

Parallel entropic auto-thresholding

M. Fleury*, L. Hayat*, A.F. Clark

Department of E.S.E., Essex University, Wivenhoe Park, Colchester CO4 4SQ, UK

Received 5 April 1995; revised 20 June 1995

Abstract

In this paper, we examine a multi-level thresholding algorithm based on a number of phases including peak-search, fuzzy logic and entropy of the fuzzy membership function. Analysis of the algorithm is presented to show its properties and behaviours at the various cascaded stages. The fuzzy entropy function of the image histogram is computed using S-function membership and Shannon's entropy function. To establish a suitable fuzzy region bandwidth, we have used a peak-search method based on successive clipping of the image histogram. Location of the valleys in the entropy function correspond to the certainties within the fuzzy region of the image. These certainties are used to indicate an optimal segmentation pattern for multi-level image thresholding. We compare and contrast this method of thresholding with a maximum entropy method. We have implemented the technique in parallel on a transputer-based machine as well as on a cluster of SUN4 workstations, availing ourselves of the PVM communication kernel. A parallel algorithm for the maximum entropy method is given, which significantly reduces computation times. An objective method is used to evaluate the resulting images.

Keywords: Fuzzy-entropy; Thresholding; Parallelism

1. Introduction

Image thresholding is a fundamental process in applications based on image processing and computer vision such as robotic vision, medical diagnosis, and even postal code recognition. Furthermore, a multi-level thresholding technique plays an important role, especially in those imaging problems where the prime objectives are to establish crisp boundaries in order to partition the image into its meaningful regions and extract its features. The ideal situation would be when peaks and valleys in the histogram of the image data are clear, symmetrical and balanced. The bottom of each valley would then be the natural choice for segmentation into meaningful homogeneous regions. For example, if there are k clear valleys then the image can be divided into $k + 1$ homogeneous segments using the following equation:

$$g(k : x, y) = \begin{cases} k, & \text{if } f(x, y) > T(k) \\ k - 1, & T(k - 1) < f(x, y) \leq T(k) \\ \dots & \dots \\ 1, & T(0) < f(x, y) \leq T(1) \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

* Email: {fleum.hayat}@uk.ac.essex.

Here $T(k)$ is the grey-tone threshold, f returns the input image grey-tone intensity and g is the output image, with $0 \leq x < N$, $0 \leq y < M$ for image size $N \times M$. Unfortunately, in practical terms, such situations are rare. To overcome this situation, during the past few decades a wide range of thresholding techniques such as histogram modification [1], valley seeking [2], maximum-entropy thresholding [3,4] and fuzzy measures [5–7] have been reported. A comprehensive survey of such techniques can be found in [8–11].

In this paper, we investigate a multi-level thresholding algorithm which utilizes the fuzzy entropy function of the image histogram combined with other techniques to arrive at a threshold pattern. This entropic method is computationally faster, which is important when processing batches of images such as those required for industrial inspection. The algorithm is broadly divided into three different stages: peak-search, fuzzy logic and minimum fuzzy-entropy search. For a given fuzzy bandwidth, image grey levels are translated into fuzzy outputs after applying a non-linear transformation. The entropy of the fuzzy outputs is computed through Shannon's function [12]. Instead of relying on one measurement of entropy, we have constructed the entropies of the image histogram for each given fuzzy bandwidth selected by a

Table 1
Number of iterations for a multi-modal maximum entropy method

No. of threshold points	Actual, A	Predicted, P	Relative Error, $ A - P /A \times 100\%$
1	254	255	0.394
2	32131	32768	1.983
3	2699004	2796202	3.601
4	169362500	1.790×10^8	5.690

peak detector. The location of valleys in the fuzzy entropy function are taken as the threshold points.

We have also examined a maximum entropy thresholding method as a point of comparison. This method works by finding the point where the maximum information is preserved for a given segmentation. This will also be the point of greatest uncertainty. Two objective measures, uniformity and shape, were used to compare the results for bi-modal and multi-modal images, with a simple averaging method used as a reference point. The trade-off with computational times when the algorithms are parallelized is taken into account.

The rest of the paper is organized as follows. Section 2 reviews the concepts of entropy and fuzzy entropy. Section 3 gives details of the cascaded stages of the algorithm. The following section describes the parallel implementation. An evaluation of the results is given in Section 5. Section 6 looks at an alternative possibility. Finally, Section 7 draws conclusions.

2. Entropy thresholding methods

The information theoretic entropy¹, which measures the mean value of the uncertainty, is defined as

$$H = - \sum_{i=1}^n p_i \log_2 p_i, H \in [0, 1] \quad (2)$$

where the $\{p_i\}$ form a probability distribution. This has the principal properties that, with uncertainty defined as $-\log_2 p$, events of lower probability are more uncertain, and the uncertainty of two simultaneous events is the sum of their individual uncertainties. In the paper by Kapur et al. [4], an *a posteriori* entropy measure to be used for image segmentation is defined as

$$Entp(s) = -(E_a + E_b) \quad (3)$$

where

$$E_a = \sum_{i=0}^s \left(\frac{p_i}{P_s} \right) \cdot \log \left(\frac{p_i}{P_s} \right) \quad (4)$$

$$E_b = \sum_{i=s+1}^{n-1} \left(\frac{p_i}{P_s} \right) \cdot \log \left(\frac{p_i}{P_s} \right) \quad \text{for all } s \in (0, n-2) \quad (5)$$

¹ The derivation of this formula relies on axioms, as using an analogue of statistical physics entropy would require large number assumptions which are not justified in this context.

Here $P_s = \sum_{i=0}^s p_i$ and $\hat{P}_s = \sum_{i=s+1}^n p_i$, whereas $\{p_i\}$ are the grey-level probabilities, with n being the number of grey levels and s representing potential threshold points. The aim is to maximize the information content by maximizing the entropy measure, which is brought about by varying s . Now, as s moves away from its leftmost position (i.e. 0) towards the rightmost position (i.e. n) the value of E_a increases the value for E_b decreases (Fig. 1). The effect of Eqs. (3–5) is depicted in Fig. 2. A feature discernible in the figure is that, because $Entp(s)$ is a relative measure, this method responds in the first instance to distributions and not to amplitudes.

The same method can be extended to multi-modal images² by defining $Entp[s] = - \sum_{i=1}^k E_i$ in the natural way. To find the optimum partition, we have adopted a scheme shown in Fig. 3 for a three-segment partition. According to the figure, A and B are fixed at positions near to the leftmost available grey level, R_i . Then B is slid sequentially and successively towards the rightmost possible position, R_j , while keeping A fixed at its previously-defined position. When B reaches R_j , then A is slid by one position towards the right and B is reset to the position $A + 1$. B is slid again towards R_j . This process is repeated until $A = R_{j-1}$ and $B = R_j$. All values of the entropies are recorded. The partition (i.e. the positions of A and B) generating maximum value of $Entp[s]$ is used to set the threshold points for image segmentation. The same procedure could be extended for the k th partition of the image into homogeneous regions. However, as the number of partitions increases, the time complexity is asymptotically dominated as

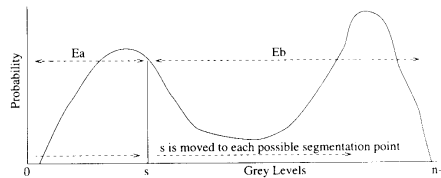


Fig. 1. Finding the optimum partition for the bi-modal maximum entropy method.

² Multi-modal images are ones with an underlying multi-modal global histogram. Though it is possible to take a different approach, in this paper a given number of multiple thresholds is imposed for practical reasons.

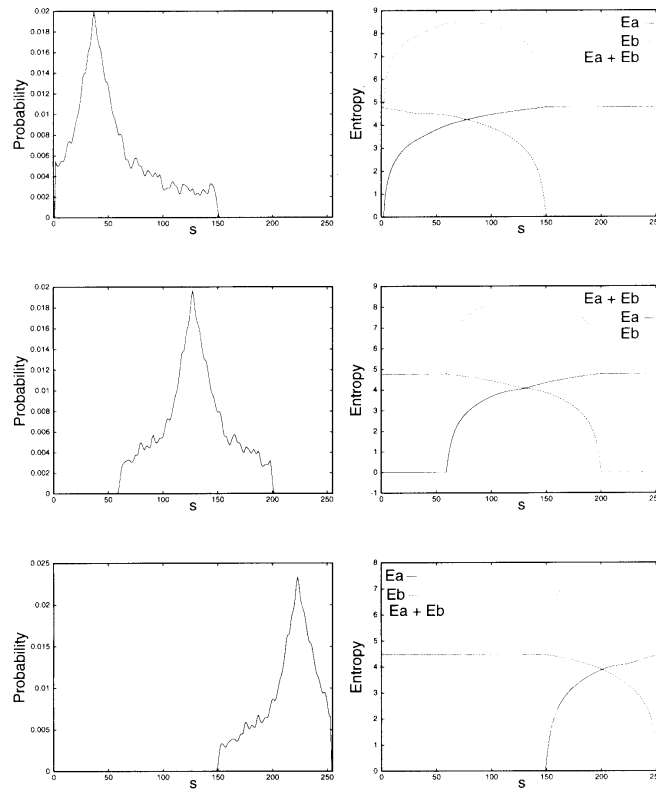


Fig. 2. Probability distributions (left) and their entropy functions (right).

$O(n^r/r!)$, in the region of interest, where n is the number grey-levels and r is the number of threshold points. In fact, the complexity is almost exactly

$$\binom{n+r}{r} = \frac{(n+r)!}{r!n!} \quad (6)$$

which is the number of possible partitions formed from r threshold points on n grey levels.³ Table 1 compares the count recorded when the computation was suppressed (so as to avoid delay) with the count predicted using $n^r/r!$.

³To see this, represent the partition system as n ones and r zeros. The zeros can be placed between the ones to form partitions. The number of ways of selecting r threshold points from the $n+r$ partition system gives the combination formula. When r is small compared to n the combinations grow approximately exponentially.

2.1. Fuzzy entropy function

Fuzzy membership is a way of grading membership of a set where there is statistical uncertainty. This is exactly the situation where the image histogram does not display a smooth set of valleys between peaks. The standard S-function [13] (Fig. 4) is typically chosen for the purpose of grading. We have taken a generalized form of this equation [14]:

$$\mu(x) = S(x, a, b, c) = \begin{cases} 0, & x \leq a \\ k \left(\frac{x-a}{c-a} \right)^2, & a \leq x \leq b, k = \frac{c-a}{b-a} \\ 1 - k \left(\frac{x-c}{c-a} \right)^2, & b \leq x \leq c, k = \frac{c-a}{c-b} \\ 1, & x \geq c \end{cases} \quad (7)$$

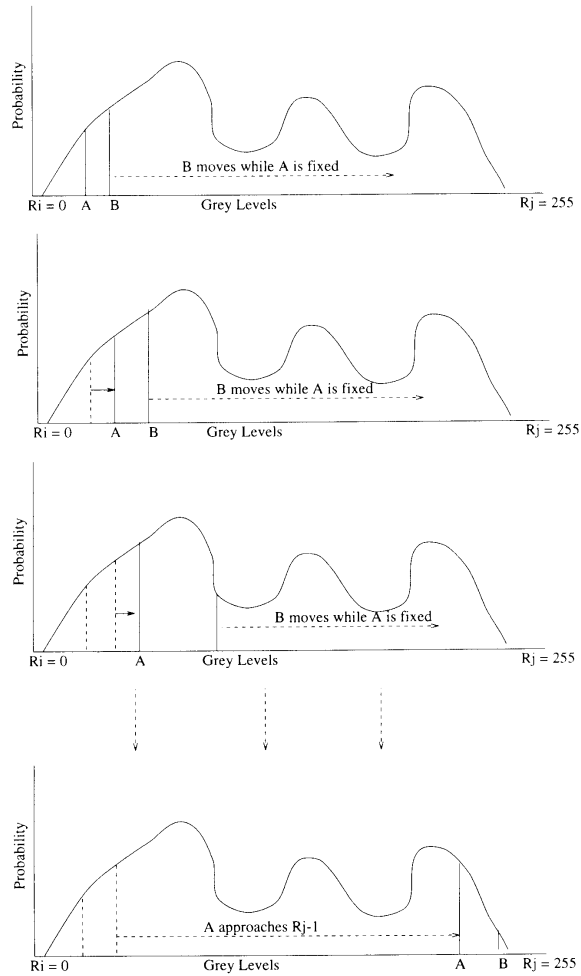


Fig. 3. Finding the optimum partition for the multi-modal maximum entropy method.

This function is symmetrical only when the cross-over point b lies exactly in the middle of the fuzzy region (a, c) (i.e. $b = a + c/2$, bandwidth $\Delta b = b - a = c - b$ and $k = 2$); see Fig. 5.⁴ From the point of view of a two-tone image, a grey-level image is fuzzy. If we had

⁴ There is no logical reason that the authors are aware of why another suitable function might not be used provided it had suitable properties.

a method of measuring fuzziness, then it might be possible to find a segmentation point that would minimize the fuzziness. Notice that this assumes a first-order Markovian model for the image field [15], i.e. we expect a smooth transition of grey-levels with the grey-level value of a particular pixel influenced only by the pixels acquired immediately before. It is possible to apply Shannon's function (Eq. (8), which is Eq. (2)

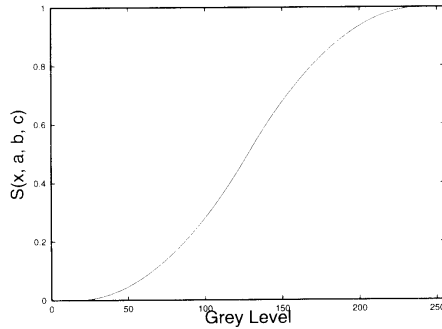


Fig. 4. Standard S-function.

with $n = 2$) to the membership function for a particular bandwidth:

$$Sn(x) = -\mu(x) \log_2 \mu(x) - (1 - \mu(x)) \log_2 (1 - \mu(x)), \quad 0 \leq x \leq n - 1 \quad (8)$$

A fuzzy entropy measure is now given by

$$H_{fuzzy}(A) = \frac{1}{n \ln 2} \sum_{i=1}^n Sn(\mu_A(x_i)) \quad (9)$$

where A is the fuzzy set of concern containing n members. The equation has been normalized because this is not a measure on a probability distribution. Because the set records 'possibilities' and not 'probabilities', it is no longer viable to use relative entropy in a manner analogous to Eq. (3), because otherwise an equalized histogram would result in maximum entropy. To apply Eq. (9) to an image histogram H , with n grey-levels within the fuzzy region g_i and with h_i pixels in the i th histogram slot, use:

$$H_{fuzzy}(H) = \frac{1}{n \ln 2} \sum_{i=1}^n Sn(\mu_H(g_i)) h_i \quad (10)$$

The maximum contribution to Eq. (10) is generated by values for $\mu_i = 0.5$, with the contribution falling off symmetrically on either side of this. We require the situation where the contribution from $\mu_i = 0.5$ is at a minimum. The result of applying this measure to a grey-level probability distribution is illustrated in Fig. 6, where the correct identification of a thresholding point is established. Fuzzy entropy is an absolute measure. Thus, bandwidth is varied to bracket the response to differing object sizes in differing images.⁵

⁵ If an object were to be reduced in area within an image, the amplitudes of the matching grey level distribution would also be reduced.

3. Multi-level thresholding algorithm

This section describes the steps in arriving at a suitable segmentation of the image. A useful preliminary is to perform histogram smoothing with a 1-D Gaussian in order to minimize the possibility of detecting spurious peaks or valleys. The next step of the algorithm is a peak search for humps in the image histogram (see Fig. 7). We have utilized a histogram clipping technique which is a common method when applied to contrast stretching [15,16]. Initially, we clip the histogram data by 5% of the highest peak available in the histogram distribution so as to eliminate small peaks. Then we scan the histogram from 0 to $n - 1$ (where n represents the maximum number of available grey levels) to locate all valleys at that clipping level, recording the valley start and end locations. This process is repeated by successively increasing the sampling level by 1% up to the maximum sampling level, say 98% of the highest peak. Once a table of peaks at each level is established, the first level at which the number of peaks is below a set maximum is utilized. Extra checks can also be carried out to isolate true and false peaks valleys using figures of merit based on the separation between any two peaks and the strength of the individual peaks, which is similar to the usage in the paper by Sahasrabudhe et al. [2]. This will help to eliminate spurious peaks by merging two peaks with a poor valley into a single clear peak. Experiments were conducted on a variety of histogram data distributions and fairly good results have been achieved. A sample histogram and the number of peaks found by our peak-search method are shown in Figs. 8 and 9.

It is now possible to switch between threshold location methods. For bi-modal images, an averaging technique⁶ may be most appropriate, in terms of computation times and subjective results. When it is possible to supply the number of segments desired, then the maximum-entropy

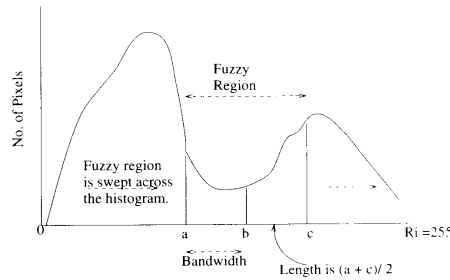


Fig. 5. Fuzzy region superimposed on an image histogram.

⁶ The averaging method finds the arithmetic mean of the histogram pixels. Its 'floor' becomes the threshold value.

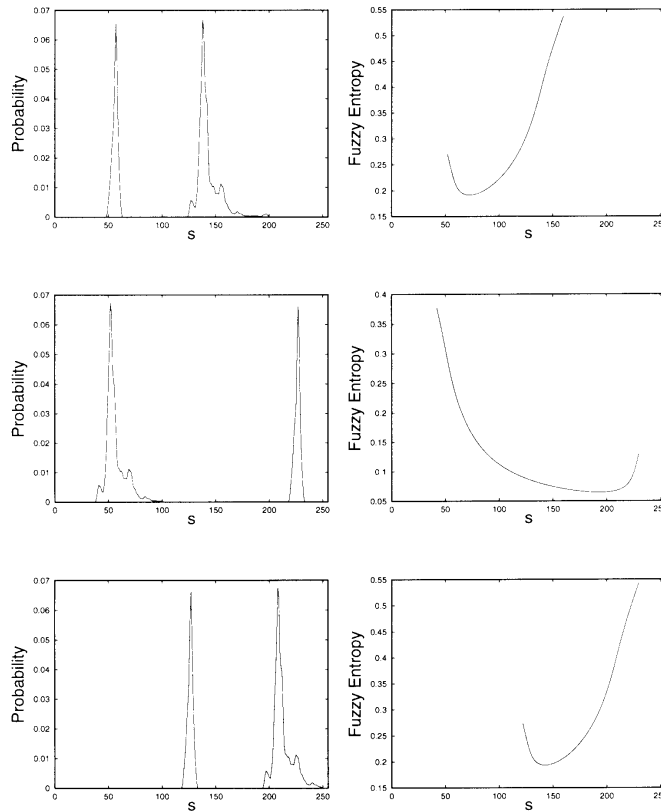


Fig. 6. Probability distributions (left) and their fuzzy entropy functions (right).

method may be selected. This method will not work in practice for segmentations above five (i.e. where there are four threshold points) because the computation time, on for instance a SUN4 workstation, is a matter of hours. The algorithm used to arrive at all possible partitions is illustrated in Fig. 10 for four segments. Points A and E in the figure act as boundary markers. Partition point D is repeatedly incremented towards marker E. The other partition points are incremented once, as required, for each occasion that D reaches E. Our first implementation recalculated the entropy (the multi-modal version of Eq. (3)) for each position of D. A second implementation only recalculated the entropy when D reached E, and then only for those segments for which the partition points had moved. When D moves otherwise, it uncovers just one grey-level probability to

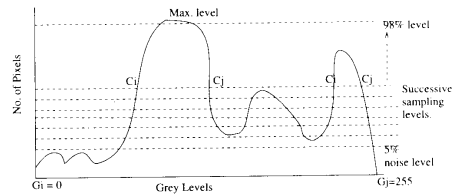


Fig. 7. Peak-search method using the image histogram. C_i, C_j – clipping points; G_i, G_j = grey level range.

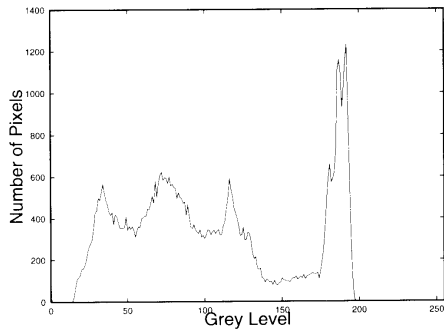


Fig. 8. Sample image histogram distribution.

its left. The segment to the right must be adjusted for the loss and the segment to the left must be adjusted for a gain. We call this algorithm the running entropy method, which is shown for the case of two partition points in Fig. 11.

The fuzzy entropy method will find a number of peaks depending on the bandwidth (Figs. 12–15).⁷ However, we first restricted the bandwidth range by consideration of the known number of peaks as given by the histogram clipping state. It is necessary to restrict the valley search region to the start position of the first peak and the end position of the last peak found at the clipping stage. This measure ensures that the fuzzy valleys match the clipping stage peaks. A heuristic is to use the arithmetic mean of the lowest and highest suitable bandwidths as the selected fuzzy entropy curve. The fuzzy entropy valleys are found simply by finding where $e(k-1) > e(k) < e(k+1)$, for successive discretely-sampled values of fuzzy entropy, $e(k)$, corresponding to each starting position of the bandwidth window. In the case of the search for a maximum (when the search procedure is reversed for the maximum entropy method), a 5% tolerance may be applied after the first maximum has been found. The intention is to avoid finding multiple maxima.

Even if we repeat for a number of bandwidths, the fuzzy-entropy algorithm has calculation complexity $O(n)$. The Shannon function transformation of the S-function is performed once for each bandwidth. The maximum entropy method is of exponential order because it does not use the extra grading information. The entropy must be re-computed for each tentative position of the thresholds and for the whole image histogram. Each re-calculation involves costly logarithmic functions. However, it is possible to reduce the

⁷ Too small a bandwidth will be influenced by histogram oscillations and too large will ignore relevant changes.

computation times by only adjusting the totals by the single value that changes for each partition point.

4. Parallel algorithms and implementation

The proposed algorithm has been implemented in parallel on two different platforms. The software was developed and tested on SUN4 workstations under the Parallel Virtual Machine (PVM) [17] inter-processor communication kernel. It was then implemented on Parsys Supernode machine (SN1000) [18]. The advantage of this procedure is that PVM on the SUNs provides a robust testing environment, while the isolated environment of the transputer machine gives good speed-ups for communication intensive applications. There is little difficulty swapping the communication primitives of PVM for those of 3L's Parallel 'C' [19].

The image is divided into equal strips and distributed to t task, where $t = 2i$ for integer i . The tasks are connected in a uni-ring topology [20] (Fig. 16). Each task performs a local grey-level histogram. So as to avoid centralizing the local histograms, local histograms are combined in the manner given in Fig. 17, where the numbers inside the task boxes represent the individual local histograms. Each task is assigned a parity derived from its identity. The parity is used to sequence the order of exchanges. While this method is potentially not as efficient as using a binary-recursive method of globally forming the complete histogram, it has the merit of being scalable on a fixed-valency processor. Where averaging or fuzzy-entropy is used, the shortness of the sequential calculation phase made it not worth attempting to parallelize; instead, each task performs an identical calculation to find the threshold(s).

Where the maximum entropy method was selected, the overhead involved in exchanging results is small in comparison to the computation time. Additionally, the

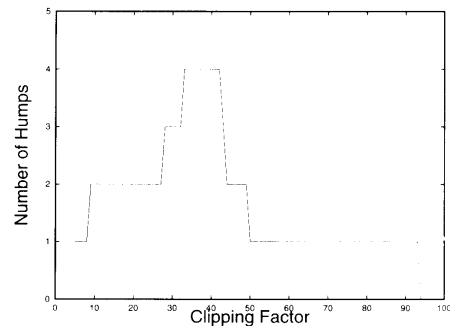


Fig. 9. Number of peaks in a sample image.

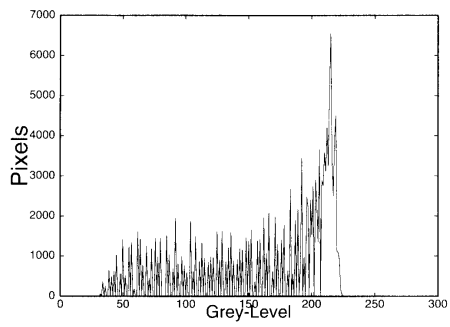


Fig. 29. DNA image (7 bits), grey-level histogram.

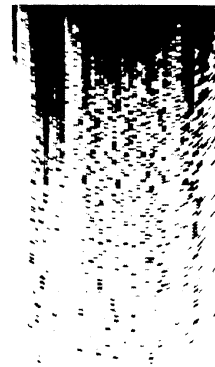


Fig. 32. DNA image, maximum-entropic method, threshold point at 134.

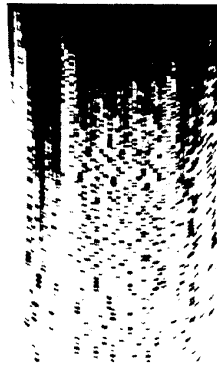


Fig. 30. DNA image, average method, threshold point at 164.



Fig. 33. Face image, size 256 x 256.

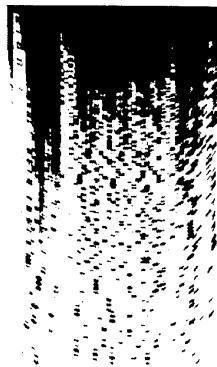


Fig. 31. DNA image, fuzzy-entropic method, bandwidth 28, threshold point at 160.

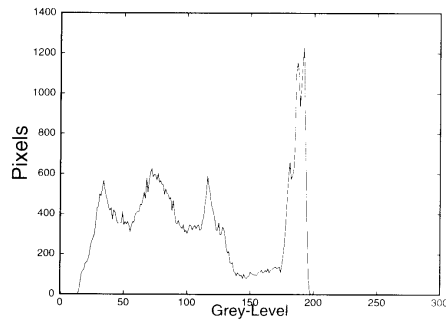


Fig. 34. Face image, grey-level histogram.



Fig. 35. Face image, average method, threshold point at 105.



Fig. 36. Face image, fuzzy-entropic method, bandwidth 27, threshold points at 38, 88 and 134.



Fig. 37. Face image, maximum-entropic method, threshold points at 75, 132 and 174.

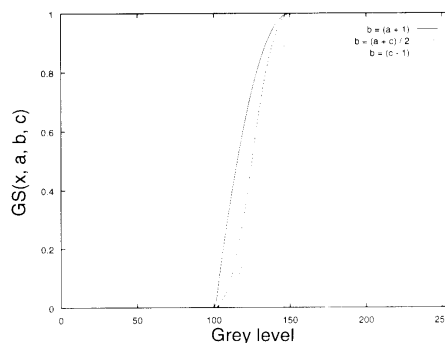


Fig. 38. Behaviour of a generalized S-function (GS).

7. Discussion and conclusion

We have examined the idea of using entropy as a method of thresholding, especially in regard to the question of whether fuzzy entropy represents an improvement on the use of information theoretic entropy. We used a maximum entropy method which is claimed as an improvement on previous entropy measures [4]. The theory of entropy, and its variant, fuzzy entropy, have been outlined: fuzzy entropy uses an extra grading step to reduce the calculation involved in forming a thresholding partition. In contrast to the maximum entropy method, which maximizes the information content provided by a suitable partition, fuzzy entropy finds the partition which makes the fuzzy set given by the partition grey-level members as 'crisp' as possible. We have provided details of robust procedures, including smoothing by a linear Gaussian filter, a clipping method for histogram

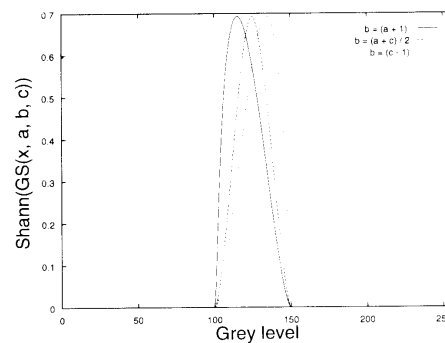


Fig. 39. Behaviour of the composition of Shannon's function with the GS function.

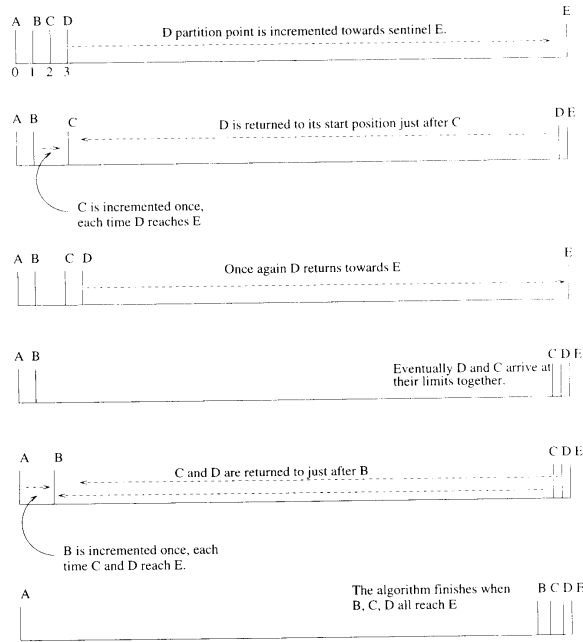


Fig. 10. Partitioning algorithm: example for four segments.

topology is already set to exchange global results. Each task finds the maximum entropy of its set of entropies, formed by taking every r th value. The global maximum entropy is now found by a similar exchange of results as used to form the global histogram. In this case, each task compares its local maximum with the new maximum and relays the larger of the two maxima. The final phase for all versions of the thresholder is to apply the segmentation in parallel to each image strip.

A suitable decomposition of the partitioning scheme in the maximum entropy algorithm is shown in Fig. 18. For simplicity of exposition, each task is shown with its first partition point, A_i (where the i index the tasks), starting at equal intervals through the range of grey-levels. While the A_i are confined in scope, further partition points must occupy all possible positions beyond A_i to the last grey-level position. With this arrangement it is possible to test all possible partition

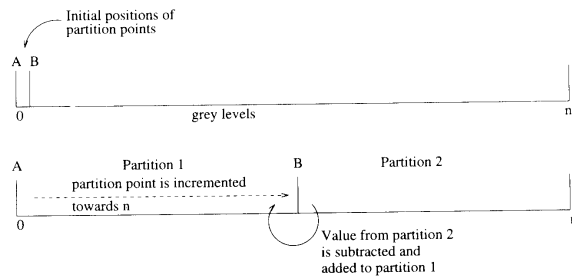


Fig. 11. The 'running' method of calculation.

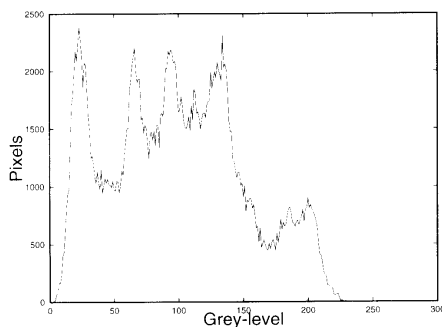


Fig. 12. Grey-level histogram of Lenna, size 512 x 512.

points. Other arrangements are possible, such as starting all but one of the partition points at the same positions as for the sequential method. The last partition points sample their range of points in interleaved fashion. The disadvantage of this other method is that it is no longer possible to use the running entropy method described in Section 3 because the partition points uncover more than one grey level when passing over the histogram.

For the method chosen for implementation, if we assign the A_i at equal intervals then the parallel algorithm is not load-balanced. To correct this a recursive load-balancing assignment method may be possible, rather than an exhaustive search. Suppose that the task with the largest workload, task 0, reduces the size of its interval by k , giving an interval length of $n/t - k$ with t the number of tasks (Fig. 19). Each of the other tasks will perform at least the same amount of calculation (in parallel). Confine attention to where two partition points are used and compare the work performed to that of a task in the most favourable position, $t - 1$. The extra computation performed by

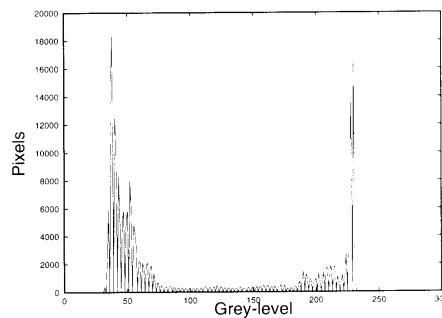


Fig. 14. Grey-level histogram of objects image, size 500 x 500.

task 0 is $(n/t - k)(n(t - 1)/t + k)$. The other tasks, considered in aggregate, will have gained work because of the extension of their region by k . The additional work performed, per task, is given by

$$\frac{1}{t-1} ((nx + k) + \dots + 1) - [(n/t - k) + \dots + 1] \quad (11)$$

where $x = (t - 1)/t$. To balance the load, the extra work done by task 0 must equal the extra work given by expression (11). Therefore, the following equation should be solved for k :

$$(n/t - k)(nx + k) - \frac{1}{t-1} ((nx + k) + \dots + 1) - [(n/t - k) + \dots + 1] = 0 \quad (12)$$

which is quadratic in k . Once we have the interval length for task 0, then subtracting this length from n allows the equation to be reused for $t - 1$ tasks. Though this formula has been extended to the case of more than two partition points, the algebra is less tractable and numerical methods are needed to find k , which is a

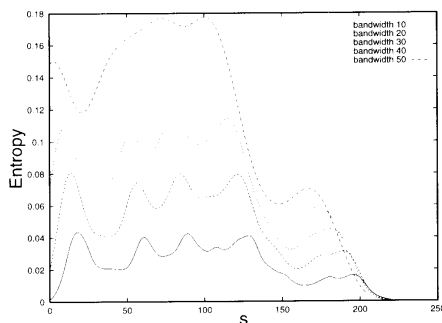


Fig. 13. Fuzzy entropy of the Lenna histogram for various bandwidths.

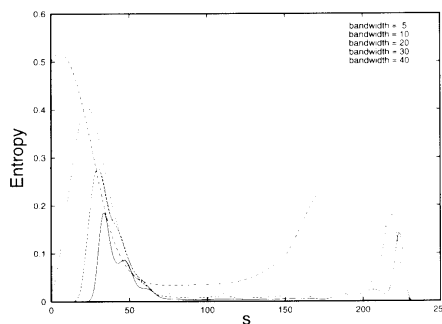


Fig. 15. Fuzzy entropy of the objects histogram for various bandwidths.

Table 2
Comparative timings for maximum-entropy algorithm, three segments

No. of processors	Transputers			PVM With load-balancing
	Without running total	With running total	With load-balancing	
4	134.564	5.134	4.400	5.264
6	91.039	3.676	3.192	4.150
8	68.902	2.977	2.534	3.651
10	55.614	2.487	2.168	3.143
12	46.736	2.198	1.882	2.861

polynomial of an order given by the number of partition points.

The performance of the maximum-entropy method for three segments is given in Table 2. When the running total method is used, the timings are reduced considerably. The timings for PVM are slightly worse than for the transputer implementation. In Fig. 20 for four segment timings, though the particular SUN4s used for PVM have a lower nominal CPU speed rating (20 MHz as returned by the Unix utility `fpversion` with Weitek 3170-based floating-point unit) than the T8 transputers (25 MHz)⁸, the timings are better for PVM because in part communication times have marginal effect on the final total. Load-balancing is now seen to have a significant effect on the outcome. It is apparent that for four segments the maximum-entropy method timings are still much in excess of those for the fuzzy-entropy multi-thresholding method.⁹

Where the fuzzy method is performed in parallel, the histogram formation and the final step of applying the segmentation are implemented in parallel, with the intermediate step being replicated by each task. As mentioned in Section 3, a range of suitable bandwidths is tested in the intermediate step. The extra communication involved in parallelizing the intermediate steps would be counter-productive in view of the overall time for four SUN4 workstations utilized under PVM

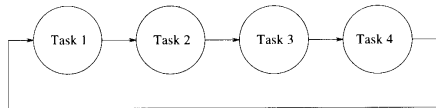


Fig. 16. The uni-ring topology used for global exchange of results.

⁸ This does not imply any disparity between the microprocessors, but merely records a fact. Considerations should be given *inter alia* to the suitability for a particular job of the instruction sets, compilers and the floating-point units, before any firm judgements are made.

⁹ It may be possible to reduce the timings for the maximum entropy method further by not examining every possible partition. For instance, every other partition could be used, when examination indicates that the final result would not necessarily be significantly altered. Another possibility is an iterative method, whereby the histogram data are initially clustered in order to find a threshold region to search in the next stage. This method would also compromise on accuracy.

of 1.44 s (averaged over five timings) for a 512×512 image. This timing assumes local I/O, since the gross averaged time was 2.97 s, compared to a single processor timing of 2.75 s.

5. Evaluation of results

While single-thresholded images can be assessed subjectively, assessment of multi-thresholding is partially dependent on the assignment of grey level to each segment. The visual appearance is also not of importance for robotic applications. To provide an objective assessment, we used two measures of the thresholded image: uniformity and shape. These measures are loosely based on those in the papers of Kapur et al. [4] and Levine et al. [21], respectively. The uniformity measured was used

$$\text{Uniformity} = 1 - \frac{\sum_{x,y \in R} (f(x,y) - \mu)^2}{B} \quad (13)$$

where R is the thresholded region of concern, f returns the grey-level, μ is the mean grey level within the region

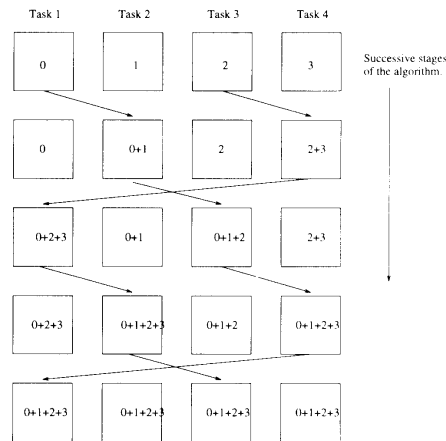


Fig. 17. Algorithm used to collect results.

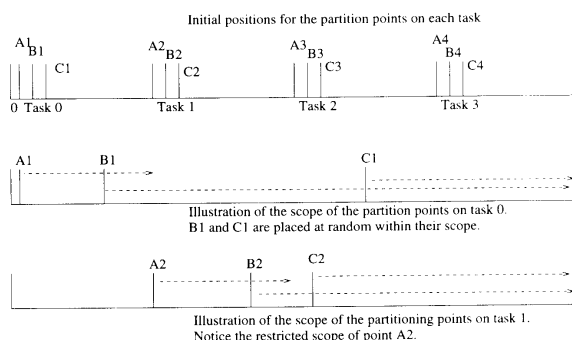


Fig. 18. Parallel algorithm for partitioning.

and B is a normalizing factor based on the region area and its grey-level range. The shape measure was:

$$\text{Shape} = 1 - \frac{\sum_{i \in R} (G + I)}{C} \quad (14)$$

where R is the threshold region of concern, G is a neighbourhood gradient measure, I is the absolute difference between the grey-level of the pixel of concern and its 8-neighbourhood grey-level mean and C is a normalizing factor based on the region area and its grey-level range. The boundary of the image is omitted in this measure as are pixels not in the interior of the thresholded region. G is given by

$$\sum_{k=1}^4 D_k^2 \quad (15)$$

with

$$D_1 = f(x + 1, y) - f(x - 1, y)$$

$$D_2 = f(x, y + 1) - f(x, y - 1)$$

$$D_3 = f(x - 1, y - 1) - f(x + 1, y + 1)$$

$$D_4 = f(x + 1, y - 1) - f(x - 1, y + 1)$$

A sample of the results of the tests are given in Table 3

for two images taken from a number of others: a multi-modal image (Fig. 21) and a bi-modal image (Fig. 22), where the latter image was captured using 7-bit sampling. As a point of comparison, in the paper by Sezan [22] thresholding points for Fig. 21 were reported as 47, 96 and 152. If the fuzzy-entropic search area is extended to include the higher grey-levels, it finds a threshold point at 155 as well as those in Table 3. Sample thresholded images are given in Figs. 23-25. Though the absolute figures are in part a result of the normalizing factor used, it might be observed that there is little to gain from using other than an averaging method on a bi-modal image (Figs. 26 and 27). Additional images, for the reader's perusal, are included in Figs. 28-37, when the histogram for Fig. 33 is already included as Fig. 8. The fuzzy-entropic method is not as good as the purely entropic method when used on the multi-modal image, but this distinction is naturally only good for the two images surveyed here. In general, shape appeared a better discriminator in separating out the better performance of the maximum-entropy method for multi-modal images. Further tests revealed a weaker response on the part of the entropic methods when Gaussian noise was added to the images. This may be because the noise has the effect of confusing the underlying distribution of grey-level

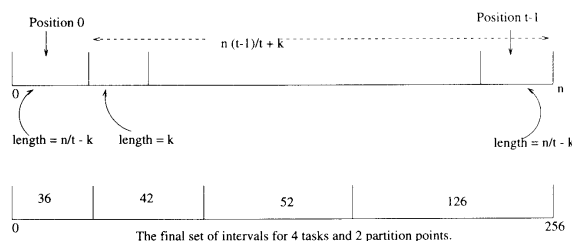


Fig. 19. Load-balancing algorithm layout.

Table 3
Comparison between various thresholding methods (bracketed numbers indicate the rank of each method, found by averaging the ratings)

Image	Threshold range	Ratings	
		Uniformity	Shape
Lena	Averaging	[3]	[2]
	0 98	0.687	0.767
	98 255	0.790	0.882
	2-segment entropic	[1]	[1]
	0 141	0.932	0.830
	141 255	0.812	0.844
	3-segment entropic	[2]	[3]
0 80	0.703	0.728	
80 149	0.681	0.795	
149 255	0.859	0.889	
Bandwidth 31	Fuzzy-entropic	[4]	[4]
	0 80	0.838	0.595
	30 63	0.594	0.488
	63 96	0.705	0.646
Objects	Averaging	[1]	[1]
	0 109	0.940	0.939
	109 255	0.812	0.934
	2-segment entropic	[2]	[2]
	0 72	0.932	0.830
	72 255	0.756	0.926
	3-segment entropic	[3]	[3]
0 76	0.933	0.923	
76 183	0.639	0.697	
183 255	0.859	0.953	
Bandwidth 40	Fuzzy-entropic	[4]	[4]
	0 120	0.931	0.940
	120 154	0.590	0.484
	154 255	0.851	0.948

probabilities. A uniform Gaussian noise distribution will mask emergent peaks.

6. Extension to the fuzzy-entropic method

Examining Eq. (7) shows that k , which is dependent on

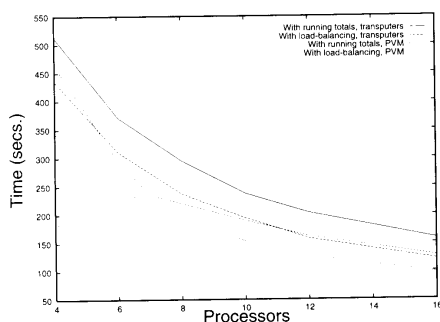


Fig. 20. Timing as the number of processors used is varied for the maximum-entropy algorithm, four segments.



Fig. 21. Original Lena image, size 512 × 512.

b , governs the symmetry of the function. Figs. 38 and 39 show the effect of varying b . When different choices of b were made and applied to actual images, the choice of b was found to make little difference to the thresholding points chosen, as Table 4 shows. This is because altering the position of b changes the effective bandwidth, by projecting more grey-levels either towards zero or one as a fuzzy grade, where they contribute little to the entropy measure. Another technique involving localized fuzzy regions was used on the histogram shown in Fig. 8. Clipping points were obtained for either side of each valley, so as to define a set of fuzzy regions for each valley. Values of fuzzy entropy were obtained by varying the crossover point across the fuzzy region. Fig. 40 illustrates the resulting local entropy curves, where the valley of each curve is close to the observed floor of each histogram valley. For Fig. 21, thresholding points were found at 46, 80 and 112. Further investigation will reveal whether the threshold points that were obtained represent a more meaningful segmentation.

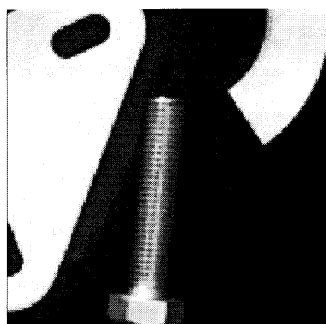


Fig. 22. Original objects image, size 500 × 500.



Fig. 23. Fuzzy-entropic thresholded Lenna image with bandwidth 31.

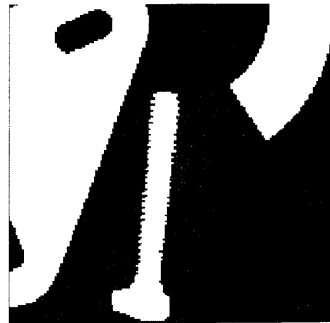


Fig. 26. Average method threshold of objects image.



Fig. 24. Maximum entropy method threshold of Lenna image using three segments.



Fig. 27. Maximum entropy method threshold of objects image using two segments.



Fig. 25. Fuzzy-entropic thresholded objects image with bandwidth 40.

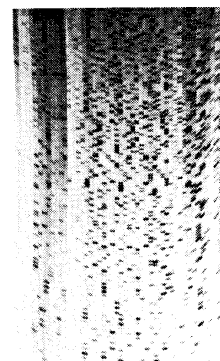


Fig. 28. DNA image, size 300 x 512 [24].

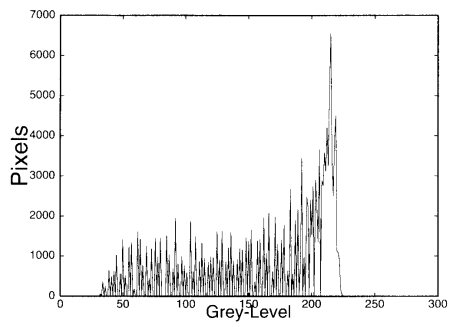


Fig. 29. DNA image (7 bits), grey-level histogram.

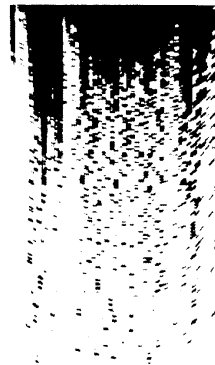


Fig. 32. DNA image, maximum-entropic method, threshold point at 134.

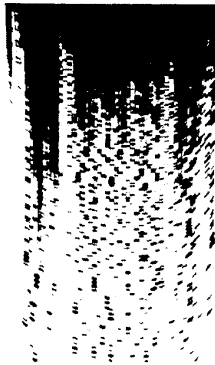


Fig. 30. DNA image, average method, threshold point at 164.



Fig. 33. Face image, size 256 x 256.

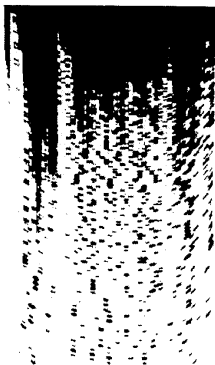


Fig. 31. DNA image, fuzzy-entropic method, bandwidth 28, threshold point at 160.

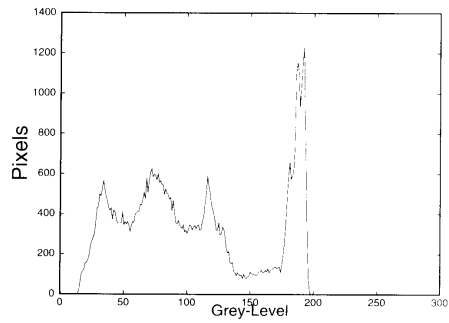


Fig. 34. Face image, grey-level histogram.



Fig. 35. Face image, average method, threshold point at 105.



Fig. 36. Face image, fuzzy-entropic method, bandwidth 27, threshold points at 38, 88 and 134.



Fig. 37. Face image, maximum-entropic method, threshold points at 75, 132 and 174.

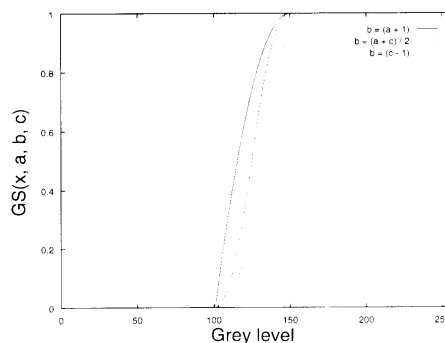


Fig. 38. Behaviour of a generalized S-function (GS).

7. Discussion and conclusion

We have examined the idea of using entropy as a method of thresholding, especially in regard to the question of whether fuzzy entropy represents an improvement on the use of information theoretic entropy. We used a maximum entropy method which is claimed as an improvement on previous entropy measures [4]. The theory of entropy, and its variant, fuzzy entropy, have been outlined: fuzzy entropy uses an extra grading step to reduce the calculation involved in forming a thresholding partition. In contrast to the maximum entropy method, which maximizes the information content provided by a suitable partition, fuzzy entropy finds the partition which makes the fuzzy set given by the partition grey-level members as 'crisp' as possible. We have provided details of robust procedures, including smoothing by a linear Gaussian filter, a clipping method for histogram

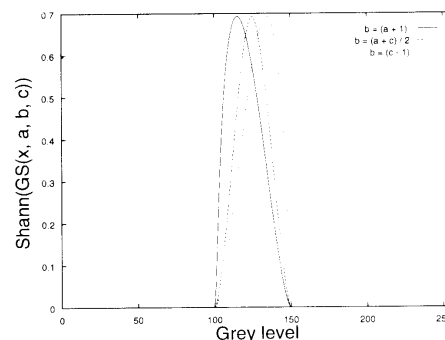


Fig. 39. Behaviour of the composition of Shannon's function with the GS function.

Table 4
Varying threshold values when using an asymmetrical membership function

Threshold point (Lemma)	Value of b relative to unit bandwidth		
	0.25	0.5	0.75
1	63	63	62
2	95	96	95
3	155	155	154

peak search and the method of selecting a suitable fuzzy bandwidth. The multi-threshold maximum entropy method will normally only be used for up to four segments because of the computational burden. We have devised a running total method for computing the maximum-entropy function, resulting in substantial computational savings. It is possible to successfully parallelize this algorithm, which is a species of combinatorial search, by means of a novel breakdown of a partitioning problem. Other multi-thresholding techniques, which use a similar partitioning strategy [23], should also benefit from this parallelization method. However, parallelization will still not enable it to compete in computational time with the fuzzy method. Both these entropic methods should be applied to images with the number of modes greater than two. Moreover, confidence in the fuzzy-logic method will only be general if the question of choice of membership function and bandwidth can be systematically resolved. In practice, so long as the choice is sensibly restricted, the precise choice is not significant. We have restricted attention to global methods because they are most suitable for automatic use, but it is possible to use local discrimination when applying to global threshold, perhaps by a fuzzy measure such as compactness, as outlined in work by Pal and Rosenfeld [6].

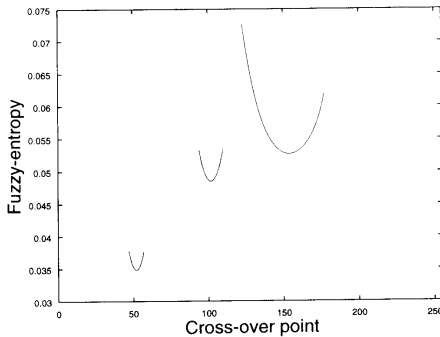


Fig. 40. Thresholds obtained for Fig. 8 by varying crossover point b .

Acknowledgement

This work was carried out as part of project IED3.12171 ('Parallel Reconfigurable Image Processing Systems').

References

- [1] Y. Nakagawa and A. Rosenfeld, Some experiments on variable thresholding, *Pattern Recognition*, 11 (1979) 191–204.
- [2] S.C. Sahasrabudhe and K.S.D. Gupta, A valley-seeking threshold selection technique, in L. Shapiro and A. Rosenfeld (eds.), *Computer Vision and Image Processing*, Academic Press, Boston, MA, 1992, pp. 55–65.
- [3] T. Pun, Entropic thresholding, a new approach, *Computer Vision, Graphics, and Image Processing*, 16 (1981) 210–239.
- [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, *Computer Vision, Graphics and Image Processing*, 29 (1985) 273–285.
- [5] S.K. Pal, R.A. King and A.A. Hashim, Automatic grey level thresholding through index of fuzziness and entropy, *Pattern Recogn. Lett.*, 11 (1983) 141–146.
- [6] S.K. Pal and A. Rosenfeld, Image enhancement and thresholding by optimization of fuzzy compactness, *Pattern Recogn. Lett.*, 7 (1988) 77–86.
- [7] S.K. Pal and A. Ghosh, Fuzzy geometry in image analysis, *Fuzzy Sets and Systems*, 48 (1992) 23–40.
- [8] P.K. Sahoo, S. Soltani, A.K.C. Wong and Y.C. Chen, A survey of thresholding techniques, *Computer Vision and Image Processing*, 41 (1988) 233–260.
- [9] S.S. Reddi, S.F. Rudin and H.R. Keshavan, An optimal multiple threshold scheme for image segmentation, *IEEE Trans. Systems, Man, and Cybernetics*, 14 (1984) 661–665.
- [10] J.G. Postaire and M. Ameziari, A pattern classification approach to multi-level thresholding for image segmentation, in L. Shapiro and A. Rosenfeld (eds.), *Computer Vision and Image Processing*, Academic Press, Boston, MA, 1992, pp. 307–328.
- [11] S.U. Lee and S.Y. Chung, A comparative performance study of several global thresholding techniques for segmentation, *Computer Vision, Graphics, and Image Processing*, 52 (1990) 171–190.
- [12] R.M. Berence, *Maximum Entropic Solutions to Scientific Problems*, Prentice Hall, Englewood Cliffs, NJ, 1993.
- [13] L.A. Zadeh, Calculus of fuzzy restrictions, in L.A. Zadeh, K.S. Fu, K. Tanaka and M. Shimura (eds.), *Fuzzy Sets and their Applications to Cognitive and Decision Processes*, pp. 1–39, Academic Press, London, 1975.
- [14] H. Li and H.S. Yang, Fast and reliable image enhancement using fuzzy relaxation technique, *IEEE Trans. Systems, Man, and Cybernetics*, 19(5) (1989) 1276–1281.
- [15] A.K. Jain, *Fundamentals of Digital Image Processing*, Prentice Hall, Englewood Cliffs, NJ, 1989.
- [16] R.C. Gonzalez and R.E. Woods, *Digital Image Processing*, Addison-Wesley, Reading, MA, 1992.
- [17] A. Geist, A. Beguelin, J. Dongarra, W. Jiang, R. Manick and P. Sunderam, *PVM 3 User's Guide and Reference Manual*, Oak Ridge National Laboratory, 1993.
- [18] Parsys Ltd., Boston Road, London, *Hardware Reference for the Parsys SN1000 Series*, 1989.
- [19] 3L Ltd., Peel House, Ladywell, Livingston, Scotland, *Parallel C Version 2.2.2*, 1991.
- [20] L. Sousa, J. Bourrios, A. Costa and M. Piedade, Parallel processing for transputer-based systems, *IEEE Int. Conf. Image Processing and its Applications*, pp. 33–36, 1992.

- [21] M.D. Levine and A.M. Nazif, Dynamic measurements of computer-generated image segmentation, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 7 (1985) pp. 155–164.
- [22] M.L. Sezan, A peak detection algorithm and its application to histogram-based image data reduction, *Computer Vision, Graphics, and Image Processing*, 49 (1990) 36–51.
- [23] J.-C. Yen, F.-J. Chang and S. Chang, A new criterion for automatic multilevel thresholding, *IEEE Trans. Image Processing*, 4(3) (1995) 370–378.
- [24] C.A. Glasbey and G.W. Horgan, *Image Analysis for the Biological Sciences*, Wiley, Chichester, UK, 1995.