COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING 34, 321–343 (1986)

Pulse and Staircase Edge Models*

MUBARAK SHAH AND ARUN SOOD

Electrical and Computer Engineering, Wayne State University, Detroit, Michigan 48202

AND

RAMESH JAIN

Electrical Engineering and Computer Science, The University of Michigan, Ann Arbor, Michigan 48109

Received April 3, 1985; accepted February 4, 1986

Present step and ramp edge models are inadequate for the edges detected by multi-resolution operators. Since the isolated edges rarely occur in the real scenes, we propose new edge models based on the pulse and staircase functions. In these models we include the effect of one edge on a neighboring edge. This effect propagates through to the higher operator sizes. Depending on the mutual polarities of the steps in the staircase and pulse functions, the edge points related to these discontinuities attract or repel each other when the operator size increases. In the case of staircase function, when the edge points attract each other at some scale they collapse into one.

1. INTRODUCTION

Edge detection is the first stage in most computer vision systems. Past experience with various edge operators has indicated that the problem of detecting edges in real scenes is extremely difficult. In an image, changes of intensity take place at many spatial scales, depending on their physical origins. Therefore, the detection of all the significant edges present in a scene requires that an edge operator be applied at various resolutions so that the discontinuities in intensities at all levels can be captured. An edge detection scheme based on the multiscale analysis performed with filters of different sizes was first introduced by Rosenfeld and Thurston [22]. Recently, Marr [18] argued on these same lines by proposing multiple-size Laplacian of Gaussian operators. In his approach, the image is convolved with the Laplacian of Gaussian operator and the edge points are detected by locating the zero-crossings in the convolved image. Operators of various sizes are used to separate the intensity changes at multiple scales.

Once the discontinuity points[1] are detected at multiple scales,[2] the next step is to manage them efficiently. This problem has been called a channel integration

*This research was jointly conducted at the RIES Computing Research Lab of Wayne State University under contract from Industrial Productivity and Products Research Institute, and at the Robotics division of CRIM at the University of Michigan under U.S. AFSOR Contract F49620-82-0089.

[1] We will use the terms discontinuity and edge points interchangeably.

[2] The use of the term "scale" has been confused in the literature. One interpretation of scale is the degree of smoothing, which essentially can be controlled by varying the variance of Gaussian. Scale has also been related to the rate of intensity changes in the gray level images. Since both of these are closely related in multi-resolution edge operators, we will use this term for both contexts without attempting to resolve the confusion.

0734-189X/86 $3.00

problem. A set of discontinuity points detected at one scale is called a channel, so the idea is to combine all channels to come up with some representation which is better than any single channel independently.

One point of view has been that, at the lower scale, many edge points are obtained because some false edges are also detected, so an effort should be made to remove the false edge points. Eklundh, Elfving, and Nyberg [8] do exactly this. They apply a threshold to the magnitude and the direction of edges in order to remove the false edge points. The removal of some edge points in the busy areas usually creates gaps in the edge contours which are otherwise closed. Next, they fill up the gaps between any two end points of a broken contour, guided by the "good continuation measure." The good continuation measure is the function of the distance between two end points and the angle between the line joining these two points. The problem with this approach is that it uses the thresholding explicitly. The determination of a threshold for the real scenes is not an easy task.

Witkin [25] presented a scale-space filtering approach to this problem. In his approach, he convolves the signal with the second derivative of multiple-size Gaussian filters and detects the zero-crossings in the output of the filters. These zero-crossings, when plotted in (x, σ) space, form the closed contours. In order to simplify the representation, he proposed a ternary tree of zero-crossings called an interval-tree. The interval-tree transforms the zero-crossings contours into a data structure which can be easily handled. By proposing the scale-space approach, Witkin not only reaffirmed the importance and complexity of this problem but also intrigued many researchers, see [28, 26, 27, 2, 3, 16, 17, 7, etc.].

Witkin's elegant approach motivated us to consider this problem in more detail in order to get a better understanding of scale-space in quantitative terms. In the long term our aim is to characterize the scene using a set of primitives. We expect that this can be more easily achieved by fitting primitives. In this way, we plan to use a priori knowledge about the behavior of the primitives in the scale-space in order to get a consistent fit of primitives at all scales.

In this paper, we consider the known edge models and study their behavior in the scale-space. We have found that modeling edges using step and ramp functions is inadequate for multi-resolution operators. It is noted that in real scenes isolated step and ramp edges are rarely encountered. In the case of edges which are present close to each other, at the bigger operator sizes one edge affects the neighboring edge. Therefore, in this case, the behavior of such edges at bigger operator sizes is not similar to the behavior of an isolated edge. We will present new edge models based on the pulse and staircase functions. The pulse and staircase functions have two discontinuities close to each other. We find that, depending on the mutual polarities of the steps in those functions, the zero-crossings attract or repel each other as the operator size increases.

In the next section, we review the related work on edge operators. In Section 3, we define the second derivative of Gaussian operator and in the process present notations utilized in the paper. Sections 4.1 and 4.2 treat the step and ramp edge models and analyze their behavior at multiple scales. We have devoted Sections 4.3 and 4.4 to the discussion of the pulse and staircase functions. Section 5 examines the zero-crossing contours related to the intensity functions considered in Section 4. The noise analysis is presented in Section 6. Finally, in Section 7 we extend our results for two-dimensional images.

2. EDGE OPERATORS

Most previous operators, e.g., those of Beaudet, Robert, and Sobel, were differential in nature [4, 21]. These operators essentially measure the first derivative in the spatial domain, which gives a rate of change of intensity values. The points where the local maxima of the first derivative occur are declared as the edge points. These maxima are located by using thresholding. Marr and Hildreth [19] proposed the Laplacian of Gaussian operator. Under certain conditions the Laplacian approximates the second derivative [15]. Therefore, the locations of zeros in the Laplacian of image signify the locations of extrema in the first derivative. The main motivation for this operator was biological vision systems, because the output of this operator resembles the response of the center-surround cells found in biological visual systems. When Marr and Hildreth used this operator at multiple resolutions in their work, it was the first time that large mask sizes were used. In comparison to the previous commonly used 3×3 masks, the smallest mask they used was 31×31. It was found that the larger operators were able to suppress the noise to some extent. The next problem was to combine the output obtained from variable-size operators. They suggested a criterion that the zero-crossings that coincide over several scales are physically significant. This claim, however, was never justified.

In the recent literature three more edge operators for detecting step discontinuities have been proposed: Haralick's zero-crossing operator [10, 11, 12, 13], Canny's edge detector [6] and Torre and Poggio's operator [23].

Haralick fits a bi-cubic polynomial to the neighborhood of a pixel. He computes the first and the second directional derivative in the direction of the gradient of the intensity function in terms of the coefficients of the polynomial. The coefficients for a given pixel location are found by using a least-square fit to the gray level neighborhood of the pixel. A given pixel is declared as an edge point if (i) the first derivative is above some threshold and (ii) the second derivative is equal to zero. There are two main differences between Haralick's operator and the Laplacian of Gaussian operator. First, Haralick's operator is directional while the Laplacian of Gaussian is not. In 2D images at the points where Laplacian does not approximate the second derivative, Haralick's operator correctly detects the gray level changes in comparison to the Laplacian of Gaussian operator. Second, Haralick's operator needs an explicit thresholding for the first derivative, while the Laplacian of Gaussian operator does not require the thresholding.

Canny [6] proposed an edge operator for detecting step edges. This operator has a shape similar to the first derivative of Gaussian. In order to detect edge points, Canny finds out the extremas in the first derivative which are essentially zero crossings in the second derivative. Canny has claimed that Haralick's operator is equivalent to his operator. He further proposed that the bi-cubic is the only polynomial which can be used in Haralick's zero-crossing edge operator in order to get an optimum edge operator for the step edges. If the degree of polynomial is changed, the operator would not be optimum. His argument is based on the comparison between the graphs of the first derivative of Gaussian operator and Haralick's operators.

Canny's first derivative of Gaussian operator is designed for detecting the step edges, but the methodology he has developed can be used to come up with an operator to detect any other function, e.g., ramp. He has also proposed that his operator should be applied at various scales. When comparing Canny's operator

with Haralick's operator, one may raise many interesting questions: Do we need to apply Haralick's operator at multiple scales? If so, what is the scale in the Facet model? Is it the degree of polynomial or the neighborhood size, or both?

Recently Torre and Poggio [23] proposed another edge operator which also has a shape similar to the first derivative of Gaussian. In their approach they use the fact that the numerical differentiation of intensity function is an ill-posed problem. They show that the edge detection scheme consists of two steps: the filtering step and the differentiation step.

It is interesting to note that these three edge detectors proposed by Haralick, Canny, and Torre and Poggio have similar shapes, even though they were designed by completely different approaches.

3. LAPLACIAN OF GAUSSIAN OPERATOR

In this section, we describe the Laplacian of Gaussian operator and introduce some notations which will be used in this paper. For simplicity, we will consider only the 1-dimensional case here. In Section 7, we will discuss the 2-dimensional case.

For one dimension the Laplacian becomes the second derivative. We denote the zero mean Gaussian density function with variance σ by $g^{\sigma}(x)$. Neglecting the multiplicative constant, $g^{\sigma}(x)$ is given as

$$g^{\sigma}(x) = e^{-x^2/2\sigma^2}. \tag{1}$$

Let ∇^2 represent the Laplacian (second derivative) for two dimensions (one dimension), and " $*$ " represent the convolution operation.

The second derivative of Gaussian is given as

$$\phi^{\sigma} = \nabla^2 g^{\sigma}(x) = \left(1 - x^2/\sigma^2\right) e^{-x^2/2\sigma^2}. \tag{2}$$

The response of this operator for an input function $f(x)$ can be computed by evaluating the following convolution integral:

$$h^{\sigma}(x) = f(x) * \nabla^2 g^{\sigma}(x)$$
$$h^{\sigma}(x) = \int_{-\infty}^{\infty} f(t-\eta) \nabla^2 g^{\sigma}(\eta)\, d\eta. \tag{3}$$

Using the linearity property of convolution, Eq. (3) can also be written as

$$h(x, \sigma) = \nabla^2 [f(x) * g^{\sigma}(x)].$$

The convolution integral for the discrete domain changes to summation as

$$h^{\sigma}(i) = \frac{1}{m} \sum_{\alpha=-(m-1)/2}^{\alpha=(m-1)/2} f(\alpha) \cdot \phi^{\sigma}(i-\alpha)$$

where m is the size of the image. Since ϕ^{σ} is circular symmetric, the above equation

can be rewritten

$$h^{\sigma}(i) = \frac{1}{m} \sum_{\alpha=-(m-1)/2}^{\alpha=(m-1)/2} f(i-\alpha) \cdot \phi^{\sigma}(-\alpha). \tag{4}$$

According to the above equation, the response at a given pixel location is the weighted sum of its neighboring pixels. The discontinuities in $f(x)$ are related to the zero-crossings in $h^{\sigma}(x)$. The positive zero-crossing can be defined as follows: if

$$\lim_{\varepsilon \to 0} h^{\sigma}(x-\varepsilon) < 0$$

and

$$\lim_{\varepsilon \to 0} h^{\sigma}(x+\varepsilon) > 0$$

then x is zero-crossing of $h^{\sigma}(x)$. Similarly the negative zero-crossing can be defined by reversing inequalities in the above expressions.

The size of the neighborhood of a pixel used by the operator[3] is called the mask size. In Eq. (4), we are performing a convolution; therefore, the mask size should be the same as the image size. The Laplacian of Gaussian operator can be defined without referring to its mask size. In an implementation of the operator, however, we always have a finite number of bits. Therefore the effective value of ϕ^{σ} becomes zero for a bigger neighborhood. Due to this fact, it is not necessary to include the elements for which ϕ^{σ} is zero. The rate of decay of Gaussian depends on σ. By knowing the word length used in an implementation, we can come up with an upper bound for a significant neighborhood size for a given σ value. This way an approximate relationship between the mask size and the variance can be derived. Hildreth [15] suggests two criteria to select the relationship between mask size and σ. First, the positive area w of the operator must satisfy: $w = \sqrt{2}\,\sigma$, which is motivated neurophysiologically. Second, the mask size should be such that the response of the operator for a uniform intensity must be zero. This last condition is a kind of check for a numerical error. The idea is to consider a neighborhood size such that the values of the operator sum to zero. The Laplacian of Gaussian has positive values in the center, surrounded by negative values. Truncating the operator by a smaller mask size would result in some loss of the negative surround. Thus, the overall sum would be positive instead of being zero (see [9, 14] for an interesting debate). In our implementation, we have used 16 bits to represent the weights of operator at each pixel and the total support of the operator is about $15 * \sigma$ pixels. In Fig. 1 we show the plots of operator having three different values of σ.

4. EDGE MODELS

Sections 4.1 and 4.2 consider step and ramp edge models. Since the gray level function changes significantly at the edges of objects, the ideal edge can be modeled by a step function. In real images, however, the gray level at the edges of an object does not change abruptly, but instead, changes gradually. These edges can be

[3] From now onwards we will use the term "operator" for the Laplacian of Gaussian operator.

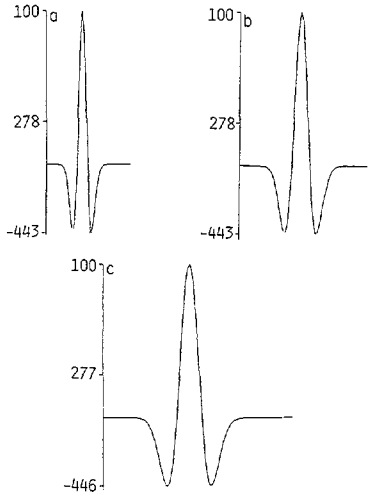

FIG. 1. The graphs of Laplacian of Gaussian operator for (a) $\sigma = 3$, (b) $\sigma = 5$, (c) $\sigma = 7$.

modeled by the ramp function. Our aim here is to study the behavior of these simple edges as the function of σ.

Next, we will consider the pulse and staircase models for intensity functions. The pulse and staircase functions contain two discontinuities which are some distance apart. We will then analyze the effects of σ for these models in Sections 4.3 and 4.4.

4.1. Step Edge

An ideal edge can be modeled by the step function. The step function, $U(x)$, is defined as

$$U(x) = \begin{cases} 0 & \text{if } x < 0 \\ c_1 & \text{if } x > 0. \end{cases}$$

If we convolve the step function with $g^\sigma(x)$ we get

$$\begin{aligned} h^\sigma(x) &= c_1 U(x) * g^\sigma(x) \\ h^\sigma(x) &= \int_{-\infty}^{\infty} c_1 U(x-\eta) g^\sigma(\eta)\, d\eta \\ &= \int_{-\infty}^{x} c_1 g^\sigma(\eta)\, d\eta. \end{aligned}$$

Now taking the second derivative, we get

$$E_1^\sigma(x) = \nabla^2 h^\sigma(x) = -c_1(x/\sigma^2) g^\sigma(x). \tag{6}$$

In Figs. 2b–d we plot E_1^σ for various values of σ. The function E_1^σ crosses zero for all values of σ. The location of the zero-crossing corresponds to the location of the discontinuity in the step function.

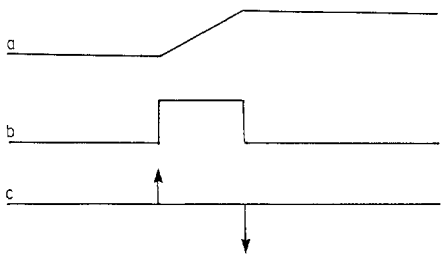

FIG. 3. (a) ramp function; (b) first derivative of ramp; (c) second derivative of ramp.

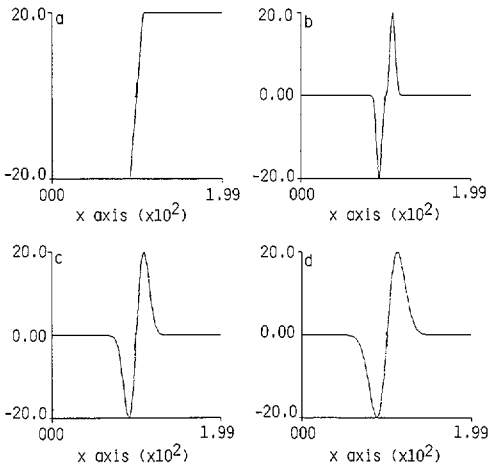

FIG. 4. The effect of σ on the ideal ramp edge. In (a) we show the ideal ramp edge, while in (b)–(d) we show the output obtained by convolving the ramp function with the second derivative of Gaussian having $\sigma = 3, 7, 11$.

which gives

$$E_2^\sigma(x) = c_2 w[g^\sigma(x) - g^\sigma(x - w)]. \tag{10}$$

Figure 4 shows the graphs of E_2^σ for various values of σ. The function E_2^σ has a zero-crossing at $x = w/2$, which corresponds to the center of the ramp function. In this case the slope c_2 and the ramp width w appear as multiplicative constants in Eq. (10) and do not effect the location of the zero-crossing for any value of σ. This implies that the operator is insensitive to the rate of intensity changes. It is emphasized that zero-crossing has the information about the location of the mid-point of the ramp function.

The actual value of an extrema of the first derivative, not the position, carries the information about the rate of intensity changes. Consider the example of ramp function $R(x)$ defined in Eq. (7). The first derivative of this function is given in Eq. (8). An extrema of $(R' * g^\sigma(x))$ is $c_2 w$, which depends on the value of slope c_2. The slope c_2 of the ramp essentially tells the rate of intensity changes.

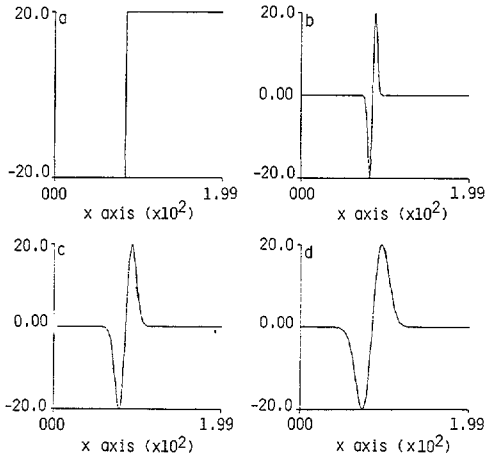

FIG. 2. The effect of σ on the ideal step edge. In (a) we show the step function, while in (b)–(d) we show the output obtained by convolving the step function with the second derivative of Gaussian having $\sigma = 3, 7, 11$.

The amplitude c_1 of the step function appears as a multiplicative constant in Eq. (6). Thus this term essentially controls the amplitude but not the shape of the graphs shown in Figs. 2b–d. A zero-crossing is obtained at the corresponding location of a discontinuity regardless of the amplitude of the step function.

4.2. Ramp Edge

Let us consider a ramp edge of width w and find out the response of the operator. The ramp function is defined as

$$R(x) = \begin{cases} 0 & \text{if } x < 0 \\ c_2 x & \text{if } 0 < x < w \\ c_2 w & \text{if } x > w \end{cases} \tag{7}$$

where c_2 is the slope of the ramp. Let R' and R'' denote the first and second derivatives of the ramp function. Then from Eq. (7) we get

$$R'(x) = c_2 w[U(x) - U(x - w)] \tag{8}$$

$$R''(x) = c_2 w[\delta(x) - \delta(x - w)]. \tag{9}$$

The ramp function is shown in Fig. 3 with its first and second derivatives. Convolution of R' and R'' with $g^\sigma(x)$ yields

$$R' * g^\sigma(x) = c_2 w\left[\int_{-\infty}^{x} c_2 g^\sigma(\eta)\, d\eta + \int_{-\infty}^{x-w} c_2 g^\sigma(\eta)\, d\eta\right].$$

And

$$E_2^\sigma(x) = R'' * g^\sigma(x)$$

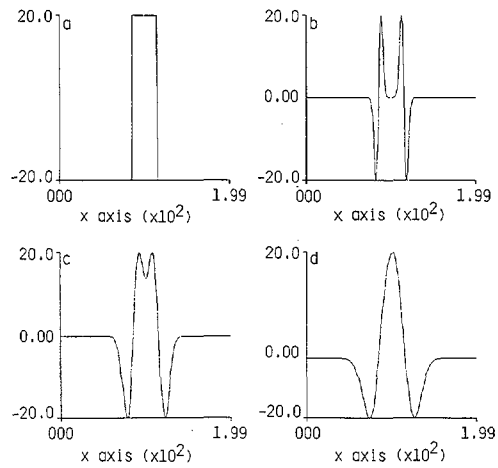

FIG. 5. The effect of σ on the ideal pulse edge. In (a) we show the pulse function, while in (b)–(d) we show the output obtained by convolving the pulse function with the second derivative of Gaussian having $\sigma = 3, 7, 13$. The zero-crossings repel each other when σ increases.

Information about the rates of intensity changes is more important for operators which have fixed mask sizes. Due to this fact, Haralick's operator and other fixed-size operators need some kind of thresholding for the first derivative of intensity function. However, we will show that in the case of pulse and staircase functions when the discontinuities are detected at the multiple scales this information is indirectly captured in the scale space of the image.

Two remarks can be made concerning the ramp discontinuity. First, the zero crossing in E_2^σ corresponding to the ramp discontinuity is solely due to the symmetric nature of the operator. Any other operator which is not symmetric will not be able to detect this discontinuity. Furthermore, in order to detect the discontinuity due to the ramp, the operator size should always be greater than the ramp width; otherwise the discontinuity will not be detected. Therefore, in order to detect the discontinuities related to the ramps of various widths we have to apply a variable-sized operator.

4.3. Pulse Function

Consider a function which is a pulse of width w, as shown in the Fig. 5a. A pulse can be represented by the summation of two step functions, i.e.,

$$f_3(x) = U(x) - PU(x - w)$$

where P is the ratio of the magnitudes of two steps.

If we apply the operator to this function we get

$$\begin{aligned} h_3^\sigma(x) &= f_3(x) * g^\sigma(x) \\ E_3^\sigma(x) &= \nabla^2 h_3^\sigma(x) \\ &= -\left(\frac{x}{\sigma^2}\right) g^\sigma(x) + P\frac{(x - w)}{\sigma^2} g^\sigma(x - w). \end{aligned} \tag{11}$$

There are two discontinuities in the pulse function as shown in Fig. 5a. We want the operator to respond at both discontinuity points.

In Eq. (11) at $x = 0$ the first term on the right-hand side is zero but the second term is non-zero. Due to this additional term the actual location of discontinuity at $x = 0$ is shifted. Similarly, at $x = w$ the second term is zero but the first term is non-zero. For a fixed value of w, the pulse width, the amount of shift depends on the value of σ. For a small value of σ the additional term does not contribute much therefore the shift is not significant. However, at bigger σ the shift is quite substantial. In Figs. 5b–d, we plot E^{σ} for three different values of σ. It is easy to see that the locations of zero-crossings, corresponding to the discontinuities, shift as σ increases.

The shifting also depends on P which is the ratio of the magnitudes of two steps. For $P = 1$ the shift is equal for both discontinuities. But, for $P > 1$ the shift is dominant for the step which has smaller magnitude.

4.4 Staircase Function

The staircase function can be represented as the summation of two steps, as follows:

$$f_4(x) = U(x) + PU(x - w)$$

where P is the ratio of magnitudes of two steps.

If we apply the operator to this function, we get

$$\begin{aligned} h_4^{\sigma}(x) &= f_4(x) * g^{\sigma}(x) \\ E_4^{\sigma}(x) &= \nabla^2 h_4^{\sigma}(x) \\ &= -\left(\frac{x}{\sigma^2}\right) g^{\sigma}(x) - P\frac{(x - w)}{\sigma^2} g^{\sigma}(x - w). \end{aligned} \tag{12}$$

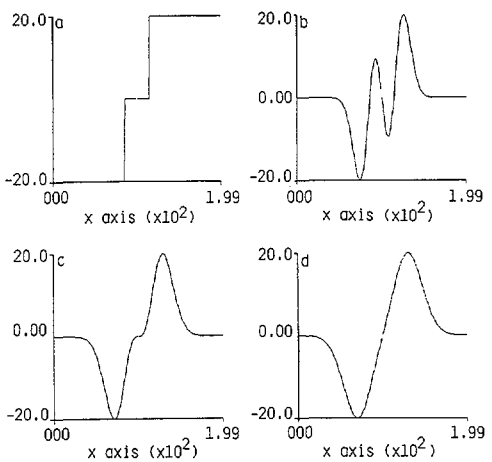

FIG. 6. The effect of σ on the ideal staircase edge. In (a) we show the pulse function, while in (b)–(d) we show the output obtained by convolving the staircase function with the second derivative of Gaussian having $\sigma = 3, 7, 11$. The zero-crossings attract each other when σ increases.

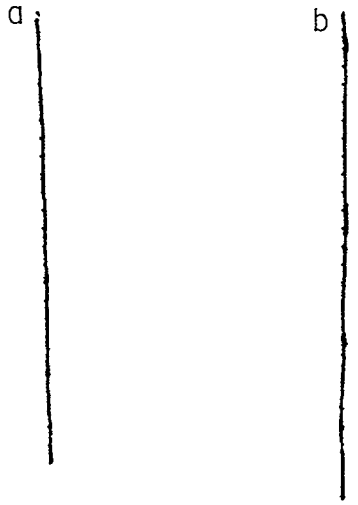

FIG. 7. The zero-crossing contours of the step (a) and the ramp (b) functions. The location of zero-crossing does not change as σ is increased. Therefore, the zero-crossing contours for the ramp and the step functions are the straight lines. These figures demonstrate that for ideal step and ramp we do not need multi-resolution operators.

The staircase function, similar to pulse function, has two discontinuities at $x = 0$ and $x = w$, as shown in Fig. 6a. However, the polarities of two steps are both positive. We plot E_4^σ for three different values of σ in Figs. 6b–d. In 6b, E_4^σ crosses zero approximately at $x = 0$ and w. The function also crosses zero at $x = w/2$. This is a false zero-crossing. While in Fig. 6d the function E_4^σ crosses zero only once.

There are two remarks to be made here. First, a false zero-crossing at $x = w/2$ is due to the symmetry of Gaussian, at $x = w/2$ Eq. (12) becomes zero regardless of the value of σ. Second, the shifting of true zero-crossings at $x = 0$ and $x = w/2$, in this case, is towards each other. This shifting is a function of σ as discussed in the previous section.

5. ZERO-CROSSING CONTOURS

The locations of zero-crossings in the function convolved with the second derivative of Gaussian can be plotted in the (x, σ) space. These zero-crossings form the contours in the (x, σ) space. In Figs. 7–9 we show the zero-crossing contours for the functions considered in the last section.

The zero-crossing contours of the step and ramp functions are the straight lines as shown in Fig. 7a and b, respectively. The straight line in the (x, σ) plane shows that the locations of zero-crossings due to the ramp and step do not change when σ increases.

In Fig. 8 we show the zero-crossing contours of the staircase function. In these contours there is a straight line and a circular arc. The zero-crossings which constitute the straight line, in this case, are the false zero-crossings, since they do not correspond to any actual discontinuities in the staircase function. They occur at this location due to the symmetric nature of the operator because for $x = w/2$ the expression for E_4 in Eq. (12) becomes zero regardless of the value of σ. Note that

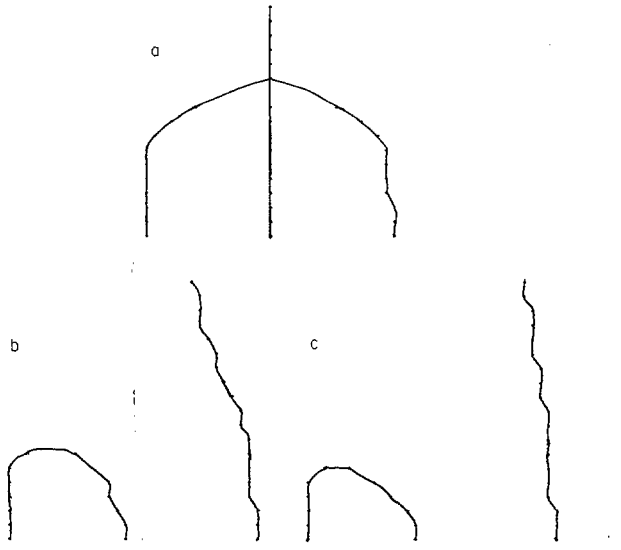

FIG. 8. The zero-crossing contours of the staircase function. The ratio of the magnitudes of two steps is 1, 2, and 4 in (a), (b), and (c), respectively. Due to the Propagation Effect the true zero-crossings move towards each other when σ increases. At some σ these zero-crossings collapse into one.

the symmetry of the operator works to our advantage in the case of a ramp. However, in this case it produces a false zero-crossing. The remaining two zero-crossings which correspond to the actual discontinuities in the staircase function make a circular arc. Since these zero-crossings shift towards each other as σ is increased, at some value of σ they collapse into one point which corresponds to the peak of the contour.

In Fig. 9 we show the zero-crossing contours for the pulse function. The zero-crossings related to the discontinuities in the pulse function make two diverging lines. As the zero-crossings shift away from each other when σ is increased, the two zero-crossings related to the pulse function would never collapse into one.

We can classify the zero-crossings into two categories. The first kind of zeros are not effected when the σ is increased. The zeros related to an isolated step or ramp fall in this class. At all scales the locations of those zeros are the same. Therefore we call these zeros *stationary zero-crossings*. The second class contains the zeros of pulse and staircase functions. These zeros are *free zero-crossings* and, by increasing σ, they can be displaced. In the pulse case, the zero-crossings move away from each other, while in the case of the staircase they move towards each other. When the zero-crossings move towards each other, they collapse into one. The direction of displacement of free zeros depends on the mutual polarities of the two step functions: if both steps have same polarities, then the zero-crossings shift towards each other; if they have opposite polarities, then they shift away from each other.

The zero-crossings shift at the higher operator size because one edge affects the neighboring edges. There are two step edges in the pulse and staircase functions, separated by a distance. Thus, at the higher operator size the two steps affect each other and the zero-crossings get displaced. In the case of the ramp and step

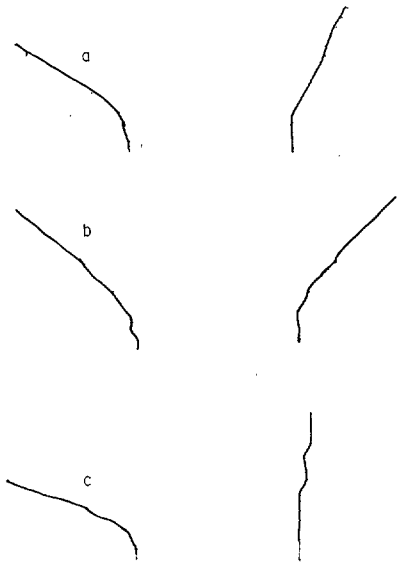

FIG. 9. The zero-crossing contours of the pulse function. The ratio of the magnitudes of two steps is 1, 2, and 8 in (a), (b), and (c), respectively. Due to the Propagation Effect the true zero-crossings move away from each other when σ increases.

functions there is only one isolated edge, so the zero-crossings related to both of them do not move when σ varies. We call this effect the *Propagation Effect*, and define it as follows.

PROPAGATION EFFECT. *The discontinuities within the field of an operator influence each other. Due to this the zero-crossings attract or repel each other, depending on the mutual polarities of steps.*

In Fig. 10, we have illustrated the Propagation Effect pictorially. In part (a), due to the smaller mask size the edge on the right side does not effect the edge on the left side. While in (b) due to the bigger mask size the right-hand edge comes within the field of the operator.

It is clear from Fig. 8 that the location in the (x, σ) plane where the two zero-crossings collapse into one is different in three cases. This location corresponds to the peak of the zero-crossing contour. The location of peak depends on the width w and the ratio of magnitudes of two steps P of the staircase function. Given the

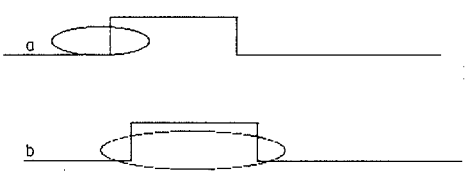

FIG. 10. Propagation Effect: (a) at smaller operator size an edge is not affected by the neighboring edge; (b) due to bigger operator size the effect of the neighboring edge propagates.

scale-space of the staircase function one can infer some qualitative information about the underlying gray level function from the location of peak. In the following we derive the equation which defines the peak analytically.

The location of the peak of curve $\sigma(x)$ can easily be found from its maxima. Differentiating Eq. (12) with respect to dummy variable η along a contour in (x, σ) space, we get

$$\frac{dE}{d\eta} = \frac{\partial E}{\partial x}\frac{dx}{d\eta} + \frac{\partial E}{\partial \sigma}\frac{d\sigma}{d\eta}.$$

Since $E = 0$ on the zero-crossings contour, so $dE/d\eta = 0$. If we choose the parameter η to be x then by *implicit function theorem* [24], we get

$$\frac{dE}{dx}\Big/\frac{dE}{d\sigma} = \frac{d\sigma}{dx}.$$

The right side of the above equation vanishes when

$$\frac{dE}{dx} = 0. \tag{13}$$

Therefore the value of x, where the above equation is satisfied is the peak of the curve $\sigma(x)$. Now, differentiating Eq. (12) with respect to x, we get

$$\frac{dE}{dx} = \frac{1}{\sigma^2}\left(1 - \frac{x^2}{\sigma^2}\right)g^{\sigma}(x) + \frac{P}{\sigma^2}\left(1 - \frac{(x-w)^2}{\sigma^2}\right)g^{\sigma}(x-w). \tag{14}$$

The location of the peak of the contour can be computed by equating the right-hand side of the above equation to zero:

$$\frac{1}{\sigma^2}\left(1 - \frac{x^2}{\sigma^2}\right)g^{\sigma}(x) + \frac{P}{\sigma^2}\left(1 - \frac{(x-w)^2}{\sigma^2}\right)g^{\sigma}(x-w) = 0. \tag{15}$$

If w is known this equation can be solved numerically to get the pair (x, σ) for the location of the peak of the contour equation. The location of the peak sets an upper bound above which the discontinuities related to the staircase function of a given width cannot be detected. On the other hand, if the location of the peak in (x, σ) space is known then the width w of the staircase can be computed from the above.

The idea of the shifting of edge points is not new. It is known qualitatively that the bigger operators dislocate the edge points and that the smaller operators detect too many edge points. In fact, Canny found an uncertainty principle between the localization and the detection which states that "for a given signal-to-noise ratio an arbitrarily good localization or detection can be obtained, but not both simultaneously." What is new is the Propagation Effect. This effect not only explains the localization quantitatively, but also describes the direction in which the zero-crossings shift. It has been proposed in the literature that the bigger operator can be used for good detection and the smaller operator for good localization. It is clear from

the Propagation Effect that this simple criterion will not work for the multi-resolution operators. For the zero-crossings which collapse into one at a particular scale, one always has to resolve the ambiguity of relating the one zero-crossing at higher scale to one of two zero-crossings present at the lower scale.

6. NOISE ANALYSIS

In the previous section, we considered the zero-crossings contours related to our edge models in the ideal situation. In this section, we study the effect of noise on these models.

In this analysis we are interested in two issues. First, we want to study the extent to which the actual shape of the contours is distorted by noise. Second, we want to know whether any false contour is formed due to the presence of the noise whose shape is similar to the contours of the pulse or staircase functions.

We define signal to noise ratio (SNR) as

$$\text{SNR} = (C_{\min}/\sigma_n)^2$$

where $C_{\min}$ is the contrast of the smaller of two steps present in the pulse and staircase functions, and σ_n is the standard deviation of the noise. This definition is close to one used by Abdou and Pratt [1], except that we use the contrast of a smaller step, since we have two steps in our models.

In Figs. 11–12 we show the results for the pulse and staircase functions with the additive uniform noise. In Fig. 11a the signal with the added noise is plotted, while in Fig. 11b we show the contour obtained by applying the operator to the noisy signal. In Fig. 11c we apply a threshold to the magnitudes of the slope of zero-crossings in order to remove some of the noisy zero-crossings. In Fig. 12 similar results are presented for the staircase function.

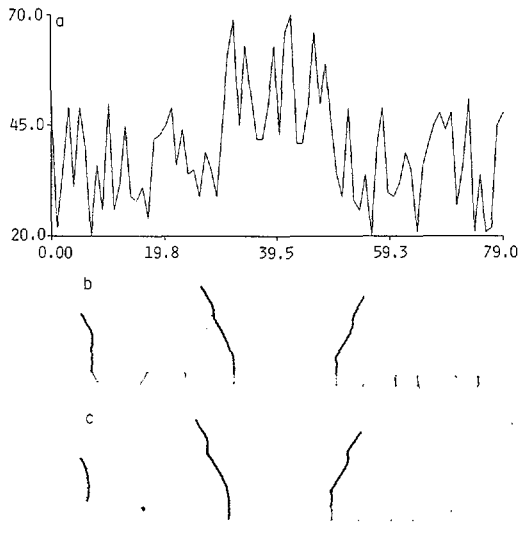

FIG. 11. Effect of Noise: (a) Pulse function with SNR = 1.5; (b) Zero-crossing contours without thresholding (c). Contours obtained by applying threshold of 10 to the slope of zero-crossings.

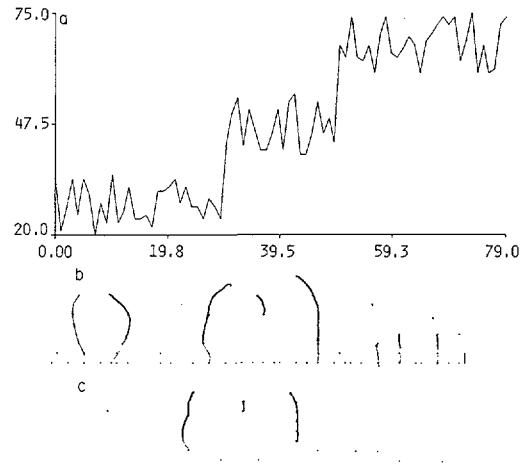

FIG. 12. Effect of Noise: (a) Staircase function with SNR = 6.25. (b) Zero-crossing contours without thresholding. (c) Contours with threshold of 10.

It is easy to see that in the case of pulse function the actual shape of the zero-crossing contour is well maintained, even in the presence of noise. Very few of the noisy zero-crossings survive after the thresholding step. In the staircase case the shape of the contour remains the same in the presence of noise, except that the straight line which passes through the center of the circular arc is distorted. As we mentioned before, the reason for getting the zero-crossings which constitute the straight line is the symmetry of the operator. In the presence of noise the two terms in Eq. (12) are not equal and hence they do not cancel each other. Therefore, in some cases the zero-crossing in the center does not appear.

7. TWO DIMENSIONS

In this section, we extend our results established previously, for two dimensions. Our intent is to verify the Propagation Effect and get an insight of the behavior of zero-crossings at multiple scales so that the theory developed in this paper can be applied to solve the edge detection problem in general. To this end, the intensity function related to isolated step edge is considered and an analytical expression describing its zero-crossings is derived. Further, the pulse and the staircase models are examined in two dimensions. Finally, we present some results for the real and noisy images to illustrate the Propagation Effect.

The intensity function of a simple step edge with an angle Φ can be described as

$$f(x, y) = cU(y - \mu x - \iota)$$
$$= \begin{cases} c & \text{if } y > \mu x + \iota \\ 0 & \text{if } y < \mu x + \iota \end{cases}$$

where $\mu = \tan \Phi$ is the slope, ι is the y-intercept, and c is the contrast of the edge (see Fig. 13).

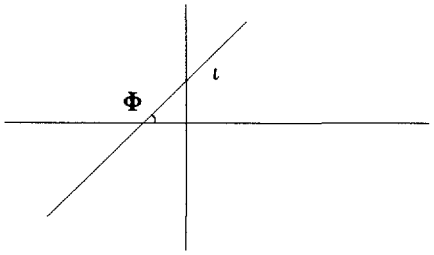

FIG. 13. Slanted edge with angle Φ and intercepts ι.

Let us find out the response of the operator for this function. Convolution of $f(x, y)$ with the bivariate Gaussian density function yields

$$\begin{aligned} h^\sigma(x, y) &= f(x, y) * g^\sigma(x, y) \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x - \xi, y - \zeta) g^\sigma(\xi, \zeta)\, d\xi\, d\zeta. \end{aligned}$$

Since the bivariate Gaussian can be decomposed into two univariate Gaussians, we can write the above expression as

$$h^\sigma(x, y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x - \xi, y - \zeta) g^\sigma(\xi) g^\sigma(\zeta)\, d\xi\, d\zeta.$$

The Laplacian of this equation can be obtained by taking the Laplacian inside the second integral and then invoking the fundamental theorem of calculus:

$$\begin{aligned} E^\sigma(x, y) &= \nabla^2 h^\sigma(x, y) \\ &= \nabla^2 \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} U(y - \zeta - \mu x + \mu\xi - \iota) g^\sigma(\xi) g^\sigma(\zeta)\, d\xi\, d\zeta \\ &= \int_{-\infty}^{\infty} g^\sigma(\xi) \nabla^2 \left[\int_{-\infty}^{(y - \mu x + \mu\xi - \iota)} c g^\sigma(\xi)\, d\zeta \right] d\xi \\ &= -\int_{-\infty}^{\infty} g^\sigma(\xi)(1 + \mu^2)(y - \mu x + \mu\xi - \iota) c g^\sigma(y - \mu x + \mu\xi - \iota)\, d\xi; \end{aligned}$$

let $A = y - \mu x - \iota$ and $\sec^2\Phi = 1 + \mu^2$. Now the above equation can be rewritten

$$E^\sigma(x, y) = -\sec^2\Phi \int_{-\infty}^{\infty} g^\sigma(\xi)(A + \mu\xi) c g^\sigma(A + \mu\xi)\, d\xi.$$

Rearranging the terms we get

$$\begin{aligned} E^\sigma(x, y) &= -\sec^2\Phi \int_{-\infty}^{\infty} g^\sigma\left(\xi \sec\Phi + \frac{A\mu}{\sec\Phi}\right) g^\sigma\left(\frac{A}{\sec\Phi}\right)(A + \mu\xi)\, d\xi \\ &= -\sec^2\Phi\, g^\sigma\left(\frac{A}{\sec\Phi}\right) \int_{-\infty}^{\infty} g^\sigma\left(\xi \sec\Phi + \frac{A\mu}{\sec\Phi}\right)(A + \mu\xi)\, d\xi. \end{aligned}$$

Let us apply the change of variable as

$$(\sec \Phi)\xi + \frac{A\mu}{\sec \Phi} = \tau,$$

therefore,

$$\xi = \frac{\tau}{\sec \Phi} - \frac{A\mu}{\sec^2 \Phi} \quad \text{and} \quad d\xi = \frac{d\tau}{\sec \Phi}.$$

Now the above equation can be rewritten

$$E^\sigma(x, y) = -c\frac{A}{\sec \Phi} g^\sigma\left(\frac{A}{\sec \Phi}\right)\int_{-\infty}^{\infty} g^\sigma(\tau)\, d\tau - cg^\sigma\left(\frac{A}{\sec \Phi}\right)\mu\int_{-\infty}^{\infty} \tau g^\sigma(\tau)\, d\tau.$$

The second integral in the above equation yields zero because the integrand is an odd function, and the first integral evaluates to unity. Finally, we have

$$E^\sigma(x, y) = -c\frac{1}{(\sec \Phi)} g^\sigma\left(\frac{y - \mu x - \iota}{\sec \Phi}\right)(y - \mu x - \iota). \tag{16}$$

The discontinuity points due to the step edge are defined by the zeros of the above equation, i.e.,

$$E^\sigma(x, y) = 0$$

$$E^\sigma(x, y) = \frac{1}{(\sec \Phi)} g^\sigma\left(\frac{y - \mu x - \iota}{\sec \Phi}\right)(y - \mu x - \iota) = 0.$$

Consider an edge positioned at the origin with angle 135°. In this case, the slope μ and the intercept ι are -1 and 0, respectively. In Fig. 14a, for this edge, we plot the locus of zero-crossings of the above equation. Four quadrants in the image show the edges detected at four different scales. Note that the edges are all parallel to each other. In Fig. 14b we show the scale-space of the step edge. In this image we have superposed the four edges, since the location of the edges is the same these four edges appear as a single edge.

Now, consider the pulse edge model which consists of two nearby edges. Assume that the edges have slopes μ_1, μ_2 and intercepts ι_1, ι_2, respectively. The response of the operator for the pulse edge can be written directly from Eq. (16) as

$$\frac{1}{(\sec \Phi_1)} g^\sigma\left(\frac{y - \mu_1 x - \iota_1}{\sec \Phi_1}\right)(y - \mu_1 x - \iota_1)$$
$$- \frac{1}{(\sec \Phi_2)} g^\sigma\left(\frac{y - \mu_2 x - \iota_2}{\sec \Phi_2}\right)(y - \mu_2 x - \iota_2) = 0. \tag{17}$$

It is obvious from the above equation that the response at a pixel location consists of a sum of two terms each corresponding to two edges in the pulse model.

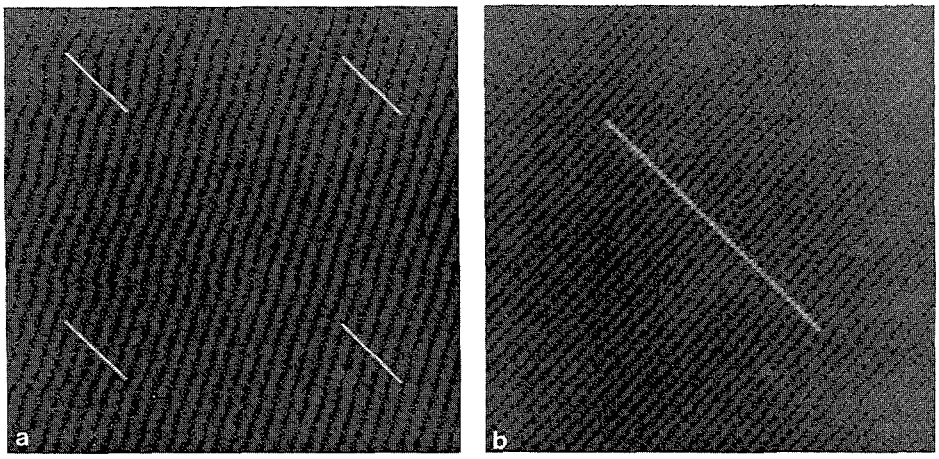

FIG. 14. (a) Step edge with slope -1 and intercept 0 at $\sigma = 3, 5, 7$, and 9. (b) Scale-space of the step edge. The location of the edge does not change when σ increases. Therefore, four edges superposed on each other in the scale-space appear as a single edge.

Therefore, the total response at one edge location in addition to its own response contains an extra term due to the neighboring edge. This extra term affects the overall behavior of a single edge in the scale space.

In order to study this effect consider an example of the pulse edge. Let the two edges have slopes μ_1 and μ_2 equal to -1 and -1, and the intercepts ι_1 and ι_2 equal to 10 and -10, respectively. In Fig. 15a we have obtained the location of edges by substituting these values in Eq. (17). Three quadrants in the image show the location of edges obtained at three different scales. In the upper left hand quadrant there is no displacement of the edges at this scale. While in the upper right-hand quadrant

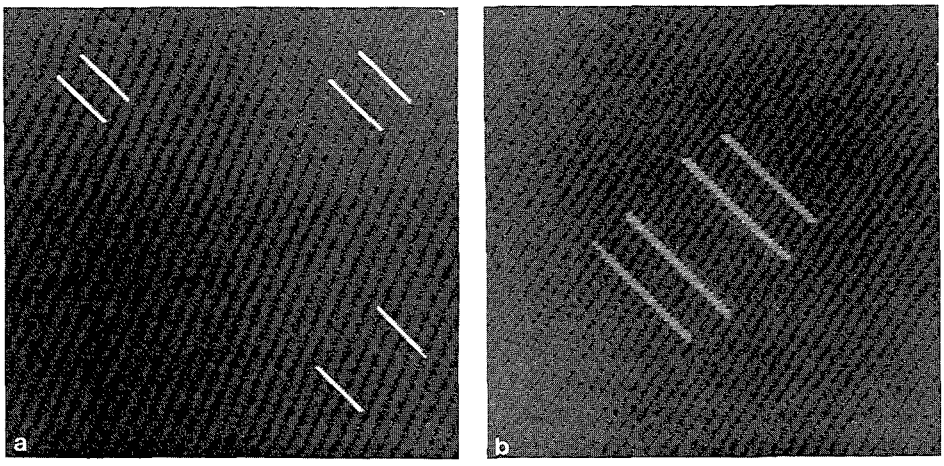

FIG. 15. (a) Pulse edge at $\sigma = 5, 7$, and 9. (b) Scale-space of pulse edge.

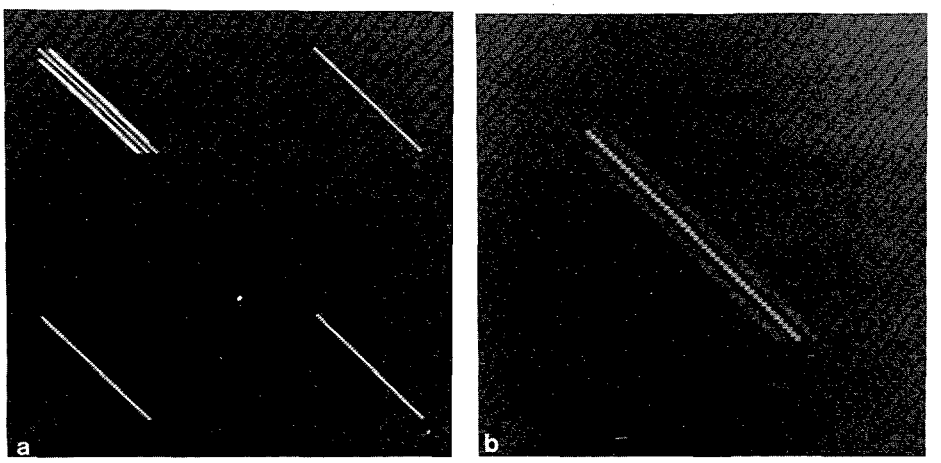

FIG. 16. (a) Staircase edge at $\sigma = 3, 5, 7$, and 9. (b) Scale space of staircase edge.

there is a slight displacement of edges. But, in the lower right-hand quadrant the displacement of edges is quite significant. The displacement of edges results due to the repulsion of the zero-crossings stated in the Propagation Effect. Since the polarities of two steps in the pulse edge are the opposite, the zero-crossings repel each other. In Fig. 15b we show the scale-space of the pulse model. In this image we have superposed the three pairs of edges shown in Fig. 15a. It is easy to notice the Propagation Effect in this image.

The response of the operator to the staircase edge can be given as

$$\frac{1}{(\sec \Phi_1)} g^{\sigma}\left(\frac{y - \mu_1 x - \iota_1}{\sec \Phi_1}\right)(y - \mu_1 x - \iota_1) + \frac{1}{(\sec \Phi_2)} g^{\sigma}\left(\frac{y - \mu_2 x - \iota_2}{\sec \Phi_2}\right)(y - \mu_2 x - \iota_2) = 0. \tag{18}$$

Now consider an example of the staircase edge. Let the two edges have slopes μ_1 and μ_2 equal to -1 and -1, and the intercepts ι_1 and ι_2 equal to 5 and -5, respectively. In Fig. 16a we have shown the location of edges by using these values in Eq. (18). Four quadrants in the image show the location of edges obtained at four different scales. In the upper left quadrant three edges are detected, two of the outer edges correspond to the actual discontinuity edges in the staircase edge model. While the third edge in the center is a false edge. This edge is detected due to the symmetric nature of the operator. In the remaining three quadrants only one edge is shown. In this case, since the edges attract each other they collapse into one at the higher scales. In Fig. 16b we show the scale-space of the staircase model. In this case also we have superposed the edges shown in part a.

Figure 17 shows the picture of a noisy pulse edge. The edge contrast for both steps is 128. We applied a uniformly distributed noise with the standard deviation equal to 64. The results are shown in Figs. 17b–d. A threshold was applied to the

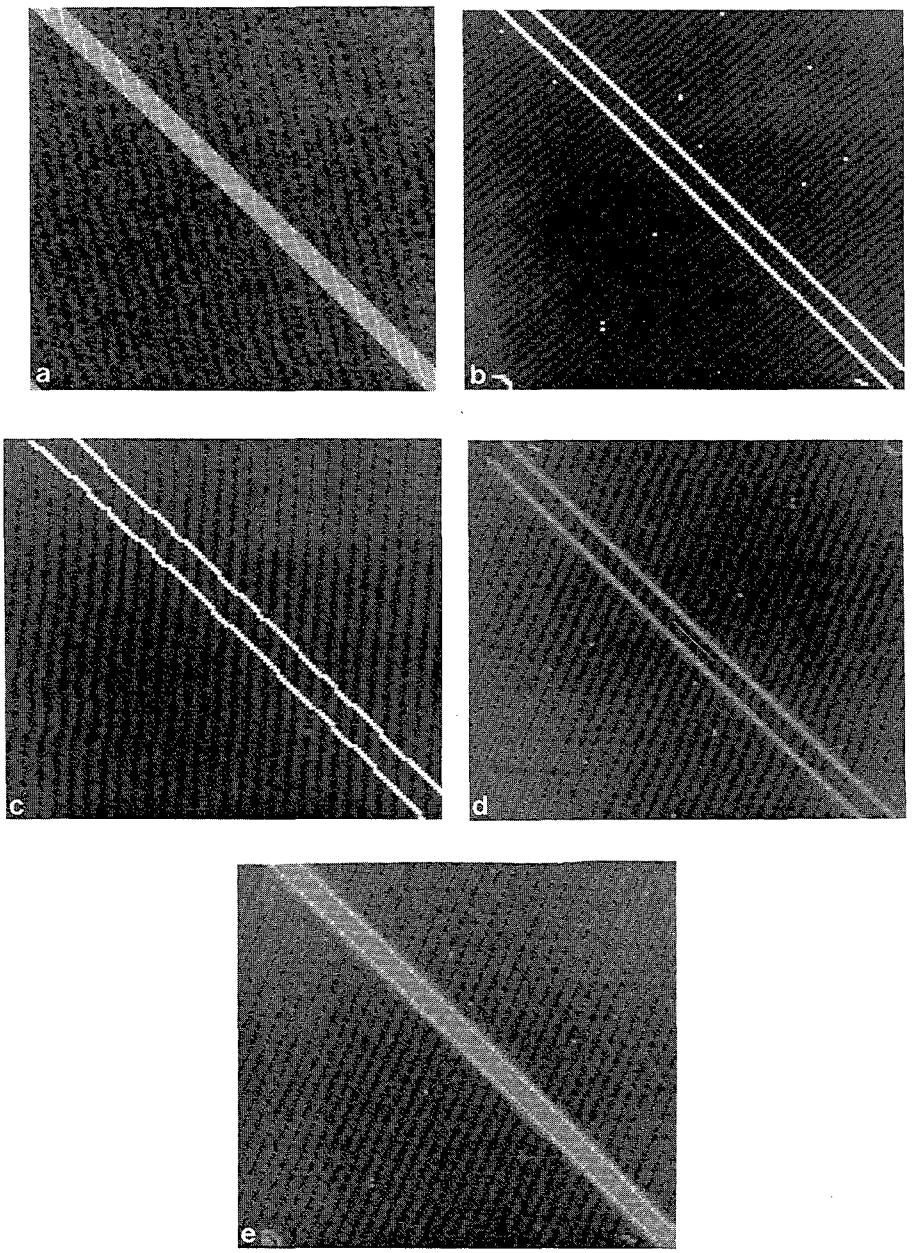

FIG. 17. (a) Image of noisy pulse edge. (b)–(c) Detected edges at $\sigma = 2$ and 4. (d) Scale-space of the noisy pulse edge. (e) Scale-space superposed on the noisy edge.

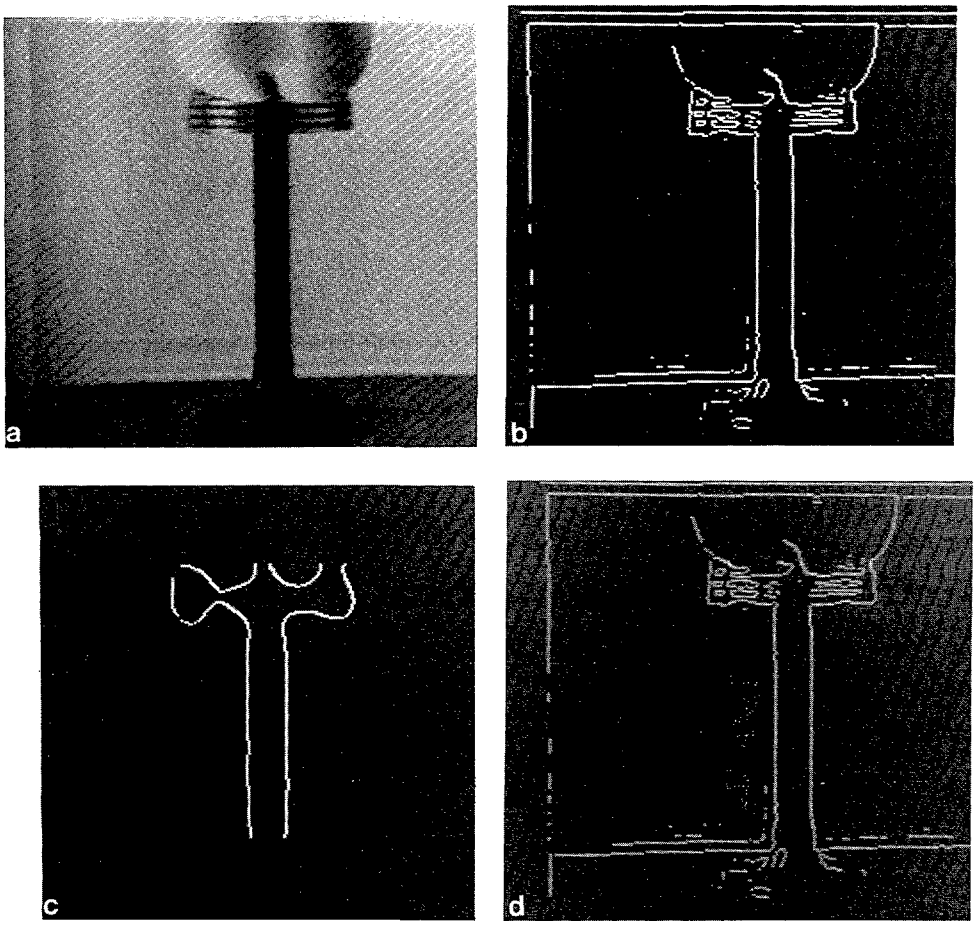

FIG. 18. (a) Bananasplit image. (b)–(c) Detected edges at $\sigma = 2$ and 4. (d) Scale space of the Bananasplit image. Reproduced from "Photographs" by Elliot McDowell. Copyright 1981 by Elliot McDowell. Reprinted by permission of David R. Godine, Publisher, Inc.

slope of zero-crossings in order to remove the noisy edges. Note that even in the presence of the noise the Propagation Effect is illustrated.

Finally, in Fig. 18 we show the results for a real scene. This is an artistic photograph of "Bananasplit" by McDowell from his book "Photographs" [20]. This photograph was digitized on 512 by 512 grid under no illumination control. In this picture the stem of the glass simulates a 2-dimensional pulse edge. Our results clearly demonstrate the Propagation Effect for this real scene..

8. SUMMARY AND CONCLUSION

In this paper we consider the problem of detecting gray level discontinuities at multiple scales. We find that the current edge models are inadequate for the edges detected by the multi-resolution operators. Since the isolated step or ramp edges rarely occur in natural scenes, we propose the pulse and staircase models. In these models we include the effect of an edge on the neighboring edge which propagates

through when the operator size increases. While analyzing the behavior of the pulse and the staircase models in the scale-space, we find that the zero-crossings attract or repel each other. When they attract each other at some value of σ they collapse into one. In principle the Propagation Effect can be modeled such that one edge is affected by more than one neighboring edge. But for simplicity in this paper we consider only the effect of one neighboring edge.

In our future work we want to fit the primitives to the discrete zero-crossings in the scale-space of the image. Using proposed models we expect to recover rich and robust information about the intensity function. In particular we are interested to extract information about the contrast, slope, and orientation of edges.

REFERENCES

1. I. Abdou and W. K. Pratt, Quantitative design and evaluation of enhancement edge detector schemes, *Proc. IEEE* **67**, 1979, 753–763.
2. H. Asada and M. Brady, *The Curvature Primal Sketch*, MIT AI Memo 758, 1984.
3. J. Babaud, A. Witkin, and R. Duda, *Uniqueness of the Gaussian Kernel for Scale-Space Filtering*, Fairchild TR 645, Flair 22.
4. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice–Hall, Englewood Cliffs, NJ.
5. V. Berzins, Accuracy of Laplacian edge detectors, *Computer Vision Graphics Image Process.* **27**, 1984, 195–210.
6. J. F. Canny, *Finding Edges and Lines in Images*, M.S. thesis, MIT, 1983.
7. M. Carlotto, Histogram analysis using a scale-space approach, in "Proc. CVPR-2," June 1985, San Francisco, pp. 334–340.
8. J. Eklundh, T. E. Elfving, and S. Nyberg, Edge detection using the Marr–Hildreth operator with different sizes, in "ICPR-6," pp. 1109–1112.
9. W. E. Grimson and E. C. Hildreth, Comments on "Digital Step Edges from Zerocrossings of Second Directional Derivative," *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-7**, 1985, 121–127.
10. R. M. Haralick, Edge and region analysis for digital image data, *Comput. Graphics Image Process.* **12**, 1980, 60–73.
11. R. M. Haralick, The digital edge in *Proc. 1981 Conf. on Pattern Recognition and Image Processing*, Dallas, Texas, 1981, pp. 285–294.
12. R. M. Haralick, Zero-crossings of second directional derivative edge operator, *Proc. Soc. Photo-Opt. Instrum. Eng. on Robot Vision*, Arlington, Virginia, 1982.
13. R. M. Haralick, Digital step edges from zero-crossings of second directional derivative, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, 1984, 58–68.
14. R. M. Haralick, Author's reply, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-7**, 1985, 127–129.
15. E. C. Hildreth, Detection of intensity changes by computer and biological vision systems, *Comput. Vision Graphics Image Process.* **22**, 1983, 1–27.
16. A. K. Mackworth and F. Mokhtarian, *Scale-Based Descriptions of Planar Curves*, Technical Report 84-1, UBC.
17. D. H. Marimont, A representation for image curves, in *Proceedings of AAAI*, 1984, pp. 237–242.
18. D. Marr, *Vision*, Freeman, San Francisco, 1982.
19. D. Marr and E. Hildreth, Theory of edge detection, *Proc. Roy. Soc. London B* **207**, 1980, 187–217.
20. E. McDowell, *Photographs*, Godine, Boston, 1981.
21. A. Rosenfeld and A. Kak, *Digital Picture Processing*, Academic Press, New York/London, 1982.
22. A. Rosenfeld and M. Thurston, Edge and curve detection for visual scene, *IEEE trans Comput.* **C-20**, 1971, 562–569.
23. V. Torre and T. Poggio, *On Edge Detection*, MIT AI Memo 768, August 1984.
24. W. I. Voxman and R. H. Goeschel, *Advanced Calculus*, Dekker, New York, 1981.
25. A. Witkin, Scale-space filtering, in *Int. Joint Conf. Artif. Intell-83.*
26. A. L. Yuille and T. Poggio, *Scaling Theorems for Zero-Crossings*, MIT Artificial Intelligence Lab Memo 722, 1983.
27. A. L. Yuille and T. Poggio, *Fingerprints Theorems for Zero Crossings*, MIT AI Memo 730, 1983.
28. S. Zucker and R. Hummel, Receptive fields and the representation of visual information, *7th ICPR*, Montreal, pp. 515–517.