Gaussian-Based Edge-Detection Methods—A Survey

Mitra Basu, *Senior Member, IEEE*

***Abstract*—The Gaussian filter has been used extensively in image processing and computer vision for many years. In this survey paper, we discuss the various features of this operator that make it the filter of choice in the area of edge detection. Despite these desirable features of the Gaussian filter, edge detection algorithms which use it suffer from many problems. We will review several linear and nonlinear Gaussian-based edge detection methods.**

***Index Terms*—Adaptive techniques, edge detection, edge localization, Gaussian filter, multiscale analysis.**

I. INTRODUCTION

THE purpose of this paper is to present a survey of Gaussian-based edge detection techniques. The detection of edges in an image has been an important problem in image processing for more than 50 years. In a gray level image, an edge may be defined as a sharp change in intensity. Edge detection is the process which detects the presence and locations of these intensity transitions. The edge representation of an image drastically reduces the amount of data to be processed, yet it retains important information about the shapes of objects in the scene. This description of an image is easy to integrate into a large number of object recognition algorithms used in computer vision and other image processing applications.

Over the years, many methods have been proposed for detecting edges in images. Some of the earlier methods, such as the Sobel and Prewitt detectors [27], [30], used local gradient operators which only detected edges having certain orientations and performed poorly when the edges were blurred and noisy. It should be mentioned here that one can combine such directional operators to approximate the performance of a rotationally invariant operator. Since then, more sophisticated operators have been developed to provide some degree of immunity to noise, to be nondirectional,[1] and to detect a more accurate location of the edge. The majority of these are linear operators that are derivatives of some sort of smoothing filter. Shen and Castan [33] used a symmetrical exponential filter in edge detection. However, since it was originally proposed by Marr and Hildreth in 1980 [23], the Gaussian filter is by far the most widely used smoothing filter in edge detection. The reasons for this are presented later in this paper.

Since the Gaussian filter is so commonly used, in this paper we review several Gaussian-based edge detection techniques. Our goal is not to give an exhaustive inventory of edge detection algorithms. Other surveys of edge detection may be found in [24], [40], and [41]. In Section II, we outline some of the major problems encountered during an edge detection process. Section III discusses the features of the Gaussian filter that have contributed to its popularity in edge detection. In Section IV, we present a discussion of the two most well-known Gaussian edge operators: 1) the Marr-Hildreth detector and 2) the Canny detector. We also review several other detectors that have developed in more recent years. A brief summary is given in Section V.

Manuscript received July 28, 2000; revised June 1, 2002 and July 1, 2002. This work was supported by a grant from the NSF CDA-9114481. This paper was recommended by Associate Editor V. Murino.

The author is with the Department of Electrical Engineering, The City College of New York, New York, NY 10031 USA (e-mail: basu@ccny.cuny.edu).

Digital Object Identifier 10.1109/TSMCC.2002.804448

[1]This is not necessarily a desirable property, since information on edge direction is usually needed by high-level processes.

II. PROBLEMS OF EDGE DETECTION

The separation of a scene into object and background is an essential step in image interpretation. This is a process that is carried out effortlessly by the human visual system, but when computer vision algorithms are designed to mimic this action, several problems can be encountered. This section describes some of the problems involved in detecting and localizing edges.

Due to the presence of noise and quantization of the original image, during edge detection it is possible to locate intensity changes where edges do not exist. For similar reasons, it is also possible to completely miss existing edges. The degree of success of an edge-detector depends on its ability to accurately locate true edges.

Edge localization is another problem encountered in edge detection. The addition of noise to an image can cause the position of the detected edge to be shifted from its true location. The ability of an edge-detector to locate in noisy data an edge that is as close as possible to its true position in the image is an important factor in determining its performance.

Another difficulty in any edge detection system arises from the fact that the sharp intensity transitions which indicate an edge are sharp because of their high-frequency components. As a result, any linear filtering or smoothing performed on these edges to suppress noise will also blur the significant transitions. However, some form of smoothing is necessary since edge detection depends on differentiating the image function and this amplifies all high-frequency components of the signal, including those of the noise. Low-pass filters are the most widely used smoothing filters. The amount of smoothing applied depends on the size or scale of the smoothing operator. In general, for a small scale, the detector extracts fine details of intensity changes from the image, but tends to be more sensitive to noise. A larger scale extracts coarse details of intensity changes, but some of the detected edges tend to have a large localization error. Selecting a single scale of smoothing which is optimal for all edges in an image is very difficult. One filter size may not be good enough to remove noise while keeping good localization.

1094-6977/02$17.00 © 2002 IEEE

Multiscale edge detection offers an alternative. This involves applying smoothing operators of different sizes to an image, extracting the edges at each scale, and combining the recovered edge information to form a more complete picture of the actual edge representation of the image. The basic premise of this technique is that different parts of an image have varying degrees of noisiness and types of edges; therefore, each part of the image needs to be smoothed differently. However, there are problems associated with this technique, namely, how many filters should be used, how to determine the scales of the filters, and how to combine the responses from each filter so as to create a single edge map.

III. Significance of the Gaussian Filter

The most widely used smoothing filters are Gaussian filters. Such filters have been shown to play an important role in edge detection in the human visual system, and to be extremely useful as detectors for edge and line detection. Marr and Hildreth [23] demonstrated that the Gaussian filter (along with the Laplacian operator) is very similar to the difference of Gaussians (DOG) filter. This is a well-known approximation to the shape of spatial receptive fields in the visual system of cats that has also been proposed for humans. By variational methods, Canny [7] derives an optimal edge detection operator which turns out to be well-approximated by the first derivative of a Gaussian function.

Babaud *et al.* [2] proved that when one-dimensional (1-D) signals are smoothed with a Gaussian filter, the scale space representation of their second derivatives shows that existing zero-crossings disappear when moving from a fine-to-coarse scale, but new ones are never created.[2] They also proved that for a wide category of signals, the Gaussian function is the only filter that has this property. This unique property makes it possible to track zero-crossings over a range of scales, and also gives the ability to recover the entire signal at sufficiently small scales. Yuille and Poggio [39] extended this work to two-dimensional (2-D) signals and proved that with the Laplacian, the Gaussian function is the only filter in a wide category that does not create zero-crossings as the scale increases. They also showed that for nonlinear directional derivatives along the gradient, there is no filter that does not create zero-crossings as the scale increases.

Another important property of the Gaussian filter is that it is the only operator that satisfies the uncertainty relation

$$\Delta x \Delta \omega \geq \frac{1}{2} \tag{1}$$

where Δx and $\Delta \omega$ are its variance in spatial and frequency domains, respectively. This property allows the Gaussian operator to give the best tradeoff between the conflicting goals of the localization in spatial and frequency domains simultaneously.

The 2-D Gaussian filter is also the only rotationally symmetric filter that is separable in Cartesian coordinates. Separability is important for computational efficiency when implementing the smoothing operation by convolutions in the spatial domain.

[2]This ensures that no new features are introduced by the edge-detection process and image integrity is maintained.

IV. Gaussian-Based Edge Detection Techniques

Marr and Hildreth [23] originally proposed the spatial coincidence assumption which led to the idea of multiscale edge analysis. Having observed that significant intensity changes occur at different scales within an image, they concluded that optimal detection of these changes would require the use of a filter that operates at several different scales. They further argued that the edge maps of the different scales each contained important information about physically significant phenomena. If an edge is present at several different scales, then it represents the presence of an image intensity change due to a single physical phenomenon. If an edge is present at only one scale, then it may be that two independent physical phenomena are operating to produce intensity changes in the same region of the image. However, this assumption is mainly based on intuition and some aspects of it still remain open for further research.

Marr and Hildreth also argued that the optimal smoothing filter for images should be localized in both spatial and frequency domains, thereby satisfying the uncertainty relation given in (1). Consider the Gaussian operator in two dimensions given by

$$g(x, y) = \frac{1}{2\pi\sigma^2} e^{-(x^2 + y2/2\sigma^2)} \tag{2}$$

where σ is the standard deviation of the Gaussian function and (x, y) are the Cartesian coordinates of the image. Marr and Hildreth suggested that by applying Gaussian filters of different scales (σ) to an image, a set of images with different levels of smoothness can be obtained. To detect the edges in these images it is necessary to find the zero-crossings of their second derivatives. Marr and Hildreth achieved this by using the Laplacian of a Gaussian (LOG) function as a filter

$$\begin{aligned}\nabla^2 g(x, y) &= \frac{d^2}{dx^2} g(x, y) + \frac{d^2}{dy^2} g(x, y) \\ &= \frac{x^2 + y^2 - 2\sigma^2}{2\pi\sigma^6} e^{-(x^2 + y^2/2\sigma^2)}.\end{aligned} \tag{3}$$

This is an orientation-independent operator and its scale is given by σ. It breaks down at corners, curves, and at locations where image intensity function varies in a nonlinear manner along an edge [6]. As a result of the breakdown, the true position of the edge is not reported by the detector.

The Marr-Hildreth operator formally introduced Gaussian filter into the edge-detection process with *so-claimed* support from the biological vision. This is a turning point in the low-level image processing research area. The newly introduced concept multiscale/multiresolution came to play an important role in the entire image processing area and injected sophisticated mathematical techniques (e.g., variational approach, nonlinear partial differential equations) into the segmentation problem. It initiated three distinct research endeavors:

1) a group of researchers directed their efforts to analyze and improve Marr-Hildreth operator within the umbrella of linear constraints;
2) a second group focused on biological vision to achieve performance comparable to that of mammalian visual system;

3) yet another group moved into nonlinear domain.

The rest of this paper summarizes efforts of researchers from each group in the order mentioned above.

It has been found that zero-crossings are only reliable in locating edges if they are well separated and the signal-to-noise ratio (SNR) in the image is high. An analytical study of the response of the 1-D second derivative of the Gaussian edge operator on ideal infinite-width step and ramp edges and finite-width step and staircase step edges is reported in [32]. It is shown that for ideal step and ramp edges, the location of the zero-crossing is exactly at the location of the edge. However, the location shifts from the true edge location for the finite-width case. This shift is a function of the standard deviation σ of the Gaussian. The other problem is the detection of false edges. The reason is that zero-crossings correspond to local maxima and minima in the first derivative of an image function, whereas only local maxima indicate the presence of real edges [8], [22]. LOG filtered images also suffer from the problem of missing edges—edges in the original image may not have corresponding edges in a filtered image. In addition, it turns out to be very difficult to combine LOG zero-crossings from different scales, primarily because of the following [28]:

1) a physically significant edge does not match a zero-crossing for more than a few and very limited number of scales;
2) zero-crossings in larger scales move very far away from the true edge position due to poor localization of the LOG operator;
3) there are too many zero-crossings in the small scales of a LOG filtered image, most of which is due to noise.

Relating image structures across scale is intrinsically problematic since, zero-crossings merge tangentially. Thus, one cannot rely upon zero-crossings to track edges. Edge-linking algorithms based on this approach will remain heuristic in nature. Furthermore, physiological experiments have found no evidence to support zero-crossings as a model for biological vision.

Haralick [14] proposed the use of zero-crossing of the second directional derivative of the image intensity function. This is theoretically the same as using the maxima in the first directional derivatives, and in one dimension is the same as the LOG filter.

Canny [7] developed an operator that has also become a standard gauge in edge detection. His technique is based on optimizing three criteria desired for any edge detection filter: good detection, good localization, and only one response to a single edge. For good detection, the probability of false detection should decrease with increasing SNR. The SNR at an edge point is given by

$$\mathrm{SNR} = \frac{\left|\int_{-W}^{W} G(-x) f(x) dx\right|}{n_0 \sqrt{\left(\int_{-W}^{W} f^2(x) dx\right)}} \tag{4}$$

where

$f(x)$ optimal operator;
$G(x)$ the edge;
n_0 root mean squared value of the noise.

It is assumed that the optimal filter has a finite impulse response bounded by $[-W, W]$.

For good localization, the expression to be maximized is

$$\text{Localization} = \frac{\left|\int_{-W}^{W} G'(-x) f'(x) dx\right|}{n_0 \sqrt{\int_{-W}^{W} f'^2(x) dx}} \tag{5}$$

where $f'(x)$ is the derivative of $f(x)$.

The criterion for eliminating multiple responses to a single edge is designed to minimize the number of spurious maxima within the operator's spatial extent. The expression for the distance between adjacent maxima of the filter output in response to noise is

$$x_{\max} = 2\pi \left(\frac{\int_{-\infty}^{\infty} f'^2(x) dx}{\int_{-\infty}^{\infty} f''^2 dx} \right)^{1/2} = kW. \tag{6}$$

Fixing the constant k to be as large as possible results in the distance between spurious maxima being as great as possible.

In order to simplify the analysis for step edges, Canny computed the function $f(x)$ so that the localization is maximized while keeping the single response criterion constant. By variational methods, he derived a set of filters corresponding to various values of the single response criterion. Canny showed that for a 1-D step edge, the derived optimal filter can be approximated by the first derivative of a Gaussian function with variance σ

$$f(x) \approx -\frac{x}{\sigma^2} e^{-(x^2/\sigma^2)}. \tag{7}$$

In two dimensions, his edge-detector uses two filters representing derivatives along the horizontal and vertical directions to optimally detect the edges in an image with additive Gaussian white noise. Canny's operator is not unique as it varies with the edge profile (e.g., step edge or ramp edge). However, for step edges, Canny's optimal operator is similar to the LOG operator because the maxima in the output of a first derivative operator corresponds to the zero-crossings in the Laplacian operator used by Marr and Hildreth.

Canny also proposed a scheme for combining the outputs from different scales. His strategy is fine-to-coarse and the method is called *feature synthesis*. It starts by marking all the edges detected by the smallest operators. It then takes the edges marked by the small operator in a specific direction and convolves them with a Gaussian normal to the edge direction of this operator so as to synthesize the large operator outputs. It then compares the actual operator outputs to the synthesized outputs. Additional edges are marked if the large operator detects a significantly greater number of edges than what is predicted by the synthesis. This process is then repeated to mark the edges from the second smallest scale that were not marked by the first, and then to mark the edges from the third scale that were not marked by either of the first two, and so on. In this way, it is possible to include edges that occur at different scales even if they do not spatially coincide.

Canny's edge-detector uses adaptive thresholding with hysteresis to eliminate streaking of edge contours. Two thresholds are involved, with the lower threshold being used for edge elements belonging to edge segments already having points above the higher threshold. The thresholds are set according to the amount of noise in the image, which is determined by a noise estimation procedure.

The problem with Canny's edge detection is that his algorithm marks a point as an edge if its amplitude is larger than that of its neighbors without checking that the differences between this point and its neighbors are higher than what is expected for random noise. His technique causes the algorithm to be slightly more sensitive to weak edges, but it also makes it more susceptible to spurious and unstable boundaries wherever there is an insignificant change in intensity (e.g., on smoothly shaded objects and on blurred boundaries).

In [31], Schunck introduces an algorithm for the detection of step edges using Gaussian filters at multiple scales. The initial steps of Schunck's algorithm are based on Canny's method. The algorithm begins by convolving an image with a Gaussian function. The gradient magnitude and gradient angle are then computed for each point in the resulting smoothed data array. Next, the gradient ridges in the results of the convolution are thinned using nonmaxima suppression. Then, the thinned gradient magnitudes are thresholded to produce the edge map.

Multiresolution edge detection is incorporated by repeating the steps above for several scales of the Gaussian filter. The gradient magnitude data at the largest scale will contain large ridges which correspond to the major edges in the image. As the scale decreases, the gradient magnitude data will contain an increasing number of ridges, both large and small. Some of these correspond to major edges, some to weaker edges, and the rest are due to noise and unwanted details. The gradient magnitudes over the chosen range of scales are multiplied to produce a composite magnitude image. Ridges that appear at the smallest scale and correspond to major edges will be reinforced by the ridges at larger scales. Those that do not will be attenuated by the absence of ridges at larger scales. Therefore, in the combined magnitude image, the ridges that correspond to major edges are much higher than the ridges that do not. Nonmaxima suppression is then performed using sectors obtained from the gradient angle of the largest filter.

Schunck's algorithm chooses the width of the smallest Gaussian filter to be around $w = 7$, and the filters that are used differ in width by a factor of two. However, he did not discuss how to determine the number of filters to use. In addition, by choosing such a large size for the smallest filter, Schunck's technique loses a lot of important details which may exist at smaller scales.

Witkin [37] was one of the first to study the property of zero-crossings across scales for 1-D signals. The signal is first smoothed by a Gaussian filter of varying σ. The idea is to examine the smoothed signal at various scales. The zero-crossings of the second derivative are marked. This representation known as the scale-space representation of a signal contains the location of a zero-crossing at all scales starting from the smallest scale to the scale at which it disappears. Witkin's work initiated the study of edge detection as a function of scale. The knowledge form these studies is used by various researchers to propose algorithms to combine edges for better edge detection. Bergholm [5] proposed an algorithm which uses the Gaussian filter and combines edge information moving from a coarse-to-fine scale. His method is called *edge focusing*, and uses a rule-based approach for detecting local features and for tracking and predicting a possible scale parameter. Both the Marr-Hildreth and Canny edge-detectors are possible schemes that can be used in edge focusing.

The image is first smoothed with a large scale Gaussian filter and then the edge detection process is performed using adaptive thresholding. Assuming that edge contours rarely move by more than two pixel for a unit change in the scale parameter, the exact location of the edges is determined by tracking them over decreasing scales. Therefore, the results from one scale of the edge-detector is used to predict the locations of edges in the next, smaller, scale.

The idea behind edge focusing is to reverse the effect of the blurring caused by the Gaussian operator. Blurring is not a desired feature as it results in poor edge localization. It is simply used as a means of removing the noise and other unnecessary features. The most obvious way of undoing the blurring process is to start with edges detected at the coarse scale and gradually track or *focus* these edges back to their original locations in the fine scale.

There are several problems associated with edge focusing, the foremost being how to determine the starting and ending scales of the Gaussian filter. Bergholm suggests the range $(3 \leq \sigma < 6)$ for the maximum scale, but did not specify a minimum scale. He also did not discuss in detail how to choose the threshold which is used at the coarsest level, and this is a parameter which is critical in determining how well the algorithm performs. If it is too high, a number of true edge points will be eliminated right from the start. If it is too low, the output of the edge focusing could be very noisy. In addition, since edge focusing is obtained at a finer resolution, some edges (i.e., the blurred ones, such as shadows) present a juggling effect at small scales. This is due to the splitting of a coarse edge into several finer edges, and tends to give rise to broken, discontinuous edges.

Lacroix [20] avoids the problem of splitting edges by tracking edges from a fine-to-coarse resolution. His algorithm detects edges using the Canny method of nonmaxima suppression of the magnitude of the gradient in the gradient direction. His method then considers three scales: 1) σ_0; 2) σ_1; and 3) σ_2. The smallest scale, σ_0, is the detection scale, and is the finest resolution at which an edgel (i.e., a short, linear edge element) first appears. The largest scale, σ_2, is the blurring scale, and is the coarsest resolution at which the edgel still remains. The intermediate resolution is computed as

$$\sigma_1 = \sigma_0 + \frac{(\sigma_2 - \sigma_0)}{3}. \tag{8}$$

An edgel is validated if: 1) it is the local maximum of a Gaussian gradient and 2) the two regions it separates are homogeneous and significantly different from one another. Only validated edgels are then tracked through the scales.

Although Lacroix avoids the problem of splitting edges, he introduces the problem of localization error as it is the coarsest resolution that is used to determine the location of the edges. He also provides no explanation as to how to decide which scales are to be used and under what conditions.

Williams and Shah [36] devised a scheme to find edge contours using multiple scales. They analyzed the movement of edge points smoothed with a Gaussian operator of different sizes, and used this information to determine how to link edge points detected at different scales. Their method, following the lead of Canny, uses a gradient of Gaussian operator to determine gradient magnitude and direction, followed by nonmaxima suppression to identify ridges in the gradient map. Since the resulting ridges are often more than one pixel wide, the gradient maxima points are thinned and then linked using an algorithm which assigns weights based on four measures: 1) noisiness; 2) curvature; 3) contour length; and 4) gradient magnitude. The set of points having the highest average weight is chosen.

The algorithm extends to multiresolution by convolving the image with the Gaussian filter at three scales: 1) σ; 2) $\sqrt{2}\sigma$; and 3) 2σ. First, the best partial edge contours are found using the largest scale. Then, the next smaller scale is used, and the regions around the end points of the contours are examined to determine if there are possible edge points at the smaller scale having similar directions to the end points of the contours. The original algorithm is then carried out for each of these possible edge points, and the best are chosen as an extension to the original edge contour. The scale is decreased to the smallest scale, and the process is repeated.

Although Williams and Shah specify the number of scales to be used and the relationship between these scales, they did not suggest the best way to choose the value of σ and under what conditions.

Goshtasby [13] proposes an algorithm that works on a *modified* scale-space representation of an image. The author creates a representation of an image by recording the signs of pixels (instead of the zero-crossings) after filtering with LOG operator. The advantage of such a representation over that of regular scale-space is that the new representation does not contain any disconnected arches. The scale-step size is determined adaptively using the image structure in the following manner. Results of convolution of an image at scales σ_1 and σ_2 are overlaid one on top of another. If more than two regions of the same sign fall on top of each other, the complete information on behavior of edges between these two scales is lacking. Therefore, one must consider an intermediate scale between σ_1 and σ_2. Otherwise, there is no new edge information between these two scales. This procedure makes a decision on step sizes as one goes along rather than choosing step sizes before the process starts. This also avoids the use of too many or too few scales. This is the crucial part of the algorithm. Once the scale-space image is constructed using the correct values for σ, tracking of edge from lowest to highest resolution is possible since there are no disconnected arches. The major problem with Goshtasby's edge focusing algorithm is the need for a considerable amount memory to store the three-dimensional (3-D) edge images.

To avoid the common problems associated with integrating edges detected at multiple scales, Jeong and Kim [16] proposed a scheme which automatically determines the optimal scales for each pixel before detecting the final edge map. To find the optimal scales for a Gaussian filter, they define an energy function that quantitatively determines the usefulness of the possible edge map. They approach the edge detection problem as finding the size of the Gaussian filter $\sigma(x, y)$ which minimizes the energy function over σ, $E(\sigma)$. The parameter σ is chosen so that:

1) it is large at uniform intensity areas, thereby smoothing out random noise;
2) it is small at locations where the intensity changes significantly, thus retrieving edges accurately;
3) it does not change sharply from pixel to pixel, therefore avoiding broken edges due to random noise.

If g and I denote the 2-D Gaussian filter and image function, respectively, the equation to be minimized is

$$E(\sigma) = \int\int \left((I * g)^2 + \lambda|\nabla\sigma^{-1}|^2\right) dxdy. \tag{9}$$

Since the Jeong-Kim algorithm is designed to adaptively find the optimal scale of the Gaussian filter for every location in the image function, it can be easily incorporated into any Gaussian-based edge-detection technique. However, this algorithm does result in reduced performance when it comes to detecting straight lines in vertical or horizontal directions. The algorithm also has the disadvantage of low-speed performance [4].

Deng and Cahill [9] also use an adaptive Gaussian filtering algorithm for edge detection. Their method is based on adapting the variance of the Gaussian filter to the noise characteristics and the local variance of the image data. Based on observations of how the human eye perceives edges in different images, they concluded that in areas with sharp edges, the filter variance should be small to preserve the sharp edges and keep the distortion small. In smooth areas, the variance should be large so as to filter out noise. They proposed that the variance of a 1-D Gaussian filter at location x is

$$\sigma^2(x) = \frac{k{\sigma_n}^2}{\sigma_f^2(x) + \sigma_n^2} \tag{10}$$

where

k	scaling factor;
σ_n^2	noise variance;
$\sigma_f^2(x)$	local variance of the signal.

The major drawback of this algorithm is that it assumes the noise is Gaussian with known variance. In practical situations, however, the noise variance has to be estimated. The algorithm is also very computationally intensive.

In [4], Bennamoun *et al.* present a hybrid detector that divides the tasks of edge localization and noise suppression between two subdetectors. This detector is the combination of the outputs from the Gradient of Gaussian and Laplacian of Gaussian detectors. The hybrid detector performs better than both the first-order and second-order detectors alone, in terms of localization and noise removal [4]. The authors extended the work to automatically determine the optimal scale σ and threshold Th, of the hybrid detector. They do this by:

1) finding the probability of detecting an edge for a signal with noise $P(A)$;
2) finding the probability of detecting an edge in noise only $P(B)$;

3) deriving a cost function which maximizes $P(A)$ and minimizes $P(B)$, thereby giving the optimal σ and Th values.

The minimization of $P(B)$ is equivalent to the maximization of its complement, therefore the cost function to be maximized is

$$CF = \eta P(A) - (1-\eta)[1 - P(B)] \tag{11}$$

where $0 \leq \eta \leq 1$. When $\eta = 0$, all the emphasis is placed on noise suppression. When $\eta = 1$, the emphasis is on edge detection. Choosing η to be around 0.5 results in σ and Th values which try to balance noise suppression and edge detection. However, as the authors' results show, their technique is still susceptible to false edge-detection, especially in the presence of high noise levels.

In [28] and [29], Qian and Huang introduce a 2-D edge detection scheme. The scheme detects edges in images with an edge-detection functional that uses the LOG zero-crossing contours as a guide to the true edge locations. The proposed edge functional is based on a parametric 2-D edge model, and was developed to be optimal in terms of SNR and edge localization accuracy (ELA).

The procedure begins by convolving an image with the LOG operator and finding the zero-crossing points. Zero-crossing contours are then segmented at points with large curvatures, and the true edge locations are determined using the 2-D edge detection functional. Adaptive thresholding based on the global noise estimation and physiological evidence from biological visual systems is then performed on the edge segments. This is followed by an edge regularization technique for continuity and smoothness of the detected edges.

Edge segments are combined from different scales using a fine-to-coarse strategy. Since the smallest scale of the detection functional gives the best ELA, and only edges with high SNR survive the edge-detection procedure described above, the final edge map always contains all the edges from the smallest scale. The results from larger scales are incorporated into the final edge map if they produce new edge segments.

The scales of the operator used by Qian and Huang were selected to span the range of filters that operate in the human fovea. They used seven scales between $\sigma_{\min} = 2.5$ and $\sigma_{\max} = 6.7$. However, this may not be the ideal range for computational methods. In addition, the range may also change depending on the type of image and the amount of noise it contains.

More recently, Lindeberg [21] suggested a framework for automatic scale selection based on maximization of two specific measures of edge strength

$$\mathcal{G}_{\gamma-\mathrm{norm}}L = \sigma^{\gamma}(L_x^2 + L_y^2) \tag{12}$$

$$\mathcal{T}_{\gamma-\mathrm{norm}}L = \sigma^{3\gamma}\left(L_x^3 L_{xxx} + L_y^3 L_{yyy} + 3L_x^2 L_y L_{xxy} + 3L_x L_y^2 L_{xyy}\right) \tag{13}$$

where scale space representation L is obtained by convolving a Gaussian with variance σ with the image. The parameter σ represents scale. The subscript(s) of L denotes derivative(s) of L taken in the direction(s) of the subscripts(s). The parameter γ provides additional control in the selection process. We will discuss the role of γ later in this paragraph. First, let us look at (12) and (13). The first equation, (12), computes the magnitude of the gradient. This is the simplest measure of edge strength. The second equation, (13), originates from the sign condition in the edge definition. An edge point is defined as a point at which the gradient magnitude assumes a maximum in the gradient direction [7]. This is the same as the second-order derivative in a given direction being zero and the third-order derivative in the same direction being negative. This equation combines the sign information with the magnitude, scale, and γ. Maximization of these measures with respect to scale yields an equation for the scale parameter with both measures producing very similar results. The parameter γ makes the scale selection method dependent on the diffuseness of the edge, i.e., fine scale is selected for sharp edges and coarse scale is selected to deal with diffused (blurred) edges. However, the authors choose $\gamma = 1$ in all their experiments. This renders the proposed approach very similar to [10] in many respects. A note of caution is required here. Automatic scale selection still requires the user to specify a scale range. For example, the authors choose 40 scales (minimum scale $= 0.1$ and maximum scale $= 256$) for an image of size 256×256. The maximization of the measures over the range detects the best scales. A major drawback of this approach is the need to compute high-order derivatives, which are known to contribute toward computational difficulties. One does not see any significant advantage in the use of such high-order derivatives from theoretical or experimental results. On the positive side, this paper uses the edge diffuseness as a parameter, which provides useful clues to the physical nature of the edge.

The paper by Elder and Zucker [10] supports the basic premise of [21] that selection of scale should be made dependent on the edge diffuseness. In contrast to Lindeberg's approach, Elder and Zucker propose a completely local method for scale selection. They also make the scale a function of the second moment of the sensor noise; information that is usually available. This definitely holds considerable appeal from a practical standpoint. The authors argue that since the distinction between *important* and *unimportant* edges is rather difficult to make at the early stage of processing, the goal of an edge-detector should be to detect all edges over the range of contrast and blur with which they occur while avoiding detection of artifactual edges. To achieve this, the authors introduce the idea of a minimum reliable scale at which and at larger scales, the possibility of detecting edges due to sensor noise is below a specified tolerance. The scale is computed locally as a function of the sensor noise and the blur scale. The computation remains simple. Furthermore, a method to compute the blur provides an estimate for the depth parameter. This can be used to separate foreground and background structures in the image. The use of a single scale as opposed to the use of multiple scales for edge-detection sets this paper apart from standard approaches. Thus, the authors argue that the problem of edge tracking over multiple scales has been avoided. One should keep in mind that the process of detecting and identifying *important* edges cannot be avoided. It can only be delayed as is done here and left to be handled by the next processing stage.

Next, we review research work that has delved more into the biological vision process. The edge-detectors were modeled after a mammalian visual system. Later, it was shown that these detectors enjoy sound mathematical support.

Kennedy and Basu [3], [17] introduced a special line-weight function (LWF) for enhancing edges in digital images. Unlike most of the works presented here, which are based either on the first- or second-order derivatives of Gaussian, this operator is a linear combination of zero- and second-order Hermite functions. This is equivalent to the combination of a Gaussian and its second derivative. The development of this operator was based on several pieces of physiological evidence. Fiorentini and Mazzatini [11] suggested that the LWF of humans are shaped like the linear combination of zero- and second-order Hermite functions. This proposition is supported by Young [38] who points out that such a function provides edge and line enhancement and noise suppression, while retaining all the information in the original image. This operator can also be derived mathematically when the contrast sensitivity experiments of psychophysics are posed as an eigenvalue problem [34]. The LWF has an excellent edge localization property and does not detect phantom edges.

The LWF enhances edges in digital images. Edges are detected by convolving an LWF filtered image with an edge operator, such as the Sobel mask. The size of the LWF is chosen to achieve the best operator performance while keeping the complexity under control, but as mentioned earlier, the use of a single scale does not produce optimal results for edge detection.

The last part of this paper looks into edge-detectors that leave the linear territory in search of better performance. Nonlinear methods based on the Gaussian filter evolved as researchers discovered the relationship between the solution to the heat equation and images convolved with Gaussian filter for a smoothing purpose. Consider a set of derived images, $I(x, y, t)$, by convolving the original image with a Gaussian filter $G(x, y, t)$ of variance t. The parameter t corresponds to time in the heat equation, whereas in the context of image it refers to the scale. This one parameter family of derived images can be viewed as the solution of the heat equation [15], [19]. However, in the case of linear heat equation as diffusion eradicates noise, it also blurs the edges isotropically. To overcome this problem, Perona and Malik [26] proposed a multiscale or scale space representation of an image based on anisotropic diffusion. In the mathematical context, this calls for nonlinear partial differential equations rather than the linear heat equation.

The essential idea here is to allow space variant blurring [26]. This is achieved by making the diffusion coefficient in the heat equation a function of space and scale. The goal is to smooth within a region and keep the boundaries sharp. A high value for the diffusion constant within the region and a very small value (possibly 0) on the boundary can produce the desired effect. Specifically, the heat diffusion coefficient is allowed to vary across the image plane and is made dependent upon the image gradient. This effectively leads to a spatially adaptive smoothing which tends to preserve the location of edges throughout the scale hierarchy. When the Perona-Malik equation is decomposed into a process across the edge and one perpendicular to it, it can be understood how smoothing and sharpening can be carried out at the same time. The entire process is a combination of forward and backward diffusion processes.[3] However, backward diffusion is well known to be an ill-posed process where the solution, if it exists at all, is highly sensitive to even the slightest perturbations of the initial data [18], [35]. In the context of image processing, the main observed instability is the so-called *staircasing effect*, where a smoothed step edge evolves into piecewise almost linear segments which are separated by jumps (see [25] for an illustration). The extent of this effect is dictated by the process of discretization. This effect is observable for fine spatial discretization and for slowly varying ramp edges. Fortunately, under practical situations, this phenomenon is hardly observed. It is an experimental fact that *reasonable* discretizations of the Perona-Malik equation are rarely unstable. Fontaine and Basu [12] suggest the use of wavelets to solve the anistropic diffusion equation. Wavelet-based multiresolution expansions are known to give more compact representations of images with regions of low contrast separated by high-contrast edges. Additionally, the use of wavelets provides a way to estimate contrast value for edges on a space-varying basis in a local or global manner as needed. The drawback of this approach is that the discretization scheme for the diffusion equation proposed in this paper cannot be directly expressed in the wavelet transform domain. This requires an iterative procedure of going back and forth between the spatial and the wavelet domains of representation and adds to the numerical complexity of the algorithm.

The most required property of scale-space representation is satisfied here—which is no new features are introduced in the derived images (i.e., in the scale-space representation of the original image) in passing from fine to coarse scale.[4]

In contrast to Perona and Malik, Aurich and Weule [1] do not use anisotropic diffusion. Their method is based on a modification of the way the solution of the heat equation is obtained by convolving the initial data with a Gaussian kernel. Thus, it avoids a large number of iterations and problems due to convergence. The method uses a nonlinear modification of Gaussian filters. The convolution of an image $I(p)$ and a nonlinear Gaussian filter is given as follows:

$$G_{\sigma_x,\sigma_z} I(p) = \frac{1}{N_p} \sum_{q \in P} g_{\sigma_x}(\|q-p\|) g_{\sigma_z}(I(q) - I(p)) \cdot I(q)$$

where image signal I is defined on a bounded discrete set of pixels P and the normalization factor in the neighborhood of pixel p, N_p, is given by

$$N_p = \sum_{q \in P} g_{\sigma_x}(\|q-p\|) g_{\sigma_z}(I(q) - I(p)).$$

Let us rewrite the above equation in the following form:

$$G_{\sigma_x,\sigma_z} I(p) = I(p) + \frac{1}{N_p} \sum_{q \in P} g_{\sigma_x}(\|q-p\|) g_{\sigma_z} \times (I(q) - I(p)) \cdot (I(q) - I(p)).$$

Intuitively, edge at location q in the neighborhood of pixel p can be preserved if $I(q)$ does not dominate the sum. Let us perform

[3]While the forward diffusion process blurs, the backward diffusion process sharpens.

[4]Note similarity with Gaussian scale space representation.

an informal analysis of the above equation to see: 1) how an edge is preserved while the image is smoothed and 2) how an edge is enhanced while the image is smoothed.

1) How an edge is preserved: Consider a pixel p. We are computing a new pixel value at p by doing a weighted average over neighborhood of p. The variance σ_x determines the size of the neighborhood.

- Case I: $I(p) - I(q)$ is small for $\forall\, q$ in the neighborhood of p. The value of $g_{\sigma_z}(I(p) - I(q))$ will be large. Thus new $I(p)$ will be more close to its neighboring pixel values.
- Case II: $I(p) - I(q)$ is small for $\forall\, q$ in the neighborhood of p except at one pixel q'. Here, $I(p) - I(q')$ will have a significantly larger value than $I(p) - I(q)$. The aim is to reduce the impact of $I(q')$. The value of $g_{\sigma_z}(I(p) - I(q'))$ will be small. Thus the weighted sum will receive only a small fraction of $I(p) - I(q')$.

We can conclude that a new pixel value of a nonedge pixel (p) will not increase unduly because of a neighboring edge pixel value. This will prevent blurring of the edge present at q'.

2) How an edge is enhanced: Consider an edge pixel p with pixel value $I(p)$. After weighted averaging is done at p we obtain a new value $I'(p)$. It is obvious that $I'(p) > I(p)$.

This shows that after each convolution, pixel values at edges are increasing. However, one must keep in mind that although pixel value is increasing due to filtering, the overall effect may not produce enhancement. The slope of the edge is a critical factor here. Enhancement is achieved if the edge is *steep*.

Aurich and Weul suggest that image should be convolved with a series of such Gaussian filters. One should start with a small value for σ_x and a large value for σ_z. In successive steps, σ_x is increased and σ_z is decreased. This continues to reduce the contrast of the fine details while sharpening the edges of the coarse structures.

Note that this method does not produce a scale-space representation of the image. The possibility of the appearance of new features in the image has not been explored mathematically or experimentally. Furthermore, unlike usual scale-space representation techniques, no order is created in the set of images produced by successive filtering. Thus, an edge point cannot be tracked from one filtered image to the next.

V. Summary

The Gaussian filter has several desirable features which accounts for its wide use in many image processing applications. However, research clearly demonstrates that edge-detection techniques involving this filter do not give satisfactory results. Linear methods presented in this paper suffer from problems associated with Gaussian filtering, namely, edge displacement, vanishing edges, and false edges. The introduction of multiscale analysis further complicates the issue by creating two major problems: 1) how to choose the size of the filters and 2) how to combine edge information from different scales. Adaptive approaches which avoid the multiresolution problem all tend to be computationally intensive. Nonlinear approaches show significant improvement in edge-detection and localization over linear methods. However, problems of computational speed, convergence, and difficulties associated with multiscale analysis remain. As it currently stands, use of the Gaussian filter requires making compromises when developing algorithms to give the best overall edge-detection performance.

Acknowledgment

The author wishes to thank the anonymous reviewers whose comments led to substantial improvements in this manuscript.

References

[1] V. Aurich and J. Weule, “Nonlinear Gaussian filters performing edge preserving diffusion,” in *Proc. 17th DAGM Symp.*, New York, 1995, pp. 538–545.
[2] J. Babaud, A. P. Witkin, M. Baudin, and R. O. Duda, “Uniqueness of the Gaussian kernel for scale-space filtering,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 26–33, Jan. 1986.
[3] M. Basu, “A Gaussian derivative model for edge enhancement,” *Pattern Recognit.*, vol. 27, pp. 1451–1461, 1994.
[4] M. Bennamoun, B. Boashash, and J. Koo, “Optimal parameters for edge detection,” in *Proc. IEEE Int. Conf. SMC*, vol. 2, 1995, pp. 1482–1488.
[5] F. Bergholm, “Edge focusing,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-9, pp. 726–741, June 1987.
[6] V. Berzins, “Accuracy of laplacian edge-detectors,” *Comput. Vis. Graph. Image Process.*, vol. 27, pp. 195–210, 1984.
[7] J. Canny, “A computational approach to edge detection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 679–698, June 1986.
[8] J. J. Clark, “Authenticating edges produced by zero-crossing algorithms,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 43–57, Jan. 1989.
[9] G. Deng and L. W. Cahill, “An adaptive Gaussian filter for noise reduction and edge detection,” in *Proc. IEEE Nucl. Sci. Symp. Med. Im. Conf.*, 1994, pp. 1615–1619.
[10] J. H. Elder and S. W. Zucker, “Local scale control for edge detection and blur estimation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 20, pp. 699–716, June 1998.
[11] A. Fiorentini and L. Mazzatini, “Neuron inhibition in the human fovea: A study of interaction between two line stimuli,” *Atti Fond G Ronchi*, vol. 21, pp. 738–747, 1966.
[12] F. L. Fontaine and S. Basu, “A wavelet-based solution to anisotropic diffusion equation for edge detection,” *Int. J. Imaging Sci. Technol.: Special Issue on Image Sequence Processing, Motion Estimation, and Compression of Image Sequences*, vol. 9, pp. 356–368, 1998.
[13] A. Goshtasby, “On edge focusing,” *Image Vis. Comput.*, vol. 12, pp. 247–256, 1994.
[14] R. M. Haralick, “Digital step edges from zero-crossing of second directional derivatives,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-6, pp. 58–68, Jan. 1984.
[15] R. A. Hummel, “Representation based on zero-crossings in scale space,” in *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, 1986, pp. 204–209.
[16] H. Jeong and C. I. Kim, “Adaptive determination of filter scales for edge-detection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 579–585, May 1992.
[17] L. M. Kennedy and M. Basu, “Image enhancement using a human visual system model,” *Pattern Recognit.*, vol. 30, no. 12, pp. 2001–2014, 1997.
[18] S. Kichenassamy, “The Perona-Malik paradox,” *SIAM J. Appl. Math*, vol. 57, no. 5, pp. 1328–1342, 1997.
[19] J. Koenderink, “The structure of images,” *Biol. Cybern.*, vol. 50, pp. 363–370, 1984.
[20] V. Lacroix, “The primary raster: A multiresolution image description,” in *Proc. 10th Int. Conf. Pattern Recognition*, 1990, pp. 903–907.
[21] T. Lindeberg, “Edge detection and ridge detection with automatic scale selection,” *Int. J. Comput. Vis.*, vol. 30, pp. 117–156, 1998.
[22] Z. Q. Liu, “Scale space approach to directional analysis of images,” *Appl. Opt.*, vol. 30, no. 11, pp. 1369–1373, 1991.
[23] D. Marr and E. Hildreth, “Theory of edge detection,” *Proc. R. Soc. Lond. A, Math. Phys. Sci.*, vol. B 207, pp. 187–217, 1980.
[24] V. S. Nawal, *A Guided Tour of Computer Vision*. Reading, MA: Addison-Wesley, 1993.
[25] M. Nitzberg and T. Shiota, “Nonlinear image filtering with edge and corner enhancement,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 826–833, July 1992.

[26] P. Perona and J. Malik, "Scale-space and edge detection using anisotropic diffusion," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 12, pp. 629–639, July 1990.
[27] W. K. Pratt, *Digital Image Processing*. New York: Wiley-Interscience, 1978.
[28] R. J. Qian and T. S. Huang, "Optimal edge detection in two-dimensional images," in *Proc. Image Understanding Workshop*, 1994, pp. 1581–1588.
[29] ——, "Optimal edge detection in two-dimensional images," *IEEE Trans. Image Processing*, vol. 5, pp. 1215–1220, July 1996.
[30] A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, 2nd ed. New York: Academic, 1982.
[31] B. G. Schunck, "Edge detection with Gaussian filters at multiple scales," in *Proc. IEEE Comp. Soc. Work. Comp. Vis.*, 1987, pp. 208–210.
[32] M. Shah, A. Sood, and R. Jain, "Pulse and stair case edge models," *Comput. Vis. Graph. Image Process.*, vol. 34, pp. 321–343, 1986.
[33] J. Shen and S. Castan, "Toward the unification of band-limited derivative operators for edge detection," *Signal Process.*, vol. 31, pp. 103–119, 1993.
[34] A. L. Stewart and R. Pinkham, "A spatial-variant differential operator for visual sensitivity," *Biol. Cybern.*, vol. 64, pp. 373–379, 1991.
[35] J. Weickert and B. Benhamouda, "A semidiscrete nonlinear scale-space theory and its relation to the Perona-Malik paradox," in *Advances in Computer Vision*, F. Solina, W. G. Kropatsch, R. Klette, and R. Bajcsy, Eds. New York: Springer-Verlag, 1997, pp. 1–10.
[36] D. J. Williams and M. Shah, "Edge contours using multiple scales," *Comput. Vis. Graph Image Process.*, vol. 51, pp. 256–274, 1990.
[37] A. P. Witkin, "Scale-space filtering," in *Proc. Int. Joint. Conf. Artificial Intelligence*, vol. 2, 1983, pp. 1019–1022.
[38] R. A. Young, "The Gaussian derivative model for spatial vision: Part I—Retinal mechanisms," *Spatial Vis.*, vol. 2, no. 4, pp. 273–293, 1987.
[39] A. L. Yuille and T. A. Poggio, "Scaling theorems for zero-crossings," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. PAMI-8, pp. 15–25, Jan. 1986.
[40] P. Zamperoni, "Image enhancements," in *Advances in Imaging and Electron Physiscs*. New York: Academic, 1995, pp. 1–76.
[41] D. Ziou and S. Tabbone, "Edge detection technique—An overview," *Int. J. Pattern Recognit. Image Anal.*, vol. 8, pp. 537–559, 1998.

Mitra Basu (S'84–M'85–SM'99) received the Ph.D. degree in electrical engineering from Purdue University, West Lafayette, IN, in 1985.

She is an Associate Professor in the Department of Electrical Engineering, The City College of New York. At present, she is on loan to the NSF, where she is a Program Director in the CISE Directorate. Her research interests include pattern recognition, biological/artificial learning systems, and bioinformatics. She is an Associate Editor of *Pattern Recognition*.

Dr. Basu is a member of the New York Center for Biomedical Engineering at the City College. She has joint appointment with the doctoral faculty of The Graduate School and the University Center's Ph.D. program in Computer Science and Electrical Engineering.