Pergamon

Pattern Recognition, Vol. 29, No. 4, pp. 575–580, 1996
Elsevier Science Ltd

0031-3203/96 $15.00 + .00

0031-3203(95)00111-5

ON MINIMUM CROSS-ENTROPY THRESHOLDING

NIKHIL R. PAL
Machine Intelligence Unit, Indian Statistical Institute, 203 Barrackpore Trunk Road, Calcutta 700035, India

(*Received* 11 *April* 1994; *in revised form* 9 *March* 1995; *received for publication* 11 *August* 1995)

Abstract—Over last few years several papers have been written on entropy-based thresholding. Some of these methods use the gray-level histogram, while others use entropy associated with the two-dimensional histogram or the co-occurrence matrix. Recently, Li and Lee proposed a thresholding scheme that is aimed to minimize the cross entropy. This note first points out a certain conceptual problem and then proposes two algorithms which are free from that problem. The proposed schemes are tested on several data sets. The results are encouraging.

Information   Cross entropy   Thresholding   Segmentation   Object extraction

1. INTRODUCTION

Segmentation is the first essential and important step of low level vision. Segmentation can be carried out in various ways such as pixel classification, iterative pixel modification, edge detection, thresholding, etc.[1–4] Of these thresholding is probably the simplest and most widely used approach. Thresholding can be carried out based on global information such as histogram or local information like the co-occurrence matrix.[4–14] Some of these thresholding methods are guided by Shannon's entropy.[8,10,12–14]

Pun[13] assumed that an image is the outcome of an L symbol source. To select the threshold he maximized an upper bound of the total *a posterior* entropy of the partitioned image. Kapur *et al.*[14] on the other hand, assumed two probability distributions, one for the object area and the other for the background area. They then maximized the total entropy of the partitioned image in order to arrive at the threshold level. Wong and Shahoo[12] maximized the *a posterior* entropy of a partitioned image subject to a constraint on the uniformity measure of Levine and Nazif[11] and a shape measure. They maximized the *a posterior* entropy over $\min(s_1, s_2)$ and $\max(s_1, s_2)$ to obtain the threshold for segmentation, where s_1 and s_2 are the threshold levels at which the uniformity and the shape measure attain the maximum values, respectively. Pal and Pal[6] modeled the image as a mixture of two Poisson distributions and developed several parametric methods for segmentation. The assumption of the Poisson distribution has been justified based on the theory of image formation.

Recently, Li and Lee[15] claimed to use the directed divergence (cross entropy) of Kullback[16] for selection of threshold. Although they used an expression similar in structure to that of Kullback's cross entropy, the objective function used in reference (15) cannot be called divergence. This note deals with this issue and proposes some methods based on the true symmetric divergence or cross entropy. The proposed methods are tried on a set of synthetic histograms and also on some gray-level images. Results are compared with those obtained by the method of Li and Lee.

2. CROSS-ENTROPY METHOD OF LI AND LEE

Kullback[16] proposed an information theoretic distance D between two probability distributions. This information theoretic distance is known as directed divergence or cross entropy. Let $P = \{p_1, p_2, \ldots, p_n\}$ and $Q = \{q_1, q_2, \ldots, q_n\}$ be two probability distributions, then the cross entropy, $D(P, Q)$ is defined as follows:

$$D(P,Q) = \sum_{i=1}^{n} p_i \log \frac{p_i}{q_i}. \tag{1}$$

Since $\sum_{i=1}^{n} p_i = \sum_{i=1}^{n} q_i = 1$, Gibbs inequality (i.e. $-\sum p_i \log p_i \le -\sum p_i \log q_i$) ensures that $D(P,Q) \ge 0$ and the equality holds only if $p_i = q_i\ \forall_i = 1, 2, \ldots, n$. Clearly $D(P,Q)$ is not symmetric, i.e. $D(P,Q) \ne D(Q,P)$.

A symmetric version can be written as:

$$D_s(P, Q) = D(P, Q) + D(Q, P). \tag{2}$$

Let $F = [f(x,y)]_{M \times N}$ be an L-level digital image, $f(x,y) \in \{1, 2, \ldots, L\}$—the set of gray levels. Let h_i be the frequency of gray level i and $p_i = h_i/(M \times N)$ be the probability of occurrence of gray level i.

Suppose s is an assumed threshold. s partitions the image into two regions, namely, the object and the background. We assume gray values in $[0 - s]$ constitute the object region, while those in $[(s+1) - L]$ constitute the background. Li and Lee considered $[0 - (s-1)]$ for the object region, while $[s - L]$ as the background region, but it really does not matter. Li and Lee defined the cross entropy of a segmented image with threshold s as:

$$\eta(s) = \sum_{j=1}^{s} jh_j \log\left(\frac{j}{\mu_1(s)}\right) + \sum_{j=s+1}^{L} jh_j \log\left(\frac{j}{\mu_2(s)}\right), \quad (3)$$

where

$$\mu_1(s) = \sum_{j=1}^{s} (jh_j) / \sum_{j=1}^{s} h_j$$

and

$$\mu_2(s) = \sum_{j=s+1}^{L} (jh_j) / \sum_{j=s+1}^{L} h_j.$$

They emphasized the intensity conservation constraint, i.e. sum of the gray level of the original image in each partition should be equal to the sum of gray values in the corresponding partition of the segmented image. This constraint does not seem to be important for a segmentation algorithm. This can only help in visual comparison of the segmented output with the original image. For the purpose of segmentation, it really does not matter how one colors the segmented regions. Further, equation (3) cannot be termed cross entropy. The reason is neither $P = \{1, 2, \ldots, s\}$ nor $Q = \{\mu_1(s), \mu_1(s), \ldots, \mu_1(s)\}$, $|Q| = s$, can be termed or interpreted as a probability distribution.

Conceptually $P = \{1, 2, \ldots, s\}$ cannot have a probabilistic interpretation as it is a collection of gray values. Mathematically also it cannot be termed a probability distribution as, except for the first one, all other values are greater than unity. Cross entropy (also known as divergence) is a type of information theoretic distance between two probability distributions.

Thus, the method proposed in reference (15) cannot be called "minimum cross-entropy thresholding". However, this does not mean that equation (3) is not a good criterion for thresholding. In fact, it is indeed a reasonable segmentation criterion.

3. PROPOSED METHODS

Equation (1) is asymmetric; we shall use the symmetric version:

$$D_s(P, Q) = \sum_i p_i \log\frac{p_i}{q_i} + \sum_i q_i \log\frac{q_i}{p_i}. \quad (4)$$

Let s be the threshold for segmentation. Then the observed probability distribution of a gray level in the object region can be defined as:

$P_O = \{p_1^o, p_2^o, \ldots, p_s^o\}$ and that of the background region as:

$P_B = \{P_{s+1}^B, p_{s+2}^B, \ldots, p_L^B\}$, where

$$p_i^o = \frac{h_i}{P_s}, \quad i = 1, 2, \ldots, s; \quad P_s = \sum_{i=1}^{s} h_i; \quad (5)$$

$$p_i^B = \frac{h_i}{MN - P_s}, \quad i = s+1, s+2, \ldots, L. \quad (6)$$

Note that $\sum_{i=1}^{s} p_i^o = \sum_{i=s+1}^{L} p_i^B = 1$.

Let Q_O and Q_B be the probability distributions of the object and background regions, respectively, based on some model. In other words, $Q_O = \{q_1^o, q_2^o, \ldots, q_s^o\}$ and $Q_B = \{q_{s+1}^B, q_{s+2}^B, \ldots, q_L^B\}$ are the probability distributions estimated based on some model.

Now the cross entropy of the object region $D_O(s) = D_s(P_o, Q_o)$ and that of the background region $D_B(s) = D_s(P_B, Q_B)$. The total cross entropy of the segmented image can then be written as:

$$D(s) = D_O(s) + D_B(s) = \sum_{i=1}^{s} p_i^o \log\left(\frac{p_i^0}{q_i^0}\right) + \sum_{i=1}^{s} q_i^o \log\left(\frac{q_i^o}{p_i^o}\right) + \sum_{i=s+1}^{L} p_i^B \log\left(\frac{p_i^B}{q_i^B}\right) + \sum_{i=s+1}^{L} q_i^B \log\left(\frac{q_i^B}{p_i^B}\right). \quad (7)$$

In order to segment an image we minimize $D(s)$ with respect to s. However, to use equation (7) one needs to obtain Q_O and Q_B. Thus, next we address the issue of obtaining Q_O and Q_B.

The gray-level histogram is often modeled as a mixture of normal distributions.[17] Recently it has been established that image histograms are more appropriately modeled by a mixture of poisson distributions.[6] In reference (6) modeling of the gray-level histogram by a mixture of poisson distributions has been derived based on the theory of formation of image. We consider here the Poisson model to estimate Q_O and Q_B. We assume that the gray value in the object region follows the Poisson distribution with parameter λ_O and the gray level in the background region follows Poisson distribution with parameter λ_B. Thus:

$$q_i^o = \frac{e^{-\lambda_O}\lambda_O^i}{i!}, \quad i = 1, 2, \ldots, s, \quad (8)$$

and

$$q_i^B = \frac{e^{-\lambda_B}\lambda_B^i}{i!}, \quad i = s+1, s+2, \ldots, L; \quad (9)$$

where λ_O and λ_B can be estimated as:

$$\lambda_O = \left(\sum_{i=1}^{s} ih_i\right) \Big/ \sum_{l=1}^{s} h_i \quad (10)$$

and

$$\lambda_B = \left(\sum_{i=s+1}^{L} ih_i\right) \Big/ \sum_{i=s+1}^{L} h_i \quad (11)$$

Thus, the threshold algorithm becomes:

Algorithm 1

Begin
  Mindiv = High-value;
  for $2 \leqslant s \leqslant L-1$ do
    Begin
      Compute p_i^O, $i = 1, 2, \ldots, s$ using equation (5)
      Compute p_i^B, $i = s+1, \ldots, L$ using equation (6)

Compute λ_O using equation (10);
Compute λ_B using equation (11);
Compute q_i^O, $i = 1, 2, \ldots, s$ using equation (8);
Compute q_i^B, $i = s+1, \ldots, L$ using equation (9);
Compute $D(s)$ using equation (7);
if (mindiv > $D(s)$);
  Begin
    mindiv = $D(s)$;
    threshold = s;
  End
 End
End.

In Algorithm 1, the Gibbs inequality may not be satisfied by Q_O and Q_B, i.e. it may not be true that $\sum_{i=1}^{s} q_i^o = \sum_{i=s+1}^{L} q_i^B = 1$. In order to satisfy Gibbs in-

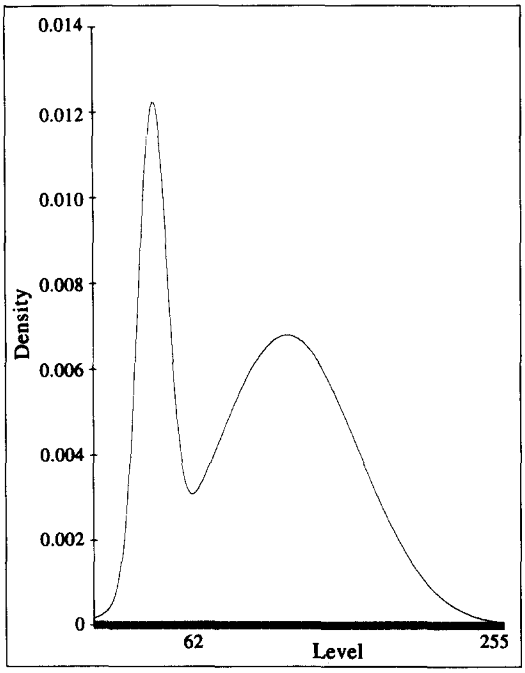


Fig. 1. Synthetic Histogram 1.

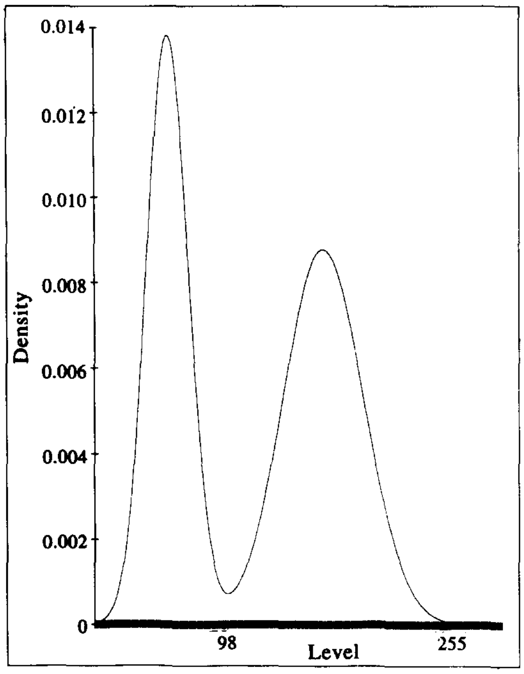


Fig. 3. Synthetic Histogram 3.

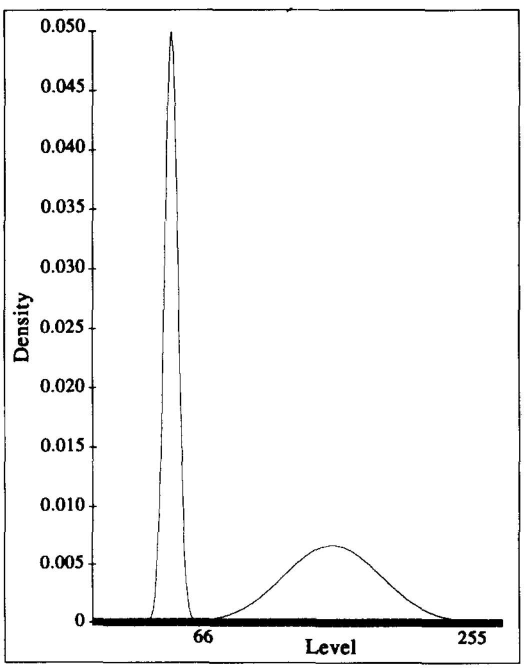


Fig. 2. Synthetic Histogram 2.

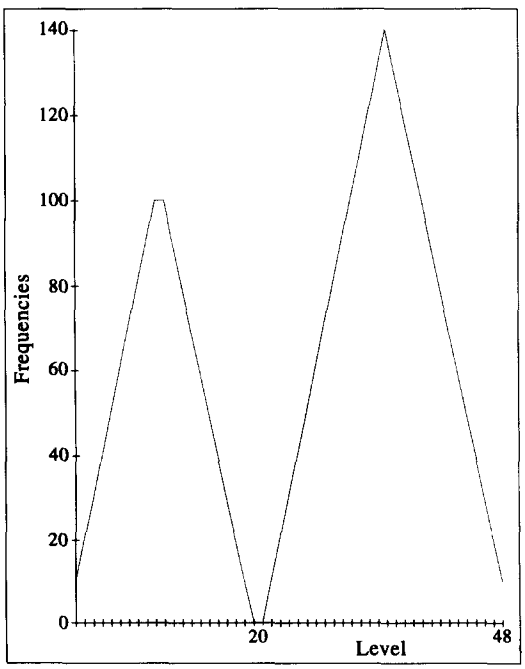


Fig. 4. Synthetic Histogram 4.

equality, we normalize Q_O and Q_B. The revised algorithm can be written as follows.

Algorithm 2

Begin
  Algorithm 1
    with Q_O normalized,
    and Q_B normalized;
    i.e. $\sum_{i=1}^{s} q_i^O = \sum_{i=s+1}^{L} q_i^B = 1.$
end.

4. RESULTS

We implemented both Algorithm 1 and Algorithm 2 and applied them on two images and four synthetically generated histograms. For the purpose of comparison we also implemented the method of Li and Lee. Three of the synthetically generated histograms [Figs 1–3] are the same as those used by Li and Lee. Each of the three is a mixture of two normal distributions. Figure 4 is a histogram for which threshold

Table 1.

Data set	Thresholds		
	Algorithm 1	Algorithm 2	Li and Lee
Histogram 1	58	49	82
Histogram 2	96	96	87
Histogram 3	58	58	92
Histogram 4	20	23	20
Biplane	13	13	14
SHU	13	13	12

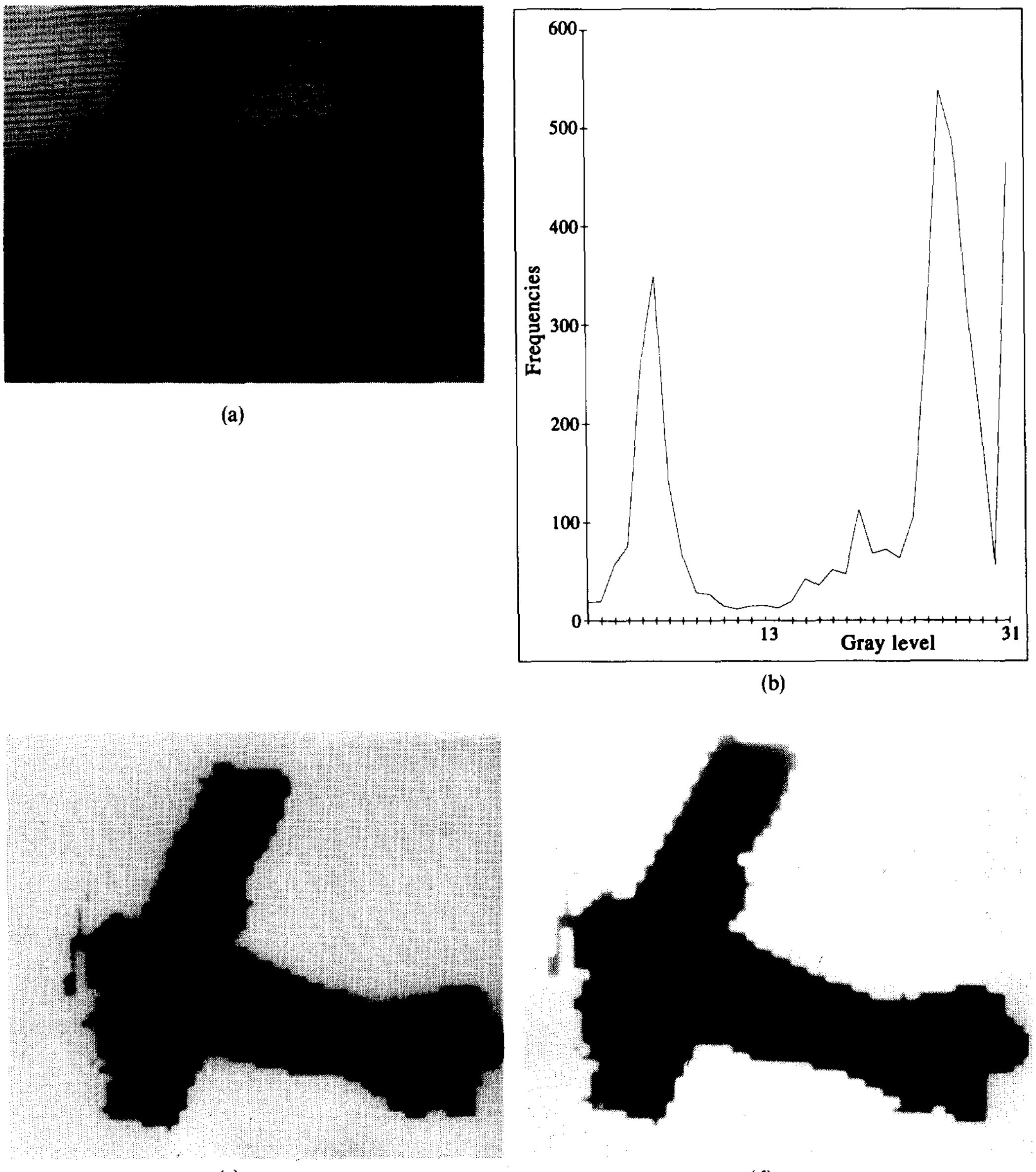


Fig. 5. Biplane image: (a) input; (b) histogram; (c) output by Algorithms 1 and 2; (d) output by Algorithm of Li and Lee.

selection is a trivial job. The purpose of this example is to check how good the proposed strategy is, when the histogram has an unambiguous valley.

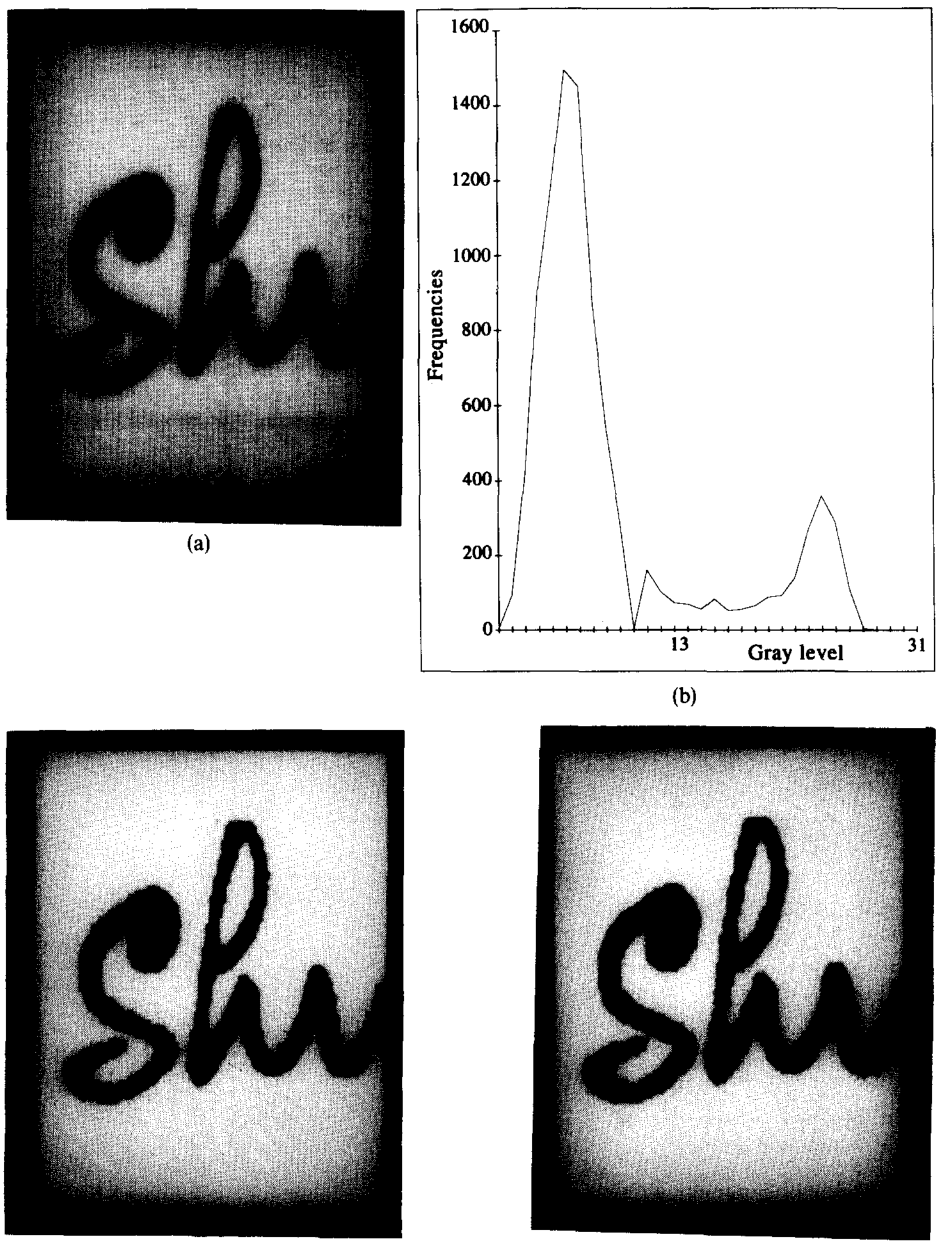
Table 1 depicts the thresholds obtained by different methods. Table 1 and Figs 1–4 clearly indicate that the proposed methods produce good thresholds. Note that here we approximated mixtures of normal distributions by mixtures of Poisson distributions, keeping in view the image segmentation problem. It is expected that for Figs 1–3, use of a normal distribution to compute Q_O and Q_B would result in more accurate thresholds.

Figures 5(a) and (b) represent an image of a biplane and its histogram, respectively, Fig. 5(c) displays the output produced by the two proposed methods, while Fig. 5(d) depicts the output obtained by the algorithm of Li and Lee. Results of all the algorithms are comparable.

For the SHU image [Fig. 6(a)], whose histogram is shown in Fig. 6(b), the proposed methods and the algorithm of Li and Lee, although producing different threshold levels, the segmented outputs [Figs 6(c) and (d)] are almost the same.

5. CONCLUSIONS

Based on Kullback's cross entropy Li and Lee proposed a method of thresholding what they termed



(a) (b) (c) (d)

Fig. 6. SHU image: (a) input; (b) histogram; (c) output by Algorithms 1 and 2; (d) output by Algorithm of Li and Lee.

"minimum cross-entropy thresholding". We pointed out why such a method cannot be called "minimum cross-entropy thresholding". Of course, the criteria used is a reasonable one for segmentation. We then proposed two algorithms based on the cross entropy (or Kullback's symmetric divergence). These algorithms are tested on a set of four synthetic histograms and two real images. The results obtained are encouraging.

REFERENCES

1. K. S. Fu and J. K. Mui, A survey on image segmentation, *Pattern Recognition* **13**, 3–16 (1981).
2. R. M. Haralick and L. G. Shapiro, Survey, image segmentation techniques, *Comput. Vis. Graphics Image Process.* **29**, 100–132 (1985).
3. P. K. Sahoo, S. Soltani and A. K. C. Wong, A survey of thresholding techniques, *Comput. Vis. Graphics Image Process.* **41**, 233–260 (1988).
4. N. R. Pal and S. K. Pal, A review on image segmentation techniques, *Pattern Recognition* **26**(9), 1277–1294 (1993).
5. N. Otsu, A threshold selection method from gray level histogram, *IEEE Trans. Syst. Man Cybernet.* **9**, 62–66 (1979).
6. N. R. Pal and S. K. Pal, Image model, Poisson distribution and object extraction, *Int. J. Pattern Recognition Artific. Intell.* **5**(3), 459–483 (1991).
7. N. R. Pal and S. K. Pal, Object-background segmentation using new definitions of entropy, *IEE Proc. E.* **136**(4), 284–295 (1989).
8. N. R. Pal and S. K. Pal, Entropic thresholding, *Signal Process.* **16**(2), 97–108 (1989).
9. N. R. Pal and S. K. Pal, Entropy: A new definition and its applications, *IEEE Trans. Syst. Man Cybernet.* **SMC-21**(5), 1260–1270 (1991).
10. N. R. Pal and D. Bhandari, Image thresholding: some new techniques, *Signal Process.* **33**(2), 139–158 (1993).
11. M. D. Levine and A. M. Nazif, Dynamic measurement of computer generated image segmentation, *IEEE Trans. Pattern Anal. Mach. Intell.* **7**, 155–164 (1985).
12. A. K. C. Wong and P. K. Sahoo, A gray level threshold selection method based on maximum entropy principle, *IEEE Trans. Syst., Man Cybernet.* **SMC-19**(4), 866–871 (July/August 1989).
13. T. Pun, A new method for gray level thresholding using the entropy of the histogram, *Signal Process.* **2**, 223–237 (1980).
14. J. N. Kapur, P. K. Sahoo and A. K. C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, *Comput. Graphics Vis. Image Process.* **29**, 273–285 (1985).
15. C. H. Li and C. K. Lee, Minimum cross entropy thresholding, *Pattern Recognition* **26**(4), 617–625 (1993).
16. S. Kullback, *Information Theory and Statistics*. Wiley, New York (1959).
17. R. C. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-wesley, Reading, Massachusetts (1987).

About the Author—NIKHIL R. PAL obtained a B.Sc. with honors in Physics and Master of Business Management from the University of Calcutta in 1979 and 1982, respectively. He obtained his M.Tech (Comp. Sc.) and Ph.D. (Comp. Sc.) from the Indian Statistical Institute in 1984 and 1991, respectively. Currently he is an associate professor in the Machine Intelligence Unit of Indian Statistical Institute, Calcutta. From September 1991 to February 1993 and July 1994 to December 1994 he was with the Computer Science Department of the University of West Florida. He was also a guest faculty of the University of Calcutta. His research interest includes image processing, fuzzy sets theory, measures of uncertainty, neural networks and genetic algorithms. He is an associate editor of the International Journal of Approximate Reasoning and IEEE Transactions on Fuzzy Systems.