COMPUTER VISION, GRAPHICS, AND IMAGE PROCESSING **29**, 273–285 (1985)

A New Method for Gray-Level Picture Thresholding Using the Entropy of the Histogram

J. N. KAPUR

Department of Mathematics, Indian Institute of Technology, Kanpur, India 208016

P. K. SAHOO

Department of Applied Mathematics, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

AND

A. K. C. WONG

Department of Systems Design, University of Waterloo, Waterloo, Ontario, Canada, N2L 3G1

Received August 27, 1982; revised April 30, 1984 and October 16, 1984

Two methods of entropic thresholding proposed by Pun (*Signal Process.* **2**, 1980, 223–237; *Comput. Graphics Image Process.* **16**, 1981, 210–239) have been carefully and critically examined. A new method with a sound theoretical foundation is proposed. Examples are given on a number of real and artifically generated histograms. © 1985 Academic Press, Inc.

1. INTRODUCTION

In picture processing, the most commonly used method in extracting objects from a picture is "thresholding." If the object is clearly distinguishable from the background, the gray-level histogram will be bimodal and the threshold for segmentation can be chosen at the bottom of the valley. However, gray-level histograms are not always bimodal. Methods other than valley-seeking are, thus, required to solve this problem.

Over the years several techniques have been proposed to overcome this difficulty. Most of them attempt to reduce the problem to the bimodal case. Weszka *et al.* [1] and Weszka and Rosenfeld [2] present methods to overcome the threshold selection problem when the peaks vary significantly in size and the valley is relatively wide. Others try to improve histograms by using second-order gray-level statistics as described in [3]. Still others attempt to make threshold selection easier to define an improved image histogram [4]. A method which uses the statistical relation of points and its neighbourhood is proposed in [5]. Besides these, there are a number of other methods such as Ostu's method [6], the iterative method [7], the quadtree method [8], and the relaxation method [9]. A survey of some of these methods may be found in [10, 11]. Recently, Pun [12, 13] has proposed two interesting algorithms based on entropy considerations for picture thresholding. But, while deriving a lower bound for the a posteriori entropy of the gray-level histogram [12] Pun made a few errors in algebraic manipulations. In Section 2, we will reformulate his algorithm (hereafter referred to as Algorithm 1) rectifying his errors. In Section 3, we will describe his other algorithm [13] (hereafter referred to as Algorithm 2). In both of these sections the strength and weakness of the two algorithms will be examined. In Section 4, we

0734-189X/85 $3.00
Copyright © 1985 by Academic Press, Inc.
All rights of reproduction in any form reserved

introduce a new algorithm (referred to as Algorithm 3) and discuss its theoretical foundations. In Section 5, we give some examples of thresholding using the newly proposed method. Finally, in Section 6, we extend Algorithm 3 to "multithresholding."

2. REFORMULATION AND ASSESSMENT OF ALGORITHM 1

Let $f_1, f_2, \ldots, f_n$ be the observed gray-level frequencies and let

$$p_i = \frac{f_i}{N}, \qquad \sum_{i=1}^{n} f_i = N, \qquad i = 1, 2, \ldots, n, \tag{1}$$

where N is the total number of pixels in the picture and n is the number of gray-levels. The a posteriori entropy is defined by

$$H_n' = -P_s \ln P_s - (1 - P_s)\ln(1 - P_s), \tag{2}$$

where

$$P_s = \sum_{i=1}^{s} p_i, \qquad 1 - P_s = \sum_{i=s+1}^{n} p_i \tag{3}$$

If we maximize H_n', we get

$$P_s = 1 - P_s = \tfrac{1}{2} \tag{4}$$

which gives an equal number of white and black pixels. To avoid this apparently trivial result we proceed as follows:

$$\begin{aligned}\sum_{i=1}^{s} p_i \ln p_i &\le \sum_{i=1}^{s} p_i \ln[\max(p_1, p_2, \ldots, p_s)] \\ &= P_s \ln[\max(p_1, p_2, \ldots, p_s)].\end{aligned}$$

Hence

$$-\sum_{i=1}^{s} p_i \ln p_i \ge -P_s \ln[\max(p_1, p_2, \ldots, p_s)].$$

This implies

$$P_s \le \frac{H_s}{-\ln[\max(p_1, p_2, \ldots, p_s)]}, \tag{5}$$

where

$$H_s = -\sum_{i=1}^{s} p_i \ln p_i$$

From (5), we get

$$-P_s \ln P_s \le \frac{H_s \ln P_s}{\ln[\max(p_1, p_2, \ldots, p_s)]}. \tag{6}$$

Similarly

$$-(1-P_s)\ln(1-P_s) \le \frac{(H_n - H_s)\ln(1-P_s)}{\ln[\max(p_{s+1}, p_{s+2}, \ldots, p_n)]}, \tag{7}$$

where

$$H_n = -\sum_{i=1}^{n} p_i \ln p_i.$$

Using (6) and (7) in (2), we obtain

$$H_n' \le H_n \left[\frac{H_s}{H_n} \frac{\ln P_s}{\ln[\max(p_1, p_2, \ldots, p_s)]} + \left(1 - \frac{H_s}{H_n}\right) \frac{\ln(1-P_s)}{\ln[\max(p_{s+1}, p_{s+2}, \ldots, p_n)]} \right].$$

Defining

$$f(s) = \left(\frac{H_s}{H_n}\right) \frac{\ln P_s}{\ln[\max(p_1, p_2, \ldots, p_s)]} + \left(1 - \frac{H_s}{H_n}\right) \frac{\ln(1-P_s)}{\ln[\max(p_{s+1}, p_{s+2}, \ldots, p_n)]}$$

and rewriting the above inequality, we obtain

$$\frac{H_n'}{H_n} \le f(s). \tag{8}$$

The function $f(s)$ is called evaluation function [12]. According to Algorithm 1 the threshold value is the s which maximizes the evaluation function $f(s)$. Due to an error in Eq. (18) of [12], Pun mistakenly obtained (8) as

$$\frac{H_n'}{H_n} \ge f(s). \tag{9}$$

Hence $H_n \cdot f(s)$ gives the upper bound and not the lower bound of H_n', as claimed by Pun [12]. Based on the inequality (8) the following comments are in order.

(a) Let $\phi(s) = H_n'/H_n$, then from (8), we get

$$\phi(s) \le f(s). \tag{10}$$

The graph of $\phi(s)$ has a fixed shape since it is related to Shannon's function through Eq. (2) whereas the graph of $f(s)$ has no fixed shape. The dotted curves on Fig. 1 represent three such arbitrary graphs of $f(s)$. By Eq. (4), the maximum of $f(s)$ is

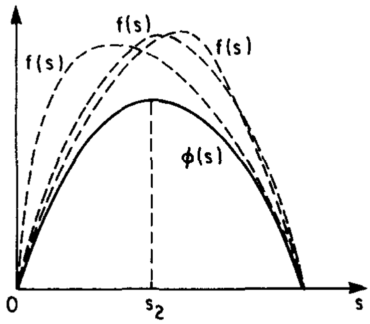

FIG. 1. The shape of $\phi(s)$ and some possible shapes of $f(s)$.

always greater than or equal to that of $\phi(s)$. However, there is no relationship between s_1 (the value of s at which the maximum of $f(s)$ occurs) and s_2 (the value of s at which the maximum of $\phi(s)$ occurs). s_2 always satisfies (4), but as is clear from Fig. 1, s_1 can be less than s_2, equal to s_2, or even greater than s_2.

Based on his numerical calculation and using image data, Pun [13] found that s_1 is nearly the same as s_2 and the closeness became pronounced as the smaller blocks of the picture are taken into account. This result is not altogether unexpected in view of the observation just made.

(b) The maximization of $f(s)$ thus does not achieve a priori maximization of a posteriori entropy H_n'.

(c) The reason why H_n' should be maximized has not been explained.

(d) Since H_n' is majorized by another function, therefore the algorithm does not always use the statistical properties of the gray-level histogram.

(e) The function $\phi(s)$ can be majorized by $sf(s)$, $s^2f(s)$, or $\alpha(s)f(s)$, where $\alpha(s)$ is any positive function which is always greater than or equal to 1. But we cannot take the values of s at which these functions attain maxima as threshold value.

3. ASSESSMENT OF ALGORITHM 2

We will first describe Algorithm 2 (for the convenience of the readers) and then critically examine its strength and weakness. Let the anisotropy coefficient α be defined by

$$\alpha = \frac{\sum_{i=1}^{m} p_i \ln p_i}{\sum_{i=1}^{n} p_i \ln p_i}, \tag{11}$$

where m is the smallest integer for which

$$\sum_{i=1}^{s} p_i = \begin{cases} 1 - \alpha & \text{if} \quad \alpha \le 0.5, \\ \alpha & \text{if} \quad \alpha > 0.5. \end{cases} \tag{12}$$

In this algorithm, the lowest integer value m which divides the total frequencies into two or nearly two equal parts is determined and then the ratio α of the entropy obtained for that partition is computed. If $\alpha < 0.5$ we choose s such that $P_s = 1 - \alpha$. If $\alpha < 0.5$, $1 - \alpha$ is greater than 0.5 and hence from (12), we see that $s > m$. Again when $\alpha > 0.5$ the threshold value s is chosen such that $P_s = \alpha$. When $\alpha > 0.5$, from (12) s must be greater than m. When $\alpha = 0.5$, s is equal to m. Hence by this algorithm the threshold value is always greater than or equal to m. In general, this algorithm produces less black than white pixels when "0" denotes the whiteness in the gray scale and thus introduces unnecessary bias.

This algorithm may give qualitatively sound results, but there is no guarantee that the results would always be satisfactory. In fact, a similar aniostropic heuristic algorithm could be obtained by considering the function

$$g(s) = \frac{\sum_{i=1}^{s} p_i \ln p_i}{\left(\sum_{i=1}^{s} p_i\right)\left(\sum_{i=1}^{n} p_i \ln p_i\right)}. \tag{13}$$

The threshold can be obtained by finding the value of s $(\leq n)$ for which $g(s)$ is nearer to unity.

4. THE NEW ALGORITHM

In this section, we will propose a new algorithm based also on the entropy concept. Let $p_1, p_2, \ldots, p_n$ be the probability distribution of gray-levels. Now we derive from this distribution two probability distributions, one defined for discrete values 1 to s and the other for values from $s + 1$ to n. The two distributions are

$$A: \frac{p_1}{P_s}, \frac{p_2}{P_s}, \ldots, \frac{p_s}{P_s} \tag{14}$$

$$B: \frac{p_{s+1}}{1 - P_s}, \frac{p_{s+2}}{1 - P_s}, \ldots, \frac{p_n}{1 - P_s}. \tag{15}$$

The entropies associated with each distribution are as follows:

$$\begin{aligned} H(A) &= -\sum_{i=1}^{s} \frac{p_i}{P_s} \ln \frac{p_i}{P_s} \\ &= -\frac{1}{P_s}\left[\sum_{i=1}^{s} p_i \ln p_1 - P_s \ln P_s\right] \\ &= \ln P_s + \frac{H_s}{P_s}. \end{aligned} \tag{16}$$

Similarly

$$\begin{aligned} H(B) &= -\sum_{i=1+s}^{n} \frac{p_i}{1-P_s}\ln\frac{p_i}{1-P_s} \\ &= -\frac{1}{1-P_s}\left[\sum_{i=s+1}^{n} p_i \ln p_i - (1-P_s)\ln(1-P_s)\right] \\ &= \ln(1-P_s) + \frac{H_n - H_s}{1-P_s}. \end{aligned} \tag{17}$$

Let us define the sum of $H(A)$ and $H(B)$ by $\psi(s)$. Hence from (16) and (17), we obtain

$$\psi(s) = \ln P_s(1-P_s) + \frac{H_s}{P_s} + \frac{H_n - H_s}{1-P_s}. \tag{18}$$

We maximize $\psi(s)$ to obtain the maximum information between the object and background distributions in the picture. The discrete value s which maximizes $\psi(s)$ is the threshold value.

If the distribution A and B are identical or similar, $\psi(s)$ is maximum. If the distributions are uniform, then

$$\begin{aligned} \psi(s) &= \ln\frac{s}{n}\left(1-\frac{s}{n}\right) + \frac{-s}{n}\ln\frac{1}{n}\Big/\frac{s}{n} + \frac{-(n-s)}{n}\ln\frac{1}{n}\Big/\frac{(n-s)}{n} \\ &= \ln s(n-s) \end{aligned} \tag{19}$$

and attains the maximum when $s = \frac{1}{2}n$. If the original gray level distribution $(p_1, p_2, \ldots, p_n)$ is symmetrical in the sense that

$$p_i = p_{n+1-i}, \qquad i = 1, 2, 3, \ldots, n \tag{20}$$

then for $s = \frac{1}{2}n$ (assume n even)

$$\ln P_s + \frac{H_s}{P_s} = \ln(1-P_s) + \frac{H_n - H_s}{1-P_s}, \tag{21}$$

i.e., $H(A) = H(B)$. And, in general, $\psi(s)$ for symmetrical distribution will be at its maximum when $s = \frac{1}{2}n$.

5. SOME EXPERIMENTAL RESULTS

First we will discuss the choice of threshold value and then we will present the threshold values of some real and artificially generated pictures. The graph of $\psi(s)$ may have one of the following shapes (Figs. 2a–e): In Fig. 2a s_0 is the obvious choice for threshold value. In Fig. 2b there is a choice between s_1 and s_2. In practice, we would like to maximize $\psi(s)$, but we also want to have neither too many, nor too few, black pixels. If the nuber of black pixels corresponding to the threshold value s_1 is sufficient, we choose s_1, otherwise we choose to have the threshold value larger than s_1. Similarly in Figs. 2c–e we choose the threshold value

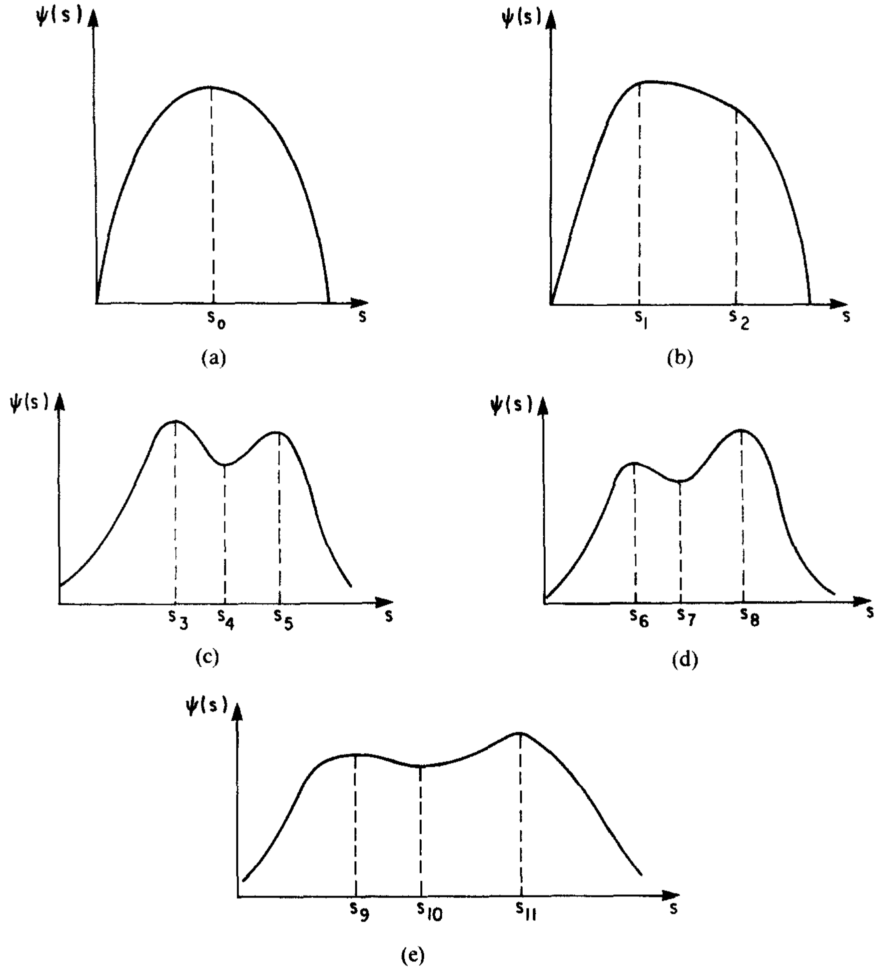

FIG. 2. Some examples of the shape of $\psi(s)$.

by means of a suitable trade-off between maximizing $\psi(s)$ and having a reasonable amount of black and white pixels for a good thresholding.

Next we will describe our experimental results on four different pictures. These four pictures are (a) a cloud cover picture [14], (b) a building picture [12], (c) a cameraman picture [12], and (d) pictures of a model. We determine the threshold for these pictures first without smoothing the gray-level frequency data, and then with smoothing using the following formula [15]:

$$F'(i) = \frac{F(i-2) + 2F(i-1) + 3F(i) + 2F(i+1) + F(i+2)}{9}. \tag{22}$$

It should be mentioned that for the following four experiments, the smoothing of the frequency data has no effect on the threshold value when Algorithm 3 is used.

(a) *Cloud cover picture*. The data for this picture was taken from [14]. Using Algorithm 3 the threshold value was found to be 26 on a 0–63 gray scale. In their

FIG. 3. Original digitized building picture.

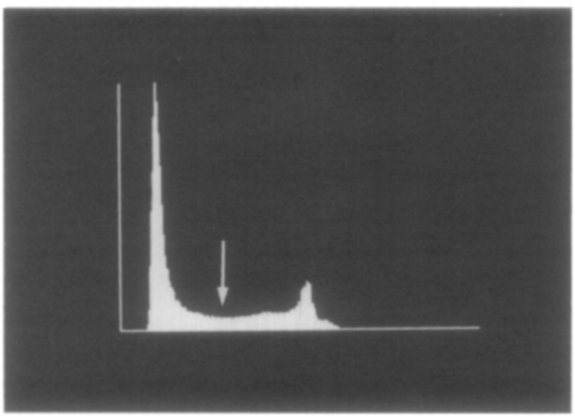

FIG. 4. Gray level histogram of building picture. The arrow indicates the position of the threshold value.

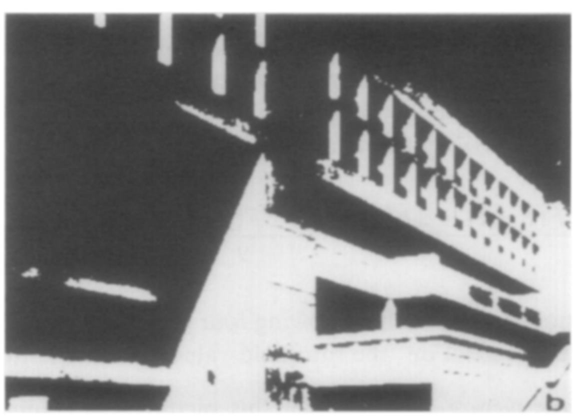

FIG. 5. Thresholded building picture.

FIG. 6. Original digitized picture of cameraman.

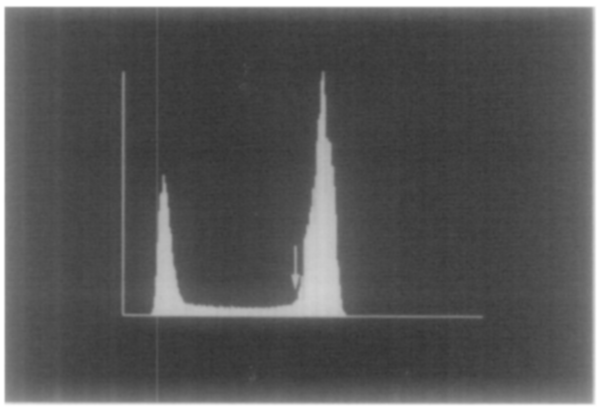

FIG. 7. Gray level histogram of picture of cameraman. The arrow indicates the position of the threshold value.

FIG. 8. Threshold picture of cameraman.

FIG. 9. Original digitized "Picture 1."

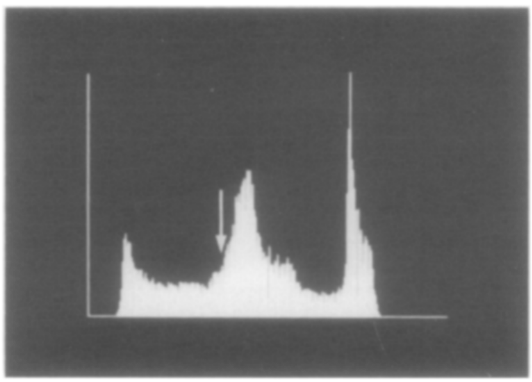

FIG. 10. Gray level histogram of "Picture 1." The arrow indicates the position of the threshold value

FIG. 11. Threshold image of "Picture 1."

FIG. 12. Original digitized "Picture 2."

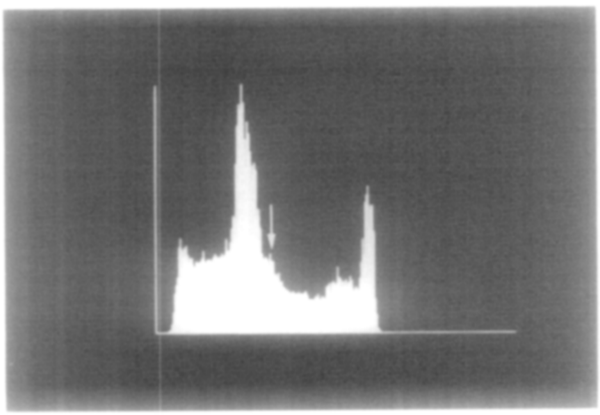

FIG. 13. Gray level histogram of "Picture 2." The arrow indicates the position of the threshold value.

FIG. 14. Threshold image of "Picture 2."

book, Rosenfeld and Kak [14] have mentioned that the good threshold for the cloud cover picture should be 29 on a 0–63 gray scale. When the data was smooth using (22), the threshold was found to be again 26.

(b) *Building picture*. This picture was taken from [12] and was digitized with a raster of 411 × 403 and quantized to 256 gray levels. Figures 3 and 4 show the original digitized image and its gray-level histogram is unimodal. Algorithm 3 is applied to this picture, a threshold value of 69 is obtained on a 0–255 gray scale with "0" denoting darkness. The threshold image is shown in Fig. 5.

(c) *Cameraman picture*. This picture was also taken from [12] and was digitized with a raster of 415 × 396 and quantized to 256 gray levels. The threshold value is found to be 123 on a 0–255 gray scale. Figures 6–8 show the original digitized image, its gray-level histogram, and the threshold image, respectively.

(d) *Pictures of a model*. Two different pictures of a Canadian model were taken and digitized with a raster of 326 × 322 and 321 × 314, respectively, and quantized to 256 gray-levels. For the convenience of description the first digitized image of the model is called Picture 1. The original digitized image of this picture is shown in Fig. 9 along with its gray-level histogram in Fig. 10. The other image is referred to as Picture 2. When Algorithm 3 is applied to Picture 1 and Picture 2 the threshold values of 96 and 81 are found, respectively. Figure 11 shows the thresholded image of Picture 1 while Figs. 12–14 show the original digitized image of Picture 2, its gray-level histogram, and the threshold image.

On the basis of a preliminary investigation with some real world images we have come to the conclusion that this new entropic method of thresholding gives a better threshold value when compared to other automatic global thresholding methods, viz. Ostu [6], Pun [12], Johansen and Bill [16], and the mutual information method [17]. The results of this investigation [18] will be reported elsewhere.

6. EXTENSION TO MULTITHRESHOLDING

If two or more objects are superimposed on the same background so that the gray-level histogram is multimodal, we calculate

$$\psi(s_1, s_2, \ldots, s_k) = \ln\left(\sum_{i=1}^{s_1} p_i\right) + \ln\left(\sum_{i=s_1+1}^{s_2} p_i\right) + \cdots + \ln\left(\sum_{i=s_k+1}^{n} p_i\right) - \frac{\sum_{i=1}^{s_1} p_i \ln p_i}{\sum_{i=1}^{s_1} p_i} - \cdots - \frac{\sum_{i=s_k+1}^{n} p_i \ln p_i}{\sum_{i=s_k+1}^{n} p_i}, \tag{24}$$

where $s_1, s_2, \ldots, s_k$ are integers and lie in the interval $[0, n]$, satisfying the condition $s_1 \leq s_2 \leq \cdots \leq s_k$. We choose $s_1, s_2, \ldots, s_k$ to maximize ψ. The number of values to be computed is equal to the number of partitions of $[0, n]$ into $k + 1$ parts. The partition which maximizes ψ gives the desired threshold vector.

7. CONCLUDING REMARKS

An algorithm (i.e., Algorithm 3) for choosing a threshold from the gray-level histogram of a picture has been derived by using the entropy concept from information theory. The advantage of this algorithm is that it uses a global and objective property of the histogram. Because of its general nature, this algorithm can be used for segmentation purposes. However, one still encounters the following problems. What happens if two different pictures have the same gray-level histogram and thus the same threshold? Will it be suitable for both? A second-order statistic or some local property with our entropic concept of thresholding might give a better insight into these problems.

ACKNOWLEDGMENTS

We are thankful to Professor A. Rosenfeld of Maryland University for providing us with some very useful image data. To Rani Brownstein, Miguel de Lascurian, Deborah Stacey, and Alexis Brown, we owe special thanks for their help and suggestions. The authors are grateful to the referee for suggestions and improvements.

REFERENCES

1. J. S. Weszka, R. N. Nagel, and A. Rosenfeld, A threshold selection technique, *IEEE Trans. Comput.* **C-23**, 1974, 1322–1326.
2. J. S. Weszka and A. Rosenfeld, Histogram modification for threshold selection, *IEEE Trans. Syst. Man Cybern.* **SMC-9**, 1979, 38–52.
3. N. Ahuja and A. Rosenfeld, A note on the use of second order gray-level statistics for threshold selection, *IEEE Trans. Syst. Man Cybern.* **SMC-8**, 1978, 895–898.
4. R. L. Kirby and A. Rosenfeld, A note on the use of (gray level, local average gray level) space as an aid in threshold selection, *IEEE Trans. Syst. Man Cybern.* **SMC-9**, 1979, 860–864.
5. L. S. Davis, A. Rosenfeld, and J. S. Weszka, Region extraction by averaging and thresholding, *IEEE Trans. Syst. Man Cybern.* **SMC-5**, 1975, 383–388.
6. N. Ostu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* **SMC-9**, 1979, 62–66.
7. T. W. Ridler and S. Calvard, Picture thresholding using an iterative selection method, *IEEE Trans. Syst. Man Cybern.* **SMC-8**, 1978, 630–632.
8. A. Y. Wu, T. Hong, and A. Rosenfeld, Threshold selection using quadtrees, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAM1-4**, 1982, 90–93.
9. A. Rosenfeld and R. C. Smith, Thresholding using relaxation, *IEEE Trans. Pattern Anal. Mach. Intell.* **PAM1-3**, 1981, 598–606.
10. J. S. Weszka, A survey of threshold selection techniques, *Comput. Graphics Image Process.* **7**, 1978, 259–265.
11. K. S. Fu and J. K. Mui, A survey on image segmentation, *Pattern Recognition* **13**, 1980, 3–16.
12. T. Pun, A new method for gray-level picture thresholding using the entropy of the histogram, *Signal Process.* **2**, 1980, 223–237.
13. T. Pun, Entropic thresholding: A new approach, *Comput. Graphics Image Process.* **16**, 1981, 210–239.
14. A. Rosenfeld and A. Kak, *Digital Picture Processing*, Academic Press, New York, 1976.
15. Y. Nakagawa and A. Rosenfeld, Some experiments on variable thresholding, *Pattern Recognition* **11**, 1979, 191–204.
16. G. Johannsen and J. Bille, A threshold selection method using information measures, in *Proc. 6th Int. Conf. on Pattern Recognition*, Oct. 1982.
17. P. K. Sahoo, Y. C. Chan, and A. K. C. Wong, A survey of thresholding methods, submitted.
18. P. K. Sahoo, Y. C. Chan, and A. K. C. Wong, *Evaluation of Some Global Thresholding Techniques*, Technical Report No. 126-R-110484, Department of Systems Design, University of Waterloo, Waterloo, Ontario.