Pergamon

Pattern Recognition, Vol. 29, No. 1, pp. 179–188, 1996
Elsevier Science Ltd

0031-3203/96 $15.00 + .00

0031-3203(95)00066-6

MINIMUM CROSS-ENTROPY THRESHOLD SELECTION

A. D. BRINK*† and N. E. PENDOCK‡

* Department of Physics, University of Pretoria, Pretoria 0002, South Africa
‡ Department of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg 2050, South Africa

(*Received* 26 *July* 1993; *in revised form* 30 *October* 1995; *accepted for publication* 15 *May* 1995)

Abstract—Thresholding is a common and easily implemented form of image segmentation. Many methods of automatic threshold selection based on the optimization of some discriminant function have been proposed. Such functions often take the form of a metric distance or similarity measure between the original image and the segmented result. A non-metric measure, the cross-entropy, is used here to determine the optimum threshold. It is shown that this measure is related to other commonly used measures of distance or similarity under special conditions, although it is in some senses more general. Some typical results using this method are presented, together with results using a metric form of the cross-entropy.

Cross-entropy Thresholding Segmentation Correlation Pearson's χ^2 Maximum entropy

1. INTRODUCTION

Thresholding is a common technique for image segmentation based on grey-level differences between various regions or features of the image (e.g. "objects" and "background"). In its simplest form, a single global threshold is selected to binarize the image into two distinct grey-levels. Such methods are usually easily extended to multi-threshold and variable (dynamic) threshold selection, hence only the binary case is dealt with in detail here.

Many techniques for the automatic selection of grey-level thresholds have been developed.[1–3] These methods usually involve optimization of a criterion function based on some property of the image or its histogram, for example, the maximum entropy approach of Kapur *et al.*[4] and Brink's[5,6] maximum correlation method. In particular, methods involving comparison of the thresholded image with the original by means of some form of distance or similarity measure have generally produced useful and/or pleasing results. The principle of cross-entropy minimization due to Kullback[7] can also be used in this way. While it is not itself a metric distance measure, it can be shown to bear relation to some commonly used metric distance measures used for threshold selection. The cross-entropy formula can also be modified to form a metric measure. The threshold selection process described here is based on the cross-entropies of the images themselves, as opposed to measures based on the histogram or other statistical distributions.

In Section 2 two common measures used for the comparison of distributions and threshold selection, the correlation coefficient and Pearson's χ^2, are derived as special cases of the cross-entropy measure. The application of this measure to threshold selection is described in Section 3 and in Section 4 some typical results are shown, together with the results of using the correlation and χ^2 methods. The evaluation and discussion of these results, and concluding remarks, appear in Section 5.

2. THE CROSS-ENTROPY BETWEEN TWO DISTRIBUTIONS AND ITS RELATIONSHIP WITH OTHER MEASURES OF DISCREPANCY

The cross-entropy distance, subject to data consistency, between *a posteriori* probability distribution $q(x)$ and *a priori* distribution $p(x)$ is defined as:[7–14]

$$H_{CE}(q,p) = \int q(x) \log \frac{q(x)}{p(x)} \mathrm{d}x. \tag{1}$$

H_{CE} is also sometimes referred to as the relative entropy, Kullback–Leibler number, discrimination information or directed divergence.

For discrete (digital) probability distributions we substitute summation for the integration above. The discrete cross-entropy distance between q_x and its prior p_x is then given by:

$$H_{CE}(q,p) = \sum_{x=1}^{N} q_x \log \frac{q_x}{p_x} \tag{2}$$

subject to

$$\sum_{x=1}^{N} p_x = \sum_{x=1}^{N} q_x \quad (=1),$$

where N is the number of discrete bins in the distribution. Since equation (2) is not symmetric, i.e.

† Author for correspondence.

$H_{CE}(q,p) \neq H_{CE}(p,q)$, it is not a metric distance measure. Symmetry can, however, be imposed by adding these together, giving a metric distance measure:

$$H_m(p,q) = H_{CE}(q,p) + H_{CE}(p,q) = \sum_{x=1}^{N} q_x \log\frac{q_x}{p_x} + \sum_{x=1}^{N} p_x \log\frac{p_x}{q_x}. \quad (3)$$

In applications such as restoration, filtering, pattern recognition and segmentation the so-called prior distribution, p, represents either such knowledge as we may have regarding the "correct" solution, desiderata such as "smoothness" or the original grey-level image itself (in the case of segmentation). The posterior distribution, q, is the result of our processing of the data. Simply stated, the principle of minimum cross-entropy, as with any discrepancy measure, is that we wish to find the solution closest to the desired or expected result by minimizing the difference (cross-entropy) between them. This is in many cases equivalent to the principle of maximum entropy, where one would attempt to *maximize* the negative of equation (1), i.e.

$$H = -\sum_{x=1}^{N} q_x \log\frac{q_x}{p_x},$$

which is the usual definition of information-theoretic entropy.[15]

2.1. *Relationship to Otsu's measure and cross-correlation*

In the case of a normal probability distribution minimization of the standard non-metric cross-entropy measure equation (2) can be shown to be related to maximization of the squared correlation coefficient as defined by Otsu[8] as follows: consider a probability distribution:

$$q(x) = \frac{1}{\sqrt{2\pi}\sigma'} \exp\frac{-(x-\mu')^2}{2\sigma'^2},$$

where μ' is the mean value of the variable x and σ' is its standard deviation, and a prior distribution (e.g. the outcome of an earlier observation or measurement of the process giving rise to the probability distribution) given by:

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\frac{-(x-\mu)^2}{2\sigma^2},$$

with mean μ and standard deviation σ. The cross-entropy between these distributions is:

$$H_{CE}(x) = \int q(x) \log\frac{q(x)}{p(x)}\,dx = \log\left(\frac{\sigma}{\sigma'}\right)\int q(x)\,dx - \int q(x)\frac{(x-\mu')^2\sigma^2 - (x-\mu)^2\sigma'^2}{2\sigma^2\sigma'^2}\,dx = \log\frac{\sigma}{\sigma'} - 0.5 + 0.5\frac{\int (x-\mu)^2 q(x)\,dx}{\int (x-\mu)^2 p(x)\,dx}.$$

If the first moment of the distribution is preserved ($\mu = \mu'$) then this becomes:

$$H_{CE}(x) = \log\frac{\sigma}{\sigma'} - 0.5 + 0.5\frac{\sigma'^2}{\sigma^2}. \quad (4)$$

If we view the prior distribution as the original grey-scale image and the modified distribution as the thresholded image (with levels given by the below- and above-threshold means), then the third term in equation (4) is simply an inverted form of Otsu's[8] measure:

$$\kappa = \frac{\sigma_T^2}{\sigma_W^2} = \frac{\sigma^2}{\sigma'^2}. \quad (5)$$

Maximization of this measure was shown by Otsu to be equivalent to maximization of his measure of separability, η, which in turn has been shown[6,9] to be the square of the correlation between the original and thresholded images. The result (4) therefore implies that use of a correlation measure for thresholding is valid (in terms of maximum entropy theory) only if the distributions are normal and the first moment is preserved.

2.2. *Relationship to the χ^2 measure of discrepancy*

Seth and Kapur[10] have shown that the minimum chi-square (χ^2) method of estimation can be derived from the principle of minimum cross-entropy, as follows. For this purpose, the distribution q_x is associated with the *observed* frequencies of occurrence Mq_x of the "events" $x = 1,\ldots,N$, where M is the total number of measurements (e.g. pixels in the image). The prior distribution p_x corresponds to the *expected* frequencies of occurrence Mp_x. We assume that the discrete elements q_x of the probability distribution q differ from those of the prior distribution p by only a small zero-mean, symmetrical random perturbation, i.e. let:

$$q_x = p_x(1+\varepsilon_x), \quad x = 1,\ldots,N,$$

where $p_x\varepsilon_x$ denotes the individual differences. Then:

$$\sum_{x=1}^{N} p_x = 1, \quad \sum_{x=1}^{N} q_x = 1 \Rightarrow \sum_{x=1}^{N} p_x\varepsilon_x = 0.$$

The cross-entropy becomes:

$$\begin{aligned} H_{CE}(q,p) &= \sum_{x=1}^{N} q_x \log \frac{q_x}{p_x} \\ &= \sum_{x=1}^{N} q_x \log(1+\varepsilon_x) \\ &= \sum_{x=1}^{N} q_x \varepsilon_x - \frac{1}{2}\sum_{x=1}^{N} q_x \varepsilon_x^2 + \sum_{x=1}^{N} q_x \varepsilon_x^3 - \cdots \\ &= \sum_{x=1}^{N} p_x(\varepsilon_x + \varepsilon_x^2) - \frac{1}{2}\sum_{x=1}^{N} p_x(\varepsilon_x^2 + \varepsilon_x^4) + \cdots \end{aligned} \tag{6}$$

Assuming that the ε_x are small enough so that ε_x^k, $k > 2$ can be ignored, equation (6) becomes:

$$H_{CE}(q,p) \simeq \sum_{x=1}^{N} p_x \varepsilon_x + \frac{1}{2}\sum_{x=1}^{N} p_x \varepsilon_x^2.$$

The first term should be close to zero (the "perturbation" being symmetric and zero-mean) and, since $\varepsilon_x = (q_x - p_x)/p_x$, we obtain:

$$\begin{aligned} H_{CE}(q,p) &\simeq \frac{1}{2}\sum_{x=1}^{N} \frac{(q_x - p_x)^2}{p_x} \\ &= \frac{1}{2M}\sum_{x=1}^{N} \frac{(Mq_x - Mp_x)^2}{Mp_x} \\ &= \frac{1}{2M}\chi^2. \end{aligned} \tag{7}$$

3. THRESHOLD SELECTION

It is indicated in the previous section that cross-entropy is a more general measure of discrepancy than either correlation, which has previously been used as a threshold selector,[5,8] and the χ^2 measure, which can easily function in such a role. In this section a simple algorithm for threshold selection using the cross-entropy, as well as the χ^2 measure, is described. Methods using correlation and related measures have been previously described by Otsu[8] and Brink.[5,6]

A digital image can be viewed as a discrete probability distribution where each pixel's grey-level represents the number (or probability) of photons hitting that particular point. This model (often referred to as the *monkey model* in articles on maximum entropy) was first proposed by Frieden[11] and has been used as a basis for maximum entropy reconstruction and restoration.[12,13] Briefly, the monkey model depicts the image as an initially empty grid or raster of cells (the digital imaging array of light-sensitive elements in a video camera), at which a team of monkeys throws a number of balls (photons) in a uniformly random fashion. The resulting digital image is a two-dimensional array of pixels (cells) whose individual grey-levels measure the amount of illumination (number of balls) reaching each cell of the imaging array. The grey-levels can be normalized to give us an approximate measure of the probability of illumination at each pixel for future images of the same scene.

For the purposes of threshold selection, we may regard the original grey-scale image as the prior distribution with each grey-level g representing a frequency (or a probability if the levels are normalized). The below- and above-threshold means, $\mu_0(T)$ and $\mu_1(T)$, of the image can be used as the two grey-levels of the binary thresholded image. These levels are given by:

$$\begin{aligned} \mu_0(T) &= \sum_{g=a}^{T} g p_g \\ \mu_1(T) &= \sum_{g=T+1}^{b} g p_g, \end{aligned} \tag{8}$$

where a and b are, respectively, the smallest and largest grey-levels present in the original image, T is the threshold level and p_g is the probability of grey-level g, given by:

$$p_g = \frac{f_g}{N},$$

where f_g is the number of pixels having grey-level g and N is the total number of pixels making up the image. By using the means in equation (8) to represent the binary image, we effectively satisfy the condition for equation (2), that is:

$$\sum_{i=1}^{N} g_i = \left[\sum_{i=1}^{N} \mu_0(T)\right]_{g_i \le T} + \left[\sum_{i=1}^{N} \mu_1(T)\right]_{g_i > T}.$$

Both the true (non-metric) cross-entropy measure [equation (2)] and the metric measure [equation (3)] have been implemented. For digital images equation (2) can be rewritten:

$$H_{CE}(T) = \left[\sum_{i=1}^{N} \mu_0(T) \log \frac{\mu_0(T)}{g_i}\right]_{g_i \le T} + \left[\sum_{i=1}^{N} \mu_1(T) \log \frac{\mu_1(T)}{g_i}\right]_{g_i > T} \tag{9}$$

and equation (3) can be similarly rewritten.

Since most digital images have a large number of pixels calculation of equation (9) and the corresponding version of equation (3) at any given threshold level, T, can be time-consuming, particularly for very large images. The grey-level frequency histogram can be used to improve computational efficiency. Each bin of such a histogram represents the frequency of occurrence, f_g, of each grey-level g, i.e. the number of pixels in the image with grey-level g. Using this equation (9)

becomes:

$$H_{CE}(T) = \sum_{g=a}^{T} f_g \mu_0(T) \log \frac{\mu_0(T)}{g} + \sum_{g=T+1}^{b} f_g \mu_1(T) \log \frac{\mu_1(T)}{g}, \tag{10}$$

and equation (3) can be rewritten:

$$H_m(T) = \sum_{g=a}^{T} f_g \left[\mu_0(T) \log \frac{\mu_0(T)}{g} + g \log \frac{g}{\mu_0(T)} \right] + \sum_{g=T+1}^{b} f_g \left[\mu_1(T) \log \frac{\mu_1(T)}{g} + g \log \frac{g}{\mu_1(T)} \right], \tag{11}$$

where, as above, a and b are the minimum and maximum grey-levels in the image and $\mu_0(T)$ and $\mu_1(T)$ are the below- and above-threshold grey-level means [equation (8)].

The simplest and most direct scheme for threshold selection would be to iterate through all possible threshold values, T, and to select the threshold, τ, corresponding to the minimum of the cross-entropy equation (10) or (11), i.e.

$$\tau_{CE} = \arg\left(\min_{a \le T < b} \{H_{CE}(T)\} \right) \tag{12}$$

and

$$\tau_m = \arg\left(\min_{a \le T < b} \{H_m(T)\} \right). \tag{13}$$

More sophisticated schemes based on optimization processes such as the isodata algorithm[16] can of course be used. However, most images have a very limited number of discrete grey-scale levels (typically 256 levels). There is thus little advantage in the additional complexity incurred by using such processes, in addition to which there is the added risk that one may incorrectly select a threshold corresponding to a local rather than a global minimum.

For comparison of results, Otsu's[8] method and a method using the χ^2 discrepancy measure were also implemented. Briefly, the χ^2 distance between the original and the thresholded images can be determined by:

$$\chi^2(T) = \left[\sum_{i=1}^{N} \frac{(\mu_0(T) - g_i)^2}{g_i} \right]_{g_i \le T} + \left[\sum_{i=1}^{N} \frac{(\mu_1(T) - g_i)^2}{g_i} \right]_{g_i > T} = \sum_{g=a}^{T} f_g \frac{(\mu_0(T) - g)^2}{g} + \sum_{g=T+1}^{b} f_g \frac{(\mu_1(T) - g)^2}{g}. \tag{14}$$

This is minimized over all possible threshold values to determine the "optical" threshold.

4. RESULTS

Typical binary segmentation results using the threshold selection methods of Section 3 are presented here (Figs 3 and 5). For comparison, the corresponding results of using a correlation-based method[5,8] and a χ^2 threshold selection method outlined in Section 3 are also shown (Figs 4 and 6). The original images and their grey-level histograms appear in Figs 1 and 2.

5. EVALUATION, DISCUSSION AND CONCLUSIONS

The histogram of the image of Magritte's painting, Fig. 1(b), is characterized by three modes, two very close together at the lower end of the grey-scale (corresponding to the dark areas of the pipe and the caption) and another much higher up (corresponding mainly to the background and the reflections on the pipe), which is separated from the lower two by a broad "valley". All the methods tested here select values in this valley, with the cross-entropy thresholds appearing to show the least bias towards either end of the valley. Visually the results are all quite pleasing, although such an evaluation depends on one's requirements: if we wished to entirely separate the region corresponding to the pipe, including reflections, from the background, none of these results would be satisfactory.

The image of Fig. 1(a) is one that is quite well suited to binary thresholding, in that it contains simple objects on a simple background whose grey-levels mostly differ from those of the objects. An image which is not so well suited to this approach is that of Fig. 2(a), where we may wish to classify parts of the image into several subranges. This would require the use of multiple thresholds: all of the methods used here can be relatively easily extended to perform such multi-threshold selection, at the cost of increased processing time: in fact, it can be shown that the number of calculations required is proportional to:

$${}_{(b-a)}C_r = \frac{(b-a)!}{r!(b-a-r)!},$$

where r is the number of thresholds to be selected. In

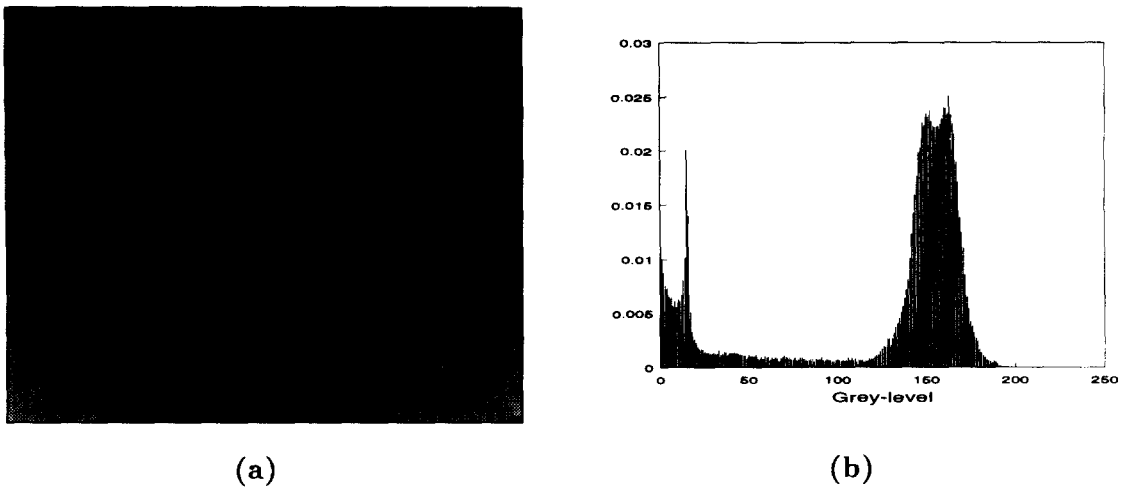


Fig. 1. The use of words I, René Magritte, 1928–1929. 1 byte per pixel, 300 × 236 pixels, grey-level range $0 \leq g \leq 200$. (a) Grey-scale image and (b) grey-level probability histogram.

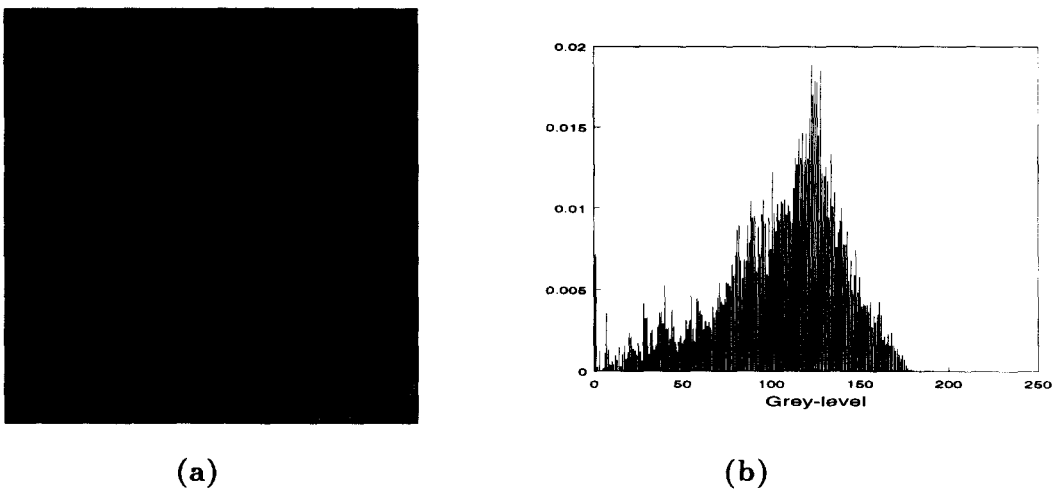


Fig. 2. Aerial photograph showing farmland and a river. 1 byte per pixel, 200 × 200 pixels, grey-level range $1 \leq g \leq 187$. (a) Grey-scale image and (b) grey-level probability histogram.

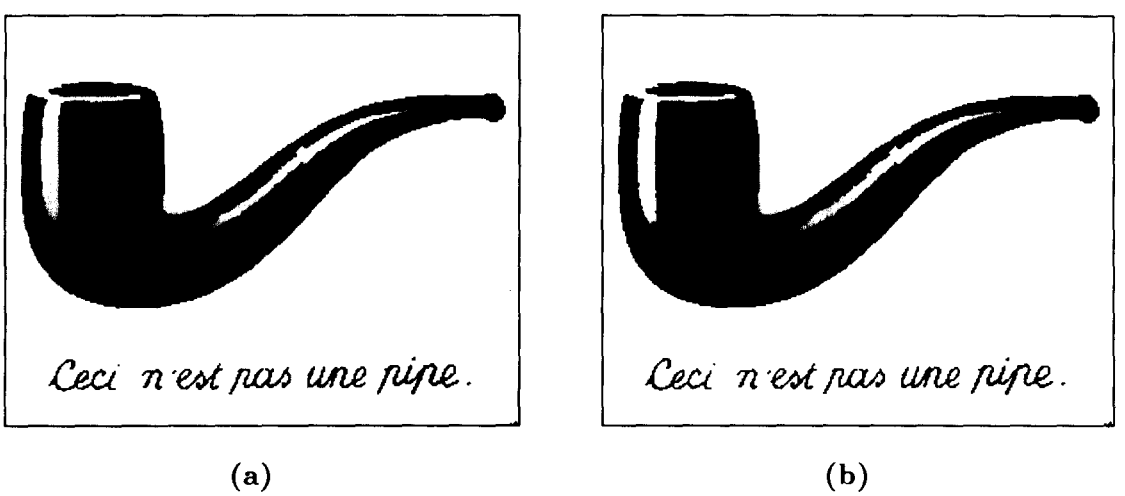


Fig. 3. Results of thresholding the image of Fig. 1 using the principle of minimum cross-entropy. (a) Standard cross-entropy (10), $\tau_{CE} = 67$ and (b) symmetric cross-entropy (11), $\tau_m = 64$.

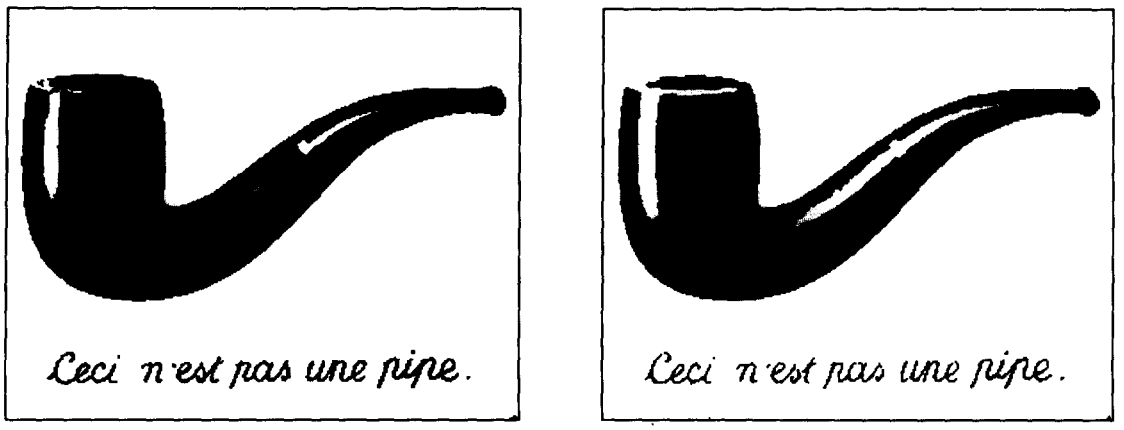


Fig. 4. Results of thresholding the image of Fig. 1 using (a) the maximum correlation between original and thresholded images:[5,8] $\tau_\rho = 87$ and (b) the minimum χ^2 distance between original and thresholded images: $\tau_{\chi^2} = 58$.

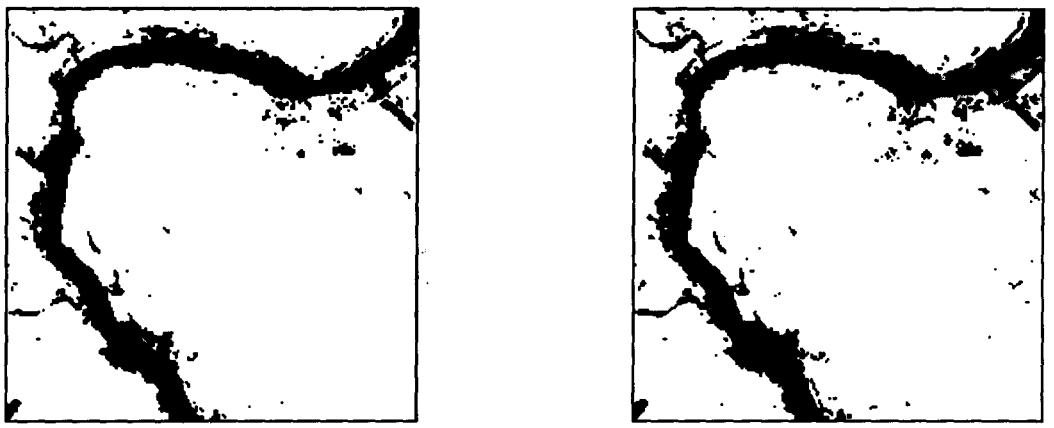
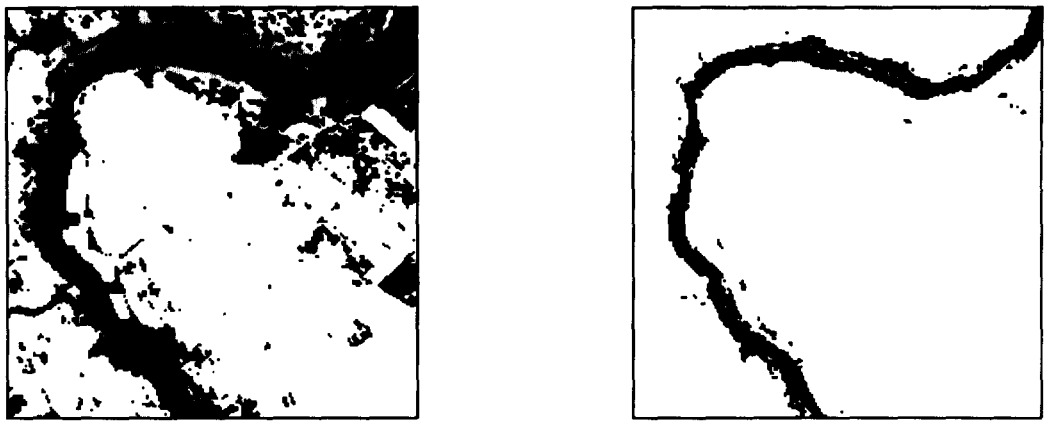


Fig. 5. Results of thresholding the image of Fig. 2 using the principle of minimum cross-entropy. (a) Standard cross-entropy (10), $\tau_{CE} = 66$ and (b) symmetric cross-entropy (11), $\tau_m = 71$.

such a situation the use of other threshold selection schemes, such as the isodata algorithm[16] mentioned above or processes involving optimization by steepest descent or conjugate gradient methods[17] would be faster than the simple approach used here.

In a binary thresholding sense, the most obvious feature apparent in the image of Fig. 2(a) appears to be the river. The two cross-entropy schemes (Fig. 5) again select threshold values between those of the correlation and χ^2 methods (Fig. 6). While the χ^2 approach has quite successfully isolated the river itself, it has lost the two small tributaries on the left, included by the cross-entropy thresholds. By comparison the correlation result looks quite messy, although it gives a better subjective impression of what the original scene looks like if we bear in mind the limitation of only two grey-levels.

Ideally we would like to quantify the evaluation of these techniques. Several approaches have been used in the past.[3,18,19] Methods which evaluate the "goodness" of the threshold level itself, such as those proposed by Weszka and Rosenfeld,[18] can of course themselves be used to select thresholds that are optimal in terms of the criteria on which the evaluation is based. This biases the evaluation in the sense that a method based on the evaluation criteria would, of course, be deemed "optimal". While one could argue that one's evaluation criteria must necessarily differ from the selection criteria in order to maintain some semblance of objectivity, one's criteria for threshold



(a) (b)

Fig. 6. Results of thresholding the image of Fig. 2 using (a) the maximum correlation between original and thresholded images:[5,8] $\tau_\rho = 93$ and (b) the minimum χ^2 distance between original and thresholded images: $\tau_{\chi^2} = 43$.

selection and evaluation should at the same time both take into account one's requirements.

Methods have been proposed whereby prior knowledge of the ideal thresholded result is used to evaluate the effectiveness of algorithms for a particular task, such as the heuristic methods adopted by Sieracki *et al.*[19] Unfortunately, such information is not usually available, leaving one in the position of having to subjectively define the "correct" result. Methods based on artificial data, where the correct result *is* known, may therefore be preferable.

The method described by Albregtsen[3] assesses the accuracy with which techniques using the histogram for threshold selection manage to divide the histogram into background and foreground distributions. To this end, a set of histograms each consisting of the superposition of two Gaussian distributions centred at different means is used. The relative sizes of the distributions and their distance apart are varied while the standard deviations are fixed. These synthetic probability histograms are given by:

$$p_g = \frac{P_0}{\sigma\sqrt{2\pi}} e^{-((g-m_0)^2/2\sigma^2)} + \frac{P_1}{\sigma\sqrt{2\pi}} e^{-((g-m_1)^2/2\sigma^2)}, \quad 0 \le g \le n-1, \quad (15)$$

where the "background" distribution is given by:

$$p_g^{(0)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-((g-m_0)^2/2\sigma^2)},$$

and the "foreground" is:

$$p_g^{(1)} = \frac{1}{\sigma\sqrt{2\pi}} e^{-((g-m_1)^2/2\sigma^2)}.$$

P_0 and P_1 are the *a priori* class proportions of the image (the relative sizes of the two Gaussian distributions), m_0 and m_1 are the respective class means and σ is the standard deviation of each class.

A histogram-based threshold selection scheme attempts to partition the histogram into its constituent background and foreground distributions. Since there is usally a certain amount of overlap, such a partitioning will inevitably result in the misclassification of some background pixels as foreground and *vice versa*. The total classification error is given by

$$E(T) = P_0 \sum_{g=T+1}^{n-1} p_g^{(0)} + P_1 \sum_{g=0}^{T} p_g^{(1)}. \quad (16)$$

The best possible threshold in terms of this measure [minimizing $E(T)$ with respect to T] is:

$$\tau = \frac{m_0 + m_1}{2} + \frac{\sigma^2}{m_0 - m_1} \log \frac{P_1}{P_0}. \quad (17)$$

Albregtsen notes that the total error in equation (16) does not necessarily tell the full story, and thus also presents separately the foreground error, defined as:

$$E_1(T) = \sum_{g=0}^{T} p_g^{(1)}. \quad (18)$$

This is simply the proportion of foreground (object) pixels incorrectly classified as background. It could be argued that background pixels misclassified as foreground can also be viewed as a "foreground error". A better approach would be to remove the mode-size bias present in equation (16). In equation (16), since each error component is scaled according to its class population size (P_0 and P_1), effectively the error associated with misclassifying pixels belonging to the smaller mode diminishes in importance with increasing mode ratio P_0/P_1. A simple solution is to remove

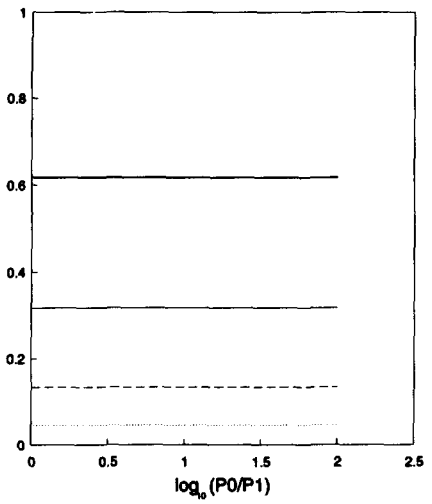


Fig. 7. Minimum possible classification error. Histogram mode size ratio (P_0/P_1) is varied from 1 to 100. The four curves shown are for the four different values of $m_1 - m_0 = D \cdot \sigma$, $D = 1$ (thick solid line), $D = 2$ (thin solid line), $D = 3$ (dashed line) and $D = 4$ (dotted line).

the bias and measure the error as:

$$E(T) = \sum_{g=T+1}^{n-1} p_g^{(0)} + \sum_{g=0}^{T} p_g^{(1)}, \tag{19}$$

with an unbiased minimum error threshold given by the midpoint between the mode means, i.e.

$$\tau = \frac{m_0 + m_1}{2}. \tag{20}$$

This means that the error associated with misclassifying a pixel belonging to a small class population is proportionally greater than that associated with one belonging to a large population, since such a pixel constitutes a proportionally larger fraction of the area of its associated class.

The unbiased classification errors in equation (19) incurred by threshold selection methods are determined for background-to-foreground ratios, P_0/P_1, over the range 1–100. Four separate graphs are obtained, corresponding to class means differing by 1, 2, 3 and 4 standard deviations, i.e. $m_1 - m_0 = D \cdot \sigma$, $D = 1, 2, 3, 4$. These are displayed as error *versus* $\log_{10}(P_0/P_1)$. Figure 7 shows the "best" threshold error obtainable, using equation (20). The classification errors due to the two cross-entropy algorithms, equations (10) and (11), appear in Fig. 8(a) and (b), and those of the correlation method[5,8] and the χ^2 method, equation (14), are shown in Figs 9(a) and (b), respectively.

The error, as one would expect, increases with increasing disparity between the histogram mode sizes. The correlation method [Fig. 9(a)] displays the best performance overall: this is easily explained if one refers to Section 2.1, where the assumption of a normal probability distribution is required for the parallel between correlation and cross-entropy to be drawn. The correlation method is in a sense tailor-made for this evaluation method. Both cross-entropy based techniques (Fig. 8) perform well and show a distinct improvement over the χ^2 method of Fig. 9(b).

An obvious shortcoming of this method of evaluation is that we are interested in correctly segmenting an image, not necessarily partitioning a histogram. It is nonetheless a useful tool for evaluation. Much work still needs to be done in this field. Similar methods based on synthetic images may also yield useful results. However, the ultimate criterion for evaluation is likely to remain the usefulness of the thresholded image for

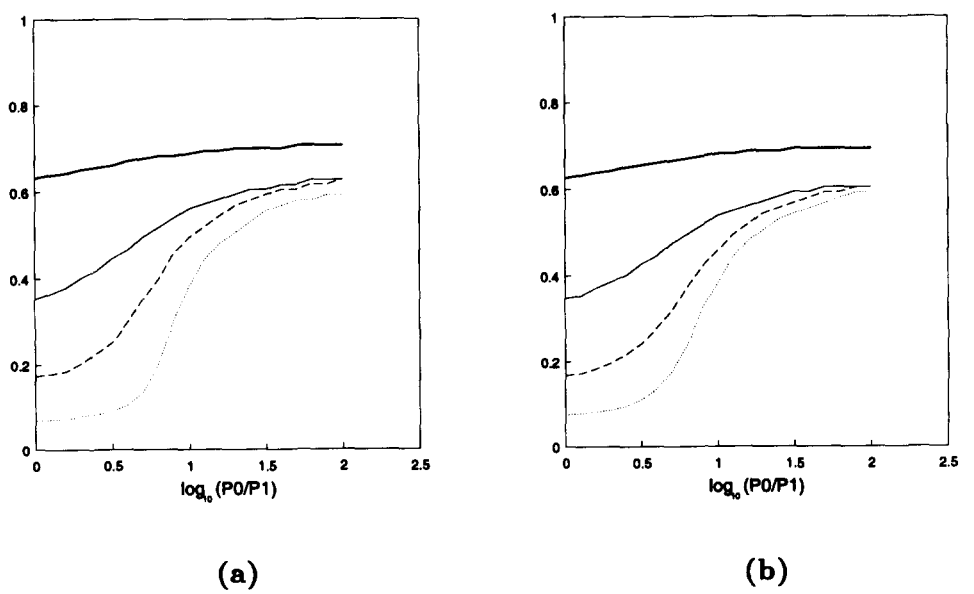


Fig. 8. Classification errors resulting from threshold selection using (a) cross-entropy equation (10) and (b) symmetric cross-entropy equation (11). The four curves shown correspond to those of Fig. 7.

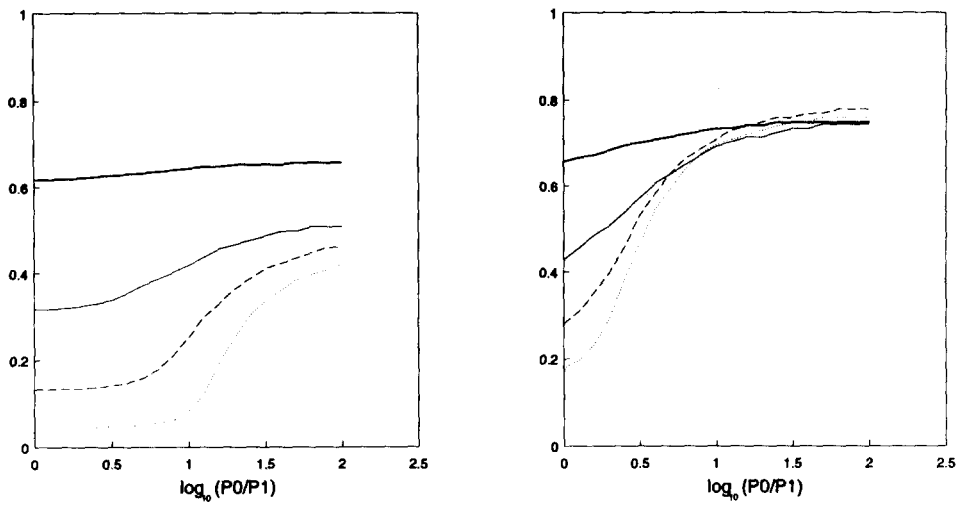


Fig. 9. Classification errors resulting from threshold selection using (a) correlation and (b) χ^2 equation (14). The four curves shown correspond to those of Fig. 7.

a given application: it is unlikely that any global quantitative evaluation can achieve this as requirements vary widely. The minimum cross-entropy threshold selection method described here appears to give relatively unbiased results which may prove useful for a wide variety of applications.

REFERENCES

1. R. M. Haralick and L. G. Shapiro, Image segmentation techniques, *Comput. Vis. Graphics Image Process.* **29**, 100–132 (1985).
2. P. K. Sahoo, S. Soltani, A. K. C. Wong and Y. C Chen, A survey of thresholding techniques, *Comput. Vis. Graphics Image Process.* **41**, 233–260 (1988).
3. F. Albregtsen, Non-parametric histogram thresholding methods—error versus relative object area, *Proc. 8th Scandinavian Conf. Image Anal.* pp. 273–280. Tromsø, Norway (1993).
4. J. N. Kapur, P. K. Sahoo and A. K. C. Wong, A new method for grey-level picture thresholding using the entropy of the histogram, *Comput. Vis. Graphics Image Process.* **29**, 273–285 (1985).
5. A. D. Brink, Grey-level thresholding of images using a correlation criterion, *Pattern Recognition Lett.* **9**, 335–341 (1989).
6. A. D. Brink, Comments on grey-level thresholding of images using a correlation criterion, *Pattern Recognition Lett.* **12**, 91–92 (1990).
7. S. Kullback, *Information Theory and Statistics.* John Wiley, New York (1959).
8. N. Otsu, A threshold selection method from grey-level histograms, *IEEE Trans. Syst. Man Cybernet.* **9**(1), 62–66 (1979).
9. I. Cseke and Z. Fazekas, Comments on grey-level thresholding of images using a correlation criterion, *Pattern Recognition Lett.* **11**, 709–710 (1990).
10. A. K. Seth and J. N. Kapur, A comparative assessment of entropic and non-entropic methods of estimation, in *Maximum Entropy and Bayesian Methods*, P.F. Fougère, ed., pp. 451–462. Kluwer Academic Publishers, The Netherlands (1990).
11. B. R. Frieden, Restoring with maximum likelihood and maximum entropy, *J. Opt. Soc. Am.* **62**, 511–518 (1972).
12. S. F. Gull and G. J. Daniell, Image reconstruction from incomplete and noisy data, *Nature* **272**, 686–690 (1978).
13. B. R. Frieden, Statistical models for the image restoration problem, *Comput. Graphics Image Process.* **12**, 40–59 (1980).
14. J. E. Shore and R. W. Johnson, Axiomatic derivation of the principle of maximum entropy and the principle of minimum cross-entropy, *IEEE Trans. Inf. Theory* **26**(1), 26–37 (1980).
15. J. Skilling, 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, U.K. (1986).
16. J. T. Tou and R. C. Gonzalez, *Pattern Recognition Principles.* Addison-Wesley, Reading, Massachusetts (1974).
17. W. H. Press, S. A. Teukolsky, W. T. Vetterling and B. P. Flannery, *Numerical Recipes in FORTRAN: The Art of Scientific Computing*, 2nd edn. Cambridge University Press, U.K. (1992).
18. J. S. Weszka and A. Rosenfeld, Threshold evaluation techniques, *IEEE Trans. Syst. Man. Cybernet.* **8**(8), 622–629 (1978).
19. M. E. Sieracki, S. E. Reichenbach and K. L. Webb, Evaluation of automated threshold selection methods for accurately sizing microscopic fluorescent cells by image analysis, *Appl. Environ. Microbiol.* **55**(11), 2762–2772 (1989).

About the Author—ANTON BRINK received an M.Sc. in image processing from Rhodes University in 1986 and recently completed a Ph.D. in image segmentation at the University of the Witwaterstr and. He is currently engaged in Post-Doctoral research in non-destructive restoration/reconstruction of San rock art in Southern Africa at the Department of Physics, University of Pretoria.

About the Author—NEIL PENDOCK is a senior lecturer in the Department of Computational and Applied Mathematics at the University of the Witwatersrand. Dr Pendock received a Ph.D. in applied mathematics from the University of the Witwatersrand in 1983. His current research interests are the applications of the Maximum Entropy formalism to problems in geophysics and applied geology.