Integrated Directional Derivative Gradient Operator

OSCAR A. ZUNIGA, MEMBER, IEEE, AND
ROBERT M. HARALICK, FELLOW, IEEE

Abstract —Accurate edge direction information is required in many image processing applications. A variety of operators for computing local edge direction have been proposed, many of them estimating a kind of gradient. These operators face two major problems. One problem is the inherent bias in their estimate of edge direction. The bias itself is a function of edge direction. Another problem is their sensitivity to the presence of noise in the image data. The second problem can be alleviated by an increase in the processing neighborhood size but usually at the expense of an increase in estimate bias and also in errors in the processing of small or thin objects. An operator based on the cubic facet model is discussed, which reduces sharply both estimate bias and noise sensitivity with no increase in computational complexity. The measure of gradient strength is the maximum value of the integral of the first directional derivative taken over a rectangular or square neighborhood, the maximum being taken over all possible directions for the directional derivative. The line direction which maximizes the integral defines the new estimate of gradient direction. Experimental results show the superiority of this operator to others such as the Roberts operator, the Prewitt operator, the Sobel operator, and the standard cubic facet gradient operator for step edges and ramp edges. Under zero-noise conditions the 7×7 integrated directional derivative gradient operator has a worst bias of less than $0.09°$, and the 5×5 integrated directional derivative gradient operator has a worst bias of less than $0.26°$ on ramp edges. For comparison purposes the 7×7 standard cubic facet gradient operator has a worst bias of about $1.2°$, and the 5×5 standard cubic facet gradient operator has a worst bias of $0.5°$. The 7×7 Prewitt operator has a worst bias of $5°$, and the 5×5 Prewitt operator has a worst bias of $4°$. This improvement in worst bias stays with the contamination of edges by additive independent zero-mean Gaussian noise.

I. INTRODUCTION

The computation of edge direction is a required step in many image processing tasks. Hough transformation techniques [3] have been used extensively together with edge direction information to detect lines [13], circles [10], and arbitrary shapes [1]. Martelli [12] and Ramer [15] each use edge direction information to perform edge linking. Kitchen and Rosenfeld [11] and Zuniga and Haralick [17] use edge direction information in schemes to detect corners.

A variety of operators for computing local edge direction have been proposed, many of them estimating a kind of gradient of the picture function. Local edge direction is then estimated as that direction which is orthogonal to the gradient direction. Knowledge of the directional derivatives D_1 and D_2 in any two orthogonal directions is sufficient to compute the directional derivative in any arbitrary direction. The gradient magnitude, which is defined as the maximum such directional derivative, is computed as $\sqrt{D_1^2 + D_2^2}$ and its direction as $\tan^{-1} D_2/D_1$. A number of local operators have been utilized to estimate these directional derivatives. Examples are the Roberts operator [16], the Prewitt operator [14], the Sobel operator [3], and the Hueckel operator [8]. Of special interest are those operators which result from a local surface fit to the graytone image data and subsequent computation of directional derivatives in two orthogonal directions from the analytic description of the estimated surface. Thus the original Roberts operator has been shown to result from a linear surface fit within a 2×2 pixel neighborhood [5]. The

original Prewitt operator was derived from a quadratic surface fit within a 3×3 pixel neighborhood. Haralick [6] and Zuniga and Haralick [17] compute edge direction information from a cubic polynomial surface fit. This fitting idea can also be extended to arbitrary sized neighborhoods [2], [5].

These gradient operators face two major problems. One problem is that their estimate of edge direction is inherently biased as a function of true edge direction and displacement of the true edge from the pixel's center. The second major problem is the sensitivity of these operators to data noise. An obvious approach to decrease noise sensitivity is to increase the neighborhood size. However, this creates problems with small or thin objects. In addition, for all the gradient operators already mentioned an increase in neighborhood size causes an increase in edge direction estimate bias.

In this correspondence we use an operator based on a cubic polynomial surface fit. This cubic facet model has been successfully used in the past to detect edges [6], topographic features [7], and corners [17]. Instead of computing directional derivatives directly from the fitted surface as in the case of the standard cubic facet gradient operator mentioned earlier, we describe an operator that measures gradient strength as the minimum value of the integral of first directional derivative taken over lines going in all possible directions. The line direction which maximizes the integral defines the new estimate of gradient direction. We present experimental evidence that this operator possesses two main characteristics.

1) Edge direction estimate bias is sharply reduced as compared with the bias of the standard cubic facet, Sobel, and Prewitt gradient operators. Noise sensitivity is comparable to that of the Sobel and Prewitt operators and much better than the standard cubic facet operator.

2) Unlike the standard cubic facet, Sobel, and Prewitt operators, increasing the neighborhood size decreases both estimate bias and noise sensitivity. For ramp edges the integrated operator is very nearly unbiased. The worst bias for the 7×7 operator is less than $0.09°$, and the worst bias for the 5×5 operator is less than $0.26°$.

Section II describes the standard cubic facet gradient operator. Section III describes the mathematical analysis necessary to derive the new gradient estimate. Section IV presents experimental results and provides a comparison of the integrated directional derivative gradient operator against the standard cubic facet gradient operator, Prewitt operator, and Sobel operator for step edges and ramp edges contaminated by zero-mean Gaussian noise. The appendix describes the mathematical analysis of the cubic facet.

II. THE STANDARD CUBIC FACET GRADIENT OPERATOR

Given a graytone intensity function f defined in the row and column coordinate system of a given pixel neighborhood, the gradient vector function ∇f is given by

$$\nabla f = \left(\frac{\partial f}{\partial r}, \frac{\partial f}{\partial c} \right). \tag{1}$$

Under the cubic facet model each surface facet centered about a given pixel is approximated by the bivariate cubic form

$$f(r, c) = K_1 + K_2 r + K_3 c + K_4 r^2 + K_5 rc$$
$$+ K_6 c^2 + K_7 r^3 + K_8 r^2 c + K_9 rc^2 + K_{10} c^3 \tag{2}$$

where the K coefficients change from neighborhood to neighborhood and are estimated using a least-square-error surface fit. A detailed analysis of this estimation procedure is provided in the Appendix. Evaluating the first row and column partial derivatives of (2) at the neighborhood center $(0,0)$ and replacing their values in (1), the gradient vector at the neighborhood center

becomes

$$\nabla f = (K_2, K_3).$$

The magnitude and direction of the standard cubic facet gradient operator are, therefore, given by $\sqrt{K_2^2 + K_3^2}$ and $\tan^{-1} K_2/K_3$, respectively, with the gradient direction being measured as a clockwise angle from the column axis.

III. THE INTEGRATED DIRECTIONAL DERIVATIVE GRADIENT OPERATOR

For a given direction vector $(\sin\theta, \cos\theta)$, it is well-known that the first directional derivative $f_\theta'(r, c)$ of f in the direction θ can be evaluated as the component of the gradient ∇f along the direction vector, that is,

$$f_\theta'(r, c) = \frac{\partial f}{\partial r} \sin\theta + \frac{\partial f}{\partial c} \cos\theta. \tag{3}$$

Let F_θ be defined as

$$F_\theta = \frac{1}{4LW} \int_{-W}^{W} \int_{-L}^{L} f_\theta'(\rho\cos\theta + \omega\sin\theta,$$

$$-\rho\sin\theta + \omega\cos\theta)\, d\rho\, d\omega \tag{4}$$

for a given $N \times N$ neighborhood.

F_θ represents the integrated first directional derivative along lines orthogonal to the direction θ forming a rectangle of length $2L$ and width $2W$ centered at the origin of the coordinate system. The proposed gradient estimate is

$$G = F_{\theta_{MAX}} u_{\theta_{MAX}} \tag{5}$$

where $F_{\theta_{MAX}} = \max_\theta F_\theta$ and $u_{\theta_{MAX}}$ is a unit vector in the direction that maximizes F_θ.

Using (2) and (3), $f_\theta'(\rho\cos\theta + \omega\sin\theta, -\rho\sin\theta + \omega\cos\theta)$ reduces to

$$f_\theta'(\rho\cos\theta + \omega\sin\theta, -\rho\sin\theta + \omega\cos\theta)$$

$$= \left[3(K_9 - K_7)\sin^3\theta + 3(K_8 - K_{10})\cos^3\theta\right.$$

$$\left. + (3K_7 - 2K_9)\sin\theta + (3K_{10} - 2K_8)\cos\theta\right]\rho^2$$

$$+ 2\left[-K_8\sin^3\theta + (3K_7 - 2K_9)\sin^2\theta\cos\theta\right.$$

$$\left. + (2K_8 - 3K_{10})\sin\theta\cos^2\theta + K_9\cos^3\theta\right]\rho\infty$$

$$+ \left[3(K_7 - K_9)\sin^3\theta + 3(K_{10} - K_8)\cos^3\theta\right.$$

$$\left. + 3K_9\sin\theta + 3K_8\cos\theta\right]\infty^2$$

$$+ \left[-K_5\sin^2\theta + 2(K_4 - K_6)\sin\theta\cos\theta + K_5\cos^2\theta\right]\rho$$

$$+ 2\left[K_4\sin^2\theta + K_5\sin\theta\cos\theta + K_6\cos^2\theta\right]\infty$$

$$+ K_2\sin\theta + K_3\cos\theta. \tag{6}$$

Substituting (6) in (4) results in

$$F_\theta = \frac{1}{4LW} \int_{-W}^{W} \int_{-L}^{L} (A\rho^2 + B\rho\omega + C\omega^2 + D\rho + E\omega + F)\, d\rho\, d\omega$$

where A, B, C, D, E, and F are the coefficients of the quadratic equation (6). Evaluating this integral results in

$$F_\theta = \frac{1}{3}AL^2 + \frac{1}{3}CW^3 + F.$$

Finally,

$$F_\theta = (K_9 - K_7)(L^2 - W^2)\sin^3\theta + (K_8 - K_{10})(L^2 - W^2)\cos^3\theta$$

$$+ \left[L^2 K_7 - \frac{1}{3}(2L^2 - 3W^2)K_9 + K_2\right]\sin\theta$$

$$+ \left[L^2 K_{10} - \frac{1}{3}(2L^2 - 3W^2)K_8 + K_3\right]\cos\theta. \tag{7}$$

Thus F_θ reduces to a trigonometric expression in $\sin\theta$ and $\cos\theta$. Notice that if $L = W$, the cubic terms $\sin^3\theta$ and $\cos^3\theta$ drop, and F_θ is maximized when

$$\theta = \theta_{MAX} = \tan^{-1} \frac{L^2 K_7 + \frac{1}{3}L^2 K_9 + K_2}{L^2 K_{10} + \frac{1}{3}L^2 K_8 + K_3}. \tag{8}$$

Then

$$F_{\theta_{MAX}} = \sqrt{D_1^2 + D_2^2} \tag{9}$$

where D_1 and D_2 are the numerator and denominator of the argument of the tangent function in (8).

The computation of the gradient strength given by (9) and gradient directional given by (8) can be simplified by precomputing "row derivative" and "column derivative" masks from the numerator and denominator of (8), respectively, using the masks for the K parameters derived in the Appendix. As a special case, when $L = W = 0$, then (8) and (9) reduce to the standard cubic facet operator defined in Section II.

If $L \neq W$, then the maximization of F_θ to obtain the estimated gradient strength $F_{\theta_{MAX}}$ and estimated gradient direction θ_{MAX} can be carried out by the following procedure. From (7), F_θ can be expressed as

$$F_\theta = u\sin^3\theta + v\cos^3\theta + w\sin\theta + z\cos\theta \tag{10}$$

where u, v, w, and z are functions of the K parameters. Then

$$F_\theta' = 3u\sin^2\theta\cos\theta - 3v\cos^2\theta\sin\theta + w\cos\theta - z\sin\theta.$$

Equating F_θ' to zero to obtain an extremum results in

$$\left[3u(1 - \cos^2\theta) + w\right]\cos\theta = (3v\cos^2\theta + z)\sin\theta. \tag{11}$$

Let $x = \cos^2\theta$; then (9) becomes

$$3u(1 - x) + w = (3vx + z)\tan\theta.$$

Squaring this expression, we obtain

$$(3u(1 - x) + w)^2 = (3vx + z)^2\tan^2\theta.$$

Replacing $\tan^2\theta$ by $(1 - x)/x$ and after some algebraic manipulation, we finally obtain the cubic expression in x:

$$9(u^2 + v^2)x^3 - (18u^2 + 6uw + 9v - 6vz)x^2$$

$$+ (9u^2 + 6uw + w^2 - 6vz + z^2)x - z^2 = 0. \tag{12}$$

Then $F_{\theta_{MAX}}$ and θ_{MAX} can be obtained by the following steps:

1) find all the roots x of (12) between 0 and 1;
2) convert these x's to all possible θ's from $x = \cos^2\theta$;
3) evaluate F_θ at all these θ's;
4) choose maximum value $F_{\theta_{MAX}}$ and θ_{MAX}.

IV. EXPERIMENTAL RESULTS

The experiments were performed with step and ramp edges contaminated by zero-mean Gaussian noise. Step edges were generated in a rectangular grid with orientations θ from zero to 90° and with random displacement from the grid's center uni-

formly distributed within the range $(-D, D)$ with the maximum displacement D given by

$$D = \begin{cases} 0.5\cos\theta, & 0 \leqslant \theta < 45 \\ 0.5\sin\theta, & 45 \leqslant \theta \leqslant 90, \end{cases}$$

assuming a unit distance between two four-neighbor pixels in the grid. A step edge passing through a pixel divides it into two parts having areas A_1 and A_2, with $A_1 + A_2 = 1$. Let the corresponding graytone intensities to each side of the edge be I_1 and I_2. The pixel is then assigned a graytone intensity I according to the rule

$$I = I_1 A_1 + I_2 A_2.$$

The experiments were performed with values for I_1 and I_2 equal to 100 and 200, respectively. That is, the edge contrast was set to 100.

Ramp edges were generated by defocusing step edges with a 3×3 averaging filter. Finally, both step and ramp edges were contaminated by adding zero-mean Gaussian noise with a given standard deviation.

Tuning Up the Integrated Directional Derivative Gradient Operator

The performance of the integrated directional derivative gradient operator depends on the choice of integration limits L and W. As seen in Section III, this operator is equivalent to the standard cubic facet gradient operator when $L = W = 0$. We expect its performance to improve as L and W move away from zero and to reach a maximum for some value of $L \neq 0$ and some value of $W \neq 0$. We are primarily interested in improving the edge direction estimate. We use two performance measurements, edge direction estimate bias and edge direction estimate standard deviation. The latter measures noise sensitivity. The estimate bias is defined as the difference between the estimate mean and the true edge direction. A single performance measurement to compare two sets of values of L and W is obtained by combining the previous two measurements into the root-mean-square error of the estimate which can be shown to be equal to the square root of the sum of the square bias and the estimate variance. It was observed that the operator achieved best performance in the root-mean-square error sense when $L = W = 1.8$ for a 5×5 neighborhood size and $L = W = 2.5$ for a 7×7 neighborhood size for both step and ramp edges and for a variety of noise levels.

Comparing Different Gradient Operators

The following gradient operators were tested: 5×5 extended Sobel [9], 5×5 and 7×7 Prewitt, 5×5 and 7×7 standard cubic facet, and 5×5 and 7×7 integrated directional derivative. Fig. 1 shows the 5×5 row derivative masks for each of the operators tested, and Fig. 2 shows the 7×7 row derivative mask for the integrated directional derivative gradient operator. The column derivative masks can be obtained from the row masks by transposition.

For a step or ramp edge of a given orientation and noise standard deviation, each operator was applied to the grid's center 10 000 times, each time with a different noise sample and a different edge displacement from the grid's center. Under zero-noise conditions the operators were applied only 100 times. Edge orientations varied from zero to 90° and noise standard deviation from zero to 100. Edge contrast was set to 100. The edge direction estimates produced by each operator were plotted as follows: for a fixed noise standard deviation, estimate bias against true edge direction, and estimate standard deviation against true edge direction; for a fixed edge direction, estimate bias against noise standard deviation, and estimate standard deviation against noise standard deviation.

We compare first the standard cubic facet gradient operator and the integrated directional derivative gradient operator. Under zero-noise conditions, Figs. 3 and 4 show estimate bias against true edge direction for step and ramp edges. Three things can be observed. First, the integrated operator is clearly superior
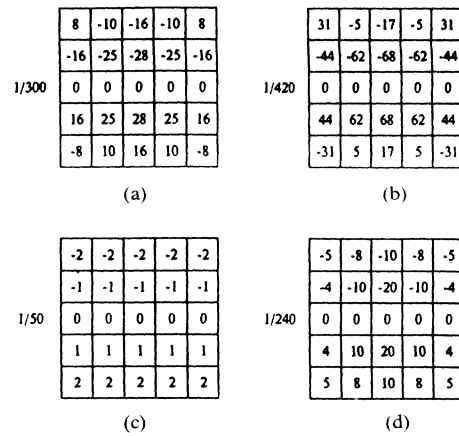

Fig. 1. Row derivative masks for gradient operators in 5×5 neighborhood size (a) Integrated directional derivative. (b) Standard cubic facet. (c) Prewitt. (d) Extended Sobel.
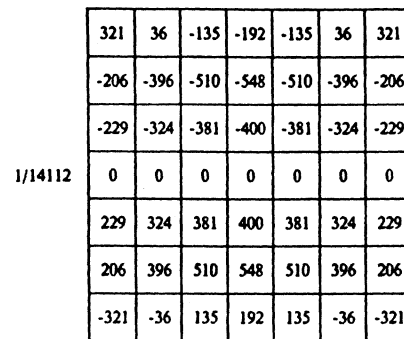

Fig. 2. Row derivative mask for integrated directional derivative gradient operator for 7×7 neighborhood size.
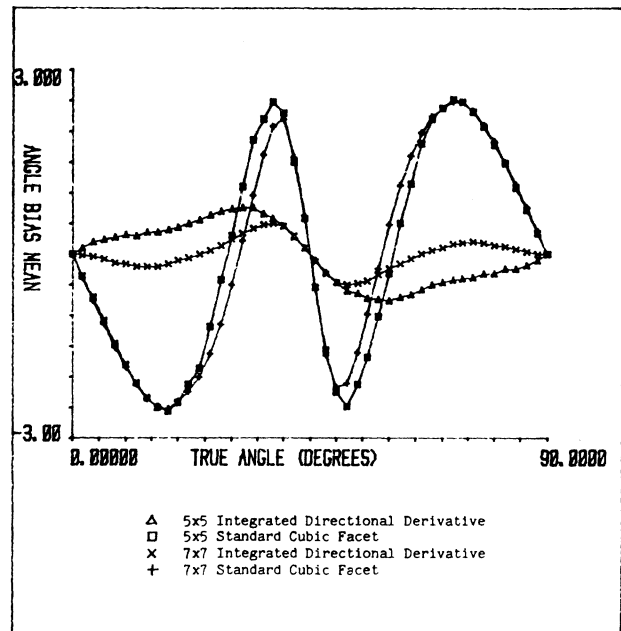

Fig. 3. Estimate bias as function of true edge direction for step edges under zero-noise conditions.

to the standard cubic facet gradient operator. Under zero-noise conditions the 7×7 integrated directional derivative gradient operator has a worst bias of less than 0.09°, and the 5×5 integrated directional derivative gradient operator has a worst bias of less than 0.26° on ramp edges. For comparison purposes, the 7×7 standard cubic facet gradient operator has a worst bias
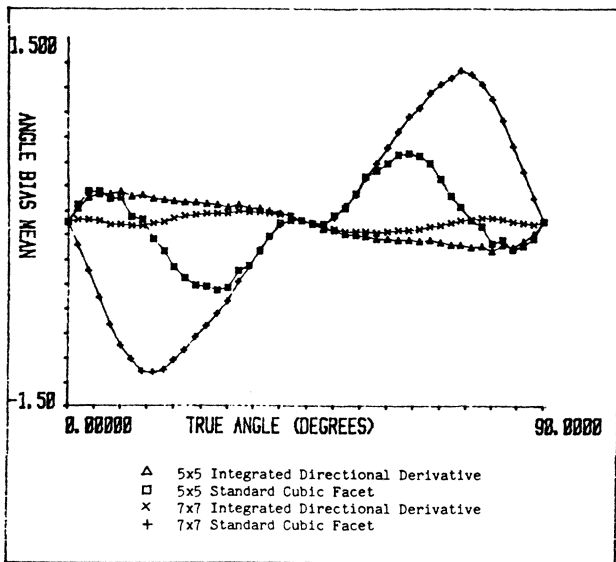
Fig. 4. Estimate bias as function of true edge direction for ramp edges under zero-noise conditions.
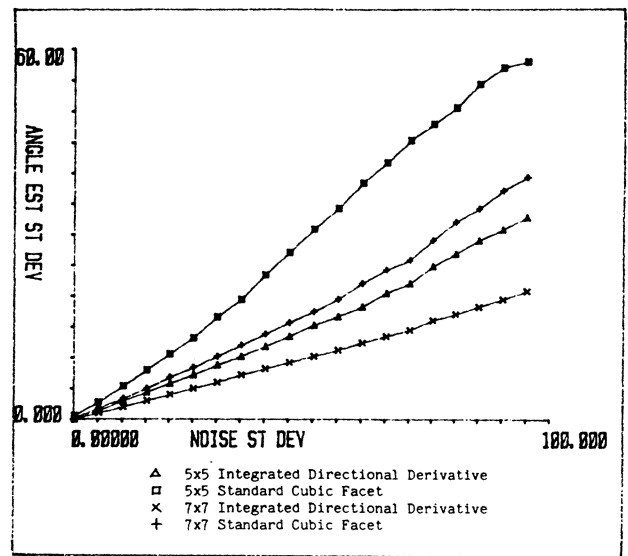
△   5x5 Integrated Directional Derivative
□   5x5 Standard Cubic Facet
✕   7x7 Integrated Directional Derivative
✛   7x7 Standard Cubic Facet

Fig. 7. Estimate standard deviation as function of noise standard deviation for step edge. Edge orientation is 22.5°. Edge contrast is 100.
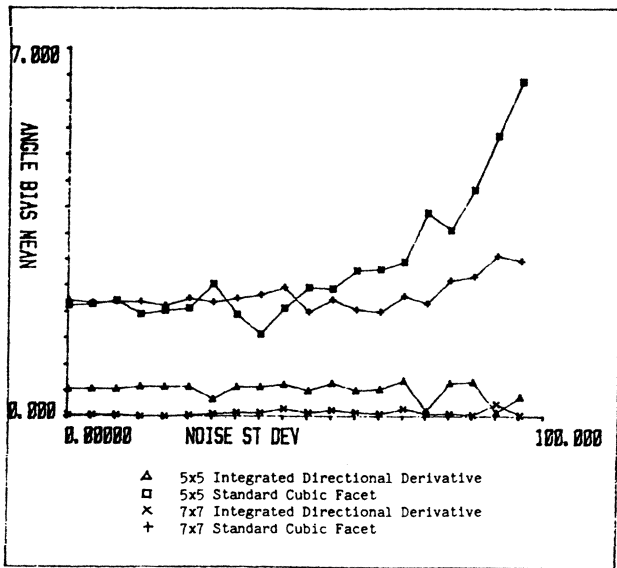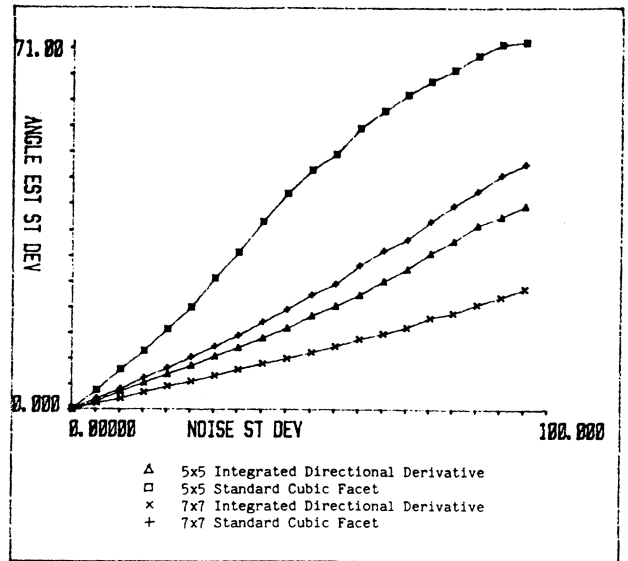
△   5x5 Integrated Directional Derivative
□   5x5 Standard Cubic Facet
✕   7x7 Integrated Directional Derivative
✛   7x7 Standard Cubic Facet

Fig. 5. Estimate bias as function of noise standard deviation for step edge. Edge orientation is 22.5°. Edge contrast is 100.

△   5x5 Integrated Directional Derivative
□   5x5 Standard Cubic Facet
✕   7x7 Integrated Directional Derivative
✛   7x7 Standard Cubic Facet

Fig. 8. Estimate standard deviation as function of noise standard deviation for ramp edge. Edge orientation is 22.5°. Edge contrast is 100.

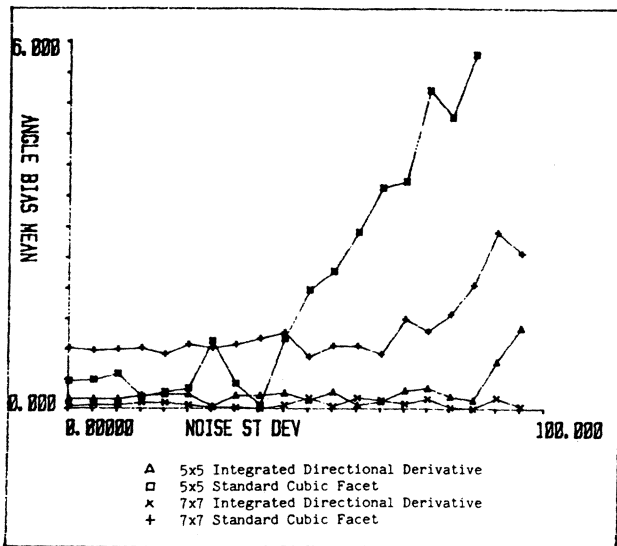△   5x5 Integrated Directional Derivative
□   5x5 Standard Cubic Facet
✕   7x7 Integrated Directional Derivative
✛   7x7 Standard Cubic Facet

Fig. 6. Estimate bias as function of noise standard deviation for ramp edge. Edge orientation is 22.5°. Edge contrast is 100.

△   5x5 Integrated Directional Derivative
□   5x5 Sobel
◇   5x5 Prewitt
✕   7x7 Integrated Directional Derivative
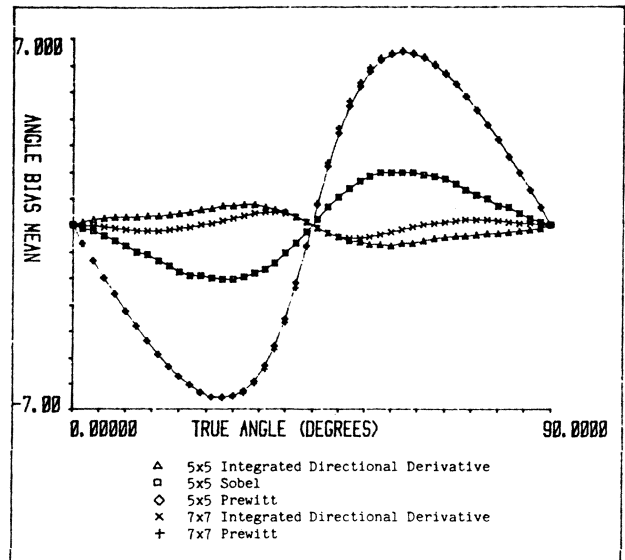✛   7x7 Prewitt

Fig. 9. Estimate bias as function of true edge direction for step edges under zero-noise conditions.
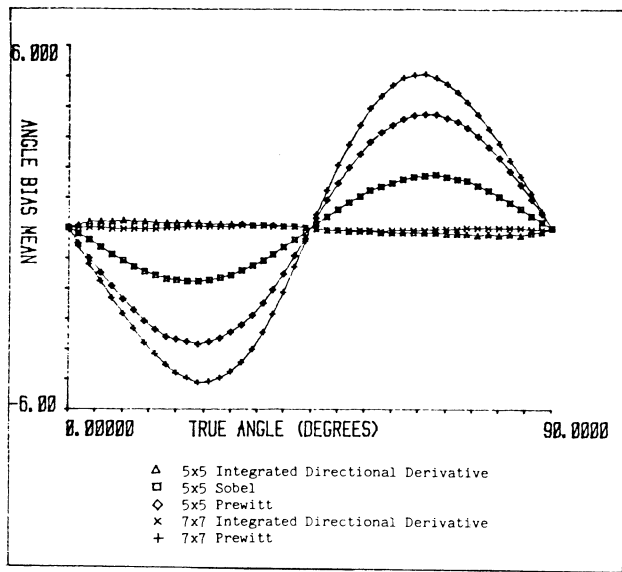
Fig. 10.   Estimate bias as function of true edge for ramp edges under zero-noise conditions.
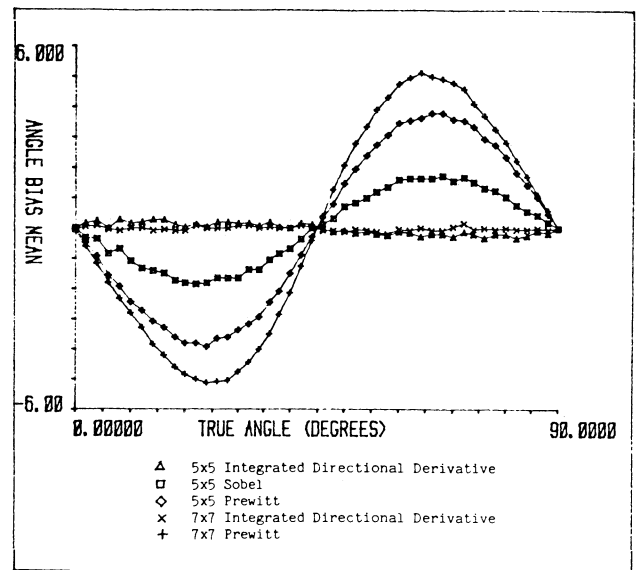

Fig. 12.   Estimate bias as function of true edge direction for ramp edge. Noise standard deviation is 25. Edge contrast is 100.
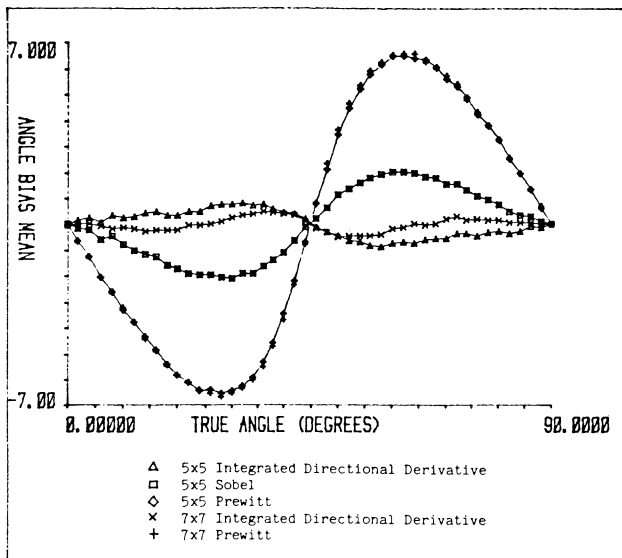

Fig. 11.   Estimate bias as function of true edge direction for step edge. Noise standard deviation is 25. Edge contrast is 100.
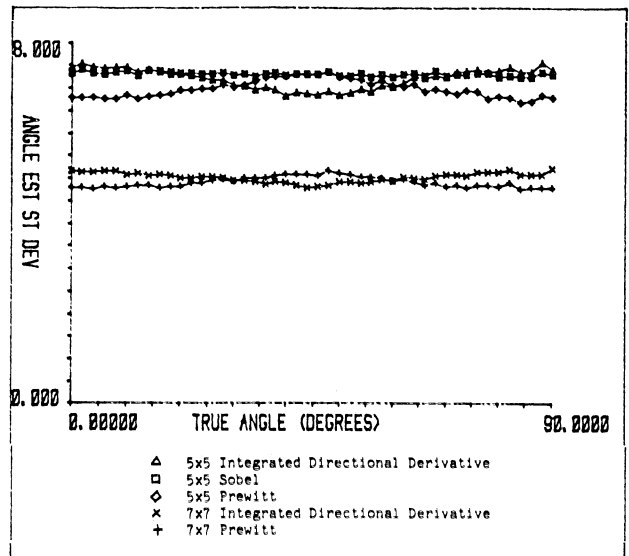

Fig. 13.   Estimate standard deviation as function of true edge direction for step edge. Noise standard deviation is 25. Edge contrast is 100.

of about 1.2°, and the 5×5 standard cubic facet gradient operator has a worst bias of 0.5°. This improvement in worst bias stays with the contamination of edges by additive independent zero-mean Gaussian noise. Second, for the integrated operator, estimate bias decreases as the neighborhood size increases while the opposite happens with the standard cubic facet gradient operator. Third, both operators perform better with ramp edges than with step edges.

Figs. 5–8 show estimate bias and estimate standard deviation against noise standard deviation for a fixed-edge orientation of 22.5° and additive independent Gaussian noise. Again, the integrated operator is uniformly superior to the standard cubic facet gradient operator for both step and ramp edges.

Next we compare the integrated directional derivative operator with the Prewitt and extended Sobel operator. Under zero-noise conditions, Figs. 9 and 10 show estimate bias as a function of true edge direction for step and ramp edges. The 7×7 integrated operator has the smallest bias followed by the 5×5 integrated operator, the 5×5 extended Sobel, and the 5×5 and 7×7 Prewitt operators. Notice that for ramp edges the response of the in-

tegrated operator is nearly flat about zero, that is, the operator is nearly unbiased. For the 7×7 integrated operator the worst bias is less than 0.09°, and for the 5×5 integrated operator the worst bias is less than 0.26°. For comparison purposes the worst bias in the 7×7 Prewitt operator is about 5°, and the worst bias in the 5×5 Prewitt operator is about 4°. Again, the integrated operator is the only one for which bias decreases as neighborhood size increases. Only the 5×5 Sobel operator is shown, but as previously demonstrated by Iannino and Shapiro [9], the 3×3 Sobel operator has a smaller bias than the 5×5 extended Sobel, but it is still significantly larger than the bias of the integrated operator and with a much worse noise sensitivity. Iannino and Shapiro [9] also show results with 3×3 and 5×5 iterated Sobel operators. The bias for these iterated operators is still larger than the bias of the integrated operator, and they are more expensive computationally.

Figs. 10–14 show estimate bias and estimate standard deviation as a function of true edge direction for step and ramp edges when the noise standard deviation is equal to 25. The bias for all the operators shown is nearly identical to the bias under zero-noise
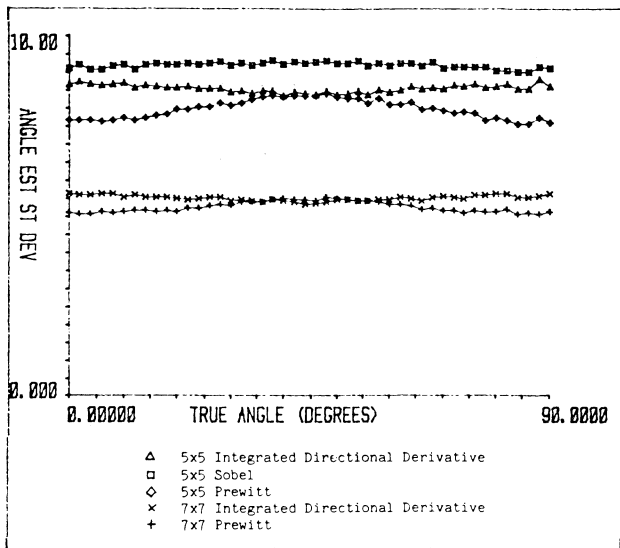
Fig. 14. Estimate standard deviation as function of true edge direction for ramp edge. Noise standard deviation is 25. Edge contrast is 100.
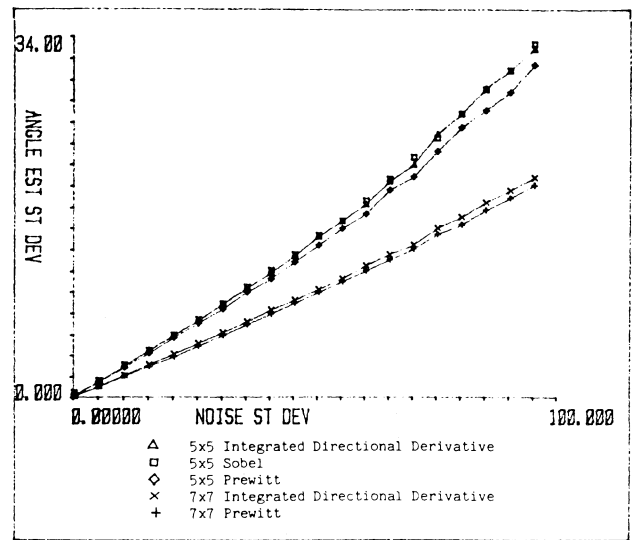

Fig. 17. Estimate standard deviation as function of noise standard deviation for step edge. Edge orientation is 22.5°. Edge contrast is 100.
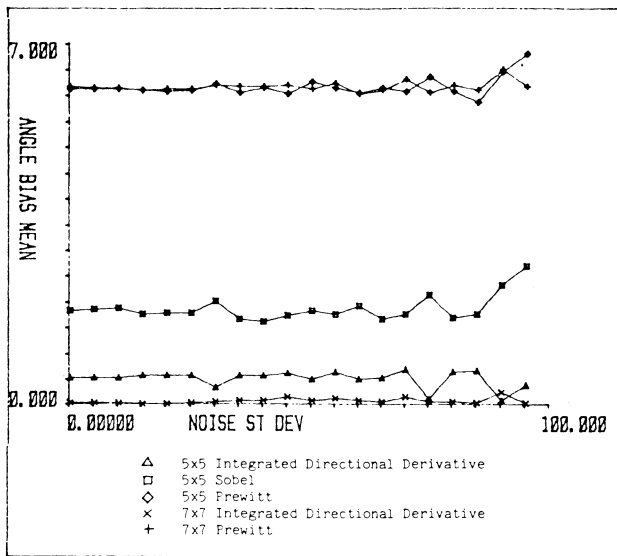

Fig. 15. Estimate bias as function of noise standard deviation for step edge. Edge orientation is 22.5°. Edge contrast is 100.
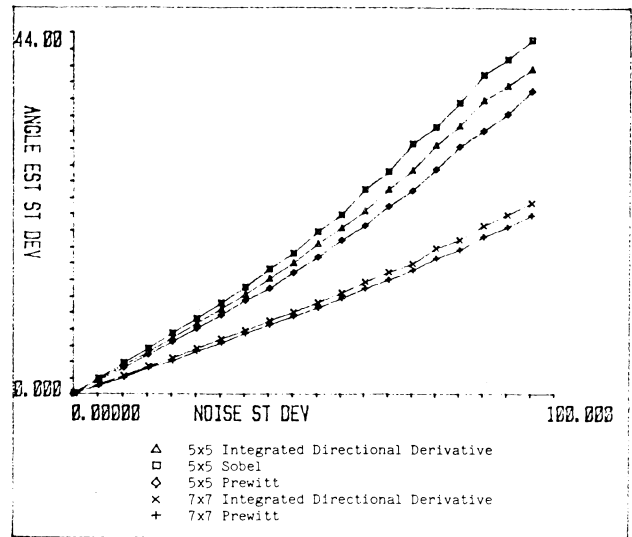

Fig. 18. Estimate standard deviation as function of noise standard deviation for ramp edge. Edge orientation is 22.5°. Edge contrast is 100.
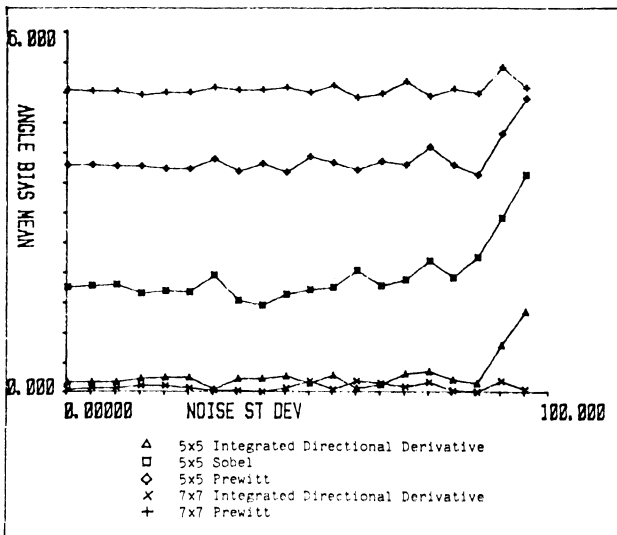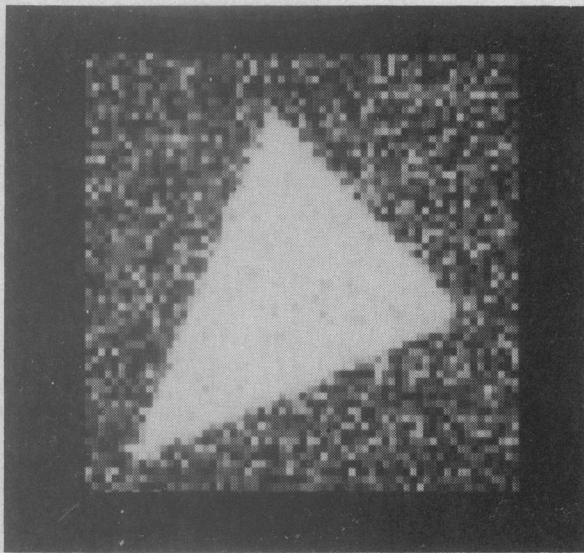

Fig. 16. Estimate bias as function of noise standard deviation for ramp edge. Edge orientation is 22.5°. Edge contrast is 100.
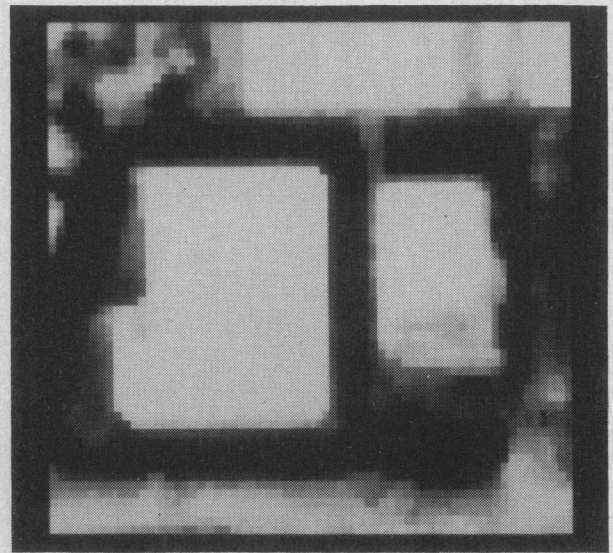
conditions. It can be seen from the plots of estimate standard deviation that, as expected, the 7×7 operators are less sensitive to noise than the 5×5 operators. The estimate standard deviations for the integrated operator and the Prewitt operator are about the same. The Sobel operator has a slightly larger estimate standard deviation.

Figs. 15–18 show estimate bias and estimate standard deviation as a function of noise standard deviation for a fixed edge orientation of 22.5°. Several things can be observed from these plots. First, estimate bias for all the operators remains nearly flat as the noise level increases up to about a standard deviation of 90. Some of the operators show an increase in estimate bias at this point. The smallest bias corresponds to the integrated operator followed by the Sobel. The Prewitt operator shows the largest bias. Second, the estimate standard deviation for all operators increases linearly with an increase in noise standard deviation. The Prewitt operator has the smallest estimate standard deviation, followed closely by the integrated operator and the Sobel operator. The 7×7 operators have a much smaller standard deviation than the 5×5 operators. For all the operators, ramp edges produce a smaller estimate bias than step edges, while step edges produce a smaller estimate standard deviation.

Finally, we show the gradient strength response of each of the 5×5 operators on the two images shown in Fig. 19. The first
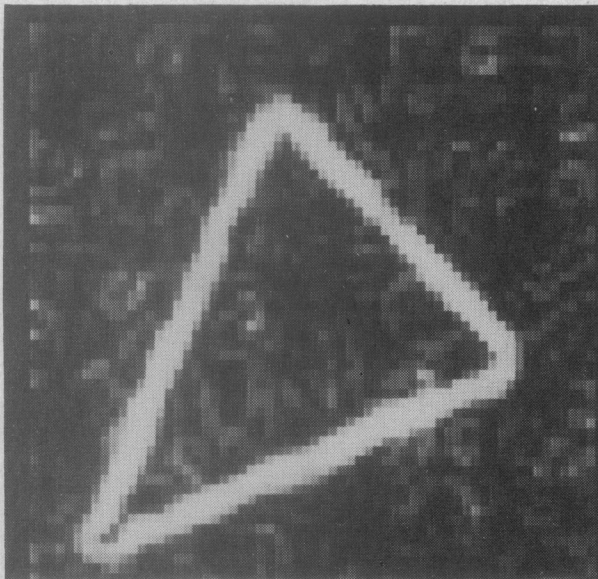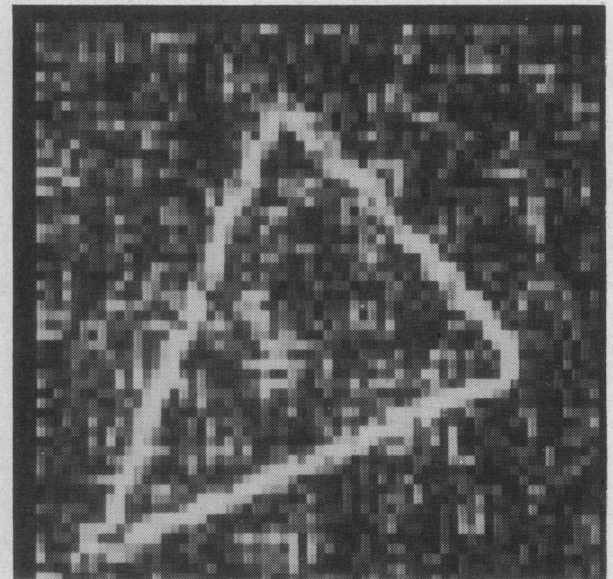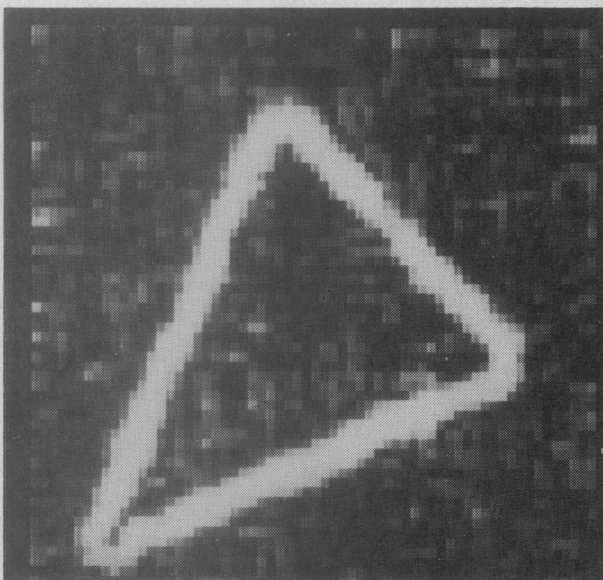
(a)

(b)

Fig. 19.   Images used to compare gradient operators. (a) Synthetic image. (b) Aerial scene.
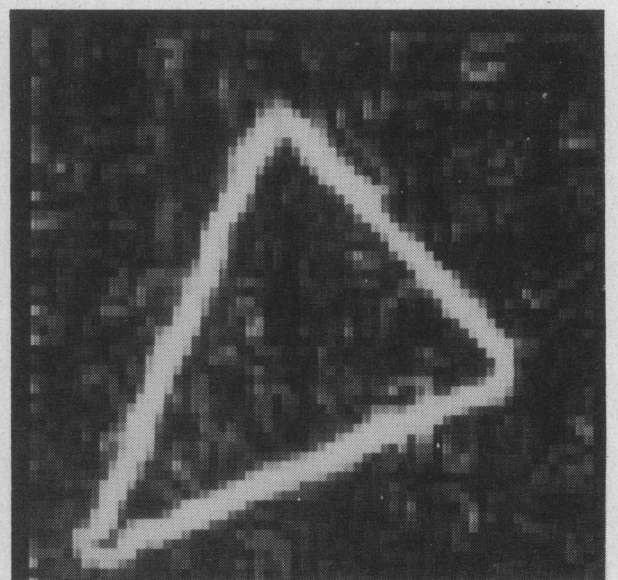

(a)

(b)

(c)

(d)

Fig. 20.   Gradient strength response for 5×5 operators on synthetic image. (a) Integrated directional derivative. (b) Standard cubic facet. (c) Prewitt. (d) Extended Sobel.
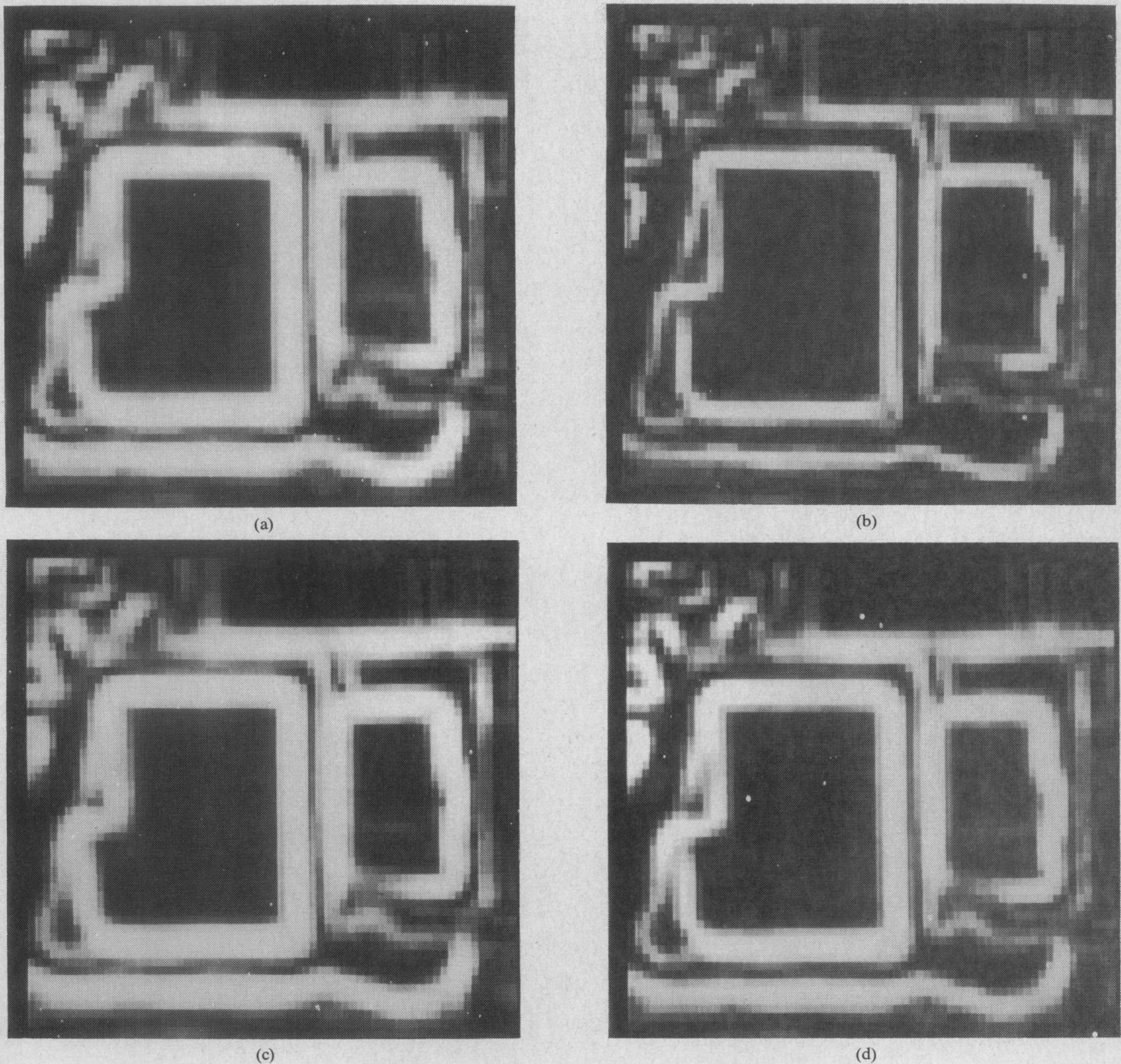
Fig. 21.   Gradient strength response for 5×5 operators on aerial scene. (a) Integrated directional derivative. (b) Standard cubic facet. (c) Prewitt. (d) Extended Sobel.

image is a 64×64 synthetic image consisting of a bright triangle on a dark background with zero-mean Gaussian noise added to it. The background gray level is 50, the object gray level is 150, and the standard deviation of the noise is 30. The second image is a 64×64 aerial scene. Figs. 20 and 21 show the gradient strength response for each of the 5×5 operators. The integrated directional derivative gradient operator and the Sobel operators yield edges with similar amounts of blur, and both operators produce a good performance in the presence of noise. The Prewitt operator has also a good performance in the presence of noise but produces a larger amount of edge blur. The standard cubic facet gradient operator yield edges with the least amount of blur but has a poor performance in the presence of noise.

V. Conclusion

A gradient operator based on an integrated directional derivative on a cubic facet has been investigated. Experimental results with step and ramp edges contaminated by zero-mean Gaussian noise show that this operator possesses the following characteristics.

1) Edge direction estimate bias is sharply reduced as compared with the bias of the standard cubic facet, Sobel, and Prewitt gradient operators. Noise sensitivity is comparable to that of the Sobel and Prewitt operators and much better than the standard cubic facet gradient operator.

2) Unlike the standard cubic facet, Sobel, and Prewitt gradient operators, increasing the neighborhood size decreases both estimate bias and noise sensitivity. For ramp edges the integrated operator is very nearly unbiased. The worst bias for the 7×7 operator is less than 0.09°, and the worst bias for the 5×5 operator is less than 0.26°. In comparison, the worst bias for the 7×7 Prewitt operator is about 5°, and the worst bias for the 5×5 Prewitt operator is about 4°.

3) Edge strength response in the presence of noise is as good as that of the Sobel operator and better than the response of the Prewitt and standard cubic facet gradient operators.

4) Computational complexity is the same as the complexity of the Sobel, Prewitt, and standard cubic facet gradient operators since it only involves the application of precomputed row and column derivative masks.

APPENDIX
THE CUBIC FACET

The cubic facet is described by the two-dimensional polynomial surface function

$$f(r,c) = K_1 + K_2 r + K_3 c + K_4 r^2 + K_5 rc$$

$$+ K_6 c^2 + K_7 r^3 + K_8 r^2 c + K_9 rc^2 + K_{10} c^3.$$

This polynomial is fit to the graytones $f(r,c)$ inside a given image neighborhood. We will assume a rectangular-shaped neighborhood whose row index set is R, whose column index set is C, and whose center is taken to be at $(0,0)$. Notice that for an even-sized neighborhood the center falls at the point where the four center pixels meet. The squared fitting error over this neighborhood is given by

$$e^2 = \sum_{r \in R} \sum_{c \in C} \big(K_1 + K_2 r + K_3 c + K_4 r^2 + K_5 rc + K_6 c^2$$

$$+ K_7 r^3 + K_8 r^2 c + K_9 rc^2 + K_{10} c^3 - f(r,c) \big)^2.$$

Taking the partial derivatives of e^2 with respect to the parameters $K_1, \cdots, K_{10}$ results in

$$\begin{bmatrix} \partial e^2/\partial K_1 \\ \partial e^2/\partial K_2 \\ \partial e^2/\partial K_3 \\ \partial e^2/\partial K_4 \\ \partial e^2/\partial K_5 \\ \partial e^2/\partial K_6 \\ \partial e^2/\partial K_7 \\ \partial e^2/\partial K_8 \\ \partial e^2/\partial K_9 \\ \partial e^2/\partial K_{10} \end{bmatrix} = \sum_{r \in R} \sum_{c \in C} 2\big(K_1 + K_2 r + K_3 c + K_4 r^2 + K_5 rc + K_6 c^2 + K_7 r^3 + K_8 r^2 c + K_9 rc^2 + K_{10} c^3 - f(r,c) \big) \begin{bmatrix} 1 \\ r \\ c \\ r^2 \\ rc \\ c^2 \\ r^3 \\ r^2 c \\ rc^2 \\ c^3 \end{bmatrix}$$

Because the sum is between symmetric limits, a considerable amount of cancellation occurs when carrying out the summation:

$$\begin{bmatrix} \partial e^2/\partial K_1 \\ \partial e^2/\partial K_2 \\ \partial e^2/\partial K_3 \\ \partial e^2/\partial K_4 \\ \partial e^2/\partial K_5 \\ \partial e^2/\partial K_6 \\ \partial e^2/\partial K_7 \\ \partial e^2/\partial K_8 \\ \partial e^2/\partial K_9 \\ \partial e^2/\partial K_{10} \end{bmatrix} = 2 \sum_{r \in R} \sum_{c \in C} \begin{bmatrix} K_4 r^2 + K_6 c^2 + K_1 - f(r,c) \\ K_7 r^4 + K_9 r^2 c^2 + K_2 r^2 - rf(r,c) \\ K_8 r^2 c^2 + K_{10} c^4 + K_3 c^2 - cf(r,c) \\ K_4 r^4 + K_6 r^2 c^2 + K_1 r^2 - r^2 f(r,c) \\ K_5 r^2 c^2 - rcf(r,c) \\ K_4 r^2 c^2 + K_6 c^4 + K_1 c^2 - c^2 f(r,c) \\ K_7 r^6 + K_9 r^4 c^2 + K_2 r^4 - r^3 f(r,c) \\ K_8 r^4 c^2 + K_{10} r^2 c^4 + K_3 r^2 c^2 - r^2 cf(r,c) \\ K_7 r^4 c^2 + K_9 r^2 c^4 + K_2 r^2 c^2 - rd^2 f(r,c) \\ K_8 r^2 c^4 + K_{10} c^6 + K_3 c^4 - c^3 f(r,c) \end{bmatrix}$$

Setting the partials to zero and solving, we immediately obtain the least-squares estimate K_5' of K_5:

$$K_5' = \frac{\displaystyle\sum_{r \in R} \sum_{c \in C} rcf(r,c)}{\displaystyle\sum_{r \in R} \sum_{c \in C} r^2 c^2}.$$

The least-squares estimates for the remaining parameters are obtained by solving the following three systems of equations:

$$\begin{bmatrix} \sum_r \sum_c 1 & \sum_r \sum_c r^2 & \sum_r \sum_c c^2 \\ \sum_r \sum_c r^2 & \sum_r \sum_c r^4 & \sum_r \sum_c r^2 c^2 \\ \sum_r \sum_c c^2 & \sum_r \sum_c r^2 c^2 & \sum_r \sum_c c^4 \end{bmatrix} \begin{bmatrix} K_1 \\ K_4 \\ K_6 \end{bmatrix} = \begin{bmatrix} \sum_r \sum_c f(r,c) \\ \sum_r \sum_c r^2 f(r,c) \\ \sum_r \sum_c c^2 f(r,c) \end{bmatrix}$$

$$\begin{bmatrix} \sum_r \sum_c r^2 & \sum_r \sum_c r^4 & \sum_r \sum_c r^2 c^2 \\ \sum_r \sum_c r^4 & \sum_r \sum_c r^6 & \sum_r \sum_c r^4 c^2 \\ \sum_r \sum_c r^2 c^2 & \sum_r \sum_c r^4 c^2 & \sum_r \sum_c r^2 c^4 \end{bmatrix} \begin{bmatrix} K_2 \\ K_7 \\ K_9 \end{bmatrix} = \begin{bmatrix} \sum_r \sum_c rf(r,c) \\ \sum_r \sum_c r^3 f(r,c) \\ \sum_r \sum_c rd^2 f(r,c) \end{bmatrix}$$

$$\begin{bmatrix} \sum_r \sum_c c^2 & \sum_r \sum_c r^2 c^2 & \sum_r \sum_c c^4 \\ \sum_r \sum_c r^2 c^2 & \sum_r \sum_c r^4 c^2 & \sum_r \sum_c r^2 c^4 \\ \sum_r \sum_c c^4 & \sum_r \sum_c r^2 c^4 & \sum_r \sum_c c^6 \end{bmatrix} \begin{bmatrix} K_3 \\ K_8 \\ K_{10} \end{bmatrix} = \begin{bmatrix} \sum_r \sum_c cf(r,c) \\ \sum_r \sum_c r^2 cf(r,c) \\ \sum_r \sum_c c^3 f(r,c) \end{bmatrix}$$

For $n = 0, 1, 2,$ or 3, let R_n and C_n be defined as

$$R_n = \sum_{r \in R} r^{2n} \qquad C_n = \sum_{c \in C} c^{2n}.$$

Furthermore, let

$$G = R_0 R_2 C_0 C_2 - R_1^2 C_1^2$$

$$A = R_1 R_3 C_0 C_2 - R_2^2 C_1^2$$

$$B = R_0 R_2 C_1 C_3 - R_1^2 C_2^2$$

$$Q = C_0 \left(R_0 R_2 - R_1^2 \right)$$

$$T = R_0 \left(C_0 C_2 - C_1^2 \right)$$

$$U = C_0 \left(R_1 R_3 - R_2^2 \right)$$

$$V = C_1 \left(R_0 R_2 - R_1^2 \right)$$

$$W = R_1 \left(C_0 C_2 - C_1^2 \right)$$

$$Z = R_0 \left(C_1 C_3 - C_2^2 \right).$$

The solution of the foregoing systems of equations is then given by

$$K_1' = \frac{1}{QT} \sum_r \sum_c \left(G - TR_1 r^2 - QC_1 c^2 \right) f(r, c)$$

$$K_2' = \frac{1}{UW} \sum_r \sum_c \left(A - WR_2 r^2 - UC_1 c^2 \right) rf(r, c)$$

$$K_3' = \frac{1}{VZ} \sum_r \sum_c \left(B - ZR_1 r^2 - VC_2 c^2 \right) cf(r, c)$$

$$K_4' = \frac{1}{Q} \sum_r \sum_c \left(R_0 r^2 - R_1 \right) f(r, c)$$

$$K_6' = \frac{1}{T} \sum_r \sum_c \left(C_0 c^2 - C_1 \right) f(r, c)$$

$$K_7' = \frac{1}{U} \sum_r \sum_c \left(R_1 r^2 - R_2 \right) rf(r, c)$$

$$K_8' = \frac{1}{V} \sum_r \sum_c \left(R_0 r^2 - R_1 \right) cf(r, c)$$

$$K_9' = \frac{1}{W} \sum_r \sum_c \left(C_0 c^2 - C_1 \right) rf(r, c)$$

$$K_{10}' = \frac{1}{Z} \sum_r \sum_c \left(C_1 c^2 - C_2 \right) cf(r, c).$$

REFERENCES

[1] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, pp. 11–122, 1981.

[2] M. Brooks, "Rationalizing edge detectors," *Comput. Graphics Image Processing*, vol. 8, pp. 277–285, 1978.

[3] R. Duda and P. Hart, "Use of Hough transformation to detect lines and curves in pictures," *Commun. Ass. Comput. Mach.*, vol. 15, p. 11, 1972.

[4] ——, *Pattern Classification and Scene Analysis*. New York: Wiley, 1973.

[5] R. M. Haralick, "Edge and region analysis for digital image data," *Comput. Graphics Image Processing*, vol. 12, pp. 60–73, 1980.

[6] ——, "Digital step edges from zero-crossing of second directional derivatives." *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-6, pp. 58–68, Jan. 1984.

[7] R. M. Haralick, L. T. Watson, and T. Laffey, "The topographical primal sketch," to be published, 1983.

[8] M. Hueckel, "An operator which locates edges in digital pictures," *J. Ass. Comput. Mach.*, vol. 18, pp. 113–125, 1971.

[9] A. Iannino and S. D. Shapiro, "An iterative generalization of the Sobel edge detection operator," in *Proc. IEEE Pattern Recognition and Image Processing Conf.*, 1979, pp. 130–137.

[10] C. Kimme, D. Ballard, and J. Sklansky, "Finding circles by an array of accumulators," *Commun. Ass. Comput. Mach.*, vol. 18, no. 2, pp. 120–122, 1975.

[11] L. Kitchen and A. Rosenfeld, "Gray level corner detection," Computer Science Center, Univ. Maryland, College Park, MD 20742, Tech. Rep. 887, Apr. 1980.

[12] A. Martelli, "Edge detection using heuristic search methods," *Comput. Graphics Image Processing*, vol. 1, pp. 169–182, 1972.

[13] F. O'Gorman and M. B. Clowes, "Finding picture edges through collinearity of feature points," *IEEE Trans. Comput.*, vol. C-25, pp. 449–456, 1976.

[14] J. M. S. Prewitt, "Object enhancement and extraction," in *Picture Processing and Psychopictorics*, B. S. Lipkin and A. Rosenfeld, Eds. New York: Academic, 1970, pp. 75–149.

[15] E. Ramer, "Transformation of photographic images into stroke arrays," *IEEE Trans. Circuits Syst.*, vol. CAS-22, pp. 363–373, 1975.

[16] L. G. Roberts, "Machine perception of three dimensional solids," in *Optical and Electro-optical Information Processing*, J. T. Tippet et al., Eds. Cambridge, MA: MIT Press, 1965, pp. 159–197.

[17] O. A. Zuniga and R. M. Haralick, "Corner detection using the facet model," in *Proc. IEEE Computer Vision and Pattern Recognition Conf.*, 1983, pp. 30–37.

Entropy and Correlation: Some Comments

TARALD O. KVÅLSETH, MEMBER, IEEE

Abstract—For measuring the degree of association or correlation between two nominal variables, a measure based on informational entropy is presented as being preferable to that proposed recently by Horibe [1]. Asymptotic developments are also presented that may be used for making approximate statistical inferences about the population measure when the sample size is reasonably large. The use of this methodology is illustrated using a numerical example.

INTRODUCTION

In a recent article in this TRANSACTIONS, Horibe [1] proposed a measure of association (correlation) between two nominal variables based on information-theoretic metrics. The purpose of this correspondence is essentially threefold. First, this author wants to take issue with Horibe's measure, specifically because of improper norming used to rescale the measure to the [0, 1] interval. Second, a more reasonable informational measure will be identified which is also symmetric and is simply the weighted average of asymmetric measures for the two nominal variables. Third, asymptotic (large-sample) methods of statistical inference will be outlined since, in addition to obtaining point estimates for the association measure, it is generally of interest to be able to construct confidence intervals and test hypotheses about the population measure. The results will be exemplified using a numerical example.

MEASURE FORMULATION

To define the notation to be used, consider that p_{ij} is the population probability of that cell in the $I \times J$ contingency (cross-classification) table which corresponds to category i of the row variable X and category j of the column variable Y. Furthermore, let p_{i+} and p_{+j} denote the marginal totals for row i and column j, respectively. The information conveyed about X by Y, $I(X; Y)$, and vice versa, $I(Y; X)$, also called transinformation, is given by the well-known formula

$$I = I(X; Y) = H(X) - H(X|Y)$$

$$= I(Y; X) = H(Y) - H(Y|X) \quad (1)$$