IMAGE THRESHOLDING BASED ON ALI–SILVEY DISTANCE MEASURES

LIANE C. RAMAC and PRAMOD K. VARSHNEY*
ECE Department, Syracuse University, Syracuse, NY 13244, U.S.A.
(Received 27 November 1995; received for publication 15 October 1996)

Abstract—A relative entropy based approach to image thresholding was proposed recently. It was demonstrated that this method was successful for image thresholding. Relative entropy is a member of the class of Ali–Silvey distance measures. In this paper we generalize the relative entropy based approach and present image thresholding algorithms based on the class of Ali–Silvey distance measures. A number of members of this class are selected and used for implementation in image thresholding algorithms. Performance is evaluated by applying these algorithms to several images and comparing them to a few histogram based thresholding methods. © 1997 Pattern Recognition Society. Published by Elsevier Science Ltd.

1. INTRODUCTION

Digital image processing consists of several steps beginning with problem definition and ending with the results. One of the fundamental steps of digital image processing is segmentation. In general, this involves dividing an image into its constituent regions or classes. Segmentation may involve finding the edges of a region or finding all the pixels in a region. Thresholding is an important tool for segmentation. The goal of thresholding is to divide an image into two or more regions or classes based on a given criterion function. In this paper we only examine the case of two regions or bilevel thresholding where the images are divided with one gray level for the entire image.

Several thresholding techniques are available today and these techniques fall into different categories. Surveys are available which compare a wide range of these algorithms. Some methods only require the gray level histogram of the image to separate an image. When the histogram of an image is bimodal, the task of finding the optimal threshold is a relatively simple matter. Choosing the valley of the histogram will usually result in good segmentation. Other algorithms utilize spatial information to determine an optimal threshold. In many cases the image’s gray level histogram is not bimodal. Therefore, other characteristics of the image may contribute to better threshold selection.

Recently, thresholding algorithms based on information theoretic approaches have received much attention. A method proposed by Chang et al. uses relative entropy, also known as the Kullback–Leibler number, as a criterion measure. Relative entropy belongs to a generalized class of functions known as the Ali–Silvey distance measures. Since relative entropy has been found useful in the area of image thresholding, an extension to the general class of Ali–Silvey distance measures appears to be a logical next step.

The terms used in representing images and their gray level distributions are defined in Section 2. In Section 3, we define the class of Ali–Silvey distance measures and then present a general Ali–Silvey distance based algorithm for image thresholding. Then the relative entropy approach and a few other Ali–Silvey distance measures are presented with derivations of their criterion functions. Next, three of the histogram-based thresholding algorithms are considered for performance comparisons with the Ali–Silvey distance measures. Section 4 contains the results of testing all the methods on several images. The section also has an analysis for the individual images and a performance comparison among the images. Finally, we present a summary and discuss results in Section 5.

2. BACKGROUND

This paper only deals with the binary thresholding of gray level images. Here we will present the notation used throughout the paper. A new definition of the thresholded image’s co-occurrence matrix will also be presented. One method for calculating the probabilities in the co-occurrence matrix for a thresholded image was defined by Chang et al. Our method is a modification of their definition.
In equation form, a pixel and its gray level are represented by one element in a matrix. Given an $M \times N$ image with $l$ gray levels $G = \{0, 1, 2, \ldots, l-1\}$, the matrix is represented by

$$F = [f(x,y)]_{M \times N},$$  
(1)

where $(x,y)$ is the spatial coordinate of the pixel and $f(x,y) \in G$ denotes the gray level of the pixel corresponding to coordinate $(x,y)$. The darkest gray level is 0 and the lightest is $l-1$.

After a threshold $t$ is selected, it is used to separate the image into two classes, $C_0 = \{0, 1, 2, \ldots, t\}$ and $C_1 = \{t+1, t+2, \ldots, l-1\}$. The threshold is $t \in G$, and $g_0, g_1 \in G$ are the two gray levels into which the image is thresholded. The gray level of each pixel in the thresholded image is assigned a value such that

$$f_t(x,y) = \begin{cases} g_0 & \text{if } f(x,y) \leq t, \\ g_1 & \text{if } f(x,y) > t, \end{cases}$$  
(2)

where $f_t(x,y)$ represents the image function of the thresholded image.

### 2.1. Co-occurrence matrix

The histogram provides point-by-point information about an image and does not contain any spatial information. The spatial information can be specified by defining the so-called co-occurrence matrix. Its elements are transition probabilities that rely on the values of the neighboring pixels. For two given gray levels, $i$ and $j$, the $(ij)$th entry of the matrix represents the relative frequency with which the two gray levels occur next to each other where the coordinates $i$ and $j$ need not be distinct. Note that these transition probabilities are joint probabilities and not conditional probabilities. The term transition probability often refers to a conditional probability but we continue to use this term for the sake of consistency with its usage in image thresholding literature. Most of the transition probabilities will lie on or near the matrix diagonal.

The information about the image is given in the form of a co-occurrence matrix, $T$. If an $M \times N$ image contains $l$ gray levels, the co-occurrence matrix has size $l \times l$ and the $(ij)$th entry is the transition probability of that gray level pair in the original image. They are defined in reference (7) and represented as follows:

$$P_{ij} = \frac{t_{ij}}{\sum_{i=0}^{l-1} \sum_{j=0}^{l-1} t_{ij}},$$  
(3)

where

$$t_{ij} = \sum_{m=1}^{M} \sum_{n=1}^{N} \delta(m,n),$$  
(4)

and

$$\delta(m,n) = \begin{cases} 1 & \text{if } f(m,n) = i, f(m+1,n) = j \\
0 & \text{otherwise}.
\end{cases}$$

The $p_{ij}$s are the transition probabilities corresponding to pixel pairs with gray levels $i$ and $j$. While computing these probabilities, only the pixels below and to the right of $(x,y)$ are considered here. This reduces computational complexity. Also, it has been found that adding the pixels above and to the left does not significantly improve the results. (7)

The co-occurrence matrix can then be divided into quadrants which represent within-class transitions and cross-class transitions. The threshold $t$ divides the matrix into quadrants and quadrant probabilities are defined in reference (5) as

$$P_A(t) = \sum_{i=0}^{t} \sum_{j=0}^{t} p_{ij},$$  
(5a)

$$P_B(t) = \sum_{i=0}^{t} \sum_{j=t+1}^{l-1} p_{ij},$$  
(5b)

$$P_C(t) = \sum_{i=t+1}^{l-1} \sum_{j=0}^{t} p_{ij},$$  
(5c)

$$P_D(t) = \sum_{i=t+1}^{l-1} \sum_{j=t+1}^{l-1} p_{ij}.$$  
(5d)

Quadrant $A$ contains the transition probabilities within class $C_0$ and quadrant $C$ contains the probabilities representing transitions within class $C_1$. Quadrants $B$ and $D$ represent the transition probabilities across the boundaries between the two classes. Note that $P_A(t)+P_B(t)+P_C(t)+P_D(t)=1$.

After the binary thresholding of the image, the elements of the thresholded co-occurrence matrix are functions of $t$ and are determined from the quadrant probabilities. The quadrant probabilities are re-distributed in the thresholded co-occurrence matrix among the elements whose gray level pairs $(ij)$ are non-zero in the co-occurrence matrix for the original image. So, for a given gray level $t$, the probabilities are given as

$$p_{ij}^{(k)}(t) = q_k(t) \cdot s_{ij}, \quad k = A, B, C, D,$$  
(6)

where

$$q_k(t) = P_k(t)/n_k(t)$$

and

$$n_A(t) = \sum_{i=0}^{t} \sum_{j=0}^{t} s_{ij}, \quad n_B(t) = \sum_{i=0}^{t} \sum_{j=t+1}^{l-1} s_{ij},$$

$$n_C(t) = \sum_{i=t+1}^{l-1} \sum_{j=0}^{t} s_{ij}, \quad n_D(t) = \sum_{i=t+1}^{l-1} \sum_{j=t+1}^{l-1} s_{ij},$$

$$s_{ij} = \begin{cases} 1 & \text{if } p_{ij} \neq 0 \\ 0 & \text{if } p_{ij} = 0.
\end{cases}$$

Since there are only two gray levels in a thresholded image, the transition probabilities within the quadrants are constant. But the quadrant probabilities, $P_A(t)$, $P_B(t)$, $P_C(t)$, and $P_D(t)$, used to calculate these values are taken...
from the original image. The values \( q_A(t), q_B(t), q_C(t), \) and \( q_D(t) \) defined above are modifications of the thresholded transition probabilities given in the paper by Chang et al. (5).

Comparing the transition probabilities defined in equation (6) and those in reference (5), for a given threshold \( t \), the transition probabilities defined in equation (6) will be greater than those defined in reference (5). Also, the transition probabilities we defined maintain more information from the original image in that they maintain the spatial relationships of the gray levels by assigning non-zero values to the thresholded co-occurrence matrix elements which correspond to non-zero values for the original image. The probabilities defined in reference (5) estimate the thresholded probability distribution by assigning non-zero values to gray level transitions even when they do not occur in the original image.

3. THRESHOLDING METHODS

Over the years, various methods for image thresholding have been proposed in the literature. (2-4) One group of algorithms uses entropic methods to determine the threshold for a given image. Earlier entropic thresholding methods took advantage of the image’s gray level histogram as an estimate of a probability distribution for calculating the thresholding criterion functions. Some recent methods use the co-occurrence matrix for the probability distribution. One algorithm was proposed by Pal and Pal (7) which used local and joint entropy as criterion functions. Another method that used the co-occurrence matrix was proposed by Chang et al. (5) whose criterion function is based on relative entropy.

Relative entropy is a member of a more general class of distance measures, namely Ali–Silvey distance measures (also known as \( f \)-divergences). In this section, we propose a generalization of the threshold finding algorithm of Chang et al. (5). Specifically, in Section 3.1 we present a methodology based on Ali–Silvey distance measures. A number of specific distance measures are proposed in the paper by Chang et al. (5) which used local and joint entropy as criterion functions. Another method that used the co-occurrence matrix was proposed by Chang et al. (5) whose criterion function is based on relative entropy.

Relative entropy is a member of a more general class of distance measures, namely Ali–Silvey distance measures (also known as \( f \)-divergences). In this section, we propose a generalization of the threshold finding algorithm of Chang et al. (5). Specifically, in Section 3.1 we present a methodology based on Ali–Silvey distance measures. A number of specific distance measures are proposed in the paper by Chang et al. (5) which used local and joint entropy as criterion functions. Another method that used the co-occurrence matrix was proposed by Chang et al. (5) whose criterion function is based on relative entropy.

### 3.1. Thresholding based on Ali–Silvey distance measures

Before presenting our algorithms, we define the class of Ali–Silvey distance measures. Our criterion functions use equation (6) for defining the thresholded image probabilities \( p' \) instead of the probabilities defined by Chang et al. (5). The general expression for the distance measures, given two probability distributions, is

\[
D(p, p') = E_0[C(L)],
\]

where \( p \) and \( p' \) are two distributions in the same space, \( D(p, p') \) is the distance between \( p \) and \( p' \), \( f \) is an increasing function, \( E_0 \) denotes expectation with respect to \( p \), \( C \) is a convex function and \( L \) is the likelihood ratio. There are several different distance measures in this class. Some members of the Ali–Silvey class that we consider here are

**Relative entropy or Kullback–Leibler number:**

\[
I = E_0[\ln(L)],
\]

**J-Divergence (Symmetric divergence):**

\[
J = E_0[(L - 1)\ln(L)],
\]

**Bhattacharyya distance:**

\[
B = -\ln(E_0[\sqrt{L}]),
\]

**Matsusita distance:**

\[
M = (E_0[(\sqrt{L} - 1)^2])^{1/2},
\]

**Chernoff Distance:**

\[
C_\alpha = -\ln(E_0[L^\alpha]), \quad 0 < \alpha < 1.
\]

Note that the Chernoff distance will reduce to the Bhattacharyya distance when \( \alpha = \frac{1}{2} \); and the Matsusita distance and the Bhattacharyya distance are increasing functions of each other.

Chang et al. used relative entropy as a thresholding method based on the idea that little information should be lost between the original and thresholded images. So, minimizing the distance with respect to \( t \) will result in the optimal threshold \( t' \). One may employ any member of the class of Ali–Silvey distance measures to determine the threshold. Here we consider five members of the class as examples. In general, the threshold \( t' \) can be obtained as

\[
t' = \text{Arg Min}_{0 < t < 1} D(t),
\]

where \( D \) is any Ali–Silvey distance defined in equation (7).

#### 3.1.1. Relative entropy approach (5)

Chang et al. proposed the relative entropy approach as another entropic thresholding method. The joint relative entropy of the probability distributions of the original \( p \) and thresholded \( p' \) images is defined as

\[
I(p; p') = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \ln(p_{ij}/p'_{ij}).
\]

It is a distance measure between the probability distributions of the original and thresholded images. So finding the optimal threshold, \( t' \), is a matter of minimizing \( I(p; p') \) with respect to \( t \). Expanding equation (10) results in

\[
I(p; p') = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \ln p_{ij} - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p'_{ij} \ln p'_{ij},
\]

Then, in order to use this measure as a criterion function, it can be further expanded with respect to the quadrants. If \( I(p; p') \) is divided into sums over the quadrants, then the result will vary depending on the value of \( t \). The first term of equation (11) is constant for
all values of \( t \), so only the second term needs to be considered. Using equations (5a)-(d) and (6),

\[
I'(t) = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln p_{ij} = \sum_{A} p_{ij} \ln q_A(t) + \sum_{B} p_{ij} \ln q_B(t) + \sum_{C} p_{ij} \ln q_C(t) + \sum_{D} p_{ij} \ln q_D(t)
\]

\[
+ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln q_C(t) + \sum_{D} p_{ij} \ln q_D(t).
\]

Using equations (10) and (12),

\[
I'(t) = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln p_{ij} + \sum_{C} p_{ij} \ln q_A(t) + \sum_{D} p_{ij} \ln q_B(t) + \sum_{B} p_{ij} \ln q_C(t) + \sum_{C} p_{ij} \ln q_D(t).
\]

Finding the threshold is a matter of minimizing equation (11) or maximizing equation (12):

\[
t^* = \text{ArgMax}(I'(t)).
\]

3.1.2. Formulations using other Ali-Silvey distance measures. Relative entropy as defined in equation (10) can be interpreted as an Ali-Silvey distance measure by recognizing the following in equation 8(a):

\[
L = p'_{ij}/p_{ij}
\]

and

\[
E_0[C(L)] = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} C(L).
\]

Substituting these into the expressions for other Ali-Silvey distance measures provides new criterion functions for determining an image’s threshold value. Substituting equation 14(a) and (b) into equation 8(b)-(e) results in the following:

\[
J = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} \left\{ p'_{ij} \ln \left( p'_{ij}/p_{ij} \right) + p_{ij} \ln \left( p_{ij}/p'_{ij} \right) \right\}
\]

\[
= I + \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p'_{ij} \ln \left( p_{ij}/p_{ij} \right),
\]

\[
B = -\ln \left\{ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij}p'_{ij})^{1/2} \right\},
\]

\[
C_h = -\ln \left\{ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij}^{1-h})(p_{ij}')^{h} \right\}
\]

and

\[
M = \left[ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} \left( p'_{ij}^{1/2} - p_{ij}^{1/2} \right)^2 \right]^{1/2}
\]

Now these equations may be expanded with respect to the quadrants. Since they are all distance measures, they should be minimized with respect to \( t \) to find the optimal threshold \( t^* \):

\[
J = I + \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p'_{ij} \ln \left( p_{ij}/p_{ij} \right)
\]

\[
= \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln p_{ij} - \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln p_{ij}'
\]

\[
+ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij}' \ln p_{ij}' - \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln p_{ij}'.
\]

Studying the equation, the first term is constant for all \( t \) (it is the same term as in the relative entropy criterion function) and the second and third terms cancel each other. This may not be apparent, but if the second and third terms are expanded over the four quadrants’ the result is

\[
J'(t) = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln p_{ij} + \sum_{C} p_{ij} \ln q_A(t) + \sum_{D} p_{ij} \ln q_B(t) + \sum_{B} p_{ij} \ln q_C(t) + \sum_{C} p_{ij} \ln q_D(t).
\]

Therefore, minimizing \( J \) is the same as maximizing the last term in equation 16(a). The optimal threshold using the J-Divergence criterion function is found as

\[
t^* = \text{ArgMax}(J'(t)),
\]

where

\[
J'(t) = \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} p_{ij} \ln \left( p_{ij}/p_{ij} \right)
\]

\[
= q_A(t) \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} \ln(p_{ij}) + q_B(t) \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} \ln(p_{ij})
\]

\[
+ q_C(t) \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} \ln(p_{ij}) + q_D(t) \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} \ln(p_{ij}).
\]

In terms of the quadrants, the Bhattacharyya distance measure expands to the following:

\[
B(t) = -\ln \left[ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij}p'_{ij})^{1/2} \right]
\]

\[
= -\ln \left[ \sum_{A} (p_{ij}p'_{ij})^{1/2} + \sum_{B} (p_{ij}p'_{ij})^{1/2} + \sum_{C} (p_{ij}p'_{ij})^{1/2} + \sum_{D} (p_{ij}p'_{ij})^{1/2} \right]
\]

\[
= -\ln \left[ \left\{ q_A(t) \right\}^{1/2} \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij})^{1/2} + \left\{ q_B(t) \right\}^{1/2} \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij})^{1/2} + \left\{ q_C(t) \right\}^{1/2} \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij})^{1/2} + \left\{ q_D(t) \right\}^{1/2} \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij})^{1/2} \right]\].

The threshold \( t^* \) is found by minimizing \( B \):

\[
t^* = \text{ArgMin}(B(t)).
\]
The Chernoff distance is expanded with respect to the quadrants as follows:

\[ C_h(t) = -\ln \left[ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij})^{(1-\alpha)} \right] \]

\[ = -\ln \left[ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij})^{(1-\alpha)} \right] + \{q_B(t)\}^{(1-\alpha)} + \{q_C(t)\}^{(1-\alpha)} \]

Minimizing the distance function to find \( t^* \):

\[ t^* = \text{Arg Min}_{0 < t < l-1} (C_h(t)). \]  

(18)

The Matsusita distance is expanded with respect to the quadrants as follows:

\[ M(t) = \left[ \sum_{i=0}^{l-1} \sum_{j=0}^{l-1} (p_{ij})^{1/2} - (p_{ij})^{1/2} \right]^{1/2} \]

\[ = \left[ 2 - 2 \left\{ q_A(t) \right\}^{1/2} + q_B(t)^{1/2} + q_C(t)^{1/2} + q_D(t)^{1/2} \right]^{1/2} \]

Minimizing the distance function to find \( t^* \):

\[ t^* = \text{Arg Min}_{0 < t < l-1} (M(t)). \]  

(19)

3.2. Histogram-based thresholding

In order to evaluate the performance of Ali–Silvey distance measures, three common thresholding methods are employed for comparison. They are point dependent because each pixel in the image is classified based only on its original gray level. The only information needed to implement these algorithms is the gray level histogram of the image.

The histogram-based thresholding methods tested here are the Otsu thresholding method, \(^{(14)}\) the Moment Preserving method, \(^{(15)}\) and the Minimum Error method. \(^{(16)}\) The Otsu algorithm is based on maximizing the separability of the two classes in the image. The threshold value for the Moment Preserving algorithm is computed deterministically in a way to preserve the moments of the original, unthresholded image. It may be interpreted as an image transformation which recovers the original image from a blurred version. The Minimum Error algorithm is based on the idea that the gray level distributions of the classes are Gaussian and the distribution of the total image is bimodal.

4. PERFORMANCE METRICS

A number of measures have been used to evaluate the performance of different thresholding techniques. The two measures used here to evaluate thresholding techniques are uniformity and shape measures. These measures are not completely satisfactory and a better method of evaluating the threshold choice is needed. In addition to the uniformity and shape, resulting binary images for visual analysis are also presented.

Uniformity. \(^{(17)}\) The uniformity measure is inversely proportional to the variances of the two classes. The measure is based on the assumption that when an image is correctly thresholded, the variances in the separate classes will be minimized. The uniformity of an image is calculated as follows:

\[ U(t) = 1 - \frac{\sigma_0^2 + \sigma_1^2}{\sigma_0^2 + \sigma_1^2}, \]  

(21)

where

\[ \sigma_i^2 = \sum_{(x,y) \in C_i} (f(x,y) - \mu_i)^2, \]

\[ C_i = \text{Segmented region i}, \]

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

\[ \mu_i = \left\{ \sum_{(x,y) \in R_i} f(x,y) \right\} / A_i, \]

\[ A_i = \text{number of pixels in C_i}, \]

\[ i = 0, 1, \]

and

\[ N_i = \text{a normalization factor}. \]

Since the uniformity measure is derived from the variances of the image, it is expected that the Otsu algorithm will perform well in this category.

Shape. \(^{(1)}\) The shape measure evaluates an image based on how each pixel \((x,y)\) in the image compares to the average gray level of its neighborhood. The measure is determined as follows:

If \( f(x,y) \geq f_{avg} \) and \( f(x,y) \geq t \) or \( f(x,y) < f_{avg} \) and \( f(x,y) < t \) then add to \( S \), otherwise subtract from \( S \).

The shape measure is defined as

\[ S = \sum_{(x,y)} \text{Sgn}[f(x,y) - f_{avg}] \Delta(x,y) \text{Sgn}[f(x,y) - t], \]  

(22)

where \( f_{avg} \) is the average gray level in the neighborhood of \((x,y)\), \( t \) is the threshold value of the image, \( N_i \) is a normalization factor, and

\[ \text{Sgn}(x) = \begin{cases} 
1 & \text{if } x \geq 0, \\
-1 & \text{if } x < 0.
\end{cases} \]
The generalized gradient value $\Delta(x,y)$ of the pixel $(x,y)$ is calculated as follows:

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

(23)

where $D_1 = f(x+1,y) - f(x-1,y)$, $D_2 = f(x,y+1) - f(x,y-1)$, $D_3 = f(x+1,y+1) - f(x-1,y-1)$, $D_4 = f(x+1,y-1) - f(x-1,y+1)$.

The best shape measure with respect to $t$ for a specific image is used to normalize the values of all shape measures. One problem with judging a thresholding algorithm based only on these two metrics is that they are not necessarily consistent with each other. Unless there is a distinct object on a background, they may produce results that are contradictory. This is shown in the performance results.

5. EXPERIMENTAL RESULTS

Several different images with varying sizes and gray levels are employed to evaluate the performance of different thresholding techniques. The shape of the histogram and the distribution of the co-occurrence matrix are two additional image characteristics under considera-

Table 1. Performance for elephant image

<table>
<thead>
<tr>
<th>Method</th>
<th>Threshold</th>
<th>Uniformity</th>
<th>Shape</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative entropy, Chernoff (0.25)</td>
<td>227</td>
<td>0.7783</td>
<td>-0.0283</td>
<td>0.3750 [6]</td>
</tr>
<tr>
<td>J-Divergence, Bhattacharyya, Matsusita</td>
<td>188</td>
<td>0.8507</td>
<td>0.4833</td>
<td>0.6670 [5]</td>
</tr>
<tr>
<td>Chernoff (0.75)</td>
<td>182</td>
<td>0.8651</td>
<td>0.5434</td>
<td>0.7067 [4]</td>
</tr>
<tr>
<td>Otsu</td>
<td>122</td>
<td>0.9359</td>
<td>0.9690</td>
<td>0.9525 [1]</td>
</tr>
<tr>
<td>Moment preserving</td>
<td>140</td>
<td>0.9258</td>
<td>0.9098</td>
<td>0.9178 [2]</td>
</tr>
<tr>
<td>Minimum error</td>
<td>55</td>
<td>0.9248</td>
<td>0.6727</td>
<td>0.7987 [3]</td>
</tr>
</tbody>
</table>

Table 2. Performance for Lenna image

<table>
<thead>
<tr>
<th>Method</th>
<th>Threshold</th>
<th>Uniformity</th>
<th>Shape</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ali-Silvey</td>
<td>84</td>
<td>0.9421</td>
<td>0.9994</td>
<td>0.9707 [1]</td>
</tr>
<tr>
<td>Otsu</td>
<td>92</td>
<td>0.9411</td>
<td>0.9945</td>
<td>0.9678 [2]</td>
</tr>
<tr>
<td>Moment preserving</td>
<td>101</td>
<td>0.9378</td>
<td>0.9505</td>
<td>0.9441 [3]</td>
</tr>
<tr>
<td>Minimum error</td>
<td>41</td>
<td>0.9373</td>
<td>0.6354</td>
<td>0.7864 [4]</td>
</tr>
</tbody>
</table>

Table 3. Performance for San Francisco image

<table>
<thead>
<tr>
<th>Method</th>
<th>Threshold</th>
<th>Uniformity</th>
<th>Shape</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative entropy, Chernoff (0.25)</td>
<td>98</td>
<td>0.9442</td>
<td>0.9477</td>
<td>0.9460 [4]</td>
</tr>
<tr>
<td>J-Divergence, Bhattacharyya, Matsusita</td>
<td>87</td>
<td>0.9403</td>
<td>0.8784</td>
<td>0.9094 [6]</td>
</tr>
<tr>
<td>Chernoff (0.75)</td>
<td>97</td>
<td>0.9439</td>
<td>0.9432</td>
<td>0.9436 [5]</td>
</tr>
<tr>
<td>Chernoff (0.75)</td>
<td>79</td>
<td>0.9371</td>
<td>0.8207</td>
<td>0.8789 [7]</td>
</tr>
<tr>
<td>Otsu</td>
<td>108</td>
<td>0.9469</td>
<td>0.9924</td>
<td>0.9697 [1]</td>
</tr>
<tr>
<td>Moment preserving</td>
<td>105</td>
<td>0.9462</td>
<td>0.9856</td>
<td>0.9659 [3]</td>
</tr>
<tr>
<td>Minimum error</td>
<td>128</td>
<td>0.9482</td>
<td>0.9877</td>
<td>0.9679 [2]</td>
</tr>
</tbody>
</table>

Table 4. Performance for Castle image

<table>
<thead>
<tr>
<th>Method</th>
<th>Threshold</th>
<th>Uniformity</th>
<th>Shape</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative entropy, Chernoff (0.25)</td>
<td>65</td>
<td>0.9662</td>
<td>0.9890</td>
<td>0.9776 [2]</td>
</tr>
<tr>
<td>J-Divergence, Bhattacharyya, Matsusita, Chernoff</td>
<td>64</td>
<td>0.9656</td>
<td>1.0000</td>
<td>0.9828 [1]</td>
</tr>
<tr>
<td>Otsu</td>
<td>74</td>
<td>0.9693</td>
<td>0.9564</td>
<td>0.9628 [4]</td>
</tr>
<tr>
<td>Moment preserving</td>
<td>69</td>
<td>0.9679</td>
<td>0.9712</td>
<td>0.9697 [3]</td>
</tr>
<tr>
<td>Minimum error</td>
<td>78</td>
<td>0.9698</td>
<td>0.9482</td>
<td>0.9590 [5]</td>
</tr>
</tbody>
</table>
Image thresholding based on Ali–Silvey distance measures

Table 5. Performance for Manhattan image

<table>
<thead>
<tr>
<th>Method</th>
<th>Threshold</th>
<th>Uniformity</th>
<th>Shape</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>Relative entropy</td>
<td>1</td>
<td>0.8588</td>
<td>0.1424</td>
<td>0.5006 [4]</td>
</tr>
<tr>
<td>J-Divergence, Bhattacharyya, Matsusita, Chernoff</td>
<td>9</td>
<td>0.8556</td>
<td>0.1201</td>
<td>0.4878 [5]</td>
</tr>
<tr>
<td>Otsu</td>
<td>73</td>
<td>0.9433</td>
<td>0.8875</td>
<td>0.9154 [2]</td>
</tr>
<tr>
<td>Moment preserving</td>
<td>87</td>
<td>0.9367</td>
<td>0.7948</td>
<td>0.8658 [3]</td>
</tr>
<tr>
<td>Minimum error</td>
<td>53</td>
<td>0.9408</td>
<td>0.9955</td>
<td>0.9682 [1]</td>
</tr>
</tbody>
</table>

The choice of images is based on trying to test a diverse group of images versus the performances of the different algorithms. In some cases, the image does not contain a distinct object for separation from the background. Such cases should produce a binary image where the details and edges of the image are recognizable.

The performance metrics of all the previously discussed thresholding methods are listed in Tables 1–5. The chosen threshold, uniformity measure, shape measure, and an overall ranking are listed. The numbers in the

Fig. 1. Elephant, size 242 × 382 with 244 gray levels. (a) Original image; (b) histogram; (c) co-occurrence matrix.

Fig. 2. Elephant thresholded images for (a) Relative entropy, Chernoff (0.25), t=227; (b) J-Divergence, Bhattacharyya, Matsusita, t=188; (c) Chernoff (0.75), t=182; (d) Otsu, t=122; (e) Moment preserving, t=140; (f) Minimum error, t=55.
last column provide the average measure assuming that uniformity and shape are equally important. Rankings for the overall measures are shown in the brackets after the metrics. We also comment on the results from a visual standpoint and compare whether or not the automatic evaluation metrics agree with the observations.

The Ali–Silvey distances implemented using the transition probabilities of equation (6) are compared to the histogram-based thresholding methods. The following images are grouped into three categories according to their histogram shapes.

5.1. **Group I: bimodal histograms**

Images with bimodal histograms are the simplest to threshold. The dark and light gray levels are usually well separated so that the threshold is chosen as a point in the valley of the histogram. The image tested here is not strictly bimodal, but there are two well-separated areas in the histogram.

5.1.1. **Elephant image.** The elephant histogram [Fig. 1(b)] has widely separated modes. The co-occurrence matrix has values on the diagonal at the extreme corners. The original image also consists of a distinct object on a background where the object occupies about half of the entire image. The results in Table 1 show that the Otsu and Moment Preserving thresholding methods perform significantly better than

---

![Fig. 2. (Continued).](image)

![Fig. 3. Lenna, size 240 × 256 with 207 gray levels. (a) Original image; (b) histogram; (c) co-occurrence matrix.](image)
the others and relative entropy gives the worst results. Also, the uniformity and shape measures are consistent with each other, meaning if the uniformity measure is high so is the shape measure.

![Lenna thresholded images](image1)

Fig. 4. Lenna thresholded images for (a) all Ali–Silvey distance measures, $t=84$; (b) Otsu, $t=92$; (c) Moment preserving, $t=101$; (d) Minimum error, $t=41$.

![San Francisco image](image2)

Fig. 5. San Francisco, size 216 × 256 with 227 gray levels. (a) Original image; (b) histogram; (c) co-occurrence matrix.
The binary images in Fig. 2 appear to agree with the performance measures. Other than the two methods mentioned above, the resulting images either choose too much of the background or not enough of the elephants.

5.2. Group II: mountain-like histograms

These images have histograms with no distinct modes. The histogram values are more evenly distributed over the whole range of gray levels in the image.
5.2.1. Lenna image. The histogram [Fig. 3(b)] is multimodal where none of the modes are well separated. The plot of the co-occurrence matrix [Fig. 3(c)] shows that the transitional probabilities, $p_{ij}$, are concentrated all along the diagonal ($i=j$) and have a wide spread. Also, all the values of the $p_{ij}$ are less than 0.002, and the image does not contain a distinct object to be

![Lenna image](image1)

![Histogram](image2)

![Co-occurrence matrix](image3)

Fig. 6. (Continued).

![Composite image](image4)

Fig. 7. Castle, size 196 x 253 with 210 gray levels. (a) Original image; (b) histogram; (c) co-occurrence matrix.

![Castle images](image5)

Fig. 8. Castle thresholded images for (a) relative entropy, Bhattacharyya, Matsusita and Chernoff $t=65$; (b) J-Divergence, $t=64$; (c) Otsu, $t=74$; (d) Moment preserving, $t=69$; (e) Minimum error, $t=78$. 
5.3. Group III: spiked histograms

The images in this group have histograms with a spike at one end of the histogram. This means that the images have high concentrations of pixels around one or a small range of gray levels.

5.3.1. Castle image. This castle image is dark when compared to the previous examples. The histogram [Fig. 7(b)] is skewed toward the dark end but still appears to have two modes. When looking at the chosen thresholds, we see that most of the algorithms chose values in the valley between the 50 and 100 gray levels. The co-occurrence matrix [Fig. 7(c)] is similar to the elephant matrix, the probabilities are in the same range, but the two groups of $p_{ij}$ are not as widely separated.

5.2.2. San Francisco image. The histogram of the San Francisco image [Fig. 5(b)] is similar to the Lenna image, it does not have well-separated modes. The entries of the co-occurrence matrix [Fig. 5(c)] concentrate all along the diagonal but the spread from $(i=j)$ is not as wide. Note that the $p_{ij}$ are about a factor of two larger than those for the Lenna image. Table 3 shows that the Otsu method performs best, but none of the methods had particularly poor results. The best uniformity measure corresponds to Minimum Error and the best shape measure is from the Otsu method. Again, the uniformity and shape measures are consistent with each other.

The resulting binary images in Fig. 6 are all very similar. In this case there is so much detail in the image that it is difficult to tell by sight which thresholding method works best.

Fig. 8. (Continued).

Fig. 9. Manhattan, size 153 × 233 with 183 gray levels. (a) Original image; (b) histogram; (c) co-occurrence matrix.
The J-Divergence method produced the best results with the rest of the Ali–Silvey distance measures ranking second overall (see Fig. 8).

5.3.2. Manhattan image. The Manhattan [Fig. 9(b)] and Castle images have similar histograms, but the modes for the Manhattan image are more widely spread. The co-occurrence matrix [Fig. 9(c)] is also similar to the Castle, but higher transition probabilities concentrate near the (0,0) coordinate.

Table 5 shows that the Minimum Error method performed best overall. The Ali–Silvey distance measures produced the worst results. But if we examine the thresholded image in Fig. 10(a), the buildings are better separated from the sky and water. So, interpreting the performance metrics depends on what characteristics are important. The binary image for the relative entropy method is not shown.

6. CONCLUSIONS

This paper presented the class of Ali–Silvey distance measures as another method for image thresholding. In their paper, Chang et al. showed that image thresholding could be accomplished using relative entropy and the probabilities of the co-occurrence matrix. We generalized their approach to the entire class of Ali–Silvey distance measures and demonstrated image thresholding based on a number of members of this class.

The image thresholding method presented in this paper is more general than Chang et al.'s relative entropy approach. For some of the images, the Ali–Silvey distance measures have been shown to produce better or similar results when compared to the histogram-based algorithms. Also, the approach presented in this paper could be useful when implemented in conjunction with other entropic measures. Therefore, thresholding based on the class of Ali–Silvey distance measures is another alternative for image thresholding. In general, a thresholding algorithm's performance is image dependent and a method needs to be developed to determine which algorithms threshold better for specific types of images. Since the criterion functions for the Ali–Silvey distance measures were based on the probabilities of the co-occurrence matrix, perhaps these distance measures could be used for image thresholding based on other characteristics of the image. Also, the distance measures might be useful in image segmentation problems other than thresholding.\(^\text{17}\)

REFERENCES


**About the Author** — LIANE RAMAC received her B.S. degree in Electrical Engineering from the University of Washington, Seattle, Washington, in 1992, and the M.S. degree in Electrical Engineering from Syracuse University, Syracuse, New York, in 1995. She is currently working as a research assistant and pursuing a Ph.D. in Electrical Engineering at Syracuse University. Her research interests include digital image processing, signal processing, and communications.

**About the Author** — PRAMOD K. VARSHNEY was born in Allahabad, India, on 1 July 1952. He received the B.S. degree in Electrical Engineering and Computer Science (with highest honors), and the M.S. and Ph.D. degrees in Electrical Engineering from the University of Illinois at Urbana-Champaign in 1972, 1974, and 1976, respectively. During 1972–1976, he held teaching and research assistantships at the University of Illinois. Since 1976 he has been with the Electrical and Computer Engineering Department at Syracuse University, Syracuse, New York, where he is currently a Professor and the Associate Chairman of the Department. His current research interests are in distributed sensor networks and data fusion, detection and estimation theory, communication networks, communication theory, and parallel algorithms. He has supervised 18 Ph.D. dissertations, authored or co-authored 40 journal papers and over 100 conference papers. He has consulted for General Electric, Hughes, Kaman Sciences Corp., SCIEE, Philips Broadband and Digicom Research. While at the University of Illinois, Dr Varshney was a James Scholar, a Bronze Tablet Senior, and a Fellow. He is a member of Tau Beta Pi and is the recipient of the 1981 ASEEE Dow Outstanding Young Faculty Award. He is an Associate Editor for *Information Sciences.* He is listed in Who's Who in Technology Today and Outstanding Young Men of America.