A Gray-Level Threshold Selection Method Based on Maximum Entropy Principle

Andrew K. C. Wong
P. K. Sahoo

Reprinted from
IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS
Vol. 19, No. 4, July/August 1989
A Gray-Level Threshold Selection Method Based on Maximum Entropy Principle

ANDREW K. C. WONG, MEMBER, IEEE, AND P. K. SAHOO

Abstract — The gray-level threshold selection method for image segmentation presented here is based on the maximum entropy principle. The optimal threshold value is determined by maximizing the posteriori entropy subject to certain inequality constraints which are derived by means of special measures characterizing uniformity and the shape of the

Manuscript received November 3, 1987; revised December 22, 1988. This work was supported in part by the Natural Science and Engineering Research Council of Canada.

A. K. C. Wong is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada N2L 3G1.

P. K. Sahoo is with the Department of Mathematics, University of Louisville, Louisville, KY 40292.

IEEE Log Number R927237.
Image thresholding is essentially a pixel classification problem. Its basic objective is to classify the pixels of a given image into two classes: those pertaining to an object and those pertaining to the background. While one includes pixels with gray values that are below or equal to a certain threshold, the other includes those with gray values above the threshold. Thresholding is a popular tool for image segmentation. It is widely used in halftone reproduction [12], [14], [27], automatic target recognition [3], design of visual navigation system for autonomous land vehicles [25], [29], [30], industrial applications of computer vision [9], and biomedical image analysis [4], [6], [19], [23]. Over the years many threshold selection techniques have been proposed. For a survey of threshold selection techniques, readers may refer to [26] and [31].

In general, threshold selection techniques can be broadly divided into two groups, namely, global and local thresholding. A global technique may be point-dependent or region-dependent. The thresholding method is point-dependent if the threshold value is determined solely from the pixel gray tone as represented by gray-level histogram and is independent of the gray tone of the neighborhood of a pixel. On the other hand, a method is called region-dependent if the threshold value is determined from the local property within a neighborhood of a pixel. A global thresholding technique is one that thresholds the entire image with a single threshold value, whereas a local thresholding technique is one that partitions a given image into subimages and determines a threshold for each of these subimages.

Though point-dependent global thresholding methods are comparatively simple, they use only a small part of the independent information a provided by the image. Therefore, these techniques are not always useful. For example, in halftone reproduction, global point-dependent methods rarely produce good pictures.

In this correspondence we propose a global threshold selection method which uses both gray-level distribution and spatial information. The work is organized as follows. In Section II, we briefly review some recent methods on thresholding. These methods have not been considered in the recent survey article [26] or in [31]. Section III describes the proposed thresholding method and its theoretical justifications. Section IV reports the effectiveness of our thresholding method when applied to some real-world images. In Section V, we show how this method is extended for segmenting chromatic images. Finally, in Section VI, we provide some concluding remarks regarding the method proposed.

II. Review of Recent Works

Over the years many threshold selection methods have been proposed. All methods optimize criterion functions based on information obtained from either gray-level histogram or spatial distribution. Some of these methods use information from both sources to select a threshold. A recent paper [26] attempts to survey many thresholding methods that have appeared in recent literature. Unfortunately, a few techniques known prior to 1986 were not included in [26]. In this section, we review only the most important recent works that had not been covered.

To relate the reviewed work in [26] to the proposed scheme, we first introduce some basic notations and definitions. Let $Z^+$ be the set of all positive integers, $(x, y)$ be the spatial coordinate of a pixel in a digitized image, $Z_+ = \{0, 1, \ldots, 1\}$ be the set of gray levels, where $l$ is the total number of quantization levels. Then, an image function can be defined as the mapping $h: Z^+ \times Z^+ \rightarrow Z^+$. The brightness (i.e., the gray level) of a pixel at $(x, y)$ is denoted as $h(x, y)$.

Let $t \in Z^+$. After thresholding an image by the threshold $t$ we obtain a bilevel image. The image function $h(x, y)$ at gray level $t$ is a binary image function $h_t(x, y)$. The explicit representation of $h_t(x, y)$ is expressed as

$$h_t(x, y) = \begin{cases} a_1, & \text{if } h(x, y) \leq t \\ a_2, & \text{if } h(x, y) > t. \end{cases}$$

The constant $a_1$ is usually 0 and $a_2$ is $l - 1$.

Let $f_0, f_1, \ldots, f_{l-1}$ be the observed gray-level frequencies of an image. The probability of the gray level $i$ in an image can be estimated as

$$p_i = \frac{f_i}{N_0}, \quad N_0 = \sum_{i=0}^{l-1} f_i, \quad i = 0, 1, \ldots, l-1$$

where $N_0$ is the total number of pixels in the image and $i$ is the total number of gray levels. Dunn et al. [5] proposed a uniform error threshold selection method which equalizes the probability of misclassification between the object and the background pixels in an image. The optimal threshold value in uniform error method is obtained by finding a threshold $t \in Z^+$ such that

$$b(t) = b(t+1) - (a^2(t) + c(t)) = 0$$

where

$$a(t) = \text{Pr} \{ \text{a pixel has gray value } > t \}$$

$$b(t) = \text{Pr} \{ \text{two adjacent pixels have gray values } > t \}$$

$$c(t) = \text{Pr} \{ \text{four neighboring pixels have gray values } > t \}$$

The probabilities $a(t), b(t),$ and $c(t)$ are estimated by examining $2 \times 2$ neighborhoods in the image for fixed $t$. The uniform error method is a global region-dependent thresholding method. Aviad and Loomskii [1] proposed a thresholding method based on minimizing the ambiguity which results from the different contexts in which gray levels are classified as object or background. Pavlidis and Wolberg [22] have given an algorithm based on a model of the image distortion. Forte and Sahoo [7] introduced a measure for the loss of information due to reduction on the range of a real random variable and proposed a thresholding method based on the loss of information criterion. Their method belongs to the class of point-dependent global techniques. Another information theoretic method that recently appeared in the literature is due to Beattie [2]. Other recent publications related to automatic threshold selection include [10], [13], and [17].

III. Maximum Entropy Thresholding

The success of the maximum entropy method for solving problems in image reconstruction [32], [33] and other areas [13] has encouraged us to apply this method to threshold selection. As is well-known, the maximum entropy principle serves as a criterion to select $a$ priori probability distributions when very little or nothing is known. It states that, for a given amount of information, the probability distribution which best describes our knowledge is the one that maximizes the Shannon entropy subject to the given evidence as constraints.

Let $t \in Z^+$. After thresholding an image by the threshold $t$ we obtain a bilevel image. The $a$ posteriori probability of the pixels with gray values less than $t$ is given by

$$F(t) = \sum_{i=0}^{t} p_i.$$  

(1)

Similarly, the $a$ posteriori probability of all those pixels with values greater than or equal to $t$ is $1 - F(t)$. Thus the Shannon entropy of the bilevel picture is

$$H(F(t)) = -F(t) \log F(t) - (1 - F(t)) \log (1 - F(t)).$$

(2)

where the logarithm is taken with respect to base 2 and $\log_2$ is assumed to be zero. If nothing else is known regarding the image,
we would maximize (2) to obtain the optimal threshold value for the image. However, unconstrained maximization of (2) with respect to \( F(t) \) yields \( F(t) = 0.5 \). Thus, if nothing else is known, then the maximum entropy principle suggests that one should choose the threshold value such that the percentage of white and black pixels should be equal. This is precisely the case if the histogram is bimodal with even contribution from each mode. However, the histogram of images is not always bimodal and other information regarding the image should be incorporated into the thresholding method in such cases. Moreover, information obtained from the spatial distribution of gray levels plays an important role in many image-processing tasks such as edge detection, texture analysis, image restoration, etc. Two measures are derived from the spatial information and using these measures we find some constraints for the probability of the white pixels as well as those of the black pixels. The two measures are a) uniformity measure and b) shape measure.

Levine and Nazif [18] introduced the uniformity measure which serves to evaluate the performance of a segmentation method. The uniformity of a feature (e.g., gray value) over a region is inversely proportional to the variance of the values of that feature evaluated at every pixel belonging to that region. For detail regarding this measure, readers may refer to Levine and Nazif [18]. Suppose that an image is segmented into two regions by thresholding at a gray level \( t \). Then the uniformity measure [18] is

\[
U(t) = 1 - \frac{\sigma^2 + \sigma_i^2}{C_i}
\]

where

\[
\sigma^2 = \sum_{(x, y) \in R,} (h(x, y) - \mu)^2,
\]

\[
R_i = \text{segmented region } i
\]

\[
h(x, y) = \text{gray level of pixel } (x, y)
\]

\[
\mu_i = \frac{\sum_{(x, y) \in R_i} h(x, y)}{A_i}
\]

and where \( A_i \) is the number of pixels in \( R_i, i = 1, 2, \) and \( C_i \) is a positive constant for normalization.

Since region uniformity is an important factor in image segmentation, we would like to segment the image based on maximum uniformity. Hence we first determine a threshold

\[
t_1 = \arg \max_{t \in Z} U(t)
\]

If we threshold the image at the gray level \( t_1 \), we would obtain two regions with maximum uniformity.

The shape of an object in an image is also an important factor. In fact, the problem of finding optimal threshold for an image is a dual to the problem of finding edges or features related to the shape of an object in that image. It is the characteristic related to the shape of an object that makes the object distinguishable from the background. A measure that provides information regarding the shape (e.g., edges) of an object can be constructed as follows:

a) assign a generalized gradient value \( \Delta(x, y) \) to each pixel \( (x, y) \); b) if the pixel \( (x, y) \) has a gray value higher than the average of its neighbors, then assign the plus sign to the generalized gradient value \( \Delta(x, y) \); else, assign the minus value.

c) compute the shape measure \( S \) using the formula

\[
S(t) = \frac{\sum_{(x, y)} \text{sgn}(h(x, y) - \bar{h}_{N(x, y)}) \Delta(x, y) \text{sgn}(h(x, y) - t)}{C_2}
\]

where \( \bar{h}_{N(x, y)} \) is the average gray value in the neighborhood \( N(x, y) \), \( t \) is the threshold value of the image, \( C_2 \) is a positive normalization factor, and

\[
\text{sgn}(x) = \begin{cases} 1, & \text{if } x > 0 \\ -1, & \text{if } x < 0 \end{cases}
\]

The computation of the generalized gradient value \( \Delta(x, y) \) of the pixel \( (x, y) \) is carried out using

\[
\Delta(x, y) = \sum_{k=1}^{4} D_k^2 + \sqrt{2} D_1(D_3 + D_4) - \sqrt{2} D_2(D_3 - D_4)
\]

where

\[
D_1 = h(x+1, y) - h(x-1, y)
\]

\[
D_2 = h(x, y+1) - h(x, y-1)
\]

\[
D_3 = h(x+1, y+1) - h(x-1, y-1)
\]

\[
D_4 = h(x+1, y-1) - h(x-1, y+1).
\]

According to the shape measure \( S \) (see [26]), if the image is segmented at the gray level \( t_2 \) such that

\[
t_2 = \arg \max_{t \in Z} S(t),
\]

then the image obtained after thresholding will retain the maximum amount of information regarding edges, as the measure \( S \) encourages segmentation of pixels with high gradient value.
Since $F$ is an increasing function of $t$ (see Fig. 1), we obtain from (5)
\[
\min(F(t_1), F(t_2)) \leq F(t) \leq \max(F(t_1), F(t_2)).
\] (6)

The inequalities in (6) provide constraints for the a posteriori entropy maximization problem mentioned earlier. It was shown that the maximization of $H(F(t))$ subject to the inequality (6) always yields unique solution [8]. Thus the optimal threshold $t^*$ can be determined from
\[
t^* = \arg \max_{t \in \mathbb{Z}_+} H(F(t))
\]
such that
\[
\min(t_1, t_2) \leq t^* \leq \max(t_1, t_2).
\]

To solve the constrained maximization problem, we assume without loss of generality that $\min(t_1, t_2) = t_1$. Furthermore, assume that $F(t_1)$ and $F(t_2)$ are in $[0, 0.5]$. Then $F(t_1) > F(t_2)$ since $F$ is an increasing function. Hence the Shannon entropy is maximized at $F(t_1)$ (see Fig. 2(a)). If $F(t_1)$ and $F(t_2)$ are in $[0.5, 1]$ then $H(F(t))$ is maximum at $F(t_1)$ (see Fig. 2(b)). However, if $F(t_1) \in [0, 0.5]$ and $F(t_2) \in [0.5, 1]$, then $H$ is maximum at $F(t = 0.5$ (see Fig. 2(c)). Thus the maximum entropic optimal threshold $t^*$ can be obtained from the equation
\[
t^* = \begin{cases} 
\max(t_1, t_2), & \text{if } F(t_1), F(t_2) \in [0, 0.5] \\
\min(t_1, t_2), & \text{if } F(t_1), F(t_2) \in [0.5, 1] \\
t, & \text{if } F(t_1) \in [0, 0.5], F(t_2) \in [0.5, 1] \\
t, & \text{if } F(t_2) \in [0, 0.5], F(t_1) \in [0.5, 1],
\end{cases}
\]
where $t$ is a gray-level such that $F(t) = 0.5$.

IV. SOME EXPERIMENTAL RESULTS

In this section, we present the outcome of the new thresholding method when applied to three real-world images, a cameraman, a building, and a model. The cameraman image is digitized with a raster of $415 \times 395$ and quantized to 256 gray levels. The digitized image and its gray level histogram are shown in Fig. 3. The
building image is also digitized with a raster of $411 \times 403$ and quantized to 256 gray levels. The image of the model is digitized with a raster of $321 \times 314$ and quantization is 256 gray levels. The building and the image of the model are shown in Figs. 4 and 5 along with their gray-level histograms. These three images were considered in [26] and the performance of some well-known methods were discussed relative to these test images.

For these images, the gray levels which maximize the uniformity measure and the shape measure were determined. Table I summarizes the gray-level pixel values. Using the proposed method the optimal thresholds for these three images are tabulated in Table II.

The digitized images shown in Figs. 3–5 are thresholded using the optimal threshold value $t^*$ for each of them. The binary images obtained by thresholding are shown in Figs. 6–8. On the basis of an investigation with some real-world images (other than the images considered above), we have come to the conclusion

---

**Table I**

<table>
<thead>
<tr>
<th>Measure</th>
<th>Cameraman</th>
<th>Building</th>
<th>Model</th>
</tr>
</thead>
</table>

**Table II**

<table>
<thead>
<tr>
<th>$F(t_1)$</th>
<th>$F(t_2)$</th>
<th>$t^*$</th>
</tr>
</thead>
</table>

---

Fig. 4. Digitized building image and its gray-level histogram.

Fig. 5. Digitized image of model and its gray-level histogram.

Fig. 6. Binary image after thresholding cameraman image at $t^* = 87$.

Fig. 7. Binary image after thresholding building image at $t^* = 74$.

Fig. 8. Binary image after thresholding image of model at $t^* = 66$. 
that this new method of thresholding gives a better threshold value than thresholding techniques such as Otsu [21], Kapur et al. [16], and Tsai [28]. Compare the optimal threshold values \( t^* \) with the optimal threshold values computed using different methods for these test images from Tables I and II in [26].

V. EXTENSION TO CHROMATIC IMAGES

The image function \( h \) for monochromatic images is a scalar valued function representing the intensity values at the image pixels. Color images are formed by superimposing red, green, and blue components. Thus the chromatic image function is a vector valued function \( h: Z^+ \times Z^+ \rightarrow Z^+_R \times Z^+_G \times Z^+_B \). The sets \( Z^+_R, Z^+_G, \) and \( Z^+_B \) denote the set of quantization values for the red, blue, and green components, respectively. The red, green, and blue (\( R, G, \) and \( B \)) system can be transformed into a number of other coordinate systems [20]. Each pixel in a color image is denoted as a three-dimensional (3-D) vector of the \( R, G, \) and \( B \) components. Thus the vector valued function \( h \) at the pixel \((x, y)\) can be decomposed into

\[
h(x, y) = (h_R(x, y), h_G(x, y), h_B(x, y))
\]

where \( h_R(x, y) \), \( h_G(x, y) \), and \( h_B(x, y) \) represent the intensities of the red, green, and blue components of the pixel at \((x, y)\), respectively. The length of the vector \( h(x, y) \) is

\[
|h(x, y)| = \left[ h_R^2(x, y) + h_G^2(x, y) + h_B^2(x, y) \right]^{1/2}
\]

We apply the maximum entropic thresholding method suggested in Section III to each of the functions \( h_R(x, y), h_G(x, y), \) and \( h_B(x, y) \). Let \( t_R^*, t_G^*, \) and \( t_B^* \) be the optimal maximum entropic threshold values. The magnitude of the vector \( t^* = (t_R^*, t_G^*, t_B^*) \) is

\[
|t^*| = \left[ t_R^2 + t_G^2 + t_B^2 \right]^{1/2}
\]

The result of thresholding a chromatic image \( h(x, y) \) at \( t_c = (t_R, t_G, t_B) \) is a binary (black and white) scalar valued image function \( h_c(x, y) \). The explicit form of \( h_c(x, y) \) is the following:

\[
h_c(x, y) = \begin{cases} 
    a_1, & \text{if } |h(x, y)| \leq |t_c| \\
    a_2, & \text{if } |h(x, y)| > |t_c|
\end{cases}
\]

where the constant \( a_1 = 0 \) and \( a_2 = 1 - 1 \).

VI. CONCLUSION

Many thresholding methods tend to disregard either spatial information or information related to the gray-level distribution. Our method, based upon the maximum entropy principle, uses both spatial information as well as gray-level distribution. Experimental results have shown that the proposed method is highly effective. Its effectiveness is measured by the quality of the binary pictures using both a uniformity and a shape criterion function. These advantages definitely offset the relatively higher computational time needed. This method can be further extended to multithresholding.

ACKNOWLEDGMENT

The authors gratefully acknowledge several helpful comments of the referee.

REFERENCES