Thresholding based on histogram approximation

N. Ramesh
J.-H. Yoo
I.K. Sethi

Indexing terms: Histograms, Image processing, Thresholding

Abstract: The authors propose two automatic threshold-selection schemes, based on functional approximation of the histogram. The first method is based on minimising the sum of square errors, and the second one is based on minimising the variance of the approximated histogram. Experimental results show that, on average, the latter scheme gives better results than the former one, at a small extra computational cost. A 'goodness' measure is proposed to measure the effectiveness of the two schemes, and to compare them against the entropy-based approach and the moment-based approach.

1 Introduction

In computer vision, abstractions of objects or features, which are used in high level tasks, are derived from images. For the purpose of abstraction, the pixels in an image have to be grouped into meaningful regions by a process called segmentation. One way to do segmentation is thresholding. Thresholding in its simplest form involves mapping all pixels above a threshold value to one grey value, say white, and the rest to another, say black. Since the result is an image with two grey values, the process is called bilevel segmentation. When multiple threshold values are used, the result is a multilevel image, and the process is called multilevel segmentation. For a survey of different thresholding and segmentation techniques see References 1–4.

Many automatic threshold-determination techniques use the histogram of the image to select a good threshold. The histogram of an image is the frequency distribution of grey levels in that image. If, in an image, the objects have distinctly different grey values from the background, the histogram will exhibit two different peaks with a valley between them. Such a histogram is called a bimodal histogram, and the determination of a suitable threshold value is a relatively simple matter. In a non-bimodal distribution, the selection of a good threshold can be rather difficult. There are a number of methods for threshold selection discussed in the literature, including those based on entropy [5–7], moment preservation [8], error minimisation [9] and maximum likelihood [10]. One common characteristic of these existing methods is that they view the histogram as a mixture density function, and usually treat the problem of threshold determination as a case of classification. In this paper, we take a different approach, and view the histogram as a 1-D function, whose bilevel approximation yields the threshold value for image segmentation. Done recursively, this approach results in a multilevel approximation of a histogram, and consequently leads to multilevel thresholding.

In this paper, we have described two automatic threshold-selection schemes based on the functional approximation of the histogram. The first method is based on minimising the sum of square errors. Though this method is computationally simpler, it can be biased, because the reduction in the square error is achieved by many terms in the sum [11]. This led us to come up with a better method based on minimising the variance. We also show here that this minimum variance method is equivalent to the optimum threshold value of a bimodal histogram, that is, the sum of two Gaussian distributions. Experimental results, for the suggested threshold schemes are presented. To better judge these schemes, we have defined a performance criteria, and have compared the performance of our methods with two other methods from the literature, namely the entropy method and the moment-based method. We have chosen the entropy method because it has been used as a benchmark in the literature, and the moment-based method because it is a more recent approach to the problem.

2 Approximating the histogram

2.1 The optimisation problem

If we assume that the image to be thresholded has two major brightness regions, and, further, that these brightness regions have a known probability distribution, Gaussian, for example, then the image can be optimally thresholded [12]. However, in many real-world images this assumption is unrealistic. What is required is an algorithm that can elegantly threshold images with arbitrary grey-level distributions. We present two algorithms which accomplish this. Further, we show that our algorithms select the optimal threshold [12], if the grey-level probability distribution is Gaussian.

Let the pixels in an image be represented by $L + 1$ grey levels, $[0, 1, 2, \ldots, L]$. Let h_j denote the number of pixels with grey level j. Forming a histogram $H(x)$ of the image results in an ordered set of discrete values, h_1, h_2,

Paper 2007K (E4), first received 15th February 1994 and in revised form 27th March 1995

The authors are with the Vision and Neural Networks Laboratory, Department of Computer Science, Wayne State University, Detroit, MI 48202, USA

The authors gratefully acknowledge the help of Manu Vasandani in preparing the final version of this document.

..., h_L. Our aim is to approximate the set of discrete values by a bi-level function $G(x)$, such that the h_j, $j = 0, 1, 2, \ldots, L$ can be replaced by g_j, $j = 0, 1, 2, \ldots, L$, where $\forall_j \leqslant t$, $g_j = H_1$, and $\forall_j > t$, $g_j = H_2$. The partition point t, where the approximation jumps from H_1 to H_2, provides us with a good threshold value. Thus, all the pixels less than or equal to t are mapped to one grey value, while those above t are mapped to another grey value. For example, consider the histogram shown in Fig. 1.

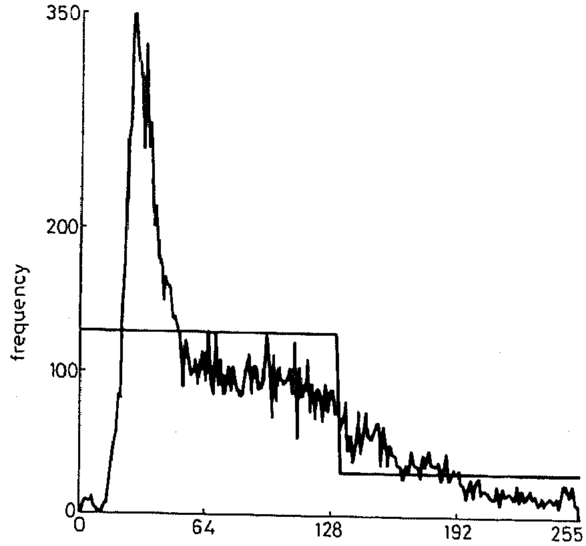

Fig. 1 *An actual histogram and its bilevel approximation*
All the pixels less than or equal to the transition point are mapped to one grey value, while those greater than the transition point are mapped to another grey value

The histogram has been approximated by a bilevel function, with a transition point at $t = 141$, which is selected as the threshold value. Such a bilevel approximation can be done recursively, using the histogram values on either side of the transition point t, which will result in a multilevel approximation of the histogram, as shown in Figs. 2 and 3.

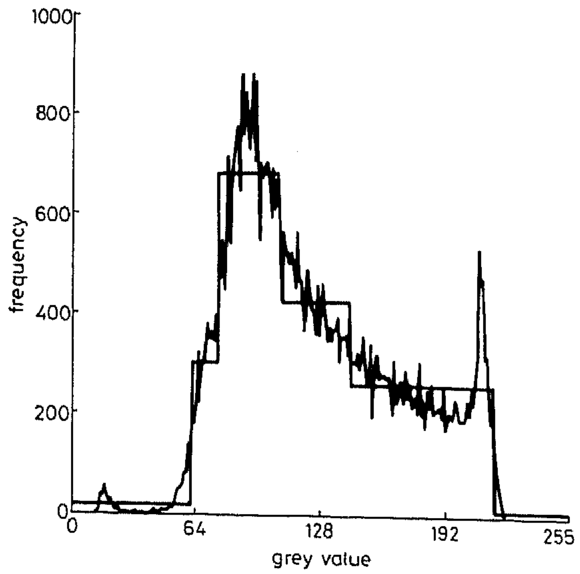

Fig. 2 *The histogram and its multilevel approximation*

In order to approximate the histogram with a bilevel function, we need an error function, minimising which will yield the transition point t. In this paper, we suggest two such error functions: one that minimises the sum of square errors and another that minimises the variance. As will be shown in Section 2.2, since there is no closed form solution for t, given these particular error functions, we have to resort to other methods to determine t. We directly search for t since the search space is small.

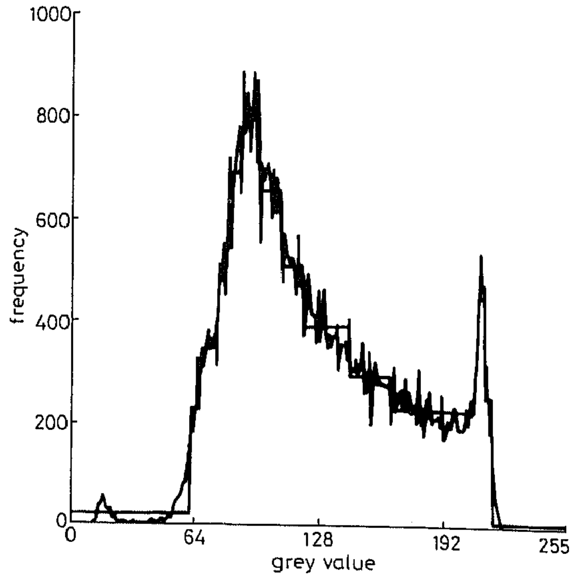

Fig. 3 *The histogram with even more levels of approximation*

The error function $E(t)$ is defined either as the sum of square errors

$$E(t) = \sum_{i=0}^{t} (i - m_1(t))^2 + \sum_{i=t+1}^{L} (i - m_2(t))^2 \tag{1}$$

or as the sum of two variances,

$$E(t) = \frac{1}{N_1 - 1} \sum_{i=0}^{t} (i - m_1(t))^2 + \frac{1}{N_2 - 1} \sum_{i=t+1}^{L} (i - m_2(t))^2 \tag{2}$$

where

$$m_1(t) = \frac{\sum_{i=0}^{t} i h_i}{\sum_{i=0}^{t} h_i} \qquad m_2(t) = \frac{\sum_{i=t+1}^{L} i h_i}{\sum_{i=t+1}^{L} h_i}$$

N_1 is the total number of pixels with grey values less than or equal to t, and N_2 is the total number of pixels with grey values greater than t. Since t can take on only a small range of values, we use direct search to find the value of t that minimises the error functions. In Section 3 we show that, on average, the threshold value determined by minimising eqn. 2 gives better results compared to the threshold value determined by minimising eqn. 1, with only a slight increase in computational cost.

2.2 Theory

Given the error functions defined by eqns. 1 and 2, we attempt to obtain a closed form solution for the value of the threshold t. The derivations are given in the following Sections.

2.2.1 Minimising the sum of square errors: We seek to minimise the sum of square errors between the grey values and the mean value, in the two regions, i.e. find $\min\{E(t)\} = \min\{E_1(t) + E_2(t)\}$, where

$$E_1(t) = \int_0^t (x - m_1(t))^2 p(x) N \, dx$$

and

$$E_2(t) = \int_t^L (x - m_2(t))^2 p(x) N \, dx$$

and $p(x)$, the p.d.f is defined as

$$p(x) \triangleq \frac{h(x)}{N} \quad N \triangleq \int_0^L h(x) \, dx$$

$$m_1 = \frac{\int_0^t x p(x) \, dx}{\int_0^t p(x) \, dx} = \frac{B}{A}$$

and

$$m_2 = \frac{\int_t^L x p(x) \, dx}{\int_t^L p(x) \, dx} = \frac{D}{C}$$

So,

$$E(t) = N \int_0^t (x^2 - 2m_1(t)x + m_1^2(t)) p(x) \, dx + N \int_t^L (x^2 - 2m_2(t)x + m_2^2(t)) p(x) \, dx$$

$$\Rightarrow \frac{E(t)}{N} = \int_0^t x^2 p(x) \, dx - 2m_1(t) \int_0^t x p(x) \, dx + m_1^2(t) \int_0^t p(x) \, dx + \int_t^L x^2 p(x) \, dx - 2m_2(t) \int_t^L x p(x) \, dx + m_2^2(t) \int_t^L p(x) \, dx$$

To find the threshold t that minimises $E(t)$, we differentiate $E(t)$ with respect to t and set the result to 0,

$$\frac{\partial E(t)}{\partial t} = 0$$

$$\frac{1}{N} \frac{\partial E(t)}{\partial t} = t^2 p(t) - 2 \frac{\partial m_1(t)}{\partial t} \int_0^t x p(x) \, dx - 2m_1(t) t p(t) + 2m_1(t) \frac{\partial m_1(t)}{\partial t} \int_0^t p(x) \, dx + m_1^2(t) p(t) - t^2 p(t) - 2 \frac{\partial m_2(t)}{\partial t} \int_0^t x p(x) \, dx + 2m_2(t) t p(t) + 2m_2(t) \frac{\partial m_2(t)}{\partial t} \int_t^L p(x) \, dx - m_2^2(t) p(t)$$

The second and fourth term (and similarly the seventh and ninth terms) reduce to zero

$$-2 \frac{\partial m_1}{\partial t} \int_0^t x p(x) \, dx + 2m_1 \frac{\partial m_1}{\partial t} \int_0^t p(x) \, dx$$

$$= -2 \frac{t p(t) A - p(t) B}{A^2} B + 2m_1(t) \frac{t p(t) A - p(t) B}{A^2} A$$

$$= -2(t p(t) m_1(t) - p(t) m_1^2(t)) + 2m_1(t) t p(t) - p(t) m_1(t)$$

$$= 0$$

Thus the dependency of $m_1(t)$ and $m_2(t)$ on t is effectively removed and

$$\frac{\partial E_1(t)}{\partial t} = (t - m_1(t)^2) p(t) N$$

and

$$\frac{\partial E_2(t)}{\partial t} = -(t - m_2(t)^2) p(t) N$$

$$\Rightarrow \frac{1}{N} \frac{\partial E(t)}{\partial t} = p(t)[m_1^2(t) - m_2^2(t) + 2m_2(t)t - 2m_1(t)t] = 0$$

$$\Rightarrow t = \frac{m_1(t) + m_2(t)}{2}$$

t is the fixed point of the mapping that can be found by successive approximation.

2.2.2 Minimising the variance: To minimise the variance of grey levels of the two regions separated by the cut-off threshold t, we minimise $E(t) = E_1(t) + E_2(t)$, where

$$E_1(t) = \frac{\int_0^t (x - m_1(t))^2 p(x) \, dx}{\int_0^t p(x) \, dx}$$

and

$$E_2(t) = \frac{\int_t^L (x - m_2(t))^2 p(x) \, dx}{\int_t^L p(x) \, dx}$$

where $m_1(t)$ and $m_2(t)$ are defined as earlier.

Differentiating $E(t)$ with respect to t, and setting the result to 0, we get

$$E(t) = \frac{\int_0^t (x^2 - 2m_1(t)x + m_1^2(t)) p(x) \, dx}{\int_0^t p(x) \, dx} + \frac{\int_t^L (x^2 - 2m_2(t)x + m_2^2(t)) p(x) \, dx}{\int_t^L p(x) \, dx}$$

Say

$$\frac{\int_0^t (x^2 - 2m_1(t)x + m_1^2(t)) p(x) \, dx}{\int_0^t p(x) \, dx} = \frac{P}{A}$$

and

$$\frac{\int_t^L (x^2 - 2m_2(t)x + m_2^2(t)) p(x) \, dx}{\int_t^L p(x) \, dx} = \frac{Q}{C}$$

$$\Rightarrow \frac{\partial E(t)}{\partial t} = \frac{P'A - PA'}{A^2} + \frac{Q'C - QC'}{C^2}$$

Using the results from Section 2.2.1, we get

$$P' = (t - m_1(t))^2 p(t)$$

and

$$Q' = -(t - m_2(t))^2 p(t)$$

$$\Rightarrow \frac{\partial E(t)}{\partial t} = \frac{(t - m_1(t))^2 p(t) A - (\int_0^t (x - m_1)^2 p(x) \, dx) p(t)}{A^2} - \frac{(t - m_2(t))^2 p(t) C - (\int_t^L (x - m_2)^2 p(x) \, dx) p(t)}{C^2}$$

$$\Rightarrow \frac{\partial E(t)}{\partial t} = p(t) \left[\frac{(t - m_1(t))^2 - s_1^2(t)}{A} \right] - p(t) \left[\frac{(t - m_2(t))^2 - s_2^2(t)}{C} \right]$$

where

$$s_1^2(t) = \frac{P}{A}$$

and

$$s_2^2(t) = \frac{Q}{C}$$

$$\Rightarrow \frac{\partial E(t)}{\partial t} = \frac{p(t)}{AC}[C(t - m_1(t))^2 - Cs_1^2(t)$$
$$- A(t - m_2(t))^2 + As_2^2(t)]$$
$$= 0$$

$$\Rightarrow 0 = (C - A)t^2 + (-2Cm_1(t) + 2Am_2(t))t$$
$$+ Cm_1^2(t) - Am_2^2(t) - Cs_1^2(t) + As_2^2(t)$$

If $C \neq A$, then

$$t = \frac{Cm_1(t) - Am_2(t) \pm \sqrt{[(Cm_1(t) - Am_2(t))^2 - (C - A)(Cm_1^2(t) - Am_2^2(t) - Cs_1^2(t) + As_2^2(t))]}}{(C - A)}$$

If $A = C, 0 < t < L, m_1(t) < t < m_2(t)$ then

$$t = \frac{m_1(t) + m_2(t)}{2} + \frac{s_1^2(t) - s_2^2(t)}{2(m_2(t) - m_1(t))}$$

Thus, t is the fixed point of the mapping that can be found by successive approximation. Comparing this to the Gaussian model [12]

$$t = \frac{\mu_1 + \mu_2}{2} + \frac{\sigma^2}{\mu_1 - \mu_2} \ln\left(\frac{P_2}{P_1}\right)$$

we observe that the *a priori* probabilities P_2 and P_1 are absorbed in the terms s_1^2 and s_2^2 under the constraint $\mu_1 < m_1(t) < m_2(t) < \mu_2$. Thus if $P_1 = P_2$ and $s_1^2(t) = s_2^2(t)$, we have

$$t = \frac{m_1(t) + m_2(t)}{2} = \frac{\mu_1 + \mu_2}{2}$$

This shows that, in the ideal case, we have a t value identical to the one obtained under the optimal threshold of the Gaussian PDF model.

3 Experimental results

Very few papers [13], among the many published on thresholding, discuss quantitative measures for the 'goodness' of a threshold, because the segmentation of an image is rather subjective. We propose a simple goodness criteria to compare our schemes with the entropy-based method and the moment-based method. We also show that our thresholding schemes work well for a particular class of images, for which histograms do not exhibit a perceivable bimodal characteristic. For a particular image, if we know the correctly segmented regions *a priori*, we can determine a goodness measure as

$$G = \frac{n_c}{n_t} \times 100$$

where n_c is the number of correctly thresholded pixels, and n_t is the total number of pixels in the image. We now present the results of three experiments that we conducted to test our algorithms.

3.1 Experiment 1

The first experiment was done on the images of a rectangle whose segmented regions were known *a priori*. The experiment was done to obtain quantitative measures of performances, to compare our schemes with the two other thresholding techiques. A number of images of a rectangle were generated (See Figs. 4–7, 12). These images had pixel correlations ranging from 0.10 to 0.90 for both

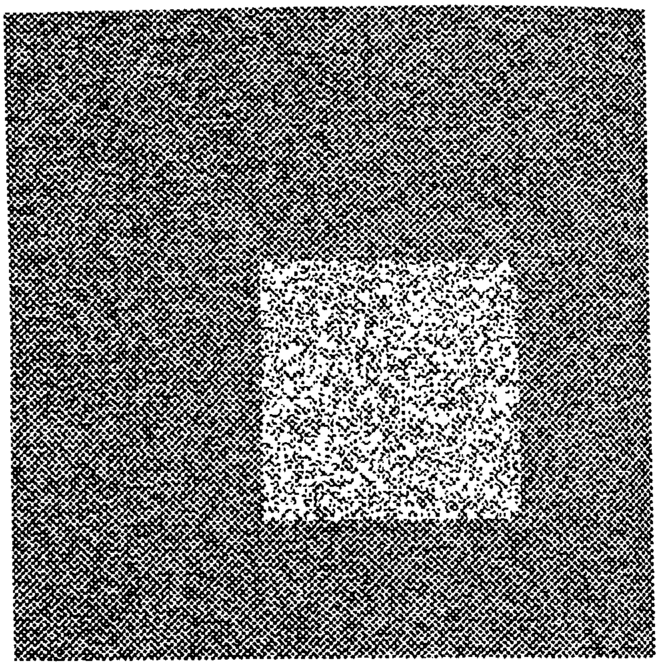

Fig. 4 *The original image with a correlation of 0.30 for background and foreground regions*

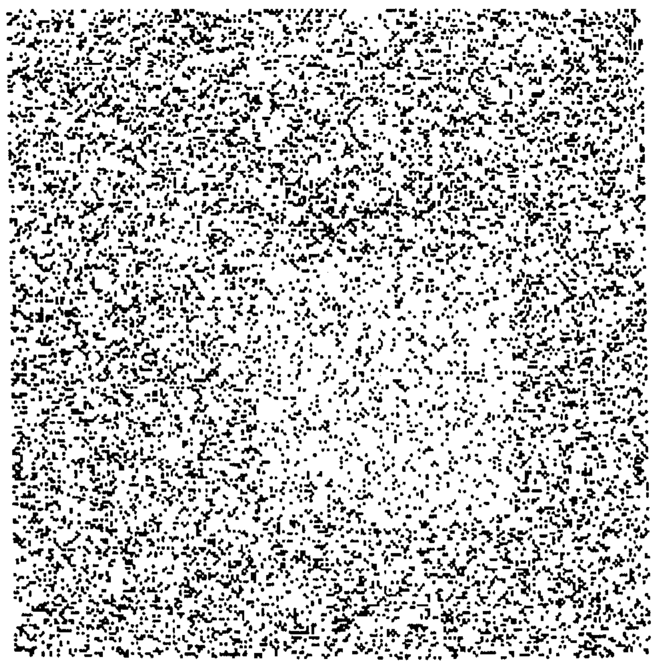

Fig. 5 *Image with correlation of 0.30 between neighbouring pixels, thresholded at 140 (entropy method)*

the foreground and background regions. The threshold values for these images were then computed using Kapur *et al.*s' [5] entropy algorithm, Pun's [6] moment preserving algorithm and our methods based on functional approximation. These threshold values and goodness measures for this experiment are shown in Table 1.

As stated earlier, a number of threshold schemes can give good results for bimodal histograms. However, for histograms that are not clearly bimodal, proper thresholding can be rather difficult. Fig. 8 shows the histogram for the image of the rectangle with a correlation of 0.80. Its histogram is bimodal. In such cases, the four thresholding schemes give nearly equal performances, measured by our goodness measure, as defined above. However, the histogram for the rectangular image with correlation 0.30 is not bimodal, as shown in Fig. 9. For such images, both the entropy approach and the moment based approach perform poorly compared to our schemes, both of which yield a goodness measure of over 90.

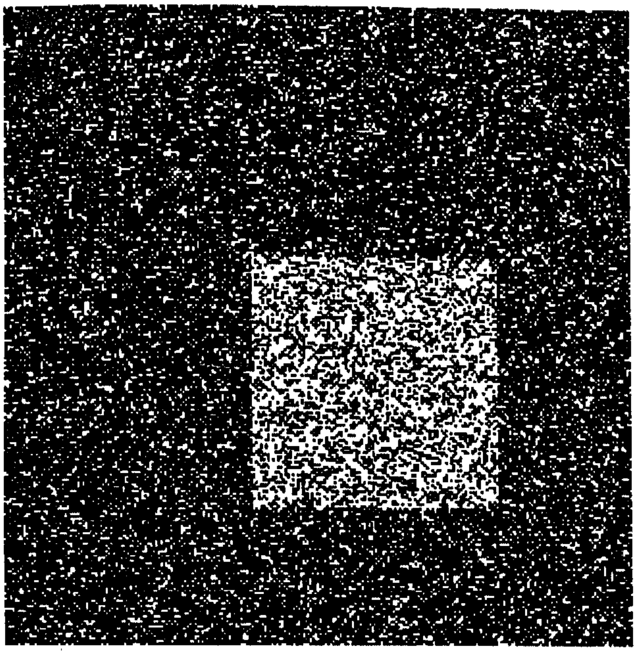

Fig. 6 *Image with correlation of 0.30 between neigbouring pixels, thresholded at 180 (moment based method)*

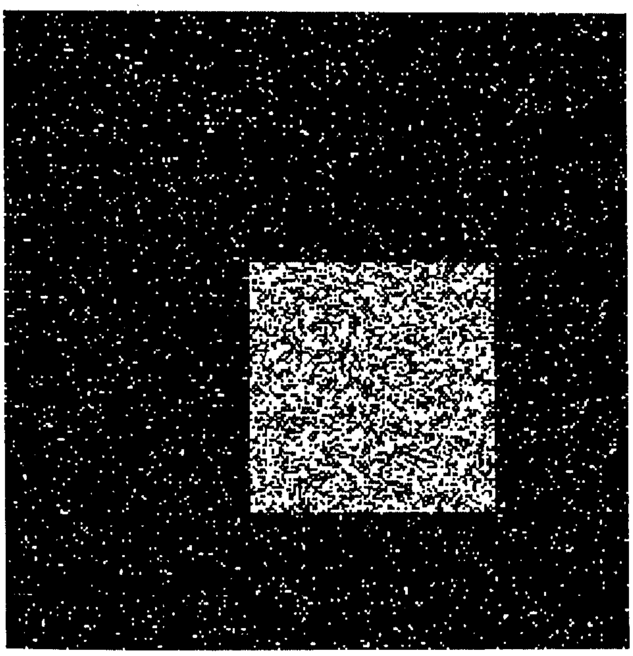

Fig. 7 *Image with correlation of 0.30 between neighbouring pixels, thresholded at 200 (sum of square-error method)*

Table 1 also shows that, on average, the performance of the sum of square-errors method, and the variance-based method of threshold determination by functional approximation, are better than the entropy approach and the moment-based approach. In the cases where the other methods do give better performances, they do so by only a small margin.

3.2 Experiment 2

To judge the performance of our schemes on real world images, the second experiment was done on the 'Lenna' image and an aerial image. Histograms are shown in

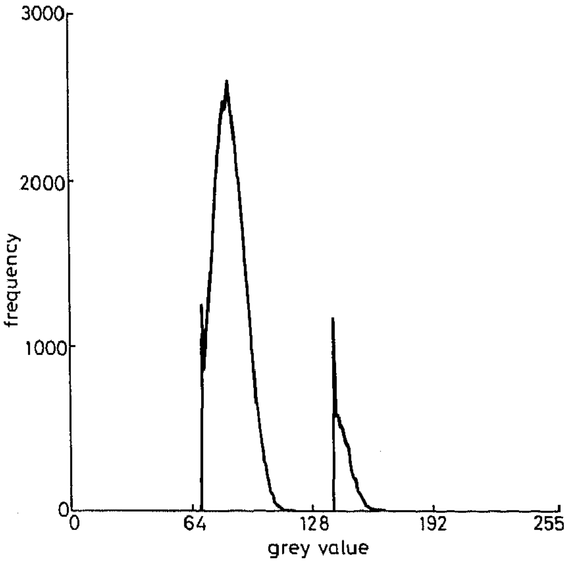

Fig. 8 *The histogram of the image of a rectangle whose pixel correlation is 0.80*

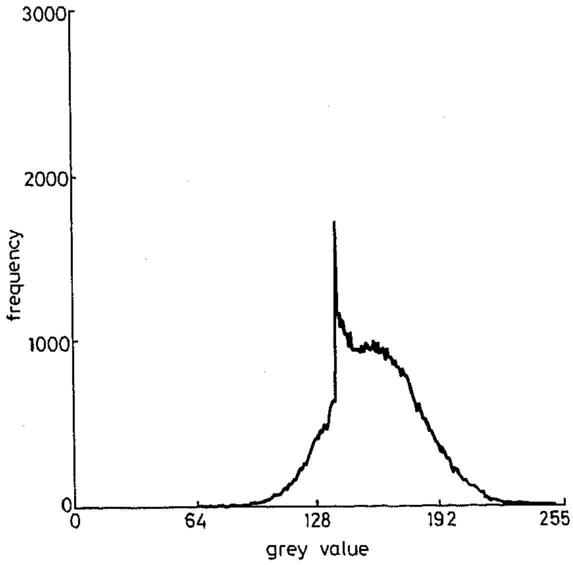

Fig. 9 *The histogram of the image of a rectangle whose pixel correlation is 0.30*

Table 1

Img. correlation	0.10	0.20	0.30	0.40	0.50	0.60	0.70	0.80	0.90
Threshold goodness	140	138	140	117	106	95	89	92	83
(Kapur *et al.*)	35.58	16.06	31.09	43.39	60.50	69.64	74.14	82.54	71.76
Threshold goodness	176	205	180	152	142	143	147	145	148
(Pun)	48.50	30.31	78.54	87.08	96.62	93.73	83.67	90.20	76.86
Sum of square-error-method	187	208	200	192	182	175	173	173	171
goodness	53.16	31.93	90.60	91.58	93.72	91.10	79.90	88.16	84.43
Variance-method	254	254	254	159	147	114	101	103	92
goodness	79.77	73.82	94.19	89.68	94.59	96.09	89.60	95.96	83.69

Figs. 10 and 11; original images in Figs. 13 and 18; and thresholding results in Figs. 14–17 and 19–22. Threshold values resulting from the use of the different methods are shown in Table 2.

Table 2

Thresholds	Rectangle	Lenna	Aerial
Entropy (Kapur *et al.*)	95	169	117
Moment preserving	143	171	97
Sum of square errors	175	161	80
Variance method	114	170	104

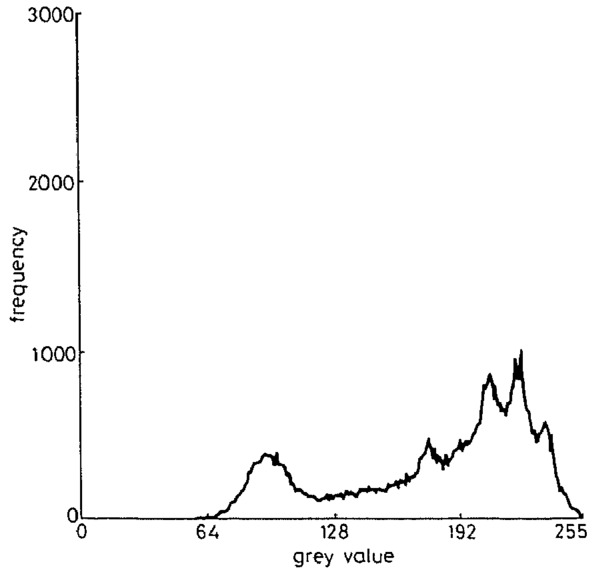

Fig. 10 *The histogram of the 'Lenna' image*

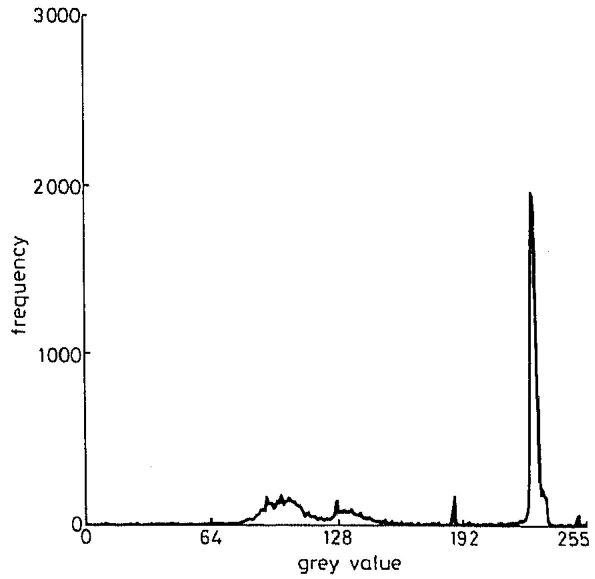

Fig. 11 *The histogram of the 'Aerial' image*

In the case of bimodal images like the 'Lenna' image, many methods should give good results. As the Table indicates, the threshold values for such an image are close to each other. The results of thresholding of the 'Lenna' image are shown in Figs. 14–17. But for a large spike in the histogram for the 'aerial' image (Fig. 11), the distribution of grey values is not bimodal. In such cases, each of the four different schemes give wisely differing threshold values. Since the 'goodness' measure cannot be obtained for these images, they have been reproduced for visual inspection in Figs. 19–22.

3.3 Experiment 3

In the third experiment, we show the result of multilevel thresholding using our variance minimisation scheme. The extension from the bilevel case is straightforward.

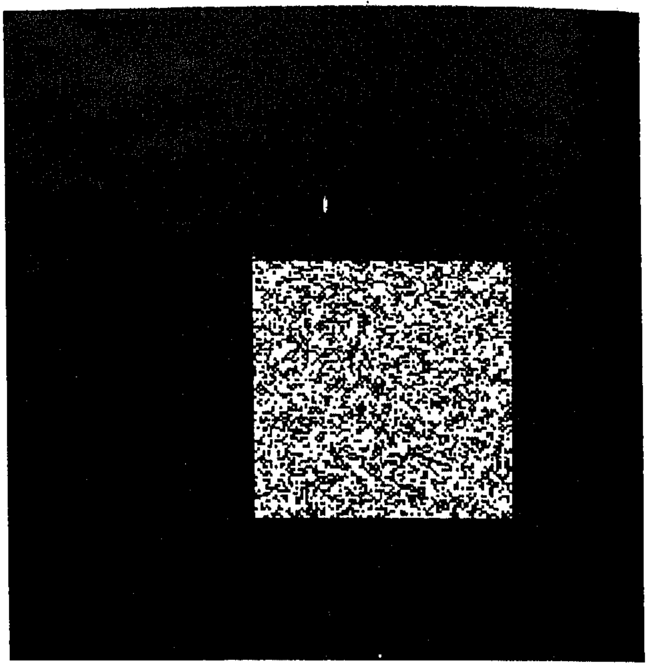

Fig. 12 *Image with correlation of 0.30 between neighbouring pixels, thresholded at 254 (variance method)*

Fig. 13 *Original 'Lenna' image*

The image histogram is first approximated by a bi-level function. The parts of the histogram to the left and right of the threshold value t are then considered separately, and each in turn is further approximated by a bilevel function. When done recursively, the result is a multilevel approximation which provides multiple threshold values. Figs. 2 and 3 show examples of a histogram approximated to multiple levels. Fig. 23 shows the 'Lenna' image approximated to four levels and Fig. 24 the approximated histogram.

4 Conclusions

In this paper, we have proposed two automatic threshold determination schemes that are based on approximating

Fig. 14 *'Lenna' image thresholded at 169 (entropy method)*

Fig. 15 *'Lenna' image thresholded at 171 (moment based method)*

Fig. 16 *'Lenna' image thresholded at 170 (variance method)*

Fig. 17 *'Lenna' image thresholded at 161 (sum of square error method)*

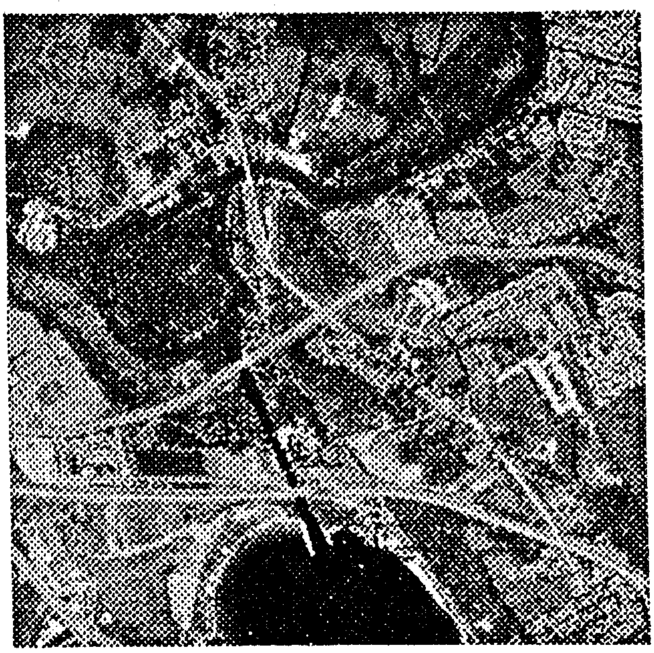

Fig. 18 *Original aerial image*

Fig. 19 *Aerial image thresholded at 97 (moment based method)*

the image histogram. The first one minimises the sum of square errors, while the second minimises the variance. We have also proposed a 'goodness' measure to quantify the performance of different thresholding algorithms, and shown that for image histograms, that do not exhibit bimodal tendencies, our schemes perform better than the entropy-based approach and the moment-based approach. For histograms that are bimodal, the performance of our schemes is comparable to the other two approaches. While the sum of square-errors method gives a good threshold value, the variance method on average gives better results at a slightly higher computational expense. These schemes can serve as viable alternatives to other automatic thresholding schemes.

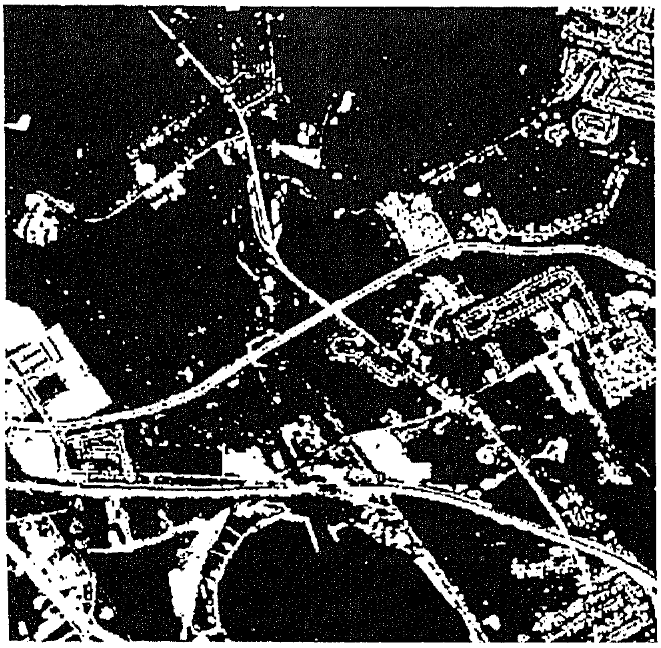

Fig. 20 *Aerial image thresholded at 117 (entropy method)*

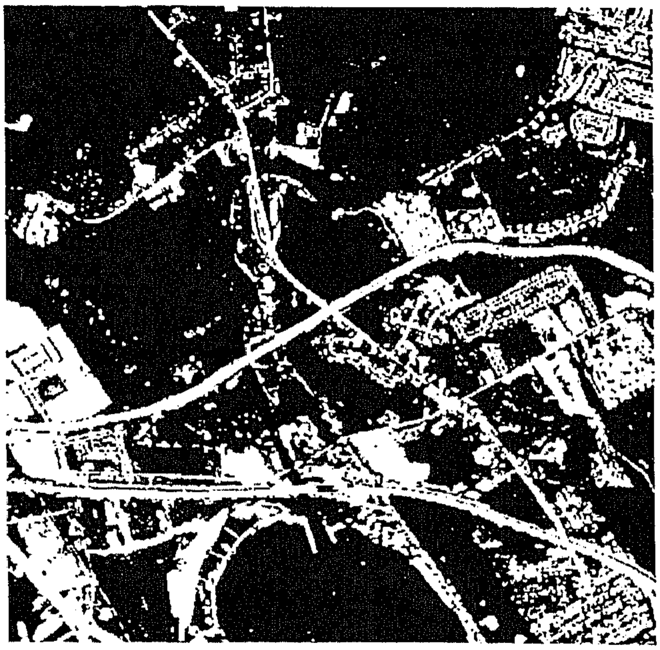

Fig. 21 *Aerial image thresholded at 104 (variance method)*

Fig. 22 *Aerial image thresholded at 80 (sum of squares method)*

Fig. 23 *Multilevel thresholding at 78, 170 and 241*

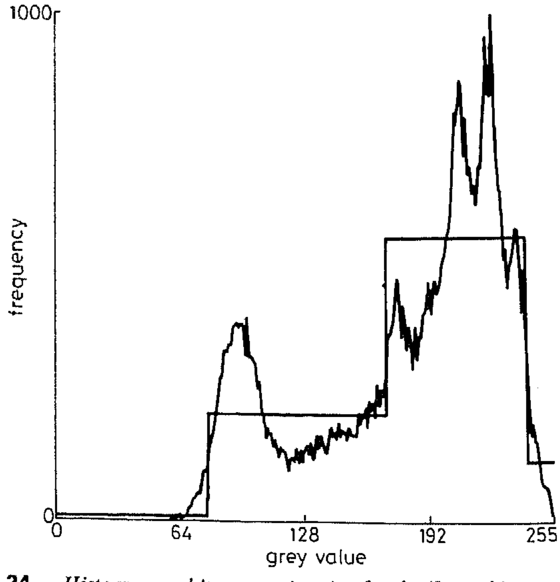

Fig. 24 *Histogram and its approximation for the 'Lenna' image*

5 References

1 SAHOO, P.K., SOLTANI, S., and WONG, A.: 'A survey of thresholding techniques', *Comput. Vis. Graph. Image Process.*, 1988, **41**, pp. 233–260

2 WESZKA, J.S.: 'A survey of threshold selection techniques', *Comp. Vis. Graph. Image Process.*, 1978, **7**, pp. 259–265

3 FU, K., and MUI, J.K.: 'A survey on image segmentation', *Pattern Recognit.*, 1980, **13**, pp. 3–16

4 HARALICK, R., and SHAPIRO, L.: 'Image segmentation techniques', *Comput. Vis. Graph. Image Process.*, 1985, **29**, pp. 100–132

5 KAPUR, J.N., SAHOO, P.K., and WONG, A.K.C.: 'A new method for gray-level picture thresholding using the entropy of the histogram', *Comput. Vis. Graph. Image Process.*, 1985, **29**, pp. 273–285

6 PUN, T.: 'A new method for gray-level picture thresholding using the entropy of the histogram', *Signal Process.*, 1980, **2**, pp. 223–237

7 Brink, A.D.: 'Thresholding of digital images using two dimensional entropies', *Pattern Recognit.*, 1992, **25**, pp. 803–808

8 TSAI, W.: 'Moment-preserving thresholding: a new approach', *Comput. Vis. Graph. Image Process.*, 1985, **29**, pp. 377–393

9 KITTLER, J., and ILLINGWORTH, J.: 'Minimum error thresholding', *Pattern Recognit.*, 1986, **19**, pp. 41–47

10 KURITA, T., OTSU, N., and ABDELMALIK, N.: 'Maximum likelihood thresholding based on population mixture models', *Pattern Recognit.*, 1992, **25**, pp. 1231–1240

11 DUDA, R., and HART, P.: 'Pattern classification and scene analysis' (John Wiley & Sons, 1973)

12 GONZALEZ, R., and WINTZ, P.: 'Digital image processing' (Addison-Wesley Publishing Co., 1986)

13 OTSU, N.: 'A threshold selection method from gray-level histogram', *IEEE Trans. Syst., Man Cybern.*, 1979, **9**, pp. 62–66