

# MINIMUM ERROR THRESHOLDING

J. KITTLER and J. ILLINGWORTH

SERC Rutherford Appleton Laboratory, Chilton, Didcot, Oxon OX11 0QX, U.K.

(Received 15 August 1984; in final form 13 June 1985; received for publication 8 July 1985)

**Abstract**—A computationally efficient solution to the problem of minimum error thresholding is derived under the assumption of object and pixel grey level values being normally distributed. The method is applicable in multithreshold selection.

Thresholding      Minimum error decision rule      Classification error      Dynamic clustering

## 1. INTRODUCTION

Thresholding is a popular tool for segmenting grey level images. The approach is based on the assumption that object and background pixels in the image can be distinguished by their grey level values. By judiciously choosing a grey level threshold between the dominant values of object and background intensities the original grey level image can be transformed into a binary form so that the image points associated with the objects and the background will assume values one and zero, respectively. Although the method appears to be simplistic, it is very important and fundamental, with wide applicability, as it is relevant not only for segmenting the original sensor data but also for segmenting its linear and non-linear image-to-image transforms.

Apart from the recently proposed direct threshold selection method,<sup>(1, 2)</sup> determination of a suitable threshold involves the computation of the histogram or other function of the grey level intensity and its subsequent analysis. For a more detailed review of various approaches to thresholding the reader is referred to Kittler *et al.*<sup>(3)</sup>

An effective approach is to consider thresholding as a classification problem. If the grey level distributions of object and background pixels are known or can be estimated, then the optimal, minimum error threshold can be obtained using the results of statistical decision theory.<sup>(4)</sup>

It is often realistic to assume that the respective populations are distributed normally with distinct means and standard deviations. Under this assumption the population parameters can be inferred from the grey level histogram by fitting, as advocated by Nagawa and Rosenfeld<sup>(5)</sup> and then the corresponding optimal threshold can be determined. However, their approach is computationally involved. In this paper we propose an alternative solution which is more efficient. The principal idea behind the method is to optimise a criterion function related to the average pixel classification error rate. The approach is devel-

oped in Section 2 and its properties discussed in Section 3. The method can be easily extended to cope with the problem of multiple threshold selection, as will be shown in Section 4. Finally, an iterative implementation of the method is described in Section 5.

## 2. METHOD OF THRESHOLD SELECTION

Let us consider an image whose pixels assume grey level values,  $g$ , from the interval  $[0, n]$ . It is convenient to summarise the distribution of the grey levels in the image in the form of a histogram  $h(g)$  which gives the frequency of occurrence of each grey level in the image. The histogram can be viewed as an estimate of the probability density function  $p(g)$  of the mixture population comprising grey levels of object and background pixels. In the following we shall assume that each of the two components  $p(g|i)$  of the mixture is normally distributed with mean  $\mu_i$ , standard deviation  $\sigma_i$  and a priori probability  $P_i$ , i.e.

$$p(g) = \sum_{i=1}^2 P_i p(g|i), \quad (1)$$

where

$$p(g|i) = \frac{1}{\sqrt{2\pi} \sigma_i} \exp\left(-\frac{(g - \mu_i)^2}{2\sigma_i^2}\right). \quad (2)$$

For given  $p(g|i)$  and  $P_i$  there exists a grey level  $\tau$  for which grey levels  $g$  satisfy (for example Devijver and Kittler<sup>(6)</sup>)

$$P_1 p(g|1) \geq P_2 p(g|2) \begin{cases} g \leq \tau \\ g > \tau \end{cases}. \quad (3)$$

$\tau$  is the Bayes minimum error threshold at which the image should be binarised. Taking the logarithm of both sides in (3), this condition can be re-expressed as

$$\begin{aligned} \frac{(g - \mu_1)^2}{\sigma_1^2} + \log \sigma_1^2 - 2 \log P_1 &\leq \frac{(g - \mu_2)^2}{\sigma_2^2} \\ &+ \log \sigma_2^2 - 2 \log P_2 \begin{cases} g \leq \tau \\ g > \tau \end{cases}. \end{aligned} \quad (4)$$

The problem of minimum error threshold selection is to determine the threshold level  $\tau$ .

The minimum error threshold can be obtained by solving the quadratic equation defined by equating the left and right hand sides of (4). However, the parameters  $\mu_i$ ,  $\sigma_i$  and  $P_i$  of the mixture density  $p(g)$  associated with an image to be thresholded will not usually be known. Nevertheless, these parameters can be estimated from the grey level histogram  $h(g)$  using fitting techniques. This approach has been advocated by Nagawa and Rosenfeld.<sup>(5)</sup> Computationally their method is very involved, as it requires optimisation of a goodness-of-fit criterion function by a hill climbing procedure.

In this paper we derive a much simpler technique for finding the optimum threshold  $\tau$ . Suppose that we threshold the grey level data at some arbitrary level  $T$  and model each of the two resulting pixel populations by a normal density  $h(g|i, T)$  with parameters  $\mu_i(T)$ ,  $\sigma_i(T)$  and *a priori* probability  $P_i(T)$  given, respectively, as

$$P_i(T) = \sum_{g=a}^b h(g), \quad (5)$$

$$\mu_i(T) = \left[ \sum_{g=a}^b h(g) g \right] / P_i(T) \quad (6)$$

and

$$\sigma_i^2(T) = \left[ \sum_{g=a}^b \{g - \mu_i(T)\}^2 h(g) \right] / P_i(T), \quad (7)$$

where

$$a = \begin{cases} 0 & i = 1 \\ T + 1 & i = 2 \end{cases} \quad (8)$$

and

$$b = \begin{cases} T & i = 1 \\ n & i = 2 \end{cases}. \quad (9)$$

Now using the models  $h(g|i, T)$ ,  $i = 1, 2$ , the conditional probability  $e(g, T)$  of grey level  $g$  being replaced in the image by a correct binary value is given by

$$e(g, T) = h(g|i, T) \cdot P_i(T) / h(g) \quad i = \begin{cases} 1 & g \leq T \\ 2 & g > T \end{cases}. \quad (10)$$

As  $h(g)$  is independent of both  $i$  and  $T$ , the denominator in (10) will for the moment be ignored in our analysis. Taking the logarithm of the numerator in (10) and multiplying the result by  $-2$  we get the quantity

$$e(g, T) = \left[ \frac{g - \mu_i(T)}{\sigma_i(T)} \right]^2 + 2 \log \sigma_i(T) - 2 \log P_i(T) \quad (11)$$

$$i = \begin{cases} 1 & g \leq T \\ 2 & g > T \end{cases},$$

which can be considered as an alternative index of correct classification performance. The average perfor-

mance figure for the whole image can then be characterised by the criterion function

$$J(T) = \sum_g h(g) \cdot e(g, T). \quad (12)$$

The role of this criterion function in finding the (approximate) minimum error threshold is rather subtle. For a given threshold  $T$  the criterion reflects indirectly the amount of overlap between the Gaussian models of the object and background populations. Note that the density functions defining the models will overlap, even though, as a result of the histogram partitioning at threshold  $T$ , the tails of the distributions are ignored in their estimates.

As threshold  $T$  is varied the models of the population distributions change. The better the fit between the data and the models, the smaller the overlap between the density functions and therefore the smaller the classification error. This behaviour is illustrated in Fig. 1 for two different values of threshold  $T$ . The corresponding overlaps of the density functions are indicated by hatching. The value of threshold  $T$  yielding the lowest value of criterion  $J(T)$  will give the best fit model and therefore the minimum error threshold.

In summary, instead of determining  $\tau$  indirectly as in Nagawa and Rosenfeld,<sup>(5)</sup> the problem of minimum error threshold selection can be formulated as one of minimising criterion  $J(T)$ , i.e.

$$J(\tau) = \min_T J(T). \quad (13)$$

It should be noted that even at the optimum value  $\tau$ , the models  $h(g|i, \tau)$ ,  $i = 1, 2$  will represent biased estimates of the true components of the mixture of normal probability distributions, due to the truncation of the tails of these distributions by histogram partitioning. However, we shall assume that the effect of this bias is small and can be ignored.

Let us express criterion  $J(T)$  as

$$J(T) = \sum_{g=0}^T h(g) \times \left\{ \left[ \frac{g - \mu_1(T)}{\sigma_1(T)} \right]^2 + 2 \log \sigma_1(T) - 2 \log P_1(T) \right\} \\ + \sum_{g=T+1}^n h(g) \times \left\{ \left[ \frac{g - \mu_2(T)}{\sigma_2(T)} \right]^2 + 2 \log \sigma_2(T) - 2 \log P_2(T) \right\}. \quad (14)$$

Substituting (5) through (7) into  $J(T)$  we find

$$J(T) = 1 + 2[P_1(T) \log \sigma_1(T) + P_2(T) \log \sigma_2(T)] \\ - 2[P_1(T) \log P_1(T) + P_2(T) \log P_2(T)]. \quad (15)$$

The criterion function  $J(T)$  in (15) can be computed easily and finding its minima is a relatively simple task, as the function is smooth.

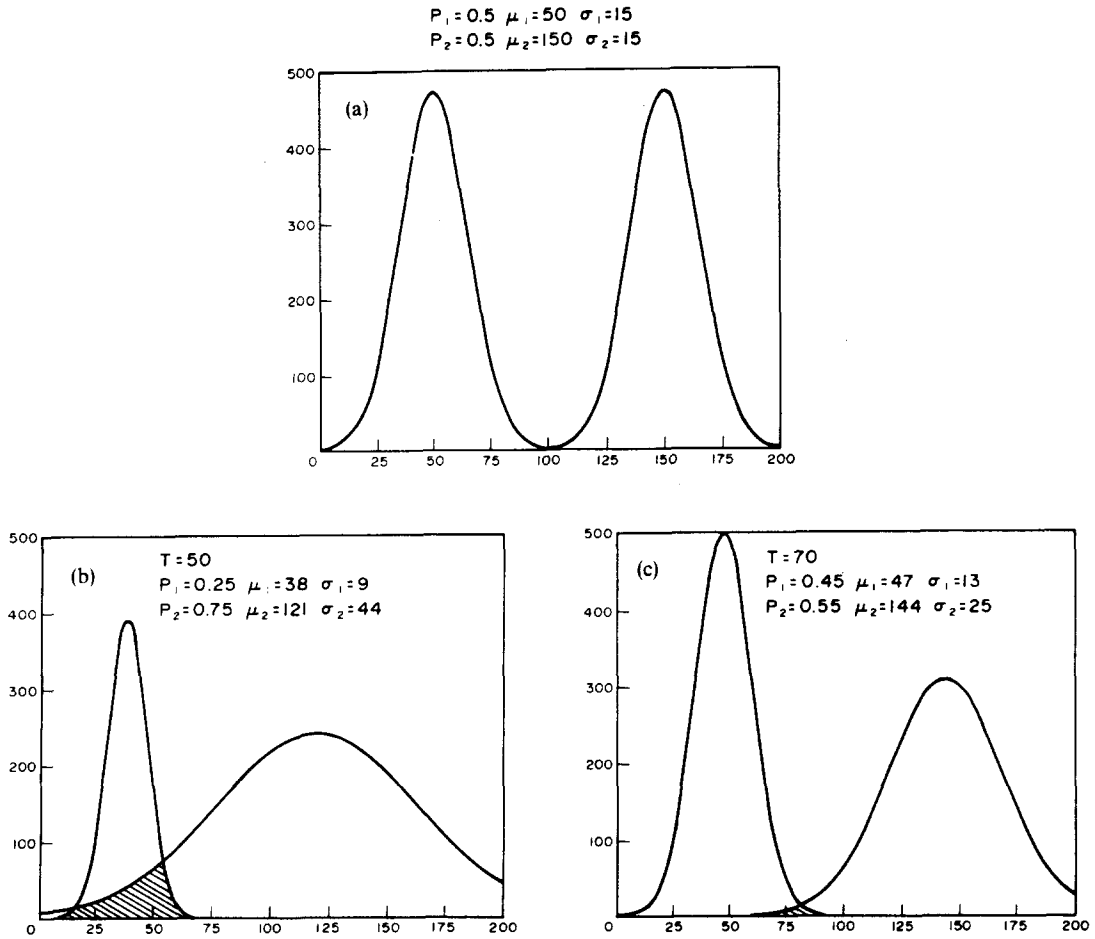


Fig. 1. (a) Grey level histogram defined by two equipopulous Gaussian distributions. Optimal threshold is at  $T = 100$ . (b, c) Population models for trial thresholds  $T = 50$  and  $T = 70$ . Binarisation error is represented by the hatched overlapping areas of the model population density functions.

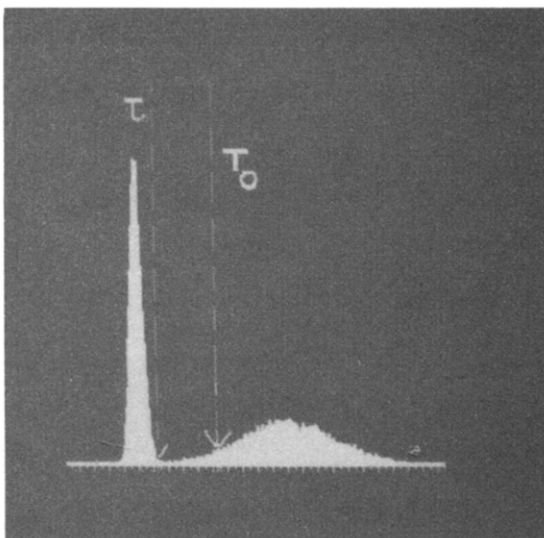


Fig. 2. Bimodal histogram.  $P_1 = 0.5, \mu_1 = 50, \sigma_1 = 4$ ;  $P_2 = 0.5, \mu_2 = 150, \sigma_2 = 30$ . Experimentally determined minimum error threshold  $\tau = 64$ . Threshold selected by standard algorithms<sup>(7-9)</sup>  $T_0 = 103$ .

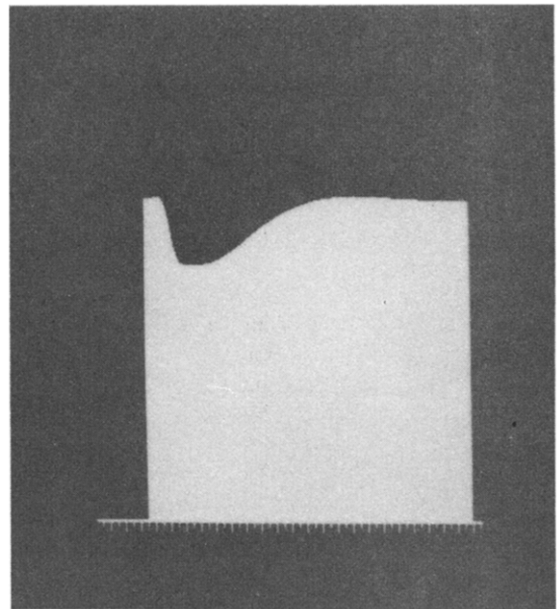


Fig. 3. Criterion function  $J(T)$  for histogram of Fig. 2.

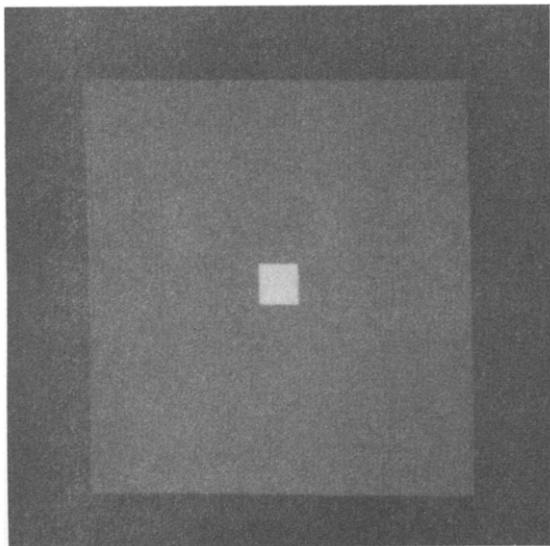


Fig. 4.  $50 \times 50$  pixel square in a  $512 \times 512$  image.  $P_1 = 0.99$ ,  $\mu_1 = 90$ ,  $\sigma_1 = 10$ ;  $P_2 = 0.01$ ,  $\mu_2 = 170$ ,  $\sigma_2 = 10$ .

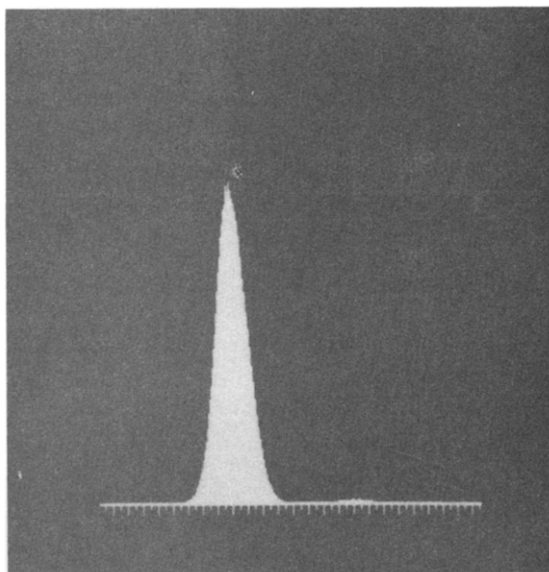


Fig. 5. Grey level histogram of Fig. 4.

### 3. PROPERTIES OF THE CRITERION FUNCTION

The criterion function may have local minima at the boundaries of the interval of grey levels. However, these minima are not of interest. A unique internal minimum implies histogram bimodality and it corresponds to the optimum (minimum error) threshold. This has been verified experimentally on the histogram shown in Fig. 2, containing object and background grey level modes of unequal variances. Figure 3 gives the corresponding criterion function. Note that the popular threshold selection methods of Otsu<sup>(7)</sup> and Ridler,<sup>(8,9)</sup> which use a less sophisticated model for the two modes, give a biased threshold indicated by  $T_0$  in Fig. 2. The reasons why the bias occurs are discussed in Kittler and Illingworth.<sup>(10)</sup>

Criterion (15) is also more appropriate for thresholding images with highly unequal population sizes exceeding the ratio of  $1:10^2$  or more, where the methods of Otsu<sup>(7)</sup> and Ridler<sup>(8,9)</sup> tend to split the larger mode into two halves.<sup>(10)</sup> Figure 4 is such an image and Fig. 5 shows the corresponding histogram of grey level values. The minimum of the criterion function, shown in Fig. 6, gives a good threshold value for segmenting the square from the background. Note that at the peak of the larger mode the criterion function attains a maximum.

The absence of an internal minimum is indicative of a unimodal histogram, which would correspond to a homogeneous image. Binarisation of such an image would be inappropriate and this case is illustrated in Figs 7 and 8, which show the histogram and the

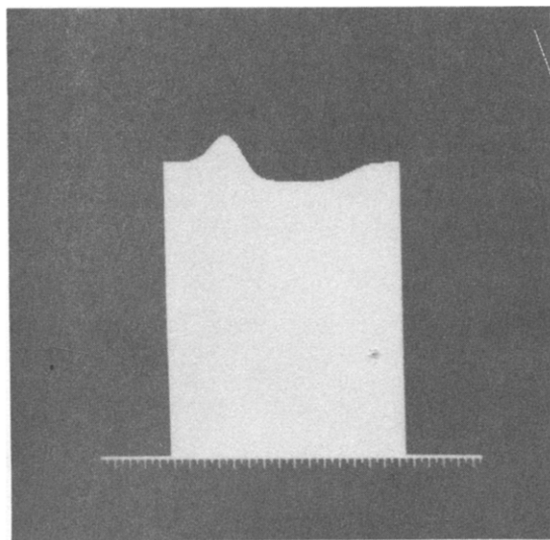


Fig. 6. Criterion function  $J(T)$  for histogram of Fig. 5.

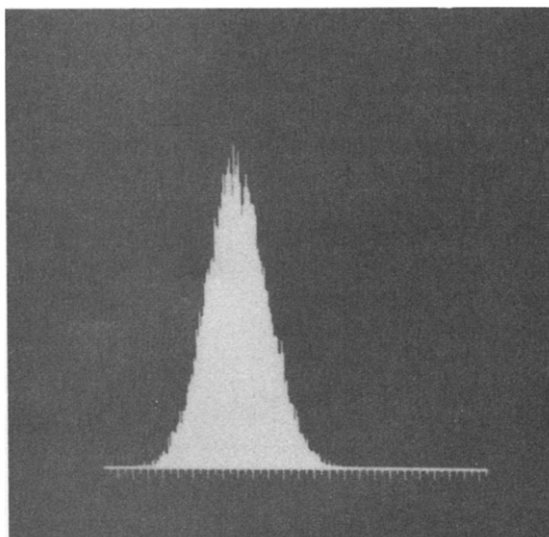


Fig. 7. Unimodal grey level histogram.

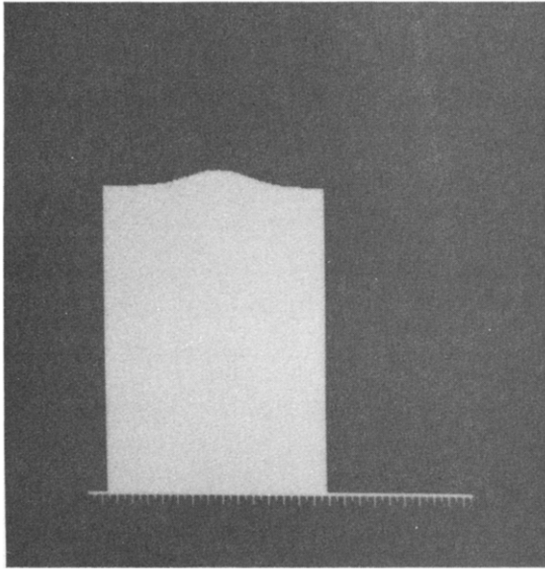


Fig. 8.  $J(T)$  for Fig. 7. No internal minima exist.

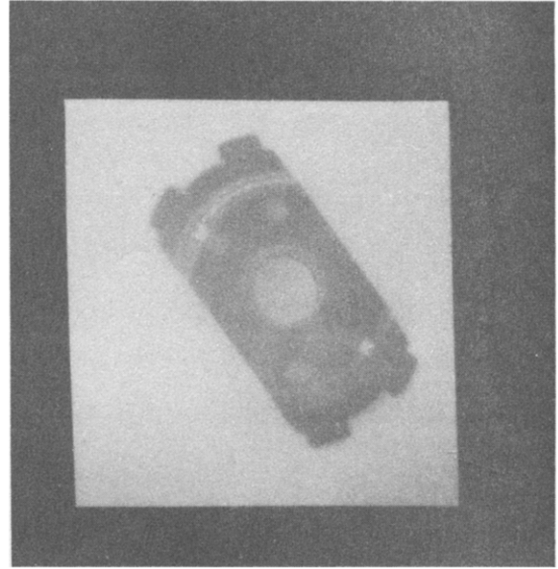


Fig. 9. Poorly illuminated image of an industrial part.

corresponding criterion function for a homogeneous image. The advocated method is again superior to those of Otsu<sup>(7)</sup> and Ridler,<sup>(8, 9)</sup> which would split the only mode in the middle and result in a 'pepper and salt' binary image.

The threshold selection approach has been applied to an image of an industrial part (Fig. 9). As the illumination was non-uniform, the image had to be binarised using a variable thresholding technique. The  $512 \times 512$  image was divided into subimages of  $32 \times 32$  pixels and an optimal threshold selected for each thresholdable window (i.e. containing both object and background pixels). A bilinear interpolation of determined thresholds defined a threshold value at every pixel of the image. The resulting binary image is shown in Fig. 10.

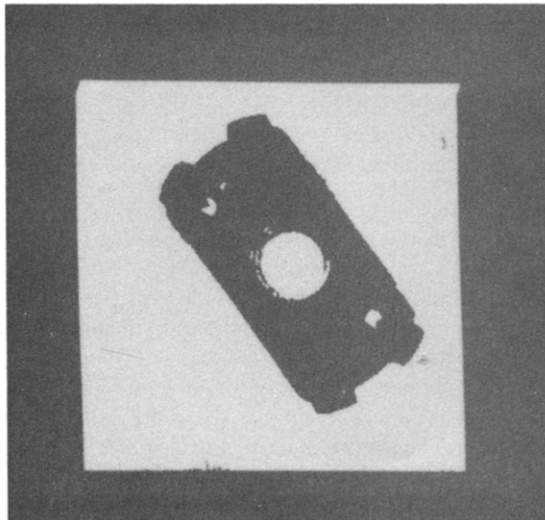


Fig. 10. Binary image derived using minimum error thresholding.

#### 4. MULTIPLE THRESHOLD SELECTION

In the presence of more than one internal minimum the grey level histogram will be at least trimodal. Although the points of minima of  $J(T)$  could be used as thresholds, they are likely to be biased. For example, Fig. 11 shows a grey level histogram consisting of 3 equipopulous modes. The means of the 3 distributions are at 50, 100 and 150 and they all have a variance of 100 grey levels. In the corresponding  $J(T)$  minima are present at grey level values of 70 and 130. A possible way of eliminating this bias is to partition the histogram at the first biased threshold level given by  $J(T)$ . The second of the two resulting grey level histograms can then be subjected to histogram analysis by the above method to determine the unbiased threshold level. This procedure is then repeated for the histogram

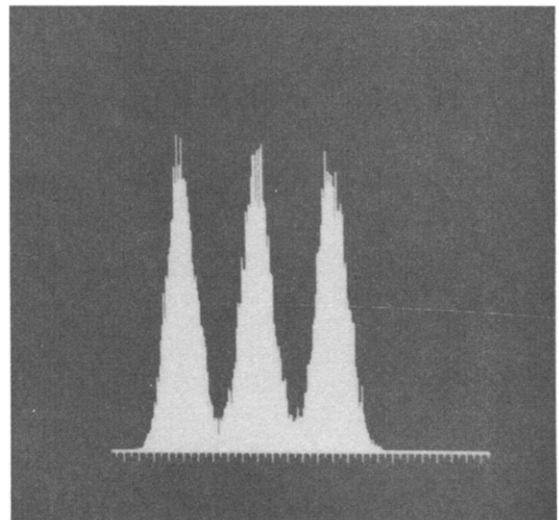


Fig. 11. Trimodal histogram.  $P_1 = P_2 = P_3$ ;  $\mu_1 = 50$ ,  $\mu_2 = 100$ ,  $\mu_3 = 150$ ;  $\sigma_1 = \sigma_2 = \sigma_3 = 10$ .

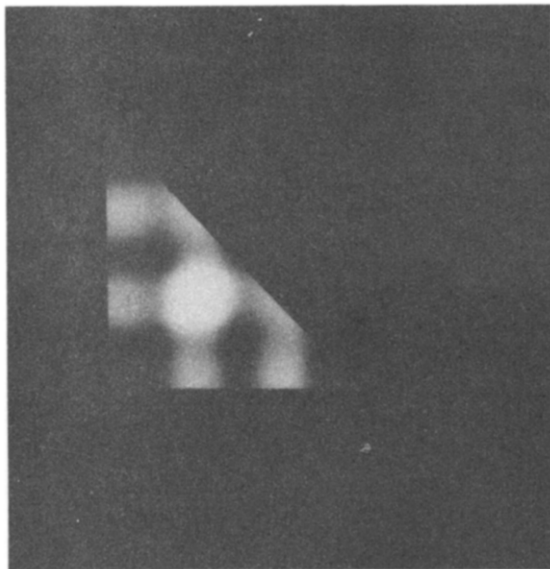


Fig. 12.  $J(T_1, T_2)$  for trimodal histogram of Fig. 11.

obtained by partitioning the original histogram at the above unbiased threshold.

Alternatively, and generally, we can use a multi-threshold extension of the proposed method. Suppose the histogram contains  $m$  modes, i.e. it is a mixture of  $m$  normal densities. By analogy, the corresponding criterion to be optimised is

$$J(T, \dots, T_{m-1}) = 1 + 2 \times \sum_{i=1}^m \{P_i(T_i) [\log \sigma_i(T_i) - \log P_i(T_i)]\}, \quad (16)$$

where

$$P_i(T_i) = \sum_{g=T_{i-1}+1}^{T_i} h(g), \quad (17)$$

$$\mu_i(T_i) = \frac{1}{P_i(T_i)} \sum_{g=T_{i-1}+1}^{T_i} g h(g), \quad (18)$$

$$\sigma_i^2(T_i) = \frac{1}{P_i(T_i)} \sum_{g=T_{i-1}+1}^{T_i} [g - \mu_i(T_i)]^2 h(g) \quad (19)$$

and

$$\begin{aligned} T_m &= n, \\ T_0 &= -1. \end{aligned} \quad (20)$$

Here the number of possible sets of candidate thresholds to be evaluated is considerably greater. Specifically, it is given by  $(n+1)!/(n+2-m)!(m-1)!$ . Thus, for instance, when  $m=3$  and  $n=255$  the number of points for which the criterion function must be computed is 32,640.

The application of the multithreshold selection procedure to the trimodal histogram of Fig. 11 yielded the criterion function space  $J(T_1, T_2)$  shown in Fig. 12, with bright grey level values corresponding to low values of  $J(T_1, T_2)$ . The only internal minimum lies in the central bright area. The coordinates of the mini-

mum correspond to the correct minimum error thresholds

$$T_1 = 75, \quad T_2 = 125.$$

## 5. ITERATIVE IMPLEMENTATION

The criterion (15) [and for that matter (16)] can be minimised iteratively using the dynamic clustering algorithm.<sup>(6)</sup> It is possible to show that for any partition of the grey level histogram using the threshold  $T$  the criterion function value can be decreased by assigning the grey level values to the classes of object and background pixels according to the Bayes minimum error rule defined in terms of parameters  $\mu_i(T)$ ,  $\sigma_i(T)$  and  $P_i(T)$ , i.e.

$$\begin{aligned} &\text{if } \left[ \frac{g - \mu_1(T)}{\sigma_1(T)} \right]^2 + 2[\log \sigma_1(T) - \log P_1(T)] \\ &\leq \left[ \frac{g - \mu_2(T)}{\sigma_2(T)} \right]^2 + 2[\log \sigma_2(T) \\ &\quad - \log P_2(T)] \text{ then } \begin{cases} g \rightarrow 1 \\ g \rightarrow 2 \end{cases} \end{aligned} \quad (21)$$

This decision rule effectively defines a new threshold which can be obtained by solving the following quadratic equation

$$\begin{aligned} &g^2 \cdot \left[ \frac{1}{\sigma_1^2(T)} - \frac{1}{\sigma_2^2(T)} \right] - 2g \left[ \frac{\mu_1(T)}{\sigma_1^2(T)} - \frac{\mu_2(T)}{\sigma_2^2(T)} \right] \\ &+ \frac{\mu_1^2(T)}{\sigma_1^2(T)} - \frac{\mu_2^2(T)}{\sigma_2^2(T)} + 2[\log \sigma_1(T) - \log \sigma_2(T) \\ &\quad - 2[\log P_1(T) - \log P_2(T)]] = 0. \end{aligned} \quad (22)$$

The procedure can be repeated for this new value of threshold, thus reducing the criterion function value even further. The algorithm is terminated when the threshold becomes stable. The algorithm can be formally stated as follows.

- Step 1. Choose an arbitrary initial threshold  $T$ ;
- Step 2. Compute  $\mu_i(T)$ ,  $\sigma_i(T)$ ,  $P_i(T)$ ,  $i = 1, 2$ ;
- Step 3. Compute the updated threshold by solving equation (22);
- Step 4. If the new threshold equals the old one then terminate the algorithm, else go to Step 2.

The threshold selection algorithm is very fast, but the user must be aware of various pitfalls that could result in a nonsensical thresholding. A suitable strategy should be adopted to counteract them. For instance, the algorithm may converge to the boundary points of the grey level range. The convergence to an internal minimum of function  $J(T)$  does not guarantee that it is a unique minimum and therefore a good threshold. The algorithm may hang at some threshold value after just one or few iterations because of the coarse quantisation of the image intensity values. The general strategy is to run the threshold selection algorithm starting from several initial thresholds and then compare the results.

Another point to note is that at some values of  $T$  the product of the *a priori* probability and the conditional density of one class exceeds that of the other. Hence the quadratic equation in (22) will have imaginary roots. If such a situation occurs the algorithm must be initialised at a new starting point.

## 6. CONCLUSIONS

A computationally efficient solution to the problem of minimum error thresholding has been derived under the assumption of object and pixel gray level values being normally distributed. The principal idea behind the method is to optimise the average pixel classification error rate directly, using either an exhaustive search or an iterative algorithm. The method is applicable in multithreshold selection.

## REFERENCES

1. J. Kittler, J. Illingworth, J. Föglein and K. Paler, An automatic thresholding algorithm and its performance, *Proc. 7th Int. Conf. on Pattern Recognition*, Montreal, pp. 287–289 (1984).
2. J. Kittler, J. Illingworth, J. Föglein and K. Paler, An automatic thresholding method for waveform segmentation, *Proc. Int. Conf. on Digital Signal Processing*, Florence, pp. 727–732 (1984).
3. J. Kittler, J. Illingworth and J. Föglein, Threshold selection based on a simple image statistic, *Comput. Vision Graphics Image Process.* **30**, 125–147 (1985).
4. A. Rosenfeld and A. C. Kak, *Digital Picture Processing*, Academic Press, New York (1976).
5. Y. Nagawa and A. Rosenfeld, Some experiments on variable thresholding, *Pattern Recognition* **11**, 191–204 (1979).
6. P. A. Devijver and J. Kittler, *Pattern Recognition: A Statistical Approach*, Prentice/Hall, Englewood Cliffs, NJ (1982).
7. N. Otsu, A threshold selection method from grey level histograms, *IEEE Trans. Syst. Man Cybernet.* **SMC-9**, 62–66 (1979).
8. T. Ridler and S. Calvard, Picture thresholding using an iterative selection method, *IEEE Trans. Syst. Man Cybernet.* **SMC-8**, 630–632 (1978).
9. H. J. Trussel, Comments on picture thresholding using an iterative selection method, *IEEE Trans. Syst. Man Cybernet.* **SMC-9**, 311 (1979).
10. J. Kittler and J. Illingworth, On threshold selection using clustering criteria, *IEEE Trans. Syst. Man Cybernet.* **SMC-15** (1985).

**About the Author**—JOSEF KITTLER was awarded a Ph.D. degree in Pattern Recognition in 1974 and since then has published a number of papers and a book (*Pattern Recognition: A Statistical Approach*, Prentice Hall, 1982) on topics in pattern recognition and image processing. Since 1980 he has been with the Rutherford Appleton Laboratory, where he is in charge of a research team working on projects in computer vision and remote sensing. He is the SERC Coordinator for Pattern Analysis.

Dr Kittler is an Associate Editor of *IEEE Transactions on Pattern Analysis and Machine Intelligence*, and is on the editorial board of *Pattern Recognition*, *Pattern Recognition Letters*, and *Image and Vision Computing*. He has been serving as a member of the BPRA Committee for several years.

**About the Author**—JOHN ILLINGWORTH received B.Sc. and D.Phil. degrees in Physics from the Universities of Birmingham and Oxford in 1978 and 1983, respectively. He is now employed as a Senior Research Associate at the Rutherford Appleton Laboratory, undertaking research in computer vision algorithms.