Journal of Computing and Information Technology - CIT 2, 1994, 2, 77–85

Maximum Entropy Segmentation Based on the Autocorrelation Function of the Image Histogram

Anton Brink

Department of Physics, University of Pretoria, Pretoria, South Africa

Most threshold selection schemes using the principle of maximum entropy regard the image or its histogram as a probability distribution. While such models can to a great extent be justified, a common assumption is that the discrete samples in these distributions (pixels or grey–levels) are independent. It is intuitively clear that this is not the case. The proposed method uses the histogram autocorrelation function as a measure of grey–level inter–dependence. The Shannon entropy of this distribution is then viewed as a measure of image grey–level entropy, where grey–level inter–dependence has implicitly been taken into account. The thresholding process splits the histogram into sub–histograms, ideally corresponding to distinct regions within the image. The entropies of the autocorrelation functions of these subranges are determined and maximized to find the optimum threshold. Two methods of maximizing the class entropies are implemented and some typical results are presented.

1. Introduction

Thresholding is a commonly used method for image segmentation based on grey–level differences between various regions or features of the image (e.g. "objects" and "background") (Pratt, 1991; Gonzalez and Woods, 1992; Haralick and Shapiro, 1985). In its simplest form, a single global threshold is selected to binarize the image into two distinct grey–levels. Such methods can usually be easily extended to multi–threshold and variable (dynamic) threshold selection (Brink, 1991a), hence only the simple case is discussed here.

Various forms of the principle of maximum entropy have in the past been applied to this problem (Kapur *et al*, 1985; Brink, 1991b, 1992b; Pal and Pal, 1989). At the root of any such approach lies the assumption, implicit or explicit, of some statistical model of the image. The most common models assume either the image itself or its histogram (in 1 or more dimensions) to represent a probability distribution. While such models can be justified (Frieden, 1980; Skilling, 1986), a shortcoming is the underlying assumption that adjacent discrete probabilities (pixels or grey–levels) are statistically independent.

The method proposed here is based on the autocorrelation function of the 1–D grey-level histogram. The autocorrelation values in this case quantify the *inter–dependence* of the histogram bins. They can also be viewed as probabilities of general grey–level cooccurrence.

2. Image and histogram entropy

The principle of maximum entropy (MaxEnt) has been successfully implemented in various image processing applications, particularly reconstruction and restoration (Frieden, 1980; Skilling, 1986; Gull and Daniell, 1978), enhancement (Daniell and Gull, 1980) and segmentation (Kapur *et al*, 1985; Abutaleb, 1989; Brink, 1992a). The most common measure of entropy used is the Shannon entropy from information theory(Shannon and Weaver, 1949). Other formulations have been proposed (Frieden, 1980; Pal and Pal, 1991) but have not been found to offer any significant advantages for this particular application. Given a set of events $f_i, i = 1, 2, \cdots, N$ with probabilities p_i, the entropy of the distribution is given by (Shannon and

Weaver, 1949)

$$H = -\sum_{i=1}^{N} p_i \log p_i \qquad (1)$$

In order to apply this to image processing, we can view the image itself as a probability distribution. The events in this case are the arrivals of photons at each pixel location. The number of photons (grey–level) at each pixel in a given image can be used as an estimate of the probability of photons reaching that particular location in a future image of the same scene. This image model was first proposed by Frieden (Frieden, 1972) and has subsequently been dubbed the "monkey model" (Skilling, 1986; Gull and Daniell, 1978; Daniell and Gull,1980) due to the analogy to a team of monkeys randomly throwing balls (photons) at a grid of boxes (pixels). Consider an image made up of N pixels $i = 1, \cdots, N$ with grey–levels g_i. Jaynes' "principle of maximum degeneracy" (Jaynes, 1968) states that, subject to *a priori* constraints, the most degenerate image formed in this manner is the most likely to occur. This is the image which can be formed in the greatest number of ways. Assuming independence of the pixels, the number of ways that the image can be formed is given by the Boltzmann law (Sears and Salinger, 1975)

$$W = \frac{G!}{g_1! g_2! \cdots g_N!} \qquad (2)$$

where $G = \sum_{i=1}^{N} g_i$. Maximizing W is clearly equivalent to maximizing $\log W$. Using Stirling's factorial approximation and dropping constant terms, it turns out that the quantity to be maximized is

$$H' = -\sum_{i=1}^{N} g_i \log g_i \qquad (3)$$

Since $g_i \propto p_i$, $p_i = g_i/G$ the probability of photons reaching pixel i, we can simply substitute these in (3). We find then that this is precisely the entropy (1) of the image.

The image grey–level histogram is one of the most commonly used distributions in image analysis and processing. It is obtained directly from the image and indicates the frequency of occurrence f_g of each discrete grey–level g in the image. These frequencies are usually normalized to give estimates of grey– level probabilities p_g. From information theory (Shannon and Weaver, 1949), the entropy of such a probability histogram is given by

$$H = -\sum_{g=0}^{n-1} p_g \log p_g \qquad (4)$$

where $g = 0, \cdots, n-1$ are the n grey–levels of the image, in this case. This definition of image entropy can be justified on the basis of an image model derived from Fermi–Dirac statistics (Sears and Salinger, 1975) as follows.

We assume the image to be analogous with a thermodynamical system: let the grey–levels g correspond to energy levels and their locations in the image (i.e. pixels) correspond to distinguishable energy states. The number of particles at level g (the number of occupied states) is given by the frequency f_g. Here the analogy becomes tenuous, since if a pixel (state) is occupied at one grey–level (energy level), it cannot be occupied at any other. There is thus some dependence between histogram bins built into this, in that we have to take account of the fact that if, say, there are N pixels in total and f_0 pixels have level $g = 0$ (the lowest grey–level), then the number of free pixels (states) available at $g = 1$ is $N - f_0$. In general we find that the number of available pixels N_{k+1} at level $g = k + 1$ is given by

$$N_{k+1} = N - \sum_{g=0}^{k} f_g \qquad (5)$$

with $N_0 = N$. Using Fermi–Dirac statistics we find the degeneracy w_g of a grey–level g, analogous to that of an energy level (Sears and Salinger, 1975), given by

$$w_g = {}_{N_g}C_{f_g} = \frac{N_g!}{(N_g - f_g)! f_g!} \qquad (6)$$

and the total degeneracy of the image (assuming statistical independence between grey–levels) is

$$W = \prod_{g=0}^{n-1} w_g \qquad (7)$$

We again want to maximize W or, equivalently,

$\log W$. Using Stirling's approximation

$$\begin{aligned}\log W &= \sum_{g=0}^{n-1} N_g \log N_g - (N_g - f_g)\times \\ &\quad \times \log(N_g - f_g) - f_g \log f_g \\ &= N \log N - (N_{n-1} - f_{n-1})\times \\ &\quad \times \log(N_{n-1} - f_{n-1}) \\ &\quad + \sum_{g=0}^{n-2} \Big[N_{g+1} \log N_{g+1} - (N_g - f_g)\times \\ &\quad \times \log(N_g - f_g) \Big] - \sum_{g=0}^{n-1} f_g \log f_g \\ &= N \log N - \sum_{g=0}^{n-1} f_g \log f_g\end{aligned}$$

since $N_{g+1} = N_g - f_g$ and $N_{n-1} - f_{n-1} = 0$. Since the first term above is constant, the quantity to be maximized is

$$H = -\sum_{g=0}^{n-1} f_g \log f_g,$$

or

$$H = -\sum_{g=0}^{n-1} p_g \log p_g,$$

in the case of the normalized histogram.

3. Autocorrelation and threshold selection

Thresholding methods based on maximizing the histogram entropy (4) of the segmented image have been used with varying degrees of success in the past (Kapur *et al*, 1985; Brink, 1991b). A common (and valid) argument against histogram–based methods is that they fail to take spatial information into account: many different images could conceivably have identical histograms. One could argue philosophically that such images should correctly have the same entropy value, but the fact remains that information is lost. To circumvent this problem, techniques based on the entropy of the image itself (1) (Brink, 1992b) and various forms of the grey–level co–occurrence matrix (Pal and Pal, 1989; Abutaleb, 1989; Brink, 1992a) have been proposed. It should however be pointed out that when using an entropy measure such spatial information is in any case not taken into account: for this reason the entropy of the image could be expressed as a 1–dimensional sum over all pixels (1), in spite of the fact that an image is clearly a 2–dimensional distribution. While some measure of the pixel inter–dependence is built into the co–occurrence matrix, these inter–dependences are limited to a small neighbourhood of each pixel. The co–occurrence frequencies are thus not entirely independent either.

For simplicity I will confine myself for the present to the image histogram. It is clear that the grey–level frequencies f_g of the histogram are inter–related, while again being assumed independent for the purposes of evaluating the histogram entropy. We can quantify the grey–level inter–dependence by determining the autocorrelation function of the histogram

$$\rho_{gg} = (r_{1-n}, \cdots, r_{-2}, r_{-1}, r_0, r_1, r_2, \cdots, r_{n-1}) \tag{8}$$

where the coefficients r_k are given by

$$r_k = \sum_{g=0}^{n-1} p_{k+g}\, p_g, \quad -(n-1) \le k \le n-1 \tag{9}$$

where n is the number of grey-levels and p_g is the probability of occurrence of grey-level g in the image. Note that $p_g = 0$ outside the range $0 \le g \le n-1$. If we normalize the autocorrelation coefficients we can interpret the resulting distribution as the probability of general grey–level inter–dependence:

$$\rho_k = \frac{r_k}{\sum_{k=1-n}^{n-1} r_k} \tag{10}$$

In other words, ρ_k quantifies the likelihood that pixels differing by k grey–levels are related. The entropy of this distribution is

$$H_r = -\sum_{k=1-n}^{n-1} \rho_k \log \rho_k \tag{11}$$

Similarly, the (2–dimensional) autocorrelation function of the actual image can be determined, in this case quantifying the inter–dependence of the actual image pixels or the likelihood that pixels a given distance apart are related. However, for the purposes of threshold selection the implementation of this approach is not apparent.

When an image is thresholded, the histogram is split into a set of smaller subranges. In the simplest case, binary thresholding, a single threshold value divides the histogram into two subranges corresponding to "background" and "object" regions of the image. Ideally these regions are distinct and homogeneous, allowing them to be regarded as separate distributions. At the optimal threshold one would then expect these distributions to be at their most uniform, resulting in a high degree of "inter–relatedness".

Given an image with a grey–level range $a \leq g \leq b$. At threshold T its histogram is split into two distributions $a \leq g \leq T$ and $T+1 \leq g \leq b$. The autocorrelation functions (8) for these distributions are determined, with coefficients given by

$$r_k^0 = \sum_{g=a}^{T} p_{k+g}^0 p_g^0, \quad -(T-a) \leq k \leq T-a$$

where

$$p_g^0 = \begin{cases} p_g/P_0, & a \leq g \leq T \\ 0, & g < a,\ g > T \end{cases}$$

and

$$r_k^1 = \sum_{g=T+1}^{b} p_{k+g}^1 p_g^1,$$
$$-(b-T-1) \leq k \leq b-T-1$$

where

$$p_g^1 = \begin{cases} p_g/P_1, & T+1 \leq g \leq b \\ 0, & g < T+1,\ g > b \end{cases}$$

P_0 and P_1 are the class probabilities

$$P_0 = \sum_{g=a}^{T} p_g,$$
$$P_1 = 1 - P_0.$$

The autocorrelation coefficients are normalized as in equation (10) and, following (11), the class entropies $H_0(T)$ and $H_1(T)$ are determined as

$$\begin{aligned} H_0(T) &= -\sum_{k=1-n}^{n-1} \rho_k^0 \log \rho_k^0, \\ H_1(T) &= -\sum_{k=1-n}^{n-1} \rho_k^1 \log \rho_k^1, \end{aligned} \qquad (12)$$

where

$$\rho_k^0 = \frac{r_k^0}{\sum_{k=1-n}^{n-1} r_k^0},$$

$$\rho_k^1 = \frac{r_k^1}{\sum_{k=1-n}^{n-1} r_k^1},$$

are the normalized class correlation probabilities.

Two methods of determining the optimum threshold τ are considered. The first follows the method of Kapur *et al* (1985) whereby a measure of total entropy is maximized: Kapur *et al* based their entropy measure directly on the grey–level histogram. Any given threshold T partitions this distribution into two classes with (histogram) entropies given by

$$\begin{aligned} H_0'(T) &= \sum_{g=a}^{T} p_g^0 \log p_g^0, \\ H_1'(T) &= \sum_{g=T+1}^{b} p_g^1 \log p_g^1. \end{aligned} \qquad (13)$$

The optimum threshold τ' is then defined as that value of T which maximizes the sum of these class entropies, i.e.

$$\tau' = \arg\{\max_{a \leq T < b} [H_0'(T) + H_1'(T)]\}$$

Simply substituting the class entropies (12) obtained from the histogram autocorrelation function thus yields the algorithm for the optimum threshold τ

$$\tau = \arg\{\max_{a \leq T < b} [H_0(T) + H_1(T)]\} \qquad (14)$$

The second method follows Brink's (1991b) modification of the above approach. Briefly, this method finds a trade–off value between the thresholds maximizing the histogram–based class entropies $H_0'(T)$ and $H_1'(T)$, respectively:

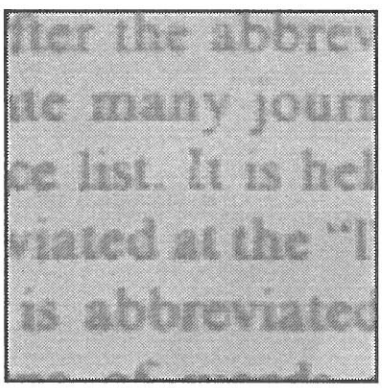
$$\tau' = \arg\{\max_{a \leq T < b} [\min[H_0'(T), H_1'(T)]]\}$$

The autocorrelation–based approach is therefore given by

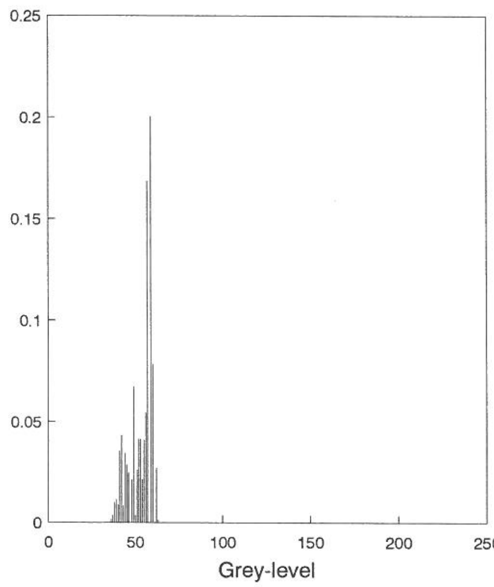
$$\tau = \arg\{\max_{a \leq T < b} [\min[H_0(T), H_1(T)]]\} \qquad (15)$$

4. Results

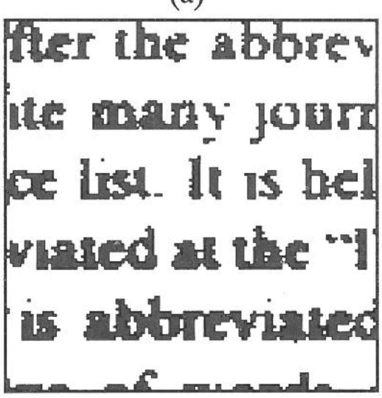
The two approaches generally select very similar threshold values, with the more subjectively pleasing results given by the "maximin" approach of (15). Some representative results appear below.



(a)



(b)



(c)

Fig. 1. **(a)** original video image of text, grey–level range $a = 33$ to $b = 63$. **(b)** Grey–level histogram. **(c)** Binary thresholded result using (14) and (15) (same threshold value), $\tau = 48$.

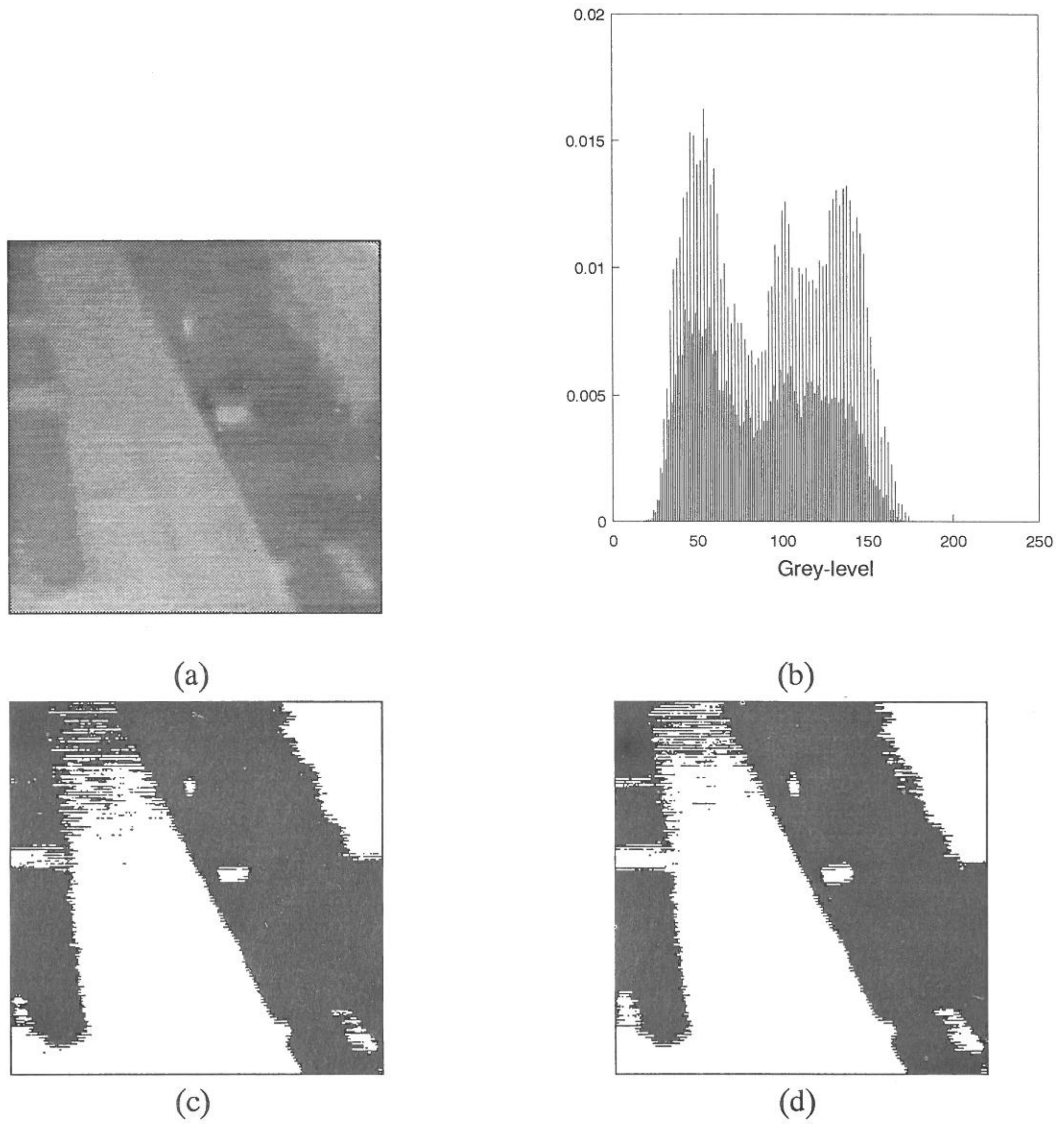


Fig. 2. (**a**) original infrared image of a warm tarred road at night, grey–level range $a = 18$ to $b = 178$. (**b**) Grey–level histogram. (**c**) Binary thresholded result using (14), $\tau = 100$. (**d**) Binary thresholded result using (15), $\tau = 95$.

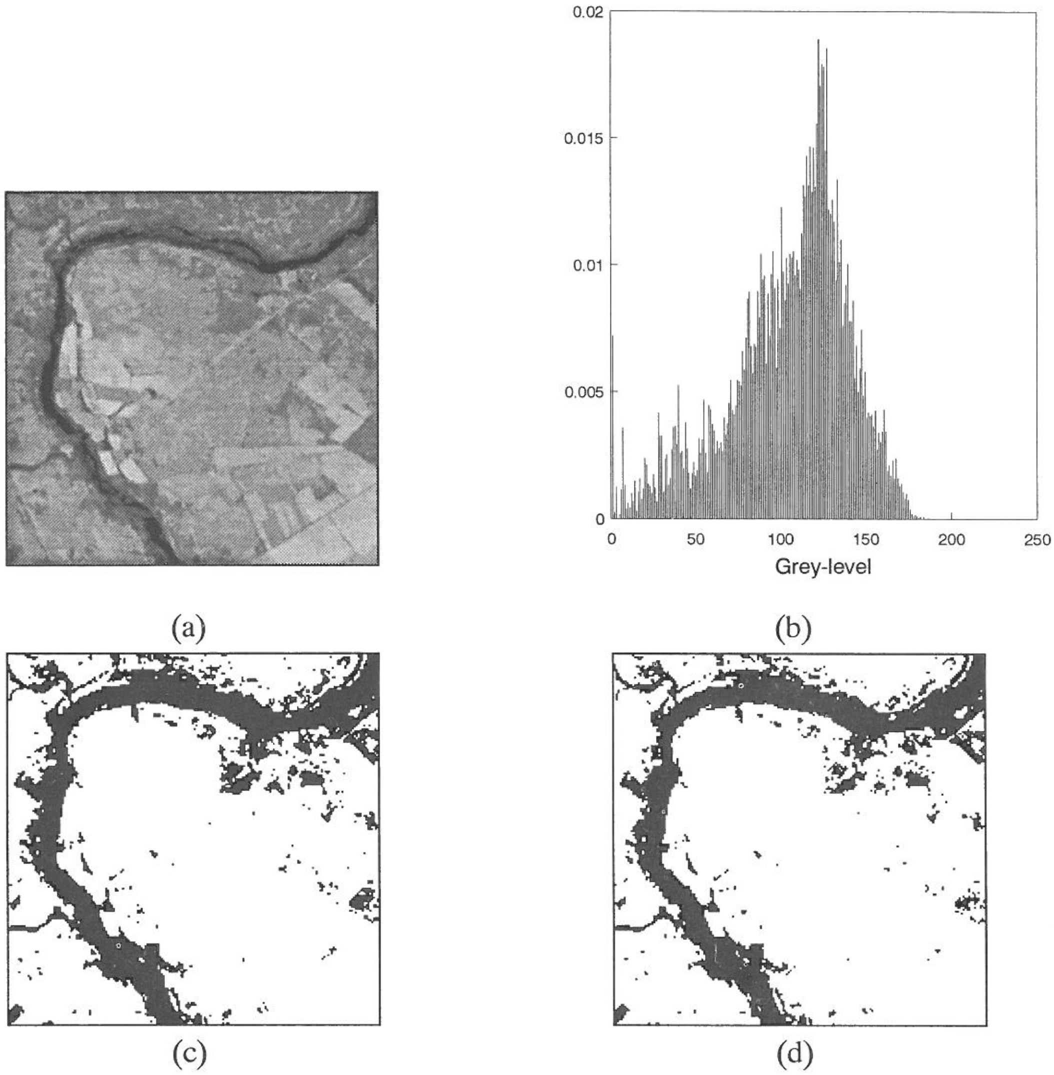


Fig. 3. **(a)** original aerial photograph showing a river feature, grey–level range $a = 1$ to $b = 187$. **(b)** Grey–level histogram. **(c)** Binary thresholded result using (14), $\tau = 81$. **(d)** Binary thresholded result using (15), $\tau = 78$.

5. Discussion and conclusions

A useful quantitative evaluation technique for image segmentation has yet to be developed. A recent method proposed by Albregtsen (1993) evaluates techniques on the basis of the accuracy with which they partition an ideal histogram consisting of the sum of two gaussian distributions corresponding to "background" and "object". A criticism of this approach is of course that we are not concerned with histogram partitioning, but with image segmentation.

Using Albregtsen's method, it was found that the technique performs very well relative to other histogram–based thresholding methods (standard techniques as well as entropic methods). Performance deteriorates when class means are close together (1 or 2 standard deviations apart), particularly when one class mode is very much larger than the other, as one would expect.

Other evaluation methods based on synthetic images (Brink and De Jager, 1987) and images where the dimensions of regions to be extracted are known *a priori* (Sieracki *et al*, 1989) can also be used to evaluate methods such as this one. Brink and De Jager (1987) proposed using a synthetic binary image which has been degraded by adding varying amounts of noise and blur to create a grey–scale image. The thresholding process then becomes one of "restoration" of the binary image. A correlation measure was used to compare the results with the original undegraded image. The design of the synthetic image and the way in which it is degraded are critical for the evaluation result to be sufficiently general. A modified version of this technique (Brink, 1994) has also been used to test the autocorrelation method described here. The results compare favourably with other methods, including those using 2–dimensional distributions (Abutaleb, 1989; Pal and Pal, 1989; Brink, 1992a). In particular, the approach using (15) for threshold selection appears to be particularly robust to increases in the severity of the image degradation. It is unfortunately beyond the scope of this paper to include the full theory and these preliminary evaluation results here.

Subjectively, the results are pleasing and useful. The need for a reliable technique to quantify such subjective evaluations of the actual image results is clear. A quantitative evaluation technique based on Albregtsen's but using a synthetic test image instead of a histogram is currently being considered.

An obvious extension to this approach is the use of the 2–dimensional autocorrelation function of the image itself. Tests indicate that this would be a useful model on which to base image enhancement and restoration methods. However, it is at present not clear how the method can be applied to image segmentation as the class regions within the image are in general not spatially continuous. Initial experiments have not yielded particularly useful results.

References

A. S. ABUTALEB (1989). Automatic thresholding of grey–level pictures using two–dimensional entropy. *Computer Vision, Graphics, and Image Processing* **47**, 22–32.

F. ALBREGTSEN (1993). Non–parametric histogram thresholding methods – error versus relative object area. *Proceedings of the 8th Scandinavian Conference on Image Analysis*, 273–280.

A. D. BRINK (1991a). Edge–based dynamic thresholding of digital images. *Proceedings of the 1991 IEEE South African Symposium on Communications and Signal Processing (COMSIG 91)*, 63.

A. D. BRINK (1991b). Thresholding of digital images using the entropy of the histogram. *Proceedings of the Second South African Workshop on Pattern Recognition*, 48–53.

A. D. BRINK (1992a). Thresholding of digital images using two–dimensional entropies. *Pattern Recognition* **25 (8)**, 803–808.

A. D. BRINK (1992b). Maximum entropy threshold selection for image segmentation. *Proceedings of the Third South African Workshop on Pattern Recognition*, 97–103.

A. D. BRINK (1994). Image models and the definition of image entropy applied to the problem of unsupervised segmentation. PhD Thesis, University of the Witwatersrand, Johannesburg, South Africa (to be submitted).

A. D. BRINK, G. J. DE JAGER (1987). Evaluation of image grey–level thresholding methods. *Proceedings of the Fifth South African Symposium on Digital Signal Processing*, 3.3.1–3.3.7.

G. J. DANIELL, S. F. GULL (1980). Maximum entropy algorithm applied to image enhancement. *IEE Proceedings* **127 E (5)**, 170–172.

B. R. FRIEDEN (1972). Restoring with maximum likelihood and maximum entropy. *Journal of the Optical Society of America* **62**, 511–518.

B. R. FRIEDEN (1980). Statistical models for the image restoration problem. *Computer Graphics and Image Processing* **12**, 40–59.

R. C. GONZALEZ, R. E. WOODS (1992). *Digital Image Processing*, Addison–Wesley, USA.

S. F. GULL, G. J. DANIELL (1978). Image reconstruction from incomplete and noisy data. *Nature* **272**, 686–690.

R. M. HARALICK, L. G. SHAPIRO (1985). Survey: Image segmentation techniques. *Computer Vision, Graphics, and Image Processing* **29**, 100–132.

E. T. JAYNES (1968). Prior probabilities. *IEEE Transactions on Systems Science and Cybernetics* **SSC–4 (3)**, 227–241.

J. N. KAPUR, P. K. SAHOO, A. K. C. WONG (1985). A new method for gray–level picture thresholding using the entropy of the histogram. *Computer Vision, Graphics, and Image Processing* **29**, 273–285.

N. R. PAL, S. K. PAL (1989). Entropic thresholding. *Signal Processing* **16**, 97–108.

N. R. PAL, S. K. PAL (1991). Entropy: a new definition and its applications. *IEEE Transactions on Systems, Man, and Cybernetics* **SMC–21 (5)**, 1260–1270.

W. K. PRATT (1991). *Digital Image Processing* 2nd ed., John Wiley & Sons, USA.

F. W. SEARS, G. L. SALINGER (1975). *Thermodynamics, Kinetic Theory and Statistical Thermodynamics* 3rd Edition, Addison Wesley, Reading.

M. E. SIERACKI, S. E. REICHENBACH, K. L. WEBB (1989). Evaluation of automated threshold selection methods for accurately sizing microscopic fluorescent cells by image analysis. *Applied and Environmental Biology* **55 (11)**, 2762–2772.

C. E. SHANNON, W. WEAVER (1949). *The Mathematical Theory of Communication*, University of Illinois Press, Urbana.

J. SKILLING (1986). Theory of maximum entropy image reconstruction. In *Maximum Entropy and Bayesian Methods in Applied Statistics* (J H JUSTICE, Ed.), pp. 156–178. Cambridge University Press, Cambridge.

Received: December, 1993
Accepted: January, 1994

Contact address:

Department of Physics
University of Pretoria
Pretoria 0002
South Africa
Tel. (+27) (12) 420 2459
Fax. (+27) (12) 342 4143
e-mail: ABRINK@scinet.up.ac.za

ANTON BRINK received an MSc in image processing from Rhodes University in 1986. He is currently employed as a lecturer in the Department of Computational and Applied Mathematics at the University of the Witwatersrand while studying towards a PhD. His research interests are in the application of the Maximum Entropy formalism to unsupervised image segmentation of digital images.