

MAXIMUM LIKELIHOOD THRESHOLDING BASED ON POPULATION MIXTURE MODELS

T. KURITA,[†] N. OTSU[†] and N. ABDELMALEK[‡]

[†] Electrotechnical Laboratory, 1-1-4, Umezono, Tsukuba 305, Japan

[‡] National Research Council of Canada, Ottawa, Canada K1A 0R6

(Received 26 February 1991; in revised form 3 February 1992; received for publication 5 March 1992)

Abstract—Maximum likelihood thresholding methods are presented on the basis of population mixture models. It turns out that the standard thresholding proposed by Otsu, which is based on a discriminant criterion and also minimizes the mean square errors between the original image and the resultant binary image, is equivalent to the maximization of the likelihood of the conditional distribution in the population mixture model under the assumption of normal distributions with a common variance. It is also shown that Kittler and Illingworth's thresholding, which minimizes a criterion related to the average classification error rate assuming normal distribution with different variances, is equivalent to the maximization of the likelihood of the joint distribution in the population mixture model. A multi-thresholding algorithm based on Dynamic Programming is also presented.

Thresholding Maximum likelihood Mixture models Dynamic programming

1. INTRODUCTION

Thresholding is a fundamental tool for segmentation of gray level images when object and background pixels can be distinguished by their gray level values. By selecting an adequate threshold, the original image can be transformed into a binary form. Many threshold selection methods have been proposed.^(1,2) Among the global thresholding methods which determine a threshold from the gray level histogram of the image, Otsu's thresholding^(3,4) and Kittler and Illingworth's thresholding⁽⁵⁾ are known as good methods.

The method proposed by Otsu is based on a discriminant criterion. The criterion is also equivalent to minimizing the mean square errors between the original image and the resultant binary image in which average gray levels of each class are assigned to each pixel. The method can be easily extended to multi-threshold problems. He presented an algorithm based on Dynamic Programming for multi-thresholding.⁽⁴⁾ Although the method selects good thresholds for many real images, it has been pointed out that it gives a biased threshold for images in which variances of the gray level of objects and background or populations of corresponding pixels are extremely different (e.g. see reference (5)).

On the other hand, Kittler and Illingworth proposed a thresholding method which optimizes a criterion function related to the average pixel classification error rate under the assumption of objects and background pixel gray level values being normally distributed. The method selects a reasonable threshold even when variances or populations of two classes are different. They also showed an iterative algorithm for multi-thresholding. It has been pointed out that the same criterion can be derived directly from the mini-

mization of equivocation, namely the average conditional entropy.⁽⁶⁾

In this paper, we present thresholding methods which maximize the likelihood based on population mixture models^(7,8) and give a unified interpretation of these two thresholding methods within this framework. Thus we will have a guideline to decide which thresholding method is suitable for a specific problem.

It is shown that the criterion used in Otsu's thresholding is equivalent to the maximization of the likelihood of the conditional distribution under the assumption of normal distributions with a common variance. This gives an explanation why such biased thresholds are selected by Otsu's thresholding for images in which variances of the gray level of objects and background or populations of corresponding pixels are extremely different. We present a modification of Otsu's thresholding for the case of unbalanced populations which is derived from the maximization of the likelihood of the joint distribution with the same assumption.

It is also shown that Kittler and Illingworth's criterion is equivalent to the maximization of the likelihood of the joint distribution under the assumption of normal distributions with different variances. We also present a multi-thresholding algorithm based on Dynamic Programming for Kittler and Illingworth's criterion, namely the maximum likelihood thresholding under the assumption of normal distributions with different variances.

In Section 2, Otsu's and Kittler and Illingworth's thresholding methods are briefly reviewed. Section 3 presents population mixture models and the maximum likelihood thresholding methods. In Section 4, the algorithm for multi-thresholding is described. A conclusion is given in Section 5.

2. PREVIOUS WORKS

Before describing the maximum likelihood thresholding methods, we will briefly review Otsu's^(3,4) and Kittler and Illingworth's⁽⁵⁾ thresholding methods. More detailed reviews on thresholding methods are found in references (1, 2, 6).

2.1. Otsu's thresholding

Let us consider an image whose pixels assume discrete gray level values, g , in the interval $[1, L]$. The distribution of the gray levels in the image can be displayed in the form of a histogram $h(g)$, $g = 1, \dots, L$ which gives the frequency of occurrence of each gray level in the image. It is convenient to normalize the gray level histogram in the form of $p(g) = h(g)/N$, where $N = \sum_{g=1}^L h(g)$ is the total number of pixels in the image.

Now suppose that we are classifying the pixels into two classes C_1 and C_2 (background and objects, or vice versa) by threshold at level k . Here C_1 denotes pixels with levels $[1, \dots, k]$ and C_2 denotes pixels with levels $[k+1, \dots, L]$. Then we have the following statistics depending on the level k :

$$\omega_1(k) = \sum_{g=1}^k p(g) \quad (1)$$

$$\omega_2(k) = \sum_{g=k+1}^L p(g) \quad (2)$$

$$\bar{g}_1(k) = \sum_{g=1}^k gp(g) \quad (3)$$

$$\bar{g}_2(k) = \sum_{g=k+1}^L gp(g) \quad (4)$$

$$\sigma_1^2(k) = \sum_{g=1}^k (g - \bar{g}_1(k))^2 p(g) \quad (5)$$

$$\sigma_2^2(k) = \sum_{g=k+1}^L (g - \bar{g}_2(k))^2 p(g). \quad (6)$$

The total mean and the total variance are given by

$$\bar{g}_T = \sum_{g=1}^L gp(g) = \sum_{j=1,2} \omega_j(k) \bar{g}_j(k) \quad (7)$$

$$\sigma_T^2 = \sum_{g=1}^L (g - \bar{g}_T)^2 p(g) = \sigma_W^2(k) + \sigma_B^2(k) \quad (8)$$

where

$$\sigma_W^2(k) = \sum_{j=1,2} \omega_j(k) \sigma_j^2(k) \quad (9)$$

$$\sigma_B^2(k) = \sum_{j=1,2} \omega_j(k) (\bar{g}_j(k) - \bar{g}_T)^2 \quad (10)$$

are the within-class variance and the between-class variance, respectively.

In order to evaluate the goodness of the threshold, Otsu used the following discriminant criterion:^(3,4)

$$\eta(k) = \sigma_B^2(k) / \sigma_T. \quad (11)$$

The optimal threshold k^* that maximizes $\eta(k)$ is selected by sequential search. From equation (8) and the fact that σ_T^2 is independent of k , this criterion is equivalent to the minimization of the within-class variance $\sigma_W^2(k)$, namely the mean square errors between the original image and the resultant binary image in which average gray levels of each class are assigned to each pixel.

2.2. Kittler and Illingworth's thresholding

It is often realistic to assume that the respective populations are distributed normally with distinct means and variances. Kittler and Illingworth⁽⁵⁾ proposed a method which optimizes a criterion related to the average pixel classification error rate. Suppose that we are classifying the pixels into two classes C_1 and C_2 by threshold at level k and that each of the distributions $f(g|C_j)$, ($j = 1, 2$) are given by a normal distribution with parameters $\bar{g}_j(k)$, $\sigma_j^2(k)$, and a priori probability $\omega_j(k)$. Then from Bayes' formula and the inequality $p \geq 1 + \log(p)$, for $0 \leq p \leq 1$, the average classification error rate can be evaluated as

$$\begin{aligned} P_e &= 1 - \sum_{g=1}^k f(C_1|g)p(g) - \sum_{g=k+1}^L f(C_2|g)p(g) \\ &\geq 1 - \sum_{g=1}^k [1 + \log f(C_1|g)]p(g) \\ &\quad - \sum_{g=k+1}^L [1 + \log f(C_2|g)]p(g) \\ &= \omega_1(k) \log \left(\frac{\sigma_1(k)}{\omega_1(k)} \right) + \omega_2(k) \log \left(\frac{\sigma_2(k)}{\omega_2(k)} \right) \\ &\quad + \frac{1}{2} (1 + \log(2\pi)) + \sum_{g=1}^L p(g) \log(p(g)). \end{aligned} \quad (12)$$

By ignoring the constant terms, we have a criterion

$$J(k) = \omega_1(k) \log \left(\frac{\sigma_1(k)}{\omega_1(k)} \right) + \omega_2(k) \log \left(\frac{\sigma_2(k)}{\omega_2(k)} \right). \quad (13)$$

The same criterion is obtained directly from the minimization of equivocation, namely the average conditional entropy:⁽⁶⁾

$$E = - \sum_{j=1,2} \sum_{g=1}^L f(C_j|g) \log(f(C_j|g)p(g)). \quad (14)$$

Thus, we can say that the criterion proposed by Kittler and Illingworth is essentially the minimization of equivocation rather than the minimization of the error rate. In fact, equivocation is known as one of the upper bounds of Bayes' mis-classification rate.

3. MAXIMUM LIKELIHOOD THRESHOLDING

It is possible to view thresholding as a clustering of the set of pixels in an image into two classes (objects and background). In the case of multi-thresholding, the number of classes may be larger than two. Here,

we will employ population mixture models^(7,8) as the underlying models. The maximum likelihood thresholding methods will be derived based on these models.

3.1. Population mixture models

Let the number of classes be K and the probability density function associated with the j th population be $f(g|C_j) = f_j(g; \beta_j)$, where β_j represents the parameters of the distribution. Suppose that the group identification parameter γ_i ($i = 1, \dots, N$) is associated with each pixel. This parameter indicates the class from which the pixel comes and is equal to j if and only if the pixel belongs to the class C_j ($j = 1, \dots, K$). The parameters γ_i are called incidental parameters, while the parameters β_j are called structural parameters.

It is convenient to reparameterize γ_i by a K -dimensional vector θ_i which consists of $K - 1$ zeros and a single 1, the position of the 1 indicating which class the pixel belongs to. Then the density given θ_i is

$$f(g_i|\theta_i) = \sum_{j=1}^K \theta_{ji} f_j(g_i; \beta_j) \quad (15)$$

where θ_{ji} is the j th component of θ_i . Note that equation (15) can be written as a product

$$f(g_i|\theta_i) = \prod_{j=1}^K f_j(g_i; \beta_j)^{\theta_{ji}}. \quad (16)$$

The marginal probability distribution function of θ is taken to be the point multinomial

$$f(\theta_i) = \prod_{j=1}^K \pi_j^{\theta_{ji}} \quad (17)$$

where $\sum_{j=1}^K \pi_j = 1$.

Then the joint probability distribution function is given by

$$f(g_i, \theta_i) = f(\theta_i) f(g_i|\theta_i) = \prod_{j=1}^K [\pi_j f_j(g_i; \beta_j)]^{\theta_{ji}}. \quad (18)$$

Now suppose that pairs (g_i, θ_i) , $i = 1, \dots, N$ are independent random samples. Then the conditional density of the values of g given the values of θ is given by

$$\begin{aligned} L(g_1, \dots, g_N | \theta_1, \dots, \theta_N) &= \prod_{i=1}^N f(g_i|\theta_i) \\ &= \prod_{i=1}^N \prod_{j=1}^K [f_j(g_i; \beta_j)]^{\theta_{ji}}. \end{aligned} \quad (19)$$

This is the likelihood of the conditional distribution of the population mixture model. On the other hand, the joint density of the values of g and θ is given by

$$\begin{aligned} L(g_1, \dots, g_N, \theta_1, \dots, \theta_N) &= \prod_{i=1}^N f(g_i, \theta_i) \\ &= \prod_{i=1}^N \prod_{j=1}^K [\pi_j f_j(g_i; \beta_j)]^{\theta_{ji}}. \end{aligned} \quad (20)$$

This is the likelihood of the joint distribution.

3.2. Maximum likelihood thresholding methods

If a threshold k is specified, then pixels are classified into two classes C_1 and C_2 . Suppose that the group identification parameters are determined by this thresholding process. Then the likelihood (19) or (20) may vary depending on the threshold. Thus, we can select the threshold whose likelihood is maximum.

Let us consider normal distributions with distinct means and a common variance, namely the probability density functions with each class are given by

$$f(g|C_j) = \frac{1}{\sqrt{(2\pi\sigma^2(k))}} \exp\left(-\frac{(g - \mu_j(k))^2}{2\sigma^2(k)}\right) \quad (j = 1, 2). \quad (21)$$

Then the likelihood (19) becomes

$$\begin{aligned} L(G|\Theta; B(k)) &= \prod_{i=1}^N \prod_{j=1}^2 \left[\frac{1}{\sqrt{(2\pi\sigma^2(k))}} \right. \\ &\quad \times \exp\left(-\frac{(g - \mu_j(k))^2}{2\sigma^2(k)}\right) \left. \right]^{\theta_{ji}(k)} \\ &= (2\pi\sigma^2(k))^{-N/2} \exp\left\{-\frac{1}{2\sigma^2(k)} \right. \\ &\quad \times \sum_{i=1}^N \sum_{j=1}^2 \theta_{ji}(g_i - \mu_j(k))^2 \left. \right\} \end{aligned} \quad (22)$$

where G , Θ , and $B(k)$ are the set of values of g , the set of values of θ , and the set of parameters $\{\mu_1(k), \mu_2(k), \sigma(k)\}$, respectively. By taking the logarithm, we have the log-likelihood

$$\begin{aligned} l(G|\Theta; B(k)) &= -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2(k)) \\ &\quad - \frac{1}{2\sigma^2(k)} \left[\sum_{i=1}^N \sum_{j=1}^2 \theta_{ji}(g_i - \mu_j(k))^2 \right]. \end{aligned} \quad (23)$$

From this log-likelihood, we can obtain the optimal parameters $\hat{\mu}_1(k)$, $\hat{\mu}_2(k)$, $\hat{\sigma}^2(k)$ as

$$\hat{\mu}_j(k) = \bar{g}_j(k) \quad (j = 1, 2) \quad (24)$$

$$\hat{\sigma}^2(k) = \sigma_w^2(k). \quad (25)$$

Thus the maximum log-likelihood is given by

$$\begin{aligned} \hat{l}(G|\Theta; \hat{B}(k)) &= -\frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma_w^2(k)) \\ &\quad - \frac{N}{2}. \end{aligned} \quad (26)$$

By ignoring the constant terms, the maximization of this log-likelihood with respect to k is equivalent to the minimization of $\sigma_w^2(k)$. This implies that the maximum likelihood thresholding based on this model is equivalent to Otsu's thresholding. We will denote by T_0 the threshold selected by this criterion.

Next, consider the maximization of the likelihood of the joint distribution assuming the same distribu-

tions. In this case, the log-likelihood is given by

$$l(G, \Theta; B(k)) = \sum_{i=1}^N \sum_{j=1}^2 \theta_{ji} \log(\pi_j(k)) - \frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma^2(k)) - \frac{1}{2\sigma^2(k)} \left[\sum_{i=1}^N \sum_{j=1}^2 \theta_{ji} (g_i - \mu_j(k))^2 \right]. \quad (27)$$

The optimal parameters $\hat{\pi}_j(k)$ are obtained as $\omega_j(k)$ and the other parameters are the same as the previous case. Thus, we have the maximum log-likelihood

$$\hat{l}(G, \Theta; \hat{B}(k)) = N \sum_{j=1}^2 \omega_j(k) \log(\omega_j(k)) - \frac{N}{2} \log(2\pi) - \frac{N}{2} \log(\sigma_w^2(k)) - \frac{N}{2}. \quad (28)$$

By ignoring the constant terms, we have a criterion

$$Q(k) = \sum_{j=1}^2 \omega_j(k) \log(\omega_j(k)) - \log(\sigma_w(k)). \quad (29)$$

The second term of the right-hand side is equivalent to Otsu's criterion and the first term depends only on the populations of each class. This can be viewed as a modification of Otsu's criterion for the case of unbalanced populations. We will denote by T_Q the threshold selected by this criterion.

Lastly, let us consider the case where the distributions of each class have distinct variances. In this case the probability density functions are given by

$$f(g|C_j) = \frac{1}{\sqrt{2\pi\sigma_j^2(k)}} \exp\left(-\frac{(g - \mu_j(k))^2}{2\sigma_j^2(k)}\right) \quad (j=1, 2). \quad (30)$$

Then the log-likelihood of the joint distribution becomes

$$l(G, \Theta; B(k)) = \sum_{i=1}^N \sum_{j=1}^2 \theta_{ji} \log(\pi_j(k)) - \frac{N}{2} \log(2\pi) - \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^2 \theta_{ji} \log(\sigma_j^2(k)) - \sum_{i=1}^N \sum_{j=1}^2 \frac{1}{2\sigma_j^2(k)} \theta_{ji} (g_i - \mu_j(k))^2. \quad (31)$$

Similarly the maximum log-likelihood is given by

$$\hat{l}(G, \Theta; \hat{B}(k)) = N \sum_{j=1}^2 \omega_j(k) \log(\omega_j(k)) - \frac{N}{2} \log(2\pi) - \frac{N}{2} \sum_{j=1}^2 \omega_j(k) \log(\sigma_j^2(k)) - \frac{N}{2}. \quad (32)$$

By ignoring the constant terms, this is equivalent to the criterion of Kittler and Illingworth's thresholding. Thus it is possible to view Kittler and Illingworth's

criterion as the maximization of the likelihood of the joint distribution in the population mixture model under the assumption of normal distribution with different variances. We will denote by T_K the threshold selected by this criterion.

These results give a unified explanation of these thresholding methods in terms of the maximum likelihood. Thus we have a guideline to decide which thresholding method is suitable for a specific problem. For example, if we can assume that the populations of each class are almost the same and the distributions of each class have a common variance, then Otsu's method is better because the number of parameters of the model to be estimated is smaller than the others and the estimator is more stable. This principle is known as *Occam's Razor*. It was shown in the survey⁽²⁾ that Otsu's thresholding method gave relatively good results for real images. This seems mainly because of the stability of the estimator. On the other hand, if the populations of each class are different and the distributions of each class have different variances, then Kittler and Illingworth's method should be used. Moreover, if you cannot assume the normal distributions, another criterion based on the maximum likelihood should be developed for the adequate distributions.

3.3. Experiments on thresholding

To compare the thresholding criteria described above, we have performed experiments using histograms generated by a normal random generator. Figure 1 shows a histogram generated with the parameters $\pi_1 = 0.5$, $\pi_2 = 0.5$, $\mu_1 = 50.0$, $\mu_2 = 150.0$, $\sigma_1 = 15.0$, and $\sigma_2 = 15.0$. For this histogram, all the thresholds T_O , T_Q , and T_K were the same. This shows that these criteria are comparable when the populations of each class are the same and the distributions of each class are normals with a common variance. Thus, Otsu's method is the best among the three because the number of parameters is smaller.

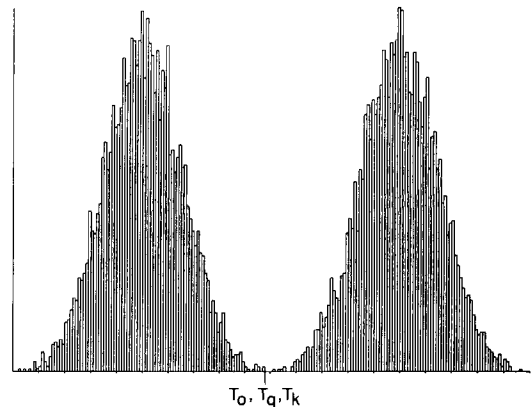


Fig. 1. A bimodal histogram generated with parameters $\pi_1 = 0.5$, $\mu_1 = 50.0$, $\sigma_1 = 15.0$; $\pi_2 = 0.5$, $\mu_2 = 150.0$, $\sigma_2 = 15.0$. T_O , T_Q , and T_K are experimentally determined thresholds.

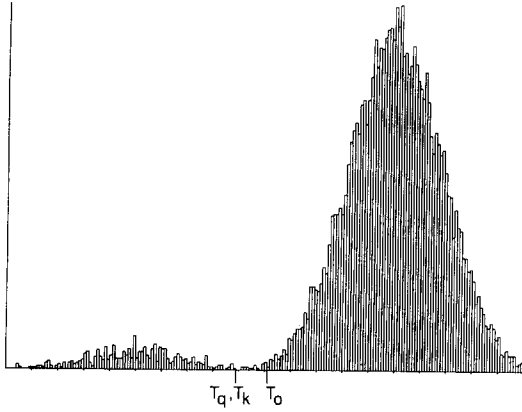


Fig. 2. A bimodal histogram generated with parameters $\pi_1 = 0.05, \mu_1 = 50.0, \sigma_1 = 18.0; \pi_2 = 0.95, \mu_2 = 150.0, \sigma_2 = 18.0$. T_O , T_Q , and T_K are experimentally determined thresholds.

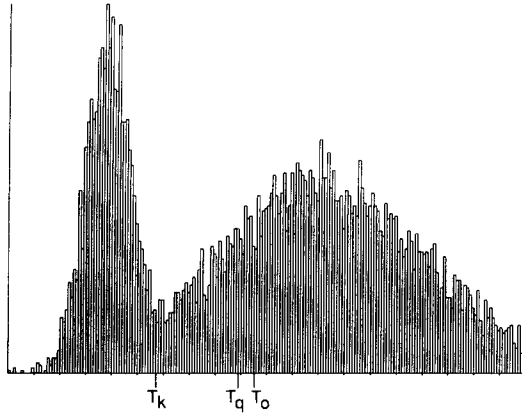


Fig. 3. A bimodal histogram generated with parameters $\pi_1 = 0.25, \mu_1 = 38.0, \sigma_1 = 8.0; \pi_2 = 0.75, \mu_2 = 121.0, \sigma_2 = 40.0$. T_O , T_Q , and T_K are experimentally determined thresholds.

Figure 2 shows a histogram generated with the parameters $\pi_1 = 0.05, \pi_2 = 0.95, \mu_1 = 50.0, \mu_2 = 150.0, \sigma_1 = 18.0$, and $\sigma_2 = 18.0$. The populations of each class are extremely different but the variances are the same. In this case, the threshold T_O was biased, while T_Q and T_K gave a good threshold. Thus, in such cases, the thresholding method using the criterion $Q(k)$ defined in equation (29) is better than Kittler and Illingworth's method because the number of parameters is smaller.

Figure 3 shows a histogram generated with the parameters $\pi_1 = 0.25, \pi_2 = 0.75, \mu_1 = 38.0, \mu_2 = 121.0, \sigma_1 = 8.0$, and $\sigma_2 = 40.0$. The populations of each class are different and the distributions of each class have different variances. For this case, only the threshold T_K was good. Thus we should use Kittler and Illingworth's thresholding in such cases because the assumptions of T_O or T_Q are not satisfied.

4. ALGORITHM FOR MULTI-THRESHOLDING

For multi-thresholding, Otsu proposed an algorithm based on Dynamic Programming.⁽⁴⁾ The algorithm finds the optimum thresholds exactly. On the other hand, Kittler and Illingworth presented an iterative algorithm for his criterion.⁽⁵⁾ The algorithm is fast, but it may be trapped at some local optimum and it heavily depends on the initial thresholds. Other divisive or agglomerative algorithms for hierarchical clustering could be applied to this problem.

In this section, we will show an algorithm based on Dynamic Programming for Kittler and Illingworth's criterion, namely the maximum likelihood thresholding under the assumption of normal distributions with different variances. We will refer to this method as MTK (multi-thresholding using Kittler and Illingworth's criterion). The algorithm is almost the same as Otsu's algorithm except that the criterion is different.

4.1. Dynamic programming

The extension to the multi-threshold problem of the maximum likelihood thresholding is straightforward. Consider a classification of pixels into K classes $\{C_1, \dots, C_K\}$ by $K-1$ thresholds $\{k_1, \dots, k_{K-1}\}$. Here C_j ($j = 1, \dots, K$) denotes pixels with levels $[k_{j-1} + 1, \dots, k_j]$, where $k_0 = 0$ and $k_K = L$. Then we can also compute the statistics of each class depending on the thresholds similar to those in Section 2. We will denote them $\omega(k_j, k_{j-1})$, $\bar{g}(k_j, k_{j-1})$, and $\sigma^2(k_j, k_{j-1})$ for $j = 1, \dots, K$.

By similar calculations, the maximum log-likelihood of the joint distribution under the assumption of the normal distributions with different variances is given by

$$\begin{aligned} \hat{l}(G, \Theta; \hat{B}(k_1, \dots, k_{K-1})) \\ = N \sum_{j=1}^K \omega(k_j, k_{j-1}) \log(\omega(k_j, k_{j-1})) - \frac{N}{2} \log(2\pi) \\ - \frac{N}{2} \sum_{j=1}^K \omega(k_j, k_{j-1}) \log(\sigma^2(k_j, k_{j-1})) - \frac{N}{2}. \end{aligned} \quad (33)$$

By ignoring constant terms, this is equivalent to the minimization of the following criterion:

$$J(k_1, \dots, k_{K-1}) = \sum_{j=1}^K \omega(k_j, k_{j-1}) \log \left(\frac{\sigma(k_j, k_{j-1})}{\omega(k_j, k_{j-1})} \right). \quad (34)$$

Again, this is equivalent to Kittler and Illingworth's criterion for multi-thresholding. According to this criterion, we can determine the optimal thresholds by searching the minimum of $J(k_1, \dots, k_{K-1})$.

To search the optimal thresholds, we can use Dynamic Programming as follows.^(4,9) Let

$$g(k, m) = \omega(k, m) \log \left(\frac{\sigma(k, m)}{\omega(k, m)} \right), \quad k > m. \quad (35)$$

The objective function to be minimized is thus

$$J_{K,L}(k_1, \dots, k_{K-1}) = \sum_{i=1}^K g(k_i, k_{i-1}) \quad (36)$$

where $0 = k_0 < k_1 < \dots < k_{K-1} < k_K = L$. Define the partial sum

$$J_{M,l}(k_1, \dots, k_{M-1}) = \sum_{j=1}^M g(k_j, k_{j-1}) \quad (37)$$

where $0 = k_0 < k_1 < \dots < k_{M-1} < k_M = l$. Let

$$J_M^*(l) = \min_{1 \leq k_1 < \dots < k_{M-1} < l} J_{M,l}(k_1, \dots, k_{M-1}). \quad (38)$$

With this notation, it is clear that the principle of optimality is satisfied as

$$J_M^*(l) = \min_{M-1 < k_{M-1} < l} \{J_{M-1}^*(k_{M-1}) + g(l, k_{M-1})\}. \quad (39)$$

Thus the optimal set of thresholds can be determined by computing the $J_M^*(l)$, $l = M, \dots, L - K + M$ for $M = 2, 3, \dots, K - 1$. Since $O(L^2)$ evaluations of g are required for each level M , the time complexity of this algorithm is $O(KL^2)$.

Similarly we can show that the multi-thresholding method proposed by Otsu (MTO) is equivalent to the maximization of the likelihood of the conditional distribution under the assumption of normal distributions with a common variance.

4.2. Experiments on multi-thresholding

Figures 4–6 show randomly generated histograms with equal populations and a common variance $\sigma = 10$. For these histograms, both algorithms selected

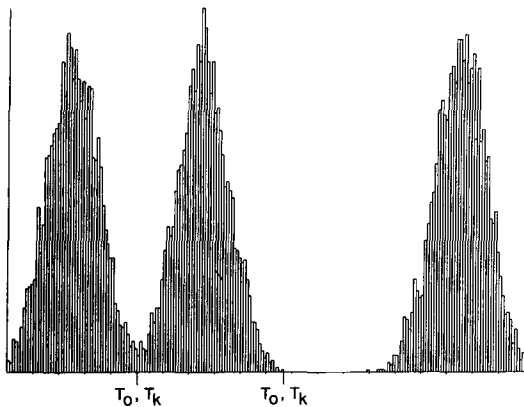


Fig. 4. A trimodal histogram generated with equal populations and a common variance $\sigma = 10.0$. Values of T_O and T_K are experimentally determined thresholds.

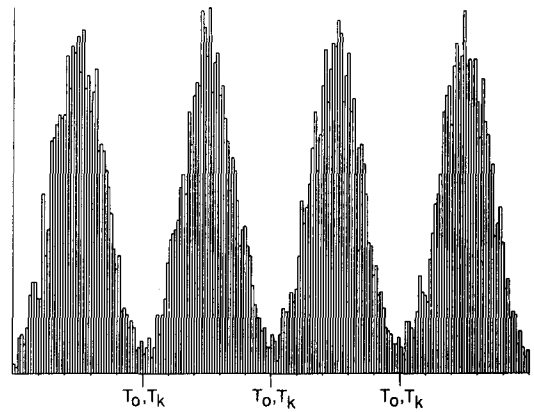


Fig. 5. A four modal histogram generated with equal populations and a common variance $\sigma = 10.0$. Values of T_O and T_K are experimentally determined thresholds.

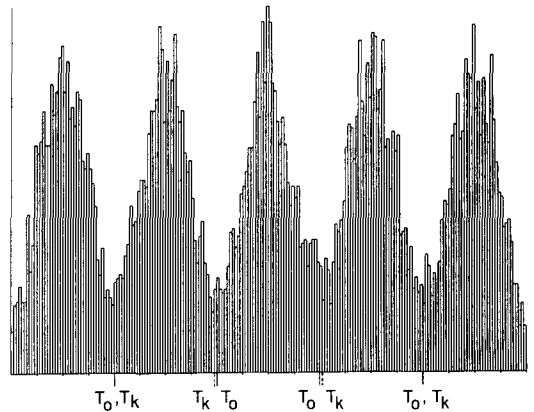


Fig. 6. A five modal histogram generated with equal populations and a common variance $\sigma = 10.0$. Values of T_O and T_K are experimentally determined thresholds.

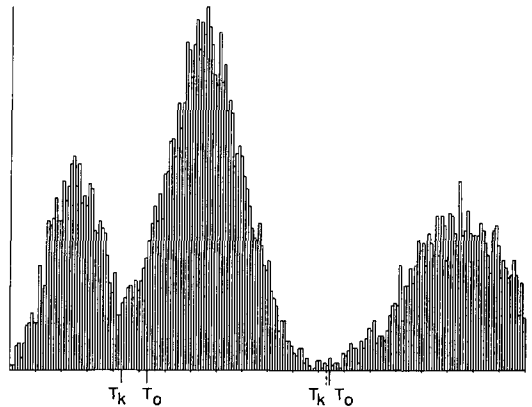


Fig. 7. A trimodal histogram generated with different populations and different variances. Values of T_O and T_K are experimentally determined thresholds.

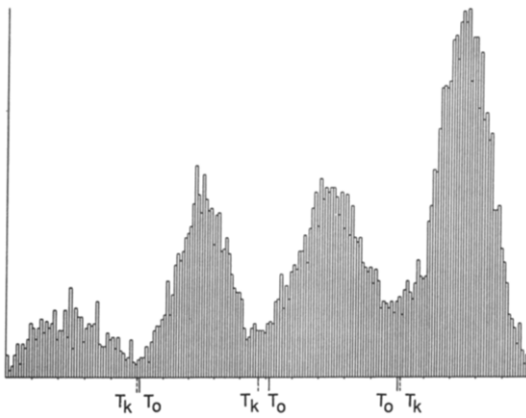


Fig. 8. A four modal histogram generated with different populations and different variances. Values of T_o and T_k are experimentally determined thresholds.

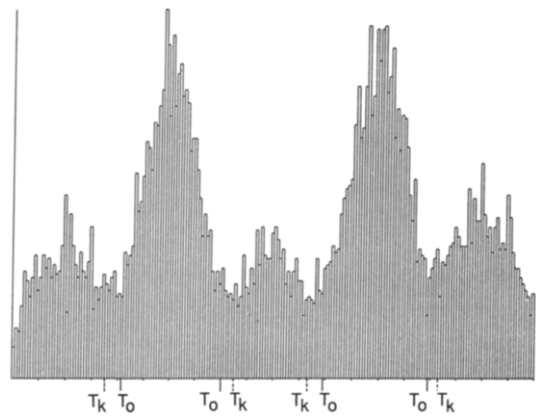


Fig. 9. A five modal histogram generated with different populations and different variances. Values of T_o and T_k are experimentally determined thresholds.

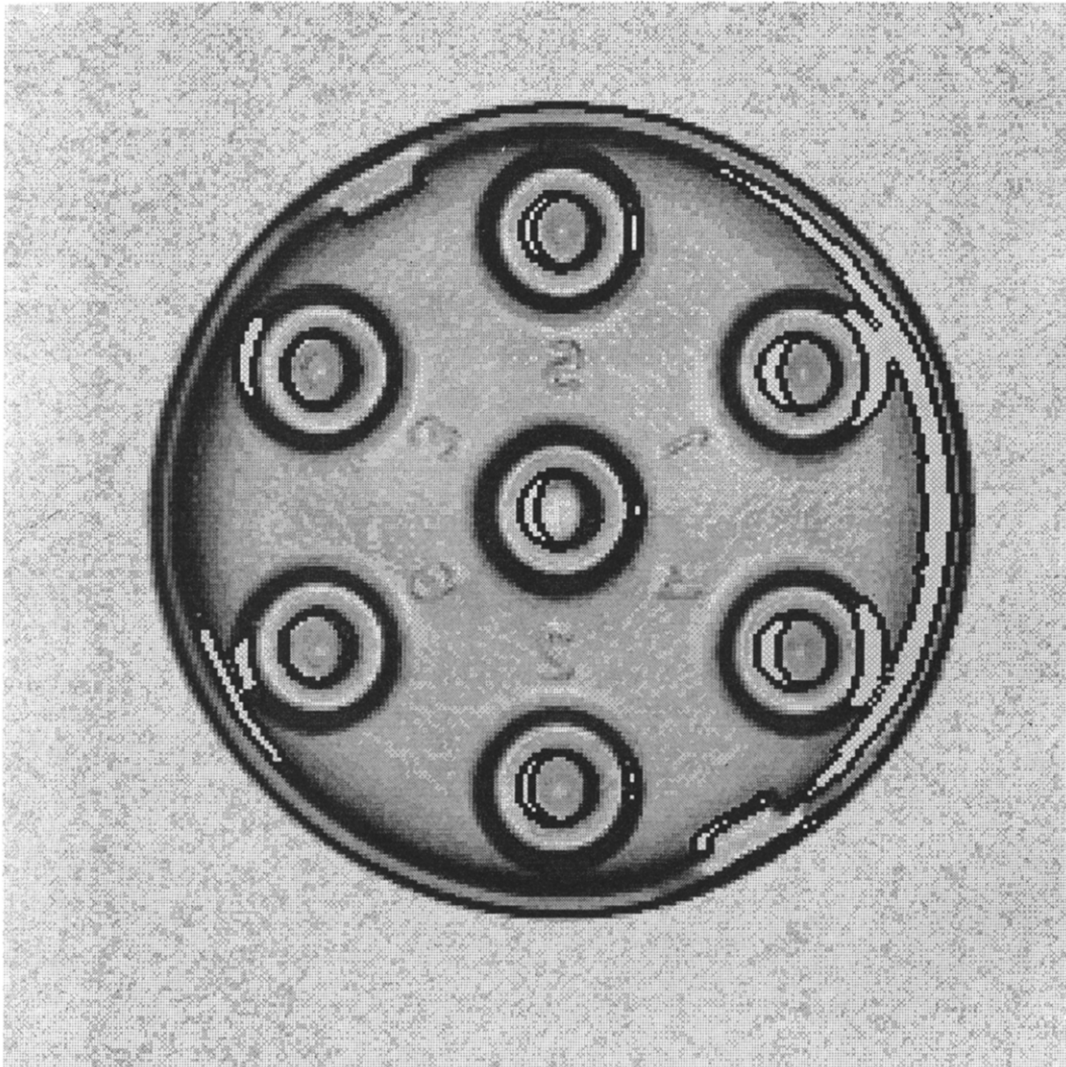


Fig. 10. An original range image (256 by 256 pixels).

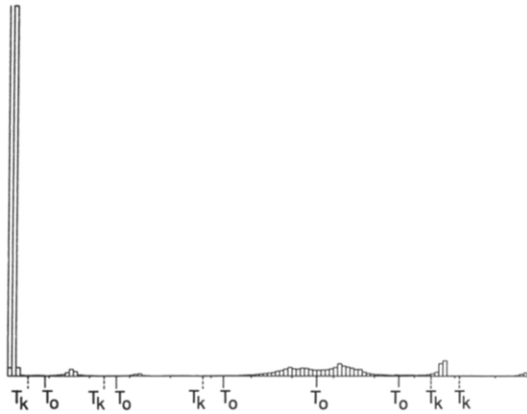
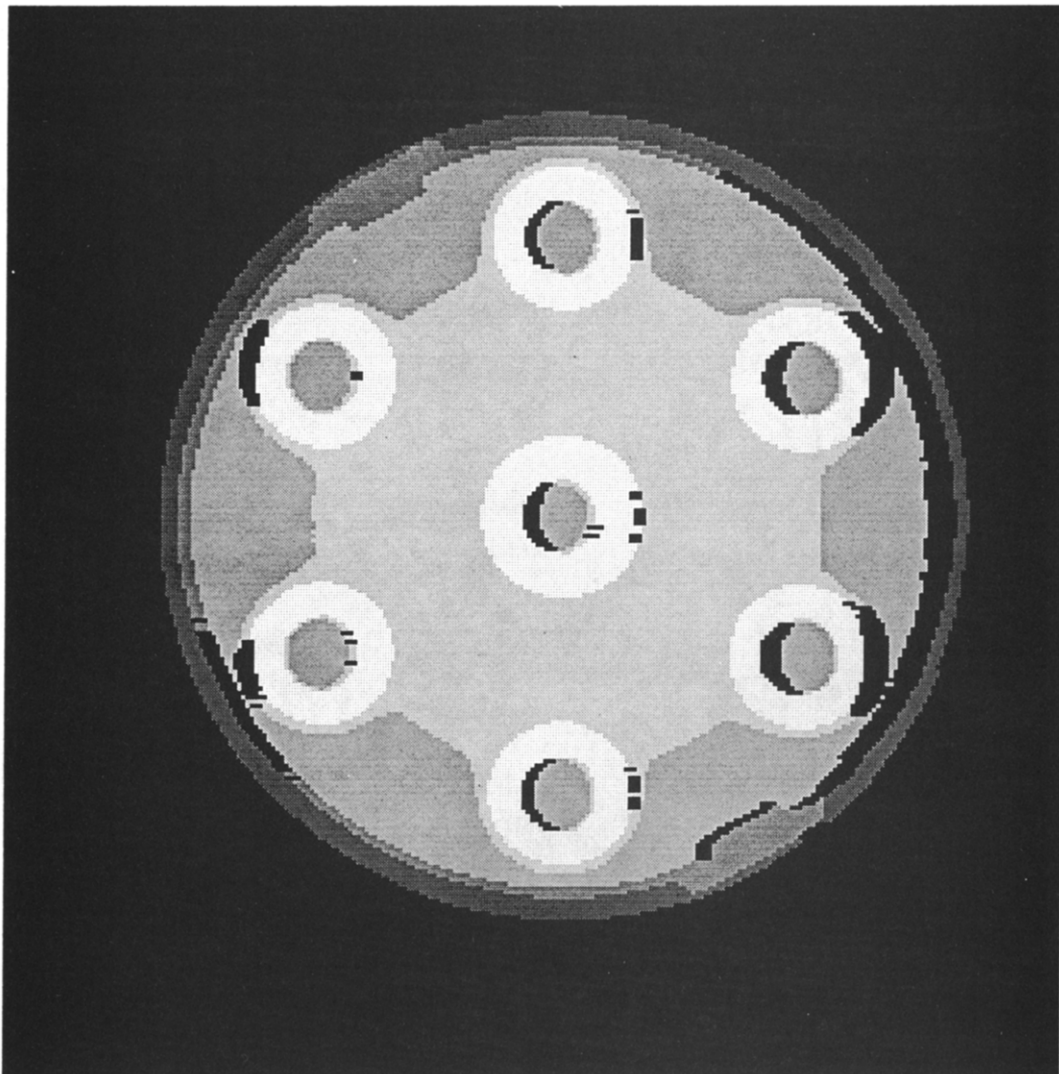


Fig. 11. The histogram of the range image in Fig. 10. Values of T_0 and T_k are experimentally determined thresholds (six classes).

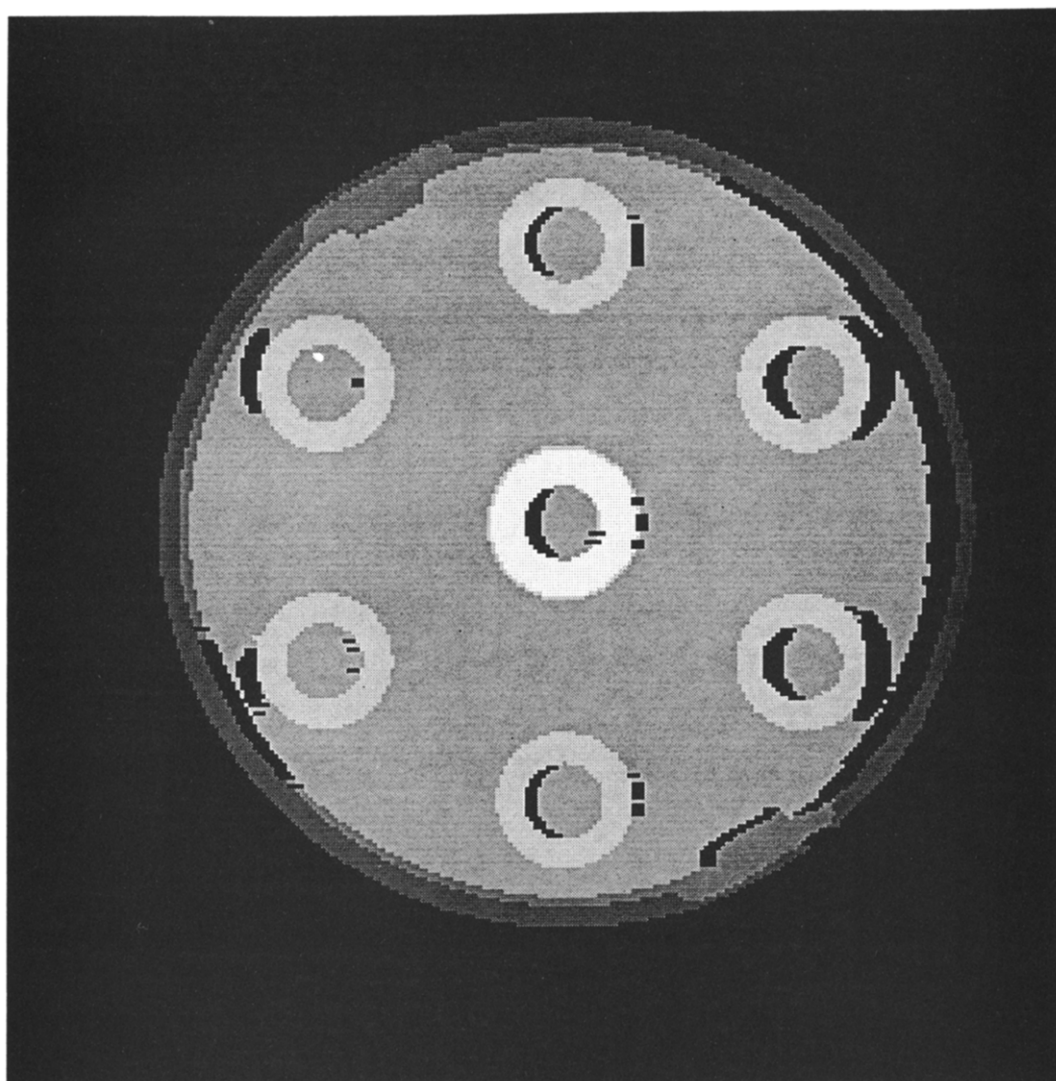
almost similar thresholds. On the other hand, for histograms with different populations and different variances (Figs 7–9), MTK gave a good result as expected while the MTO criterion selected slightly biased thresholds since the assumption of this algorithm was not satisfied.

We also applied them to segmentation of range data. Figure 10 shows an original range image (256 by 256 pixels) and Fig. 11 is the histogram of the image. Results of segmentations of the image in Fig. 10 by MTO and MTK are shown in Figs 12(a) and (b), respectively. The number of classes was set to 6 for this image. It seems that MTK gives a better result for this image.



(a)

Fig. 12. (a) The segmentation result of the range image in Fig. 10 by MTO. (b) The segmentation result of the same image by MTK.



(b)

Fig. 12. (Continued.)

5. CONCLUSION

Maximum likelihood thresholding methods have been presented on the basis of population mixture models. We showed that the standard thresholding method proposed by Otsu is equivalent to the maximization of the likelihood of the conditional distribution under the assumption of normal distributions with a common variance, and that the minimum error thresholding method proposed by Kittler and Illingworth is equivalent to the maximization of the likelihood of the joint distribution under the assumption of normal distributions with different variances. This gives a unified explanation of these thresholding methods in terms of the likelihood maximization and also gives a guideline to decide which method is suitable for a specific problem.

For multi-thresholding, similar results were obtained. We also presented a multi-thresholding algorithm using Dynamic Programming for Kittler and Illingworth's criterion.

Another criterion based on the maximum likelihood should be derived by assuming adequate distributions, when the assumption of normal distributions does not hold. For example, thresholding of edge values would be interesting.

REFERENCES

1. J. S. Weszka, A survey of threshold selection techniques, *Comput. Graphics Image Process.* **7**, 259–265 (1978).
2. P. K. Sahoo, S. Soltani, A. K. C. Wong and Y. C. Chen, A survey of thresholding techniques, *Comput. Vision Graphics Image Process.* **41**, 233–260 (1988).

3. N. Otsu, A threshold selection method from gray-level histograms, *IEEE Trans. Syst. Man Cybern.* **SMC-9**(1), 62–66 (1979).
4. N. Otsu, An automatic threshold selection method based on discriminant and least squares criteria, *Trans. IECE* **63-D**(4), 349–356 (1980).
5. J. Kittler and J. Illingworth, Minimum error thresholding, *Pattern Recognition* **19**, 41–47 (1986).
6. N. Otsu, Automatic thresholding methods from histograms, *Proc. 4th Conf. of Japanese Classification Society*, pp. 8–15 (1987).
7. S. L. Sclove, Population mixture models and clustering algorithms, *Commun. Statist.-Theor. Meth.* **A6**(5), 417–434 (1977).
8. S. L. Sclove, Application of the conditional population-mixture model to image segmentation, *IEEE Trans. Pattern Analysis Mach. Intell.* **PAMI-5**(4), 428–433 (1983).
9. R. Bellman, On the approximation of curves by line segments using dynamic programming, *Commun. ACM* **4**, 284 (1961).

About the Author—TAKIO KURITA joined Electrotechnical Laboratory, Agency of Industrial Science and Technology, MITI, Japan, in April 1981. From 1990 to 1991 he was a visiting research scientist at Institute for Information Technology, National Research Council of Canada, in Ottawa. His research interests include multivariate analysis, pattern recognition, cluster analysis and image processing. He received the B.A. degree in electronics engineering from Nagoya Institute of Technology in 1981.

About the Author—NOBUYUKI OTSU received the B. Eng., the M. Eng. and D. Eng. degrees in mathematical engineering from the University of Tokyo, Tokyo, Japan, in 1969, 1971 and 1981, respectively. Since 1971, he has been with the Electrotechnical Laboratory, Agency of Industrial Science and Technology, MITI, Japan. From 1982 to 1983 he was a visiting research scientist at the Department of Electrical Engineering, National Research Council of Canada, in Ottawa. From 1985 he was the Section Chief, Mathematical Informatics Section, Information Science Division, and has directed research projects on soft logic for the foundation of flexible information processing and on neuro-computing. Since 1991 he has been the Director of the Machine Understanding Division. His main research interests are theoretical aspects of pattern recognition, image processing and multivariate data analysis.

About the Author—NABIH NESSIM ABDELMALEK received his B.Sc. degree in electrical engineering in 1951 and his B.Sc. degree in mathematics in 1954, both from Cairo University, Egypt, and his Ph.D. degree in theoretical nuclear physics in 1958 from Manchester University, U.K. After 7 years in the Department of Mathematics in the Faculty of Engineering, Cairo University, he moved to Canada, and became a member of the scientific staff at Bell Northern Research in Ottawa. After a period of 2 years he joined the National Research Council of Canada, in Ottawa. His current interest is the application of numerical analysis techniques to problems of image processing, particularly image enhancement and restoration, data compression, pattern classification, parameter estimation and 3D robot vision.