Pattern Recognition, Vol. 21, No. 5, pp. 493–503, 1988.
Printed in Great Britain.

0031-3203/88 $3.00 + .00
Pergamon Press plc
Pattern Recognition Society

GRADIENT THRESHOLD SELECTION USING THE FACET MODEL

OSCAR A. ZUNIGA
Department of Electrical Engineering, VPI & SU, Blacksburg, VA 24061, U.S.A.

and

ROBERT M. HARALICK*
Department of Electrical Engineering, University of Washington, Seattle, WA 98195, U.S.A.

(*Received* 8 *November* 1985; *in revised form* 2 *December* 1987)

Abstract—Automatic gradient threshold selection for edge detection is a non-trivial task due to the presence of image noise. This problem is posed within a statistical framework based on a cubic facet model for the image data and a Gaussian model for the noise. Under these assumptions, two statistics which are functions of the gradient strength and facet residual error are derived. Experiments show that thresholds on these statistics produce results which are superior to those obtained by the best subjective threshold on the gradient image. A Bayes decision procedure is developed which makes threshold selection automatic.

Image analysis   Facet model   Gradient image   Threshold selection
Bayesian decision   Edge detection

1. INTRODUCTION

The computation of the gradient of a digital image is usually one step of image processing tasks such as edge or corner detection. Typically, edge candidates are taken from pixels whose gradient strength is significantly different from zero. One of the most effective ways of detecting edges and corners in digital image data is by finding zero-crossings of second directional derivative of the underlying graytone intensity surface.[3,4,7] These zero-crossings have associated with them non-zero gradient strengths. The presence of noise in real image data causes the occurrence of spurious zero-crossings. We are faced then with the problem of distinguishing between zero-crossings which arise due to the presence of a true edge or corner and zero-crossings which arise due to the presence of noise. In practice, this problem is handled by considering only zero-crossings whose gradient strength is significantly different from zero.

Our approach to this problem is based on the facet model for digital images.[2,5] The basic philosophy of this model derives from recognizing that the discrete set of values which form the digital image are the result of sampling and quantizing a real-valued function, f, defined on the domain of the image which is a bounded and connected subset of the real plane. Thus, any property associated with a pixel or a neighborhood of pixel values should be evaluated by relating it to the property of the corresponding gray tone surface f which underlies the neighborhood. This involves estimating the surface function f locally, from the neighborhood samples available to us. The most natural way of accomplishing this is by assuming a parametric form for f and then estimating its associated parameters.

In this paper we are concerned with that property of a pixel called gradient strength which is defined as the Euclidean norm of the first partial derivatives of the graytone intensity surface evaluated at the pixel position. An assumption about the nature of the noise enables us to put the problem of choosing a suitable gradient threshold in a statistical framework. We assume the noise to be Gaussian with zero mean and variance σ^2. We derive two statistics which are functions of the gradient strength and the facet residual error of fit. We show that:

(1) Thresholds on the statistics derived in this paper produce results which are superior to those obtained by the best subjective threshold on the gradient image.

(2) Threshold selection can be made automatic by applying a Bayes decision method.

The analysis presented in the following sections are valid for a facet model of arbitrary order although the experiments were performed with a cubic facet model. Section 2 shows how the statistics are derived and

* To whom correspondence should be addressed.

poses the problem of selecting a suitable threshold as a hypothesis test problem. Section 3 shows how the thresholding can be made automatic by posing the problem as a Bayesian decision problem. Section 4 presents the experimental results. The appendix provides the mathematical analysis required to derive the statistical distributions of the Euclidean norm of any subset of facet parameters or partial derivatives, and the distribution of the total facet residual error.

2. GRADIENT THRESHOLD SELECTION AS A HYPOTHESIS TEST

Let μ_r and μ_c denote the true but unknown values of the first row and column partial derivatives of the underlying graytone intensity surface at a pixel position. Let $\hat{\mu}_r$ and $\hat{\mu}_c$ denote their estimates based upon a neighborhood of K values. According to the results of the appendix $\hat{\mu}_r$ and $\hat{\mu}_c$ are normally distributed and:

$$\begin{aligned} E[\hat{\mu}_r] &= \mu_r \\ E[\hat{\mu}_c] &= \mu_c \\ V[\hat{\mu}_r] &= \sigma^2 k \\ V[\hat{\mu}_c] &= \sigma^2 k \\ E[\hat{\mu}_r \hat{\mu}_c] &= \mu_r \mu_c \end{aligned} \tag{1}$$

where k is a constant whose value depends on the neighborhood size and the basis functions used to estimate the graytone intensity surface.

Consider testing the hypothesis that $\mu_r = \mu_c = 0$ (zero gradient). This hypothesis must be rejected if there is to be a zero-crossing of second directional derivative. Under this hypothesis:

$$\frac{\hat{\mu}_r^2 + \hat{\mu}_c^2}{k\sigma^2}$$

has a χ_2^2 distribution.

Also, from the results of the appendix, the total residual error S^2 normalized by the noise variance S^2/σ^2 has a χ_{K-N}^2 distribution, where N is the number of basis functions or the number of facet parameters. Hence,

$$x_1 = \frac{(\hat{\mu}_r^2 + \hat{\mu}_c^2)/2}{kS^2/(K-N)} \tag{2}$$

has a $F_{2,K-N}$ distribution and the hypothesis for zero gradient would be rejected for suitable large values of x_1. The value of the threshold for x_1 is chosen to correspond to a given significance level of test.

Notice that the statistic defined by equation (2) may be regarded as a significance or reliability measure associated with the existence of a non-zero gradient. It is essentially proportional to the square gradient normalized by $S^2/(K-N)$ which is a random variable whose expected value is σ^2, the variance of the noise. This scaling of the gradient by a local estimate of the image noise makes optimum selection possible by a fixed threshold procedure.

If the noise variance is known to be constant everywhere throughout the image domain a better estimate for it is possible by averaging the total residual errors S^2 over M non-overlapping neighborhoods. Let this average be denoted by E^2, then ME^2/σ^2 has a $\chi_{M(K-N)}^2$ distribution and

$$x_2 = \frac{(\hat{\mu}_r^2 + \hat{\mu}_c^2)/2}{kE^2/(K-N)} \tag{3}$$

has a $F_{2,M(K-N)}$ distribution and the hypothesis for zero gradient would be rejected for suitable large values of x_2. Again, the threshold on x_2 is chosen to correspond to a given significance level of test.

3. GRADIENT THRESHOLD SELECTION AS A BAYESIAN DECISION PROBLEM

This method derives from considering any image point to be in one of two states: zero gradient strength or non-zero gradient strength. We denote these two states by z and nz, respectively. Let x be a continuous random variable test statistic whose distribution depends on the pixels' state. We define x as:

$$x = \frac{(\hat{\mu}_r^2 + \hat{\mu}_c^2)/2}{kS^2/(K-N)} \tag{4}$$

As previously seen, for pixels with zero gradient strength x has a $F_{2,K-N}$ distribution.

A simple Bayes decision rule with a unity gain function is:

$$\text{Decide} \begin{cases} \textit{zero gradient if } P(z \mid x) > P(nz \mid x) \\ \textit{non-zero gradient, otherwise} \end{cases} \tag{5}$$

and the probability of error associated with this decision is:

$$P(error \mid x) = \begin{cases} P(nz \mid x) & \text{if} \quad P(z \mid x) > P(nz \mid x) \\ P(z \mid x), & otherwise \end{cases}$$

or equivalently

$$P(error \mid x) = \min\{P(z \mid x), P(nz \mid x)\} \tag{6}$$

We can express the decision rule and probability of error in terms of the distributions of x by using Bayes rule.

$$P(z \mid x) = \frac{P(x \mid z)P(z)}{P(x)}$$

$$P(nz \mid x) = \frac{P(x \mid nz)P(nz)}{P(x)}$$

The decision rule then becomes

$$\text{Decide} \begin{cases} \textit{zero gradient if } P(x \mid z)P(z) > P(x \mid nz)P(nz) \\ \textit{non-zero gradient, otherwise} \end{cases} \tag{7}$$

and the probability of error becomes

$$P(error \mid x) = \frac{\min\{P(x \mid z)P(z), P(x \mid nz)P(nz)\}}{P(x)}. \tag{8}$$

The density function $P(x\,|\,z)$ is known to be $F_{2,K-N}$, $P(x)$ is the mixture distribution and can be estimated from a histogram of the gradient image. Then $P(x\,|\,nz)$ can be obtained as follows:

$$P(x) = P(x, z) + P(x, nz) = P(x\,|\,z)\,P(z) + P(x\,|\,nz)\,P(nz) \qquad (9)$$

and from here

$$P(x\,|\,nz) = \frac{[P(x) - P(x\,|\,z)\,P(z)]}{P(nz)}$$

$$P(x\,|\,nz) = \frac{[P(x) - P(x\,|\,z)\,P(z)]}{[1 - P(z)]}. \qquad (10)$$

The total probability of error is given by:

$$P(error) = \sum_x P(error\,|\,x)\,P(x).$$

Using relations (8) and (10):

$$P(error) = \sum_x \min\{P(x\,|\,z)\,P(z),\, P(x\,|\,nz)\,P(nz)\}$$

$$P(error) = \sum_x \min\{P(x\,|\,z)\,P(z),\, P(x) - P(x\,|\,z)\,P(z)\} \qquad (11)$$

Since the prior probability $P(z)$ of zero gradient is not known it must be user specified. For many images values of 0.9 to 0.95 are reasonable. Another method of choosing $P(z)$ is obtained by observing from equation (9) that:

$$\text{for all } x \quad P(x) \geq P(x\,|\,z)\,P(z).$$

Then

$$\text{for all } x, \quad P(x\,|\,z) > 0, \quad P(z) \leq \frac{P(x)}{P(x\,|\,z)}.$$

A suitable value for $P(z)$ is therefore:

$$P(z) = \min_{x, P(x|z)>0} \frac{P(x)}{P(x\,|\,z)}. \qquad (12)$$

Once $P(z)$ has been specified, a suitable threshold x_{th} can be chosen from (7) by solving

$$P(x_{th}\,|\,z)\,P(z) = P(x_{th}\,|\,nz)\,P(nz) \qquad (13)$$

We assume that only one intersection point x_{th} exists. The same analysis can be carried out with the statistic defined by equation (3) if the noise variance is known to be constant throughout the image domain.

4. EXPERIMENTAL RESULTS

Test images

Two artificially generated images TEST1 and TEST2 were tested. They are shown in Fig. 1. Their size is 100 × 100 pixels and they consist of four bright circles on a dark background. The diameter of the circles is 25. The edge contrasts for each of the four circles are 9, 18, 27 and 50. The edges are ramp edges whose width is 5 pixels. The noise is additive Gaussian with zero mean. The standard deviation of the noise for TEST1 is 5. The standard deviation of the noise for TEST2 changes from circle to circle and their values are 5, 10, 15, and 28 which keeps the signal to noise ratio constant at about 20 log (9/5) or approximately 5 DB.

The first partial derivatives of the graytone intensity function at every pixel position were computed by fitting a cubic polynomial surface defined in the row and column coordinates of a 9 × 9 neighborhood centered about the pixel. Figure 2 shows the gradient images computed from the Euclidean norm of the first partial derivatives and Fig. 3 shows the result of applying a constant threshold to the gradient images. The thresholds for each image were found interactively with subjective quality as the selection criteria. Notice

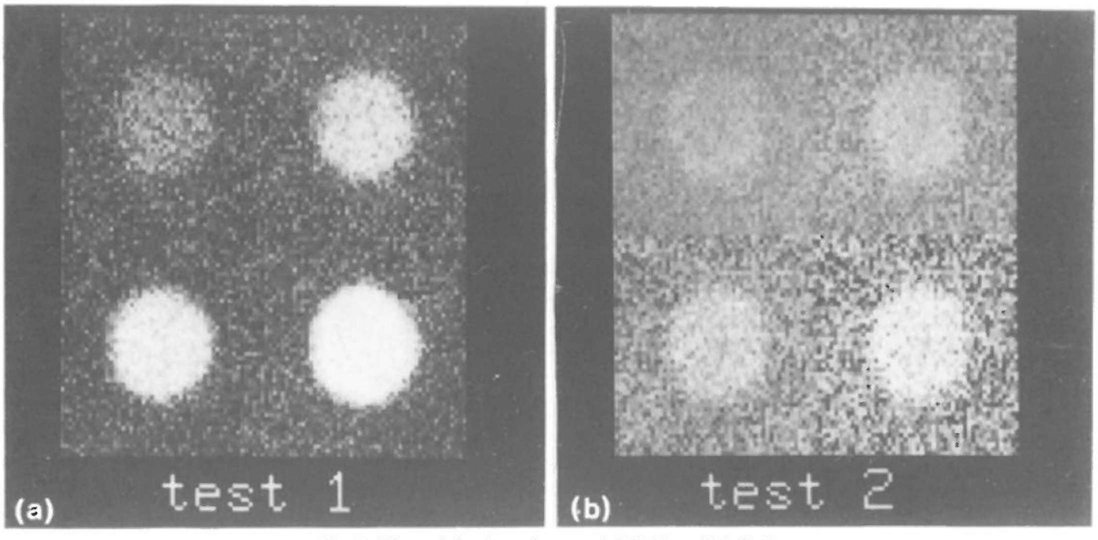

Fig. 1. The original test images TEST1 and TEST2.

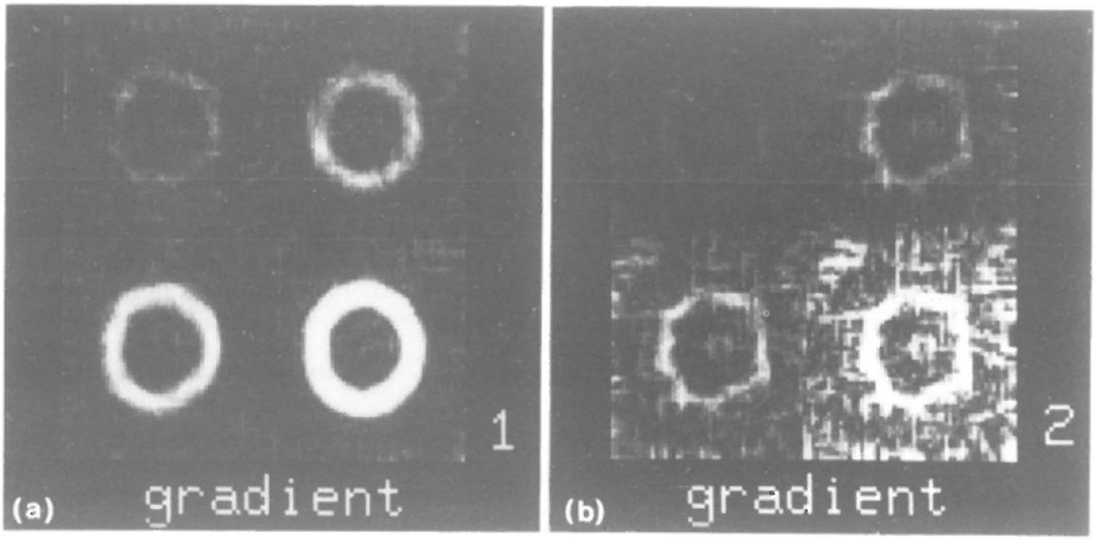

Fig. 2. The gradient images computed from a 9 × 9 cubic facet applied to the test images. (a) TEST1, (b) TEST2.

the inability of this procedure to perform well everywhere in TEST2 due to the changing nature of the noise variance.

The $F_{2,M(K-N)}$ *test statistic*

In this case the test statistic defined by equation (3) is used for both the hypothesis test method and the Bayesian decision method of automatic gradient threshold selection. This statistic was derived under the assumption that the variance of the noise remains constant throughout the image.

Figure 4 shows images of the test statistic x and Fig. 5 shows the results of a 1% and 5% significant level hypothesis test. Figure 6 shows the histograms of the mixture distribution $P(x)$, and the conditional probability distributions given zero gradient $P(x \mid z)$ and non-zero gradient $P(x \mid nz)$. The prior probability of non-zero gradient $P(nz)$ was computed automatically using equation 12. These prior probabilities were found to be 0.22 for image TEST1 and 0.45 for TEST2. We can compare these values with the true prior probability of non-zero gradient computed as $4(3.14)D(W)/10{,}000$ where D is the diameter of the circles and W the width of the non-zero gradient region. This yields a value of 0.157 for $P(nz)$. Notice that since TEST2 does not meet our assumption of constant noise variance, the error in its estimate of $P(nz)$ is much bigger than that in TEST1. The values of the computed threshold for the test statistic x were found to be 3.94 for TEST1 and 3.4 for TEST2. Finally, Fig. 7 shows the images of the threshold of the test statistics.

The $F_{2,K-N}$ *test statistic*

In this case the test statistic defined by equation (2) is used for both the hypothesis test method and the Bayesian decision method of automatic gradient threshold selection. This statistic is valid under both

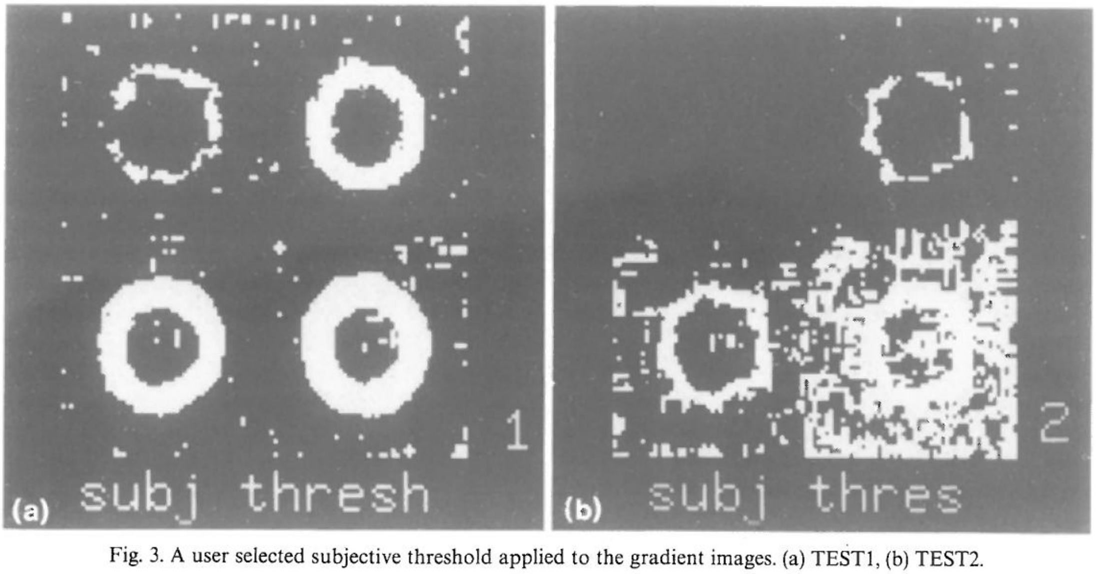

Fig. 3. A user selected subjective threshold applied to the gradient images. (a) TEST1, (b) TEST2.

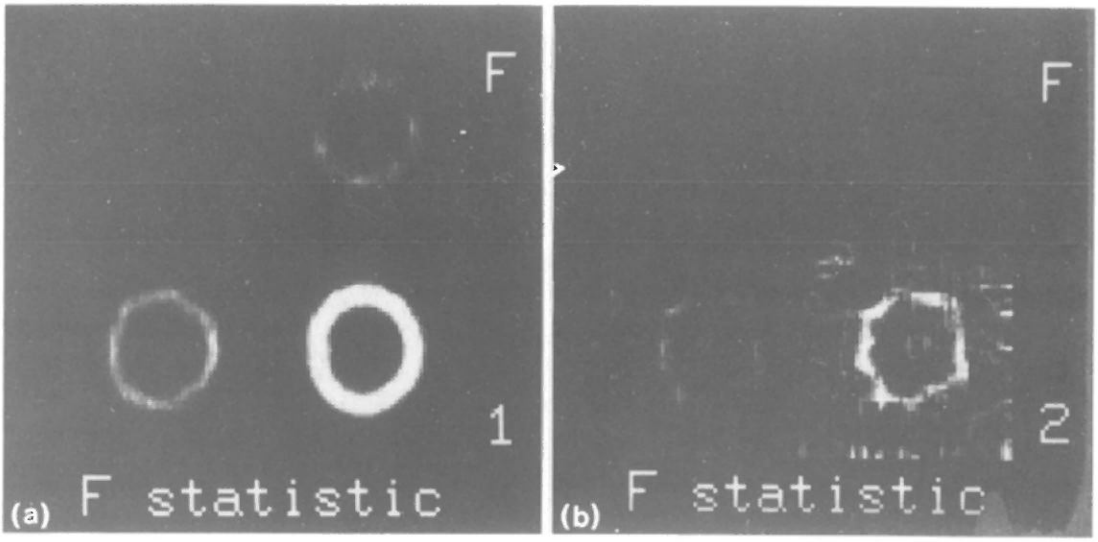

Fig. 4. The values of the test statistic $F_{2,M(K-N)}$ for each of the test images. (a) TEST1, (b) TEST2.

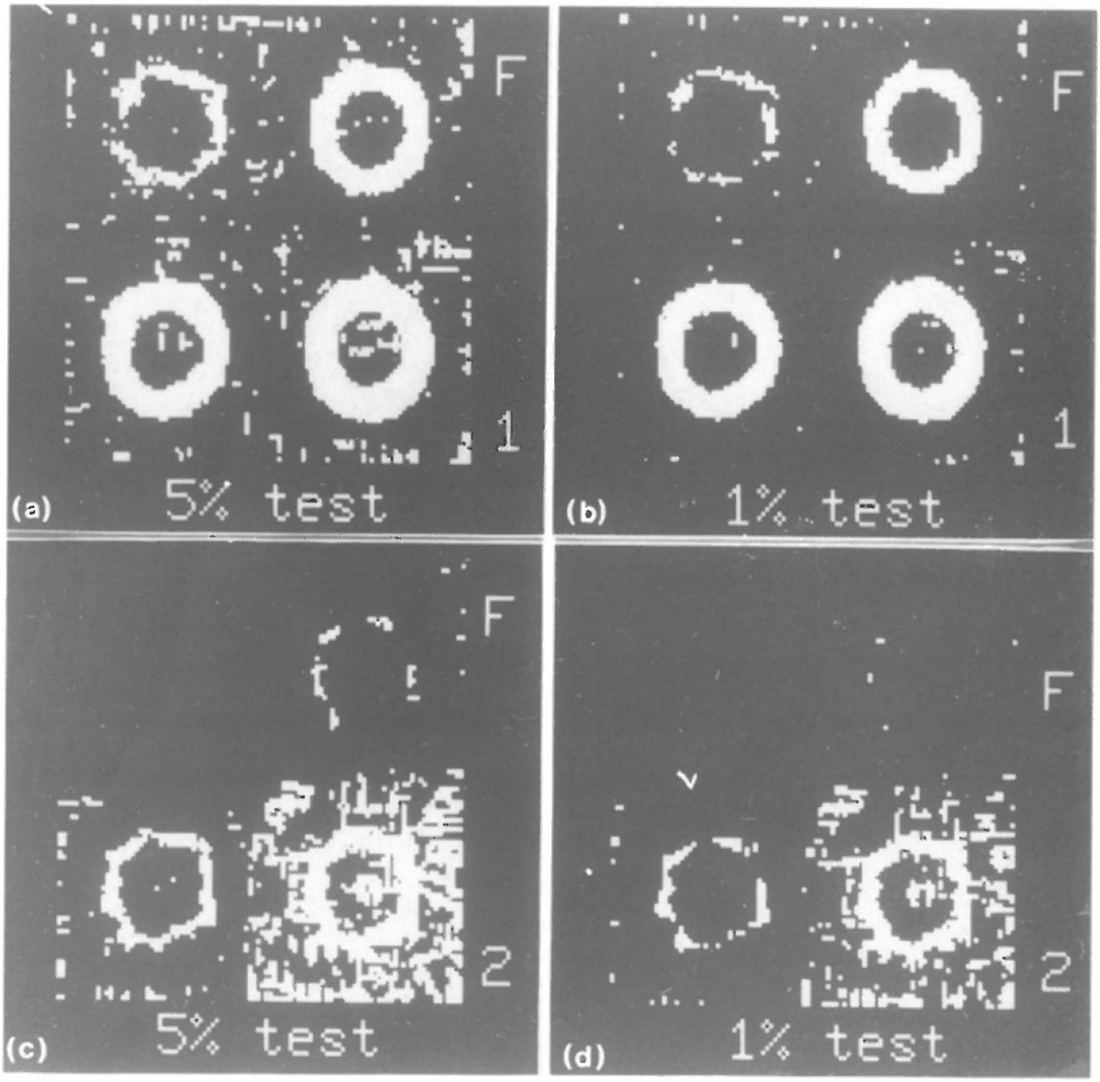

Fig. 5. The results of a 1 and 5% significant level test on the test images. (a) TEST1, 5%, (b) TEST1, 1%, (c) TEST2, 5%, (d) TEST2, 1%. Statistic used is $F_{2,M(K-N)}$.

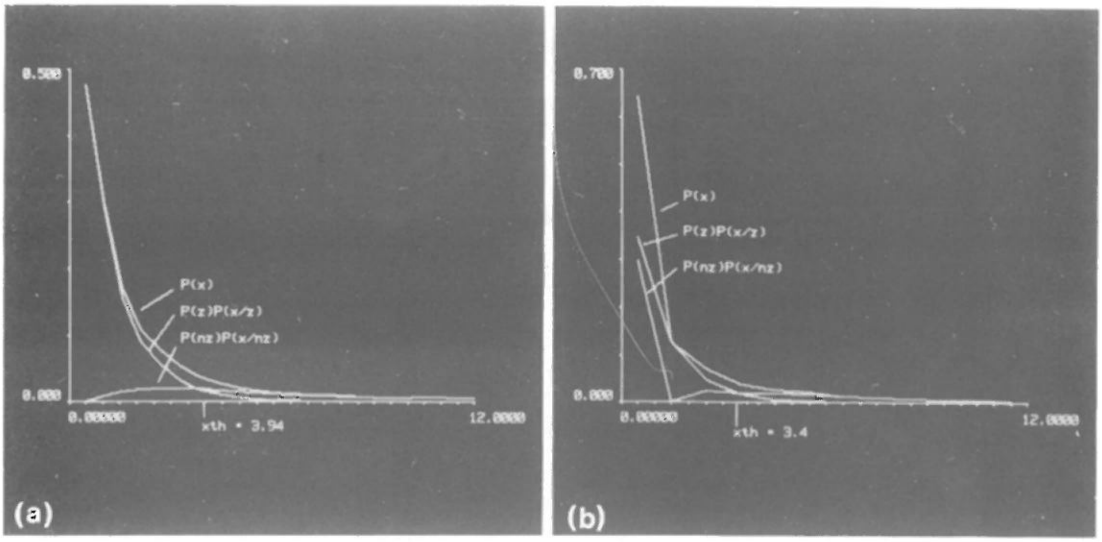

Fig. 6. The histograms $P(x)$, $P(x/z)$, and $P(x/nz)$ for the $F_{2,M(K-N)}$ statistic. (a) TEST1, (b) TEST2.

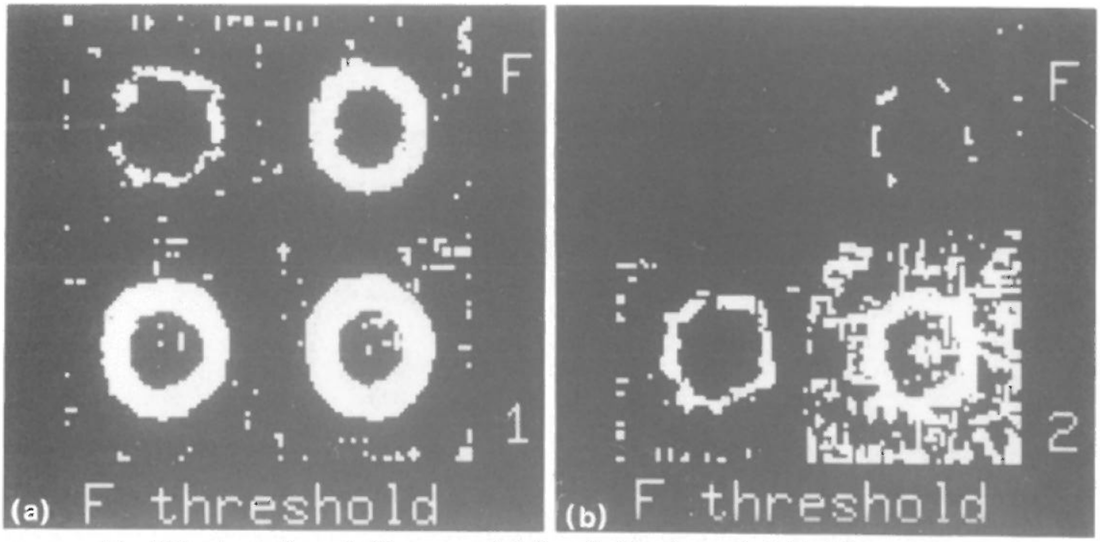

Fig. 7. The thresholds on the $F_{2,M(K-N)}$ statistic from the Bayesian method. (a) TEST1, (b) TEST2.

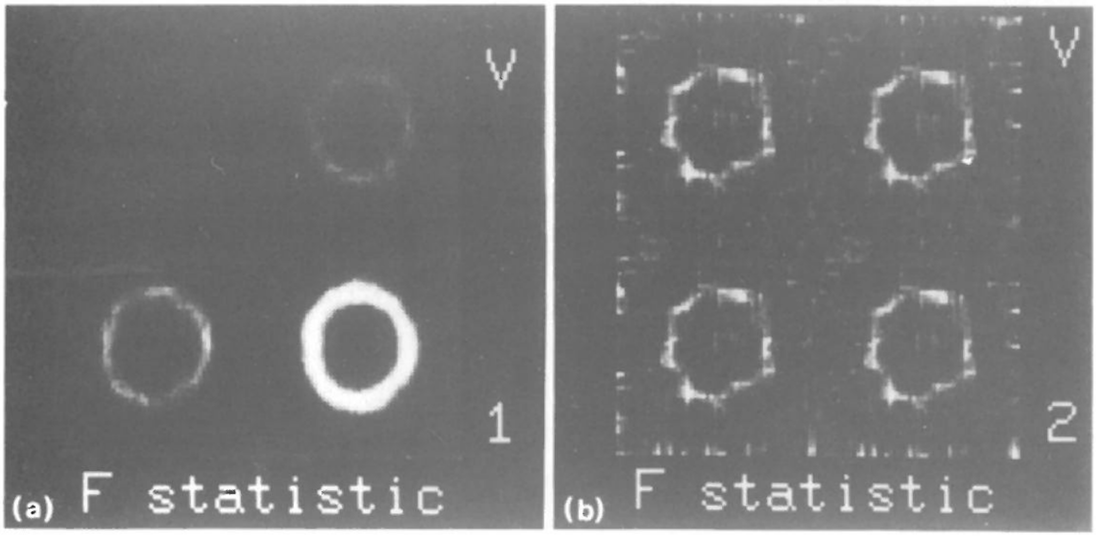

Fig. 8. The test statistics $F_{2,K-N}$. (a) TEST1, (b) TEST2.

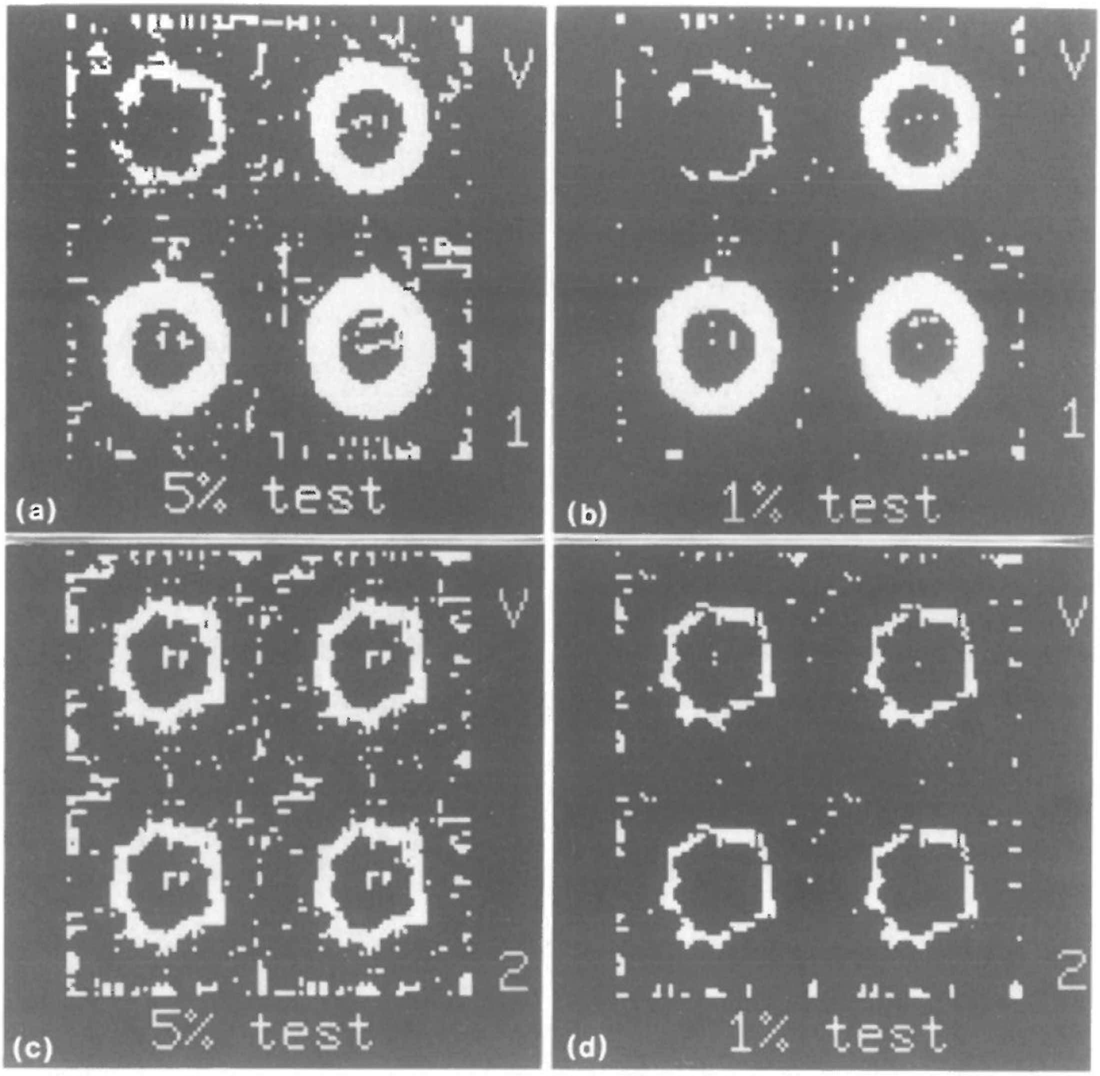

Fig. 9. The results of a 5% and 1% significant level hypothesis test on the $F_{2,K-N}$ statistics. (a) TEST1, 5%, (b) TEST1, 1%, (c) TEST2, 5%, (d) TEST2, 1%.

constant or changing noise variance conditions.

Figure 8 shows images of the test statistic x and Fig. 9 shows the results of a 1% and 5% significance level hypothesis test. Figure 10 shows the histograms of the mixture distribution $P(x)$ and the conditional probability distributions given zero gradient $P(x|z)$ and non-zero gradient $P(x|nz)$. The prior probability of non-zero gradient $P(nz)$ was comuted automatically using equation 12. These prior probabilities were found to be 0.24 for image TEST1 and 0.23 for TEST2. Notice that in this case since TEST2 does not meet the assumption on the nature of the noise variance its estimate of $P(nz)$ is in closer agreement with the true value of $P(nz)$. The values of the computed threshold for the test statistic x were found to be 3.93 for TEST1 and 3.21 for TEST2. Finally, Fig. 11 shows the images of the threshold of the test statistics.

Comparison with other methods

As a comparison with other methods of thresholding the gradient image we used the running mean method. Each pixel in the gradient image was thresholded in proportion to the mean of the gradient values on a local 20 × 20 neighborhood around the given pixel. The results are shown in Fig. 12 for a variable threshold of 1.5 times the running mean. The value 1.5 was found interactively for best subjective results. Notice that this method performs as well as the methods that use the $F_{2,K-N}$ statistic, although user interaction is required to select the best value for the constant of proportionality (1.5 for the test images) as opposed to the Bayesian decision method which requires no interaction.

5. CONCLUSIONS

Several conclusions can be drawn from the experimental results of the previous section:

(1) Thresholds on the $F_{2,K-N}$ and $F_{2,M(K-N)}$ test statistics produce similar results to that obtained by a

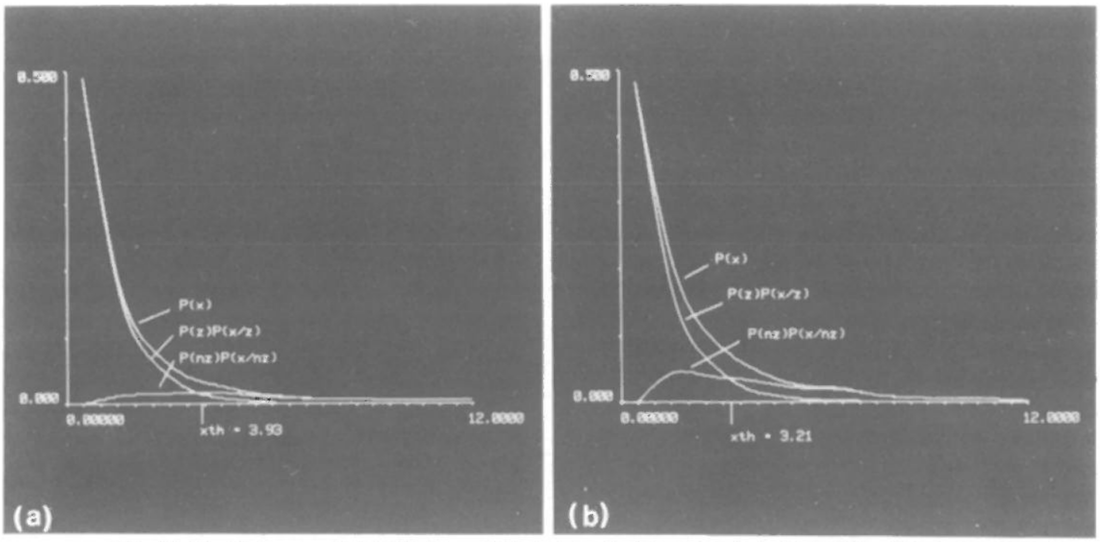

Fig. 10. The histograms $P(x)$, $P(x/z)$, and $P(x/nz)$ for the $F_{2,K-N}$ statistic. (a) TEST1, (b) TEST2.

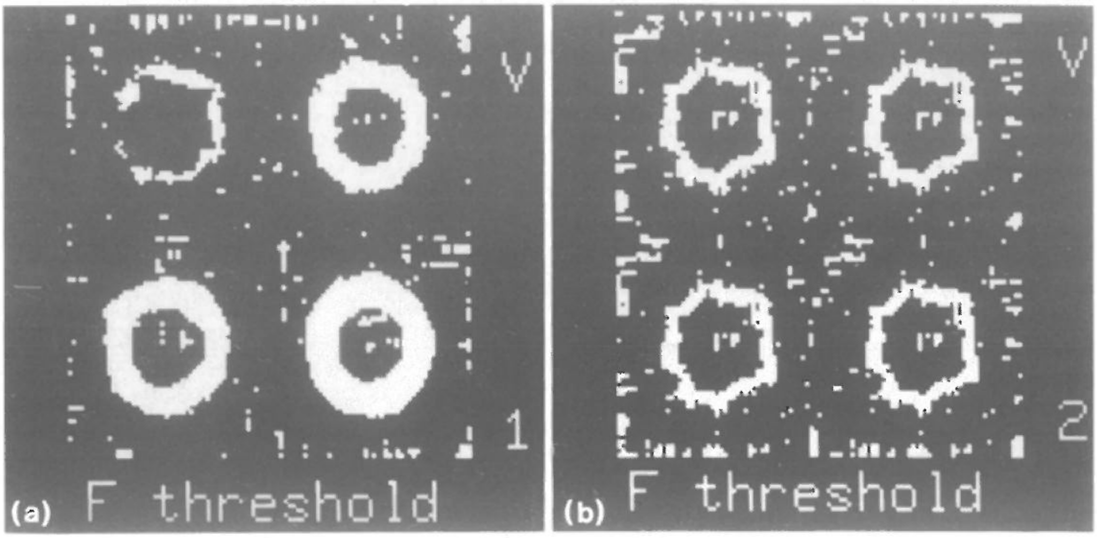

Fig. 11. The thresholds on the $F_{2,K-N}$ statistics from the Bayesian method. (a) TEST1, (b) TEST2.

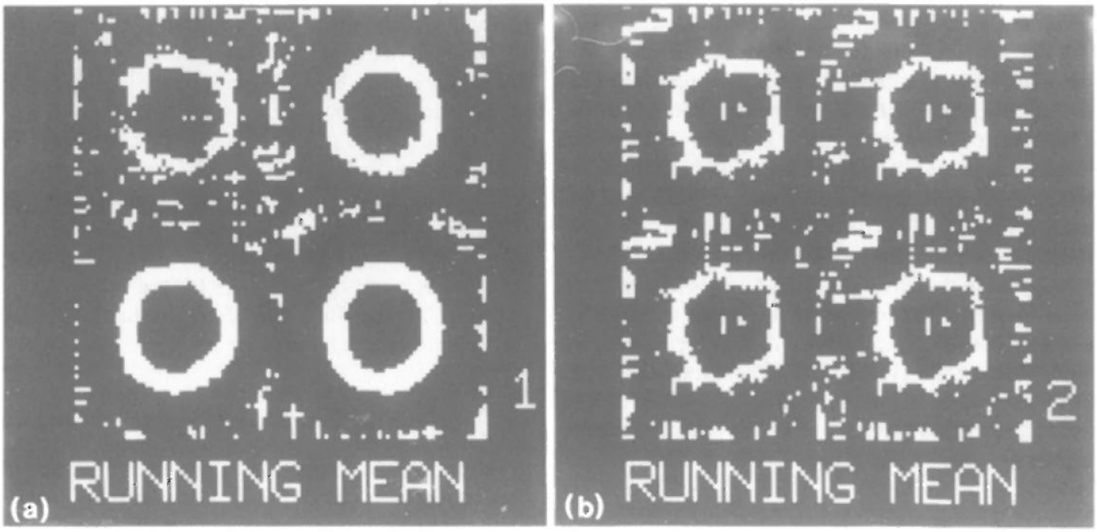

Fig. 12. The running mean threshold applied to the gradient images. (a) TEST1, (b) TEST2.

best subjective user interactive threshold on the gradient image, when the noise variance remains constant throughout the image.

(2) A threshold on the $F_{2,K-N}$ statistic produce a clearly superior result to either a threshold on the $F_{2,M(K-N)}$ statistic or to the best subjective user interactive threshold on the gradient image, when the noise variance changes from region to region in the image.

(3) The Hypothesis test method of gradient threshold selection is inferior to the Bayesian decision method in the sense that the optimal value of significance level of the test cannot be known in advance and thus user interaction is required. No user interaction is required by the Bayesian decision method, except the choice of neighborhood size which is required in all the methods tested.

These conclusions are valid under the controlled nature of our experiments where the test images meet the image and noise model assumptions. Further research is necessary to determine the sensitivity of these methods to a departure in the model assumptions.

SUMMARY

The computation of the gradient of a digital image is usually one step of image processing tasks such as edge or corner detection. Typically edge candidates are taken from pixels whose gradient strength is significantly different from zero. One of the most effective ways of detecting edges and corners in digital image data is by finding zero-crossings of second directional derivative of the underlying graytone intensity surface. These zero-crossings have associated with them non-zero gradient strengths. The presence of noise in real image data causes the occurrence of spurious zero-crossings. We are faced then with the problem of distinguishing between zero-crossings which arise due to the presence of a true edge or corner and zero-crossings which arise due to the presence of noise. In practice this problem is handled by considering only zero-crossings whose gradient strength is significantly different from zero.

Our approach to this problem is based on the facet model for digital images. The basic philosophy of this model derives from recognizing that the discrete set of values which form the digital image are the result of sampling and quantizing a real-valued function, f, defined on the domain of the image which is a bounded and connected subset of the real plane. Thus, any property associated with a pixel or a neighborhood of pixel values should be evaluated by relating it to the property of the corresponding gray tone surface f which underlies the neighborhood. This involves estimating the surface function f locally, from the neighborhood samples available to us. The most natural way of accomplishing this is by assuming a parametric form for f and then estimating its associated parameters.

In this paper we are concerned with that property of a pixel called gradient strength which is defined as the Euclidean norm of the first partial derivatives of the graytone intensity surface evaluated at the pixel position. An assumption about the nature of the noise enables us to put the problem of choosing a suitable gradient threshold in a statistical framework. We assume the noise to be Gaussian with zero mean and variance σ^2. We derive two statistics which are functions of the gradient strength and the facet residual error of fit. We show that:

(1) Thresholds on the statistics derived in this paper produce results which are superior to those obtained by the best subjective threshold on the gradient image.

(2) Threshold selection can be made automatic by applying a Bayes decision method.

REFERENCES

1. F. A. Graybill, *Theory and Application of the Linear Model*, Duxbury Press, North Scituate, MA (1976).
2. R. M. Haralick, Edge and region analysis for digital image data, *Comput. Graph. Image Processing* **12**, 60–73 (1980).
3. R. M. Haralick, Zero-crossing of second directional derivative edge operator, *Society of Photogrammetric Instrumentation Engineering Symposium on Robot Vision*, Washington, DC (1982).
4. R. M. Haralick, Digital step edge from second directional derivatives, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAMI-6**, 58–68 (1984).
5. R. M. Haralick and L. T. Watson, A facet model for image data, *Comput. Graph. Image Processing* **15**, 113–129 (1981).
6. J. S. Weszka, A survey of threshold selection techniques, *Comput. Graph. Image Processing* **7**, 259–265 (1978).
7. O. A. Zuniga and R. M. Haralick, Corner detection using the facet model, *IEEE Computer Vision and Pattern Recognition Conference*, pp. 30–37 (1983).

APPENDIX: STATISTICAL DISTRIBUTIONS OF THE NORM OF FACET PARAMETERS, PARTIAL DERIVATIVES AND FACET ERRORS

In this appendix we derive the distributions of the norm of any subset of facet parameters or partial derivatives, and the distribution of the total facet residual error. We assume the noise to be Gaussian with zero mean and known covariance matrix Σ. In order to proceed we will make use of three theorems frequently used in multivariate statistics. Proofs of these theorems can be found in Graybill.[1]

Theorem 1

Let x be a $K \times 1$ random vector with a $N(\mu, \Sigma)$ distribution, that is the elements of x have a multivariate normal (Gaussian) distribution with mean vector μ and covariance matrix Σ. Then $(x-\mu)' \Sigma^{-1}(x-\mu)$ is a chi-squared variate $\chi^2(K)$ with K degrees of freedom.

Theorem 2

Let x be a $K \times 1$ random vector with distribution $N(\mu, \Sigma)$. Consider the m linear functions on the elements of x defined by $y = Bx$, where y is an $m \times 1$ vector, $m \leq K$ and B is an

$m \times K$ real matrix of rank m. Then y has a distribution $N(B\mu, B\Sigma B')$.

Theorem 3

Let $Q = Q_1 + \ldots + Q_{L-1} + Q_L$ where $Q, Q_1, \ldots, Q_L$ are $L + 1$ random variables that are quadratic forms in any multivariate normal variables. Let Q be a $\chi^2(r)$ variate, let Q_i be $\chi^2(r_i)$, $i = 1, \ldots, L-1$ variates, and let Q_L be non-negative. Then the random variables $Q_1, \ldots, Q_L$ are mutually independent and hence Q_l is a $\chi^2(r_L = r - r_1 - \ldots - r_{L-1})$ variate.

The model

The image data model is described by

$$F\alpha + \eta = x \tag{1}$$

where F is a $K \times N$ basis matrix, α is an $N \times 1$ parameter vector, x is a $K \times 1$ observation vector, and η is a $K \times 1$ noise vector. We assume the noise to be Gaussian with zero mean and known covariance matrix Σ, that is η has a multivariate $N(0, \Sigma)$ distribution.

The minimum variance unbiased estimate $\hat{\alpha}$ of the true parameter vector α (an estimate that minimizes $(x - \hat{x})' \Sigma^{-1}(x - \hat{x})$ is known to be:[17]

$$\hat{\alpha} = Px, \tag{2}$$

where

$$P = (F'\Sigma^{-1}F)^{-1}F'\Sigma^{-1} \tag{3}$$

Distribution of the parameter vector

Replacing equation (1) in equation (2) we obtain

$$\hat{\alpha} = P(F\alpha + \eta)$$
$$= PF\alpha + P\eta$$
$$= \alpha + P\eta,$$

therefore

$$\hat{\alpha} - \alpha = P\eta. \tag{4}$$

Since η is $N(0, \Sigma)$, applying theorem 2 we conclude that $(\hat{\alpha} - \alpha)$ is $N(0, P\Sigma P')$.

Using equation (3) $P\Sigma P'$ reduces to

$$P\Sigma P' = (F'\Sigma^{-1}F)^{-1},$$

therefore

$$(\hat{\alpha} - \alpha) \text{ is } N[0, (F'\Sigma^{-1}F)^{-1}]; \tag{5}$$

that is, the parameter vector $\hat{\alpha}$ has a multivariate normal distribution with mean α and covariance matrix $(F'\Sigma^{-1}F)^{-1}$.

Distribution of the norm of the parameter vector

Using equation (5) and applying theorem 1 it follows that

$$(\hat{\alpha} - \alpha)'(F'\Sigma^{-1}F)(\hat{\alpha} - \alpha) \text{ is } \chi^2(N), \tag{6}$$

a chi-squared variate with N degrees of freedom.

This result also applies to the norm of any subset of elements of $(\hat{\alpha} - \alpha)$, as follows. Let α_m be an $m \times 1$ vector obtained by selecting m elements of α, $1 \le m \le N$. Moreover, let F_m be a $K \times m$ matrix containing the m basis vectors of the basis matrix F which correspond to the m elements selected from α. It then follows that

$$(\hat{\alpha}_m - \alpha_m)(F'_m \Sigma^{-1} F_m)(\hat{\alpha}_m - \alpha_m) \text{ is } \chi^2(m), \tag{7}$$

a chi-squared variate with m degrees of freedom.

Distribution of the total residual error

The residual error vector e is given by:

$$e = x - \hat{x}$$
$$= x - F\hat{\alpha}$$
$$= F\alpha + \eta - F\hat{\alpha}$$
$$= \eta - F(\hat{\alpha} - \alpha).$$

The total residual error is therefore:

$$e'\Sigma^{-1}e = [\eta - F(\hat{\alpha} - \alpha)]'\Sigma^{-1}(\eta - F(\hat{\alpha} - \alpha))$$
$$e'\Sigma^{-1}e = \eta'\Sigma^{-1}\eta - 2(\hat{\alpha} - \alpha)F'\Sigma^{-1}\eta + (\hat{\alpha} - \alpha)'F'\Sigma^{-1}F(\hat{\alpha} - \alpha). \tag{8}$$

From equations (3) and (4) we obtain:

$$F'\Sigma^{-1}\eta = (F'\Sigma^{-1}F)(\hat{\alpha} - \alpha).$$

Substituting this expression in equation (8) yields

$$e'\Sigma^{-1}e = \eta'\Sigma^{-1}\eta - (\hat{\alpha} - \alpha)'(F'\Sigma^{-1}F)(\hat{\alpha} - \alpha). \tag{9}$$

Since by assumption η is $N(0,\Sigma)$, applying theorem 1 it follows that:

$$\eta'\Sigma^{-1}\eta \text{ is } \chi^2(K),$$

a chi-squared variate with K degrees of freedom.

Finally, using equation (6) and applying theorem 3 we obtain the result:

$$e'\Sigma^{-1}e \text{ is } \chi^2(K - N), \tag{10}$$

that is, the total residual error is a chi-squared variate with $K - N$ degrees of freedom.

Distribution of the partial derivatives

We assume that F is a polynomial basis matrix. It then follows that each partial derivative at (0, 0) in the row and column directions is given as some linear combination of the elements of the parameter vector.

Let μ be an $m \times 1$ vector containing any m partial derivatives, $1 \le m \le N - 1$. Let B be the $m \times N$ linear combination matrix. Then,

$$\mu = B\alpha \quad \text{and} \quad \hat{\mu} = B\hat{\alpha}.$$

Also

$$\hat{\mu} - \mu = B(\hat{\alpha} - \alpha). \tag{11}$$

Using equations (11), (5) and applying theorem 2 we find that:

$$(\hat{\mu} - \mu) \text{ is } N[0, B(F'\Sigma^{-1}F)^{-1}B']. \tag{12}$$

Some special cases of interest

Independent, equally distributed noise. In this case η is $N(0, \sigma^2 I)$ and the minimum variance estimate of α becomes a least square estimate.

$$\hat{\sigma} = (F'F)^{-1}F'x.$$

Our previous results, (5), (7), (10) and (12), reduce to:

$$(\hat{\alpha} - \alpha) \text{ is } N[0, \sigma^2(F'F)^{-1}]$$
$$(\hat{\alpha}_m - \alpha_m)'(F'_m F_m)(\hat{\alpha}_m - \alpha_m)/\sigma^2 \text{ is } \chi^2(m)$$
$$e'e/\sigma^2 \text{ is } \chi^2(K - N)$$
$$(\hat{\mu} - \mu) \text{ is } N[0, \sigma^2 B(F'F)^{-1}B'].$$

Independent noise, orthonormal basis. Our previous results further simplify to

$$\hat{\alpha} = F'x$$
$$(\hat{\alpha} - \alpha) \text{ is } N(0, \sigma^2 I)$$
$$(\hat{\alpha}_m - \alpha_m)'(\hat{\alpha}_m - \alpha_m)/\sigma^2 \text{ is } \chi^2(m)$$
$$e'e/\sigma^2 \text{ is } \chi^2(K - N)$$
$$(\hat{\mu} - \mu) \text{ is } N(0, \sigma^2 BB').$$

About the Author—OSCAR A. ZUNIGA was born in Callao, Peru, on 29 September 1953. He received the B.Sc. degree in Electrical Engineering from the Universidad Nacional de Ingenieria, Lima, Peru, in 1975, and the M.Sc. degree in Electrical Engineering from the University of Kansas in 1979. He is currently a doctoral candidate at Virginia Polytechnic Institute and State University, Blacksburg, Virginia, where he has also served as an instructor. He is a member of the Institute of Electrical and Electronics Engineering and is currently involved in low-level and intermediate-level computer vision research.

About the Author—ROBERT M. HARALICK was born in Brooklyn, New York, on 30 September 1943. He received a B.A. degree in Mathematics from the University of Kansas in 1964, a B.Sc. degree in Electrical Engineering in 1966 and a M.Sc. degree in Electrical Engineering in 1967. In 1969, after completing his Ph.D. at the University of Kansas, he joined the faculty of the Electrical Engineering Department there where he last served as Professor from 1975 to 1978. In 1979 Dr Haralick joined the Electrical Engineering Department at Virginia Polytechnic Institute and State University where he was a Professor and Director of the Spatial Data Analysis Laboratory. From 1984 to 1986 Dr Haralick served as Vice President of Research at Machine Vision International, Ann Arbor, MI. Dr Haralick now occupies the Boeing Clairmont Egtvedt Professorship in the Department of Electrical Engineering at the University of Washington.

Professor Haralick has made a series of contributions in computer vision. In the high-level vision area, he has identified a variety of vision problems which are special cases of the consistent labeling problem. His papers on consistent labeling, arrangements, relation homomorphism, matching and tree search translate some specific computer vision problems to the more general combinatorial consistent labeling problem and then discuss the theory of the look-ahead operators that speed up the tree search. This gives a framework for the control structure required in high-level vision problems. Recently, he has extended the forward-checking tree search technique to propositional logic.

In the low-level and mid-level areas, Professor Haralick has worked in image texture analysis using spatial gray tone co-occurrence texture features. These features have been used with success on biological cell images, X-ray images, satellite images, aerial images and many other kinds of images taken at small and large scales. In the feature detection area, Professor Haralick has developed the facet model for image processing. The facet model states that many low-level image processing operations can be interpreted relative to what the processing does to the estimated underlying gray tone intensity surface of which the given image is a sampled noisy version. The facet papers develop techniques for edge detection, line detection, noise removal, peak and pit detection, as well as a variety of other topographic gray tone surface features.

Professor Haralick's recent work is in shape analysis and extraction using the techniques of mathematical morphology. He has developed the morphological sampling theorem which establishes a sound shape/size basis for the focus of attention mechanisms which can process image data in a multiresolution mode, thereby making some of the image feature extraction processes execute more efficiently.

Professor Haralick is a Fellow of IEEE for his contributions in computer vision and image processing. He serves on the Editorial Board of *IEEE Transactions on Pattern Analysis and Machine Intelligence*. He is the computer vision area editor for *Communications of the ACM*. He also serves as associate editor for *Computer Vision, Graphics, and Image Processing* and *Pattern Recognition*.