixels, resulting in a T of $1/3$ (for the corner). The $g(x, y)$ filter assigns the corner pixel a weight of $\sqrt{2}A$, and the outside pixels a weight of A, reducing the bias. Both the Laplacian and the $g^2(x, y)$ filters assign a weight of $2A$ to the inside corner pixel and A to the outside pixels, yielding an ideal value of T locally. Curiously, this optimimum behavior is retained when considering objects in digital images of an arbitrary number of dimensions. Consider a boundary point at a location $\mathbf{x} \in \mathbb{Z}^N$ with N the number of dimensions. Assume that (after reordering of the dimensions) the grey level f at and around $\mathbf{x}$ is given by

$$f(\mathbf{x}) = B + A$$

$$\left.\begin{aligned} f(\mathbf{x} + \hat{x}_i) &= B + A \\ f(\mathbf{x} - \hat{x}_i) &= B \end{aligned}\right\} \quad \text{if } 1 \le i \le M$$

$$f(\mathbf{x} \pm \hat{x}_i) = B + A \qquad \text{if } M < i \le N$$

and

$$f(\mathbf{x} + \mathbf{y}) = B \quad \text{for any nonunit vector } \mathbf{y} \in \mathbb{Z}^N,$$

with $\hat{x}_i$ a unit vector in the ith direction, and $0 < M \le N$. In other words, the corner is no sharper than a (hyber)cube corner. In this case the weight assigned to the corner point is

$$g^2(\mathbf{x}) = \sum_{i=1}^{N} \{ f(\mathbf{x} + \hat{x}_i) - f(\mathbf{x} - \hat{x}_i) \}^2 = MA^2$$

and

$$|\nabla^2(\mathbf{x})| = \left| N f(\mathbf{x}) - \sum_{i=1}^{N} f(\mathbf{x} + \hat{x}_i) + f(\mathbf{x} - \hat{x}_i) \right| = MA$$

which is precisely compensated by total weight assigned to the M background neighbors of the corner point, each of which have weight A^2 and A in the $g^2(x, y)$ and Laplacian cases,

respectively. Thus, corners no sharper than a (hyper) cube corner are assigned the correct threshold. They are of course assigned large weights, which may increase the variance of the T-statistic. For three dimensions this should not pose large problems.

Noise bias may be reduced either by the use of higher order filters or by the methods developed by Kittler *et al.* [2]. The use of higher order filters is less effective since it increases the variance of T. Kittler's method provided good bias reduction, with an increase in the variance of T only at very high λ. In the first derivative edge detectors tested which did not use smoothing, $\lambda \approx 5$ yields near optimal results. The first derivative edge detectors which use Sobel kernels converged sooner: at $\lambda \approx 3$. The ratio of these two values is in good agreement with the theory. In all cases, as $\lambda\sigma$ approaches the magnitude of the step itself, the variance of T increases. This is probably caused by elimination of real edge pixels; as $\lambda\sigma$ increases, fewer and fewer pixels are used to compute T.

The overall winner in this comparison of edge detectors is the quadratic Sobel filter $g_s^2(x, y)$. It shows little curvature bias, T-statistics computed from it do not show large variance, and there is a wide range of values of λ over which the bias is negligible. This implies that it is also relatively insensitive to the estimate of σ. It should be stressed that neither the data nor the theory presented prove that second-order filters, or Sobel kernels are optimal in any sense other than in comparison with the other filters tested. The methods developed to evaluate edge detectors do provide a means to test any other edge detector in the context of generalized RATS. The optimum filter proposed by Shen and Castan [6], using infinite base exponential smoothing to reduce noise while retaining good localization, is a promising candidate in this respect. The infinite base of their filter may limit its application on computers with the modest computing power prevalent in many biomedical settings.

The entire discussion has focused on global thresholding, yet it is very easy to use all forms of RATS for local thresholding. Any of the approaches given by Kittler *et al.* [1, 2] are equally applicable to any form of RATS. Our own implementation of RATS [10] uses a quadtree hierarchy or image pyramid. Using the original formulation of RATS (using $e(x, y)$), 512×480 pixel images of fluorescently stained bacteria, with differences in brightness between individual objects of a factor of 8 or more within the same field of view, could be segmented properly in just 2.5 s on a 66 MHz 80486-DX2 based personal computer [10]. Similar results have been obtained with the improvements suggested in this paper (data not shown).

ACKNOWLEDGMENTS

I thank J. B. T. M. Roerdink and N. Petkov for carefully reading the manuscript and providing useful comments and encouragement. This work, and the image analysis research programme at the Department of Medical Microbiology, has been supported by the Dutch Foundation for Micro-Morphological Systems' Development, and Symbiopharm GmbH in Herborn, Germany.

REFERENCES

1. J. Kittler, J. Illingworth, and J. Föglein, Threshold selection based on a simple image statistic, *Comp. Vision Graph. Image Proc.* **30** (1985), 125–147.
2. J. Kittler, J. Illingworth, J. Föglein, and K. Paler, An automatic thresholding algorithm and its performance, *in* "Proc. 7 Int. Conf. Pattern Recognition Montreal 1984," pp. 287–289, 1985.
3. A. Beghdadi and A. Le Negrate, Contrast enhancement technique based on local detection of edges, *Comp. Vis. Graph. Image Proc.* **46** (1989), 162–174.

4. F. Kammoun and J. P. Astruc, Optimum edge detection for object-background picture, *CVGIP: Graph. Models Image Proc.* **56** (1994), 25–28.
5. D. J. Williams and M. Shah, Edge contours using multiple scales, *Comp. Vis. Graph Image Proc.* **51** (1990), 256–274.
6. J. Shen and S. Castan, An optimal linear operator for step edge detection, *CVGIP: Graph. Models Image Proc.* **54** (1992), 112–133.
7. J. W. Weszka and A. Rosenfeld, Histogram modification for threshold selection, *IEEE Trans. Systems Man Cybern.* **SMC9** (1979), 38–52.
8. M. Nadler, A note on the coefficients of compass mask convolutions, *Comp. Vis. Graph. Image Proc.* **51** (1990), 96–101.
9. J. C. Russ, "The Image Processing Handbook," p. 58, CRC Press, Boca Raton, FL, 1992.
10. M. H. F. Wilkinson, Rapid automatic segmentation of fluorescent and phase-contrast images of bacteria, *in* "Fluorescence Microscopy and Fluorescent Probes" (J. Slavik, Ed.), pp. 261–266, Plenum Press, New York, 1996.