Proceedings of International Conference on Image
Processing and Character Recognition

Entropic and Relative Entropic Thresholding Techniques

Ching-Wen Yang[1] Pau-Choo Chung[1] Chein-I Chang[2] Jianwei Wang[2]

[1]Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan 70101, Republic of China

[2]Department of Electrical Engineering, University of Maryland Baltimore County Baltimore, MD 21228-5398, U. S. A.

Abstract

One of important features to be used for image thresholding is the gray-level co-occurrence matrix. A threshold decomposes a co-occurrence matrix into four quadrants which correspond to background-to-background (BB), background-to-foreground (BF), foreground-to-background (FB), foreground-to-foreground (FF) respectively. In this paper, thresholding techniques based on maximization of Shannon's entropy and minimization of relative entropy are studied and compared, particularly, those developed by N.R. P et al-S.K. Pal, Kittler-Illingworth and Chang et al. The former (i.e., Pal-Pal's technique) maximizes Shannon's entropies of sum of BB and FF or sum of BF and FB, whereas, the latter (i.e., Kittler-Illingworth and Chang et al's methods) minimizes the relative entropy between an original image and a thresholded image. Despite the difference in optimization, all the three approaches are indeed closely related. Conceptually, they are developed based on a general design rationale widely used in pattern classification, namely, while minimizing the differences of samples within class, the differences of samples between class are also maximized. As a result, their performances can be well explained in terms of the concepts of within class and between class.

Key Words: Thresholding Local entropy (LE), Joint entropy (JE), Global entropy (GE), Minimum error thresholding, Local relative entropy (LRE), Joint relative entropy (JRE), Global relative entropy (GRE)

1. Introduction

Entropy is an uncertainty measure first introduced by Shannon to information theory to describe information contained in a source governed by a probability law (1). The principle of maximum entropy is that when a priori knowledge is not given, the best one can do is to maximize the uncertainty, i.e., entropy. For instance, if an experiment is conducted based on an unknown probability distribution which cannot be estimated a priori, the best approach is to assume that all outcomes of the experiment are equally likely which result in the maximum entropy (uncertainty). This is intuitively appealing since if one does not have any preference among samples, the best way is not to introduce any biased knowledge into the decision process but treat all samples equally important in which case the probability distribution is assumed to be uniformly distributed, thus it yields the maximum entropy. Maximum entropy approach has been widely used in signal processing community. In image thresholding, Pun [2,3] and Kapur *et al.*[4] considered the entropy of the gray level histogram and selected a threshold value to be one which yields the maximum entropy. However, the drawback of their approaches is that the spatial correlation of the image was not taken into account and only information of gray levels was concerned. Therefore, different images with an identical histogram generate the same entropy and so, produce the same threshold. In order to alleviate this problem, Sklansky [5] used the co-occurrence matrix of gray levels as an estimate of the distribution of pairs of gray levels separated spatially by a specified displacement vector. Haralick *et al.*[1] proposed several measurements that can be employed to extract useful texture information resulting from the co-occurrence matrix. Conners and Harlow [7] concluded that from a theoretical point of view, the co-occurrence matrix of gray levels resulted in the best discrimination. Consequently, it could be used as an information criterion for the global texture analysis. More specifically, the co-occurrence matrix can be viewed as a feature space so that the measurements defined on the co-occurrence matrix should be able to provide information for feature extraction.

Owing to above works N.R. Pal and S.K. Pal [8] derived two second-order entropy definitions based on the co-occurrence matrix, called local entropy (LE) and joint entropy (JE). Assume that t is a value to be used for thresholding an image. The t decomposes the co-occurrence matrix into four quadrants which correspond to background-to-background (BB), background-to-foreground (BF), foreground-to-background (FB), foreground-to-foreground (FF) respectively. According to Pal-Pal's definitions, the local entropy is defined as the sum of entropies of BB and FF and joint entropy as the sum of entropies of BF and FB.The entropic thresholding developed by Pal and Pal is to select a threshold which maximizes either LE or JE.

Using relative entropy as a thresholding measure was first studied by Kittler and Illingworth[9] which is called minimum error thresholding (MET). The underlying assumption of their approach is that the image to be thresholded can be modeled by a mixture of two Gaussian distributions with appropriate weights where the two Gaussian distributions are used to describe the image background and foreground respectively and the weights are determined by a threshold t. The desired threshold is one which generates a Gaussian mixture that best matches the original image gray-level histogram. The criterion to measure the matching discrepancy between a Gaussian mixture and the image histogram is relative entropy. Unfortunately, Kittler-Illingworth's MET will not work well if the image cannot be well separated and approximated by a Gaussian mixture.

Instead of following Kittler-Illingworth's idea, Chang *et al.*[10] recently developed an alternative relative entropic thresholding, hereafter referred to as GRE, which also used relative entropy as a measure to describe the mismatch between an image and a thresholded image. Surprisingly, what Pal-Pal's approach is to Pun-Kapur *et al.*'s entropic method is what Chang *et al.*'s GRE is to Kittler-Illingworth's MET. Kittler-Illingworth's MET and Pun-Kapur *et al.*'s methods dealt with first-order entropies generated by a gray-level image histogram, whereas, Pal-Pal's and Chang *et al.*'s approaches consider second-order entropies generated by the gray-level co-ocurrence matrix of an image. Furthermore, a crucial difference between entropic thresholding and relative entropic thresholding is that the former finds a value which maximizes entropy and the latter finds a value minimizing relative entropy.

Since Pun-Kapur *et al.*'s method is not comparable to Pal-Pal's method, it will not be discussed in this paper but referred to references [2,3,4]. Only Pal-Pal's entropic thresholding, Kittler-Illingworth's MET and relative entropic thresholding including Chang *et al.*'s GRE will be studied and compared. In particular, a global entropy (GE) and two relative entropies[11], called local relative entropy (LRE) and joint relative entropy (JRE) are introduced so that GE, LRE and JRE are the counterparts of Chang at el.'s GRE[10] and Pal-Pal's LE and JE in entropic thresholding respectively.

In this paper, three entropies (LE, JE and GE) and four relative entropies (Kittler-Illingworth's MET, LRE, JRE, Chang *et al.*'s GRE) are considered for a comparative study. Because Pal-Pal's LE maximizes the sum of the entropies of gray-level transitions within background and foreground, it is expected that LE thresholding works best among three entropy-based thresholding techniques. On the other hand, JRE outperforms Kittler-Illingworth's MET, LRE and GRE since it minimizes the mismatch of the gray-level transitions between background and foreground. Experimental results support this justification.

2. Entropy Thresholding

Entropic thresholding is a technique using entropy as a criterion to threshold an image. The concept of entropy has been widely used in data compression to measure information content of a source. Suppose that an L-symbol source X is governed by a probability distribution $p=[p(1),...,p(L)]$ Then the information generated by source X can be described by its entropy $H(X)$ defined as follows.

$$H(x)=-\sum_{j=1}^{L} p(j)\log p(j)$$

Since an image can be viewed as a source with the probability distribution given by its gray-level image histogram, the spatial information contained in the image can be characterized by the entropy of the histogram. Pun [2,3] and Kapur *et al.* [4] used this concept to derive two thresholding algorithms. As mentioned previously, their approaches did not take the spatial correlation into consideration. As a result, two completely different images with an identical histogram will generate the same threshold. One way to fix this problem is to consider the gray-level co-occurrence matrix which contains the information of transitions between any pair of two gray levels in the histogram in such a manner that the spatial correlation will be appropriately taken care of by the co-occurrence matrix.

2.1 Gray-Level Co-occurrence Matrix

Given a digital image of size MxN with L gray levels $G=\{0, 1, 2, ..., L-1\}$, let the gray level of the pixel at the spatial location (x,y) be denoted by f(x,y). Then the image can be represented by the matrix $F=[f(x,y)]_{MxN}$. A co-occurrence matrix of an image is an LxL dimensional matrix, $=[t_{ij}]_{LxL}$, which contains information about spatial dependency of gray levels in image F and the number of transitions between two gray levels specified in a particular way. A widely used co-occurrence matrix is an asymmetric matrix which only considers the gray level transitions between two adjacent pixels, horizontally right and vertically below. More specifically, let t_{ij} be the (i,j)th entry of the co-occurrence matrix W. Following the definition[8],

$$t_{ij}=\sum_{j=1}^{M}\sum_{k=1}^{N}\delta(l,k) \quad (1)$$

where

$$\delta(l,k)=1, if \begin{cases} f(l,k)=i, & f(i,k+1)=j, \quad and/or \\ f(l,k)=i, & f(l+1,k). \end{cases}$$

$$\delta(l,k)=0, \quad otherwise$$

It should be noted that the co-occurrence matrix defined above considers only horizontally right and left transitions as well as both vertically above and below transitions since it has been found that including horizontally left and vertically above transitions does not provide significant information and improvement.

Normalizing the total number of transitions in the co-occurrence matrix, we
obtain the desired transition probability from gray level i to j as follows.

$$p(i,j) = t_{ij} / (\sum_{i=1}^{L-1} \sum_{j=1}^{L-1} t_{ij}) \tag{2}$$

2.2 Quadrants of the Co-occurrence Matrix

Let $t \in G$ be a threshold of two classes (foreground and background) in an image. The co-occurrence matrix defined by (1) partitions the matrix into four quadrants, namely, A, B, C, and D, shown in Figure 1.

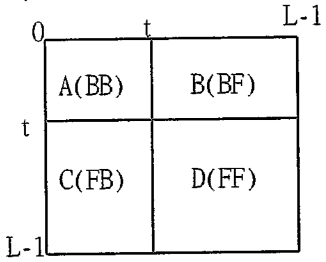

Figure 1: Quadrants of a Co-occurrence Matrix

These four quadrants may be separated into two classes. If we assume that pixels with gray levels above the threshold be assigned to the foreground (objects), and those below, assigned to the background, then, the quadrants A and C correspond to local transitions within background and foreground respectively; whereas, quadrants B and D represent transitions across the boundaries of background and foreground. So, as commonly used in pattern classification, the former can be referred to as *within-class quadrants* and the latter is referred to as *between-class quadrants*. The probabilities associated with each quadrant are then defined by

$$P_A^t = \sum_{i=0}^{t} \sum_{j=0}^{t} p(i,j),\ P_B^t = \sum_{i=0}^{t} \sum_{j=t+1}^{L-1} p(i,j),$$
$$P_C^t = \sum_{i=t+1}^{L-1} \sum_{j=t+1}^{L-1} p(i,j),\ P_D^t = \sum_{i=t+1}^{L-1} \sum_{j=0}^{t} p(i,j) \tag{3}$$

The probabilities in each quadrant can be further defined by so called "cell probabilities".

$$p_A(i,j) = \frac{p(i,j)}{P_A^t} = \frac{t_{ij} / (\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} t_{ij})}{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} (t_{ij} / \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} t_{ij})} = \frac{t_{ij}}{\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} t_{ij}}; \tag{4}$$

$$p_B(i,j) = \frac{p(i,j)}{P_B^t},\quad p_C(i,j) = \frac{p(i,j)}{P_C^t},\quad p_D(i,j) = \frac{p(i,j)}{P_D^t} \tag{5}$$

for $0 \le i \le t, 0 \le j \le t$

2.3 Local Entropy, Joint Entropy and Global Entropy Methods

Three definitions can be derived based on the cell probabilities given by (4-5), each of which yields a different method. The first two were proposed by N.R. Pal and S.K. Pal[8] which are called local entropy (LE) and joint entropy (JE) methods. The third one, GE, will be referred to as global entropy method which was defined in reference [11] but not included in reference [8].

2.3.1 Local Entropy (LE)

Since quadrant A and quadrant C contain the local transitions from background to background (BB), and objects to objects (FF), the local entropy of background and local entropy of foreground by $H_{BB}(t)$ and $H_{FF}(t)$ can be defined respectively as follows.

$$H_{BB}(t) = -\frac{1}{2} \sum_{i=0}^{t} \sum_{j=0}^{t} p_A(i,j) \log p_A(i,j) \tag{6}$$

$$H_{FF}(t) = -\frac{1}{2} \sum_{i=t+1}^{L-1} \sum_{j=t+1}^{L-1} p_C(i,j) \log p_C(i,j) \tag{7}$$

It should be noted that (6) and (7) are determined by the threshold t, thus they are function of t.

By summing up the entropies of the foreground and the background, a second-order local entropy can be obtained by

$$H_{LE}(t) = H_{BB}(t) + H_{FF}(t). \tag{8}$$

Obviously, $H_{LE}(t)$ describes the entropy of *within-class quadrants* and can be also called "*within-class entropy*" to reflect the characteristic of quadrants A and C. The LE method proposed by Pal and Pal[8] is one to select a threshold which maximizes the $H_{LE}(t)$ over t.

2.3.2 Joint Entropy (JE)

Alternatively, quadrant B and quadrant D provides edge information on transitions from background to foreground (BF) and foreground to background (FB). In analogy to the local entropy defined above, a second-order joint entropy of the background and the foreground was also defined by averaging the entropy $H_{BF}(t)$ resulting from quadrant B, and the entropy $H_{FB}(t)$ from quadrant D as follows,

$$H_{JE}(t) = H_{BF}(t) + H_{FB}(t) \tag{9}$$

where

$$H_{BF}(t) = -\frac{1}{2} \sum_{i=0}^{t} \sum_{j=t+1}^{L-1} p_B(i,j) \log p_B(i,j)$$

$$H_{FB}(t) = -\frac{1}{2} \sum_{i=t+1}^{L-1} \sum_{j=0}^{t} p_D(i,j) \log p_D(i,j)$$

Similarly, $H_{JE}(t)$ can be called "*between-class entropy*" to reflect the gray-level transition activities in quadrants B and D. The method finding the maximum of (9) is called JE method which is the second algorithm developed by N.R. Pal and S.K. Pal [8].

2.3.3 Global Entropy (GE)

The global entropy $H_{GE}(t)$ is simply defined as the sum of the local entropy $H_{LE}(t)$ and the joint entropy $H_{JE}(t)$, i.e.,

$$H_{GE}(t) = H_{LE}(t) + H_{JE}(t) \tag{10}$$

A value which maximizes $H_{GE}(t)$ is called the global entropy threshold. It should be noted that Pal and

Pal did not define the GE. However, the GE turns out to be the counterpart of Chang *et al.*'s GRE [10]. Since GE is the sum of LE and JE, it can be expected that the performance based on GE will be moderate between LE and JE. The experiments seem to justify this claim. Since LE, JE and GE are all image-dependent, when there is no preference to choose LE or JE, GE may be a good candidate for a compromise.

3. Relative Entropy-Based Thresholding

In information theory relative entropy has been used to measure the dissimilarity between two sources. The smaller the relative entropy, the more similar the two sources.

Definition of Relative Entropy

Let two sources with L symbols be described by probability distributions $p=[p(1),...,p(L)]$ and $h=[h(1),...,h(L)]$. The relative entropy between p and h (or more precisely, the entropy of p relative to h) is defined by

$$J(p;h) = \sum_{j=1}^{L} p(i)\log\frac{p(i)}{h(i)} \tag{11}$$

The definition given by (11) was first introduced by Kullback[12] as a distance measure between two probability distributions, thus it is sometimes called Kullback-Leiber's discrimination distance function. Two other synonyms are also used in the literature, cross entropy and directed divergence. As the definition implied, the smaller the relative entropy, the less the discrepancy; thus, the better the match between p and h. It should be noted that in the definition of relative entropy (11), the source given by p is considered to a nominal source and the source given by h is one trying to match the source p. Furthermore, the relative entropy is not symmetric. In this paper, the original image is always designated to be the nominal image and the thresholded image is the one which tries to match it.

3.1 Kittler-Illingworth's Minimum Error Thresholding (MET)

The concept of using relative entropy for thresholding was first suggested by Kittler and Illingworth [9] where they assumed that an image can be modelled by a mixture of two Gaussian distributions. More specifically, let $p_{true}(i)$ be the original histogram distribution of the image and t be a threshold. Assume that t segments the image into the background and foreground both of which are modelled by Gaussian distributions, $p_B^t(i)$ and $p_F^t(i)$ respectively. $p_{mix}^t(i)$ is a mixture of these two Gaussian distributions, i.e.,

$$p_{mix}^t(i) = \alpha_1^t p_B^t(i) + \alpha_2^t p_F^t(i)$$

where α_1^t and α_2^t are weights determined by the portions of the background and objects in the image. Namely,

$$p_{mix}^t(i) = \begin{cases} \alpha_1^t p_B^t(i) & \text{if } i < t \\ \alpha_2^t p_F^t(i) & \text{if } i \geq t \end{cases}$$

Kittler-Illingworth's MET is to find a threshold t^* which minimizes the mismatch between $p_{true}(i)$ and $p_{mix}^t(i)$ over t. The criterion they used to measure the matching discrepancy between these two probability distributions is relative entropy given by (11).

Using (11) Kittler and Illingworth formulated a thresholding problem as a minimization problem of $J(p_{ture}(i); p_{mix}^t(i))$ over t, called the minimum error thresholding problem. The desired threshold is one achieving the minimum of $J(p_{ture}(i); p_{mix}^t(i))$ denoted by t^*. Then $J(p_{ture}(i); p_{mix}^{t^*}(i)) = \min_t J(p_{ture}(i); p_{mix}^t(i))$ and the $p_{mix}^{t^*}(i)$ is the Gaussian mixture which best matches the true gray-level histogram, p_{true} of the original image. As a result, if the background and foreground are well separated, we may expect that the Kittler-Illingworth's MET will work very well. However, this assumption is generally not true in many practical applications. In the next section, we will present an alternative but completely different approach from MET which is also based on the concept of relative entropy.

3.2 Relative Entropy Approaches Based on Gray-Level Co-occurrence Matrix

Since the gray-level histogram does not consider the spatial correlation, the co-occurrence matrix of an image is considered. As defined previously, the co-occurrence matrix describes the frequency of transitions of one gray level to another in the image. If $p(i)$ and $h(i)$ defined in (11) are replaced by the transition probability, $p(i,j)$ generated by the co-occurrence matrix and the transition probability, $h^t(i,j)$ generated by the binary image thresholded by the threshold t respectively, then the probability distribution $h^{t^*}(i,j)$ which minimizes the relative entropy between $p(i,j)$ and $h^t(i,j)$ will be the desired co-occurrence matrix which best matches the co-occurrence matrix of the original image. Since the gray-level transitions generally reflect jumps in gray levels, the approach described above can be used to detect edges, thus, it should have good capability of thresholding.

As noticed in Kittler-Illingworth's MET, their method is based solely on the first-order gray level histogram of an image, exploiting the spatial dependency of the pixel values in the image can help to determine a better threshold. The proposed idea extends their first-order relative entropy to a second-order joint relative entropy between $p(i,j)$ and $h^t(i,j)$. Since transition probability distributions defined by the co-occurrence matrix contain the spatial information which reflects homogeneity of *within-class quadrants* A and C in Figure 1, and changes across boundaries in *between-class quadrants*, B and D in Figure 1, one can envision that a better result may be obtained if the thresholded bilevel image is chosen to be the one which has the best transition match to that of the original image in terms of relative entropy.

Let the jointly relative entropy of the probability distributions $p(i,j)$ and $h^t(i,j)$ be defined by:

$$J(p;h^t) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)\log\frac{p(i,j)}{h^t(i,j)} \quad (12)$$

where $p(i,j)$ and $h^t(i,j)$ are the transition probabilities from gray level i to gray level j of the original image and the binary image respectively. Minimizing $J(p;h^t)$ over t generally renders a binary image which best matches the original image. It should be noted that when we threshold an image, we basically assign all gray levels in an original image two gray levels (in this paper, 0 and 1 are used) which correspond to background and foreground.

3.2.1 Co-occurrence Matrix of a Thresholded Image

Assuming that t is the selected threshold. By assigning 1 to all gray levels above threshold t, $G_1=\{t+1,..., L-1\}$ and 0 to all gray levels below t, $G_2=\{0, 1,.., t\}$, we obtain a binary image where the gray levels in G_1 will be treated equally likely in probability, so are gray levels in G_2. The resulting $h^t(i,j)$ can be found as follows (see Figure 1).

$$h_A^t(i,j) = q_A^t = \frac{P_A^t}{(t+1)x(t+1)};\ \text{for } 0 \le i \le t,\ 0 \le j \le t$$

$$h_B^t(i,j) = q_B^t = \frac{P_B^t}{(t+1)x(t+1)};\ \text{for } 0 \le i \le t,\ t+1 \le j \le L-1$$

$$h_C^t(i,j) = q_C^t = \frac{P_C^t}{(t+1)x(t+1)};\ \text{for } t+1 \le i \le L-1,\ t+1 \le j \le L-1$$

$$h_D^t(i,j) = q_D^t = \frac{P_D^t}{(t+1)x(t+1)};\ \text{for } t+1 \le i \le L-1,\ 0 \le j \le t \quad (13)$$

where, P_A^t, P_B^t, P_C^t, and P_D^t are defined by (3). For each selected t, $h_A^t(i,j)$, $h_B^t(i,j)$, $h_C^t(i,j)$, and $h_D^t(i,j)$ are constants in each individual quadrant and only depend upon the quadrant to which they belong. Therefore, they can be denoted by q_A^t, q_B^t, q_C^t, and q_D^t respectively.

3.2.2 Three Relative Entropy-Based Methods

By expanding (14),

$$J(p;h^t) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)\log\frac{p(i,j)}{h^t(i,j)} = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)\log p(i,j) - \sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)\log h^t(i,j) \quad (14)$$

Because the first term in the above equation is independent of the threshold t, minimizing the relative entropy (12) is equivalent to maximizing the second term of (14).

The second term of the right side of (14) can be further simplified as follows.

$$\sum_{i=0}^{L-1}\sum_{j=0}^{L-1} p(i,j)\log h^t(i,j) = \sum_A p(i,j)\log q_A^t + \sum_B p(i,j)\log q_B^t + \sum_C p(i,j)\log q_C^t + \sum_A p(i,j)\log q_A^t$$
$$= P_A(i,j)\log q_A^t + P_A(i,j)\log q_A^t + P_C(i,j)\log q_C^t + P_D(i,j)\log q_D^t \quad (15)$$

This implies that in order to obtain a desirable threshold to segment the foreground from the background, only the last expression in (15) needs to be maximized over t.

In analogy to Section 2, three different relative entropies can be defined based on the global feature of gray-level transitions; the local feature of gray-level transitions within background and foreground; and the joint feature of gray-level transitions from background to foreground and foreground to background respectively.

$$H_{GRE}(t) = P_A(i,j)\log q_A^t + P_B(i,j)\log q_B^t + P_C(i,j)\log q_C^t + P_D(i,j)\log q_D^t$$
$$H_{LRE}(t) = P_A(i,j)\log q_A^t + P_C(i,j)\log q_C^t$$
$$H_{JRE}(t) = P_B(i,j)\log q_B^t + P_D(i,j)\log q_D^t \quad (16)$$

From (16), three relative entropy-based methods can be derived as follows.

1. Local Relative Entropic (LRE) Thresholding

The local relative entropic thresholding (LRE) method defined below has its counterpart in entropic thresholding (i.e., Pal and Pal's LE) which is to find a value t_{LRE} maximizing $H_{LRE}(t)$

$$t_{LRE} = \arg_t\{\max_t H_{LRE}(t)\} \quad (17)$$

In this case, LRE calculates the maximum relative entropy of BB and FF.

2. Joint Relative Entropic (JRE) Thresholding

In analogy to LRE, the joint relative entropic (JRE) thresholding method is also the counterpart of Pal and Pal's joint entropy (JE) method which is to find a value t_{JRE} to max $H_{JRE}(t)$, i.e., the maximum relative entropy of BF and FB.

$$t_{JRE} = \arg_t\{\max_t H_{JRE}(t)\} \quad (18)$$

3. Global Relative Entropic (GRE) Thresholding

Unlike LRE and JRE, the global relative entropic (GRE) thresholding method[10] maximizes the relative entropy of all the four quadrants. It is to find a value t_{GRE} maximizing $H_{GRE}(t)$,

$$t_{GRE} = \arg_t\{\max_t H_{GRE}(t)\} \quad (19)$$

It is very interesting to note that according to our experiments in Section 4, JRE generally performs better than LRE. This is because that JRE is designed to maximize the relative entropy of the gray level transitions between two classes, background and foreground. It is the opposite of Pal and Pal's entropy thresholding where LE method seems better than JE method in which case LE maximizes the entropy of *within-class quadrants* of background and foreground. Like the case of GE , the performance of GRE is better than LRE and comparable to JRE, but occasionally outperforms JRE as seen in Experiment 2 give below.

4. Experimental Results and Comparative Study

In this section, entropic thresholding, Kittler-Illingworth's MET and relative entropic thresholding will be examined and compared through four experiments. All figures labelled by (a) are original images along with their gray-level histograms labelled by (b). Images thresholded by entropic thresholding, GE, LE and JE are labelled by (c), (d) and (e) respectively. Images produced by Kittler-Illingworth's MET is labelled by (f) and images are generated by relative entropic thresholding, GRE, LRE and JRE are labelled by (g), (h) and (i) respectively.

4.1 Experiments

The following 4 images are selected for experiments because they can be used to demonstrate the performance of relative entropic thresholding relative to entropic thresholding.

Experiment 1: Lena-Figure 2(a)

Figures 2(c-i) were produced by $t=159$ for LE, $t=124$ for JE, $t=136$ for GE, $t=138$ for MET, $t=166$ for LRE, $t=203$ for JRE and $t=170$ for GRE . As shown in these figures, three entropic thresholding methods perform nearly the same and so do the three relative entropic thresholding methods. However, it is very clear that the relative entropic thresholding performs differently from entropic thresholding. The quality of images generated by the former seems better than that by the latter since the former describes more details of Lena's face including her mouth while the latter missing Lena's mouth.

Experiment 2: Coffee cup-Figure 3(a)

The histogram of the coffee cup image is very interesting. The images produced by relative entropic thresholding. GRE produced the threshold value $t=237$ which picks up the open edge of the cup while LE, JE and GE show the side edges of the cup and miss its open edge.

Experiment 3: Building-Figure 4(a)

Compared to the gray-level histogram of the coffee cup image, Figure 3(e), the building image has a very similar gray-level histogram distribution, Figure 4(e). So, it is not surprising that both coffee cup and building generate the same GRE $t= 237$ since GRE is calculated based on the complete co-occurrence matrix, i.e., 4 quadrants. Unlike GE, GRE and MET, LE, JE, LRE and JRE are calculated based on only two quadrants. It is anticipated that they will generate different threshold values. LE, JE, GE, MET, LRE and JRE generate close thresholds $t=166$, $t=172$, $t=172$, $t=238$, $t=213$ and $t=233$ respectively. As a result, the corresponding thresholded images, Figure 4(b) and Figure 4(c) are close. However, Figure 4(d) produced by the relative entropy using threshold $t=237$ is quite different from Figures 4(b-c). The local entropy and joint entropy seem to give a better description of the building while failing to pick up the middle edges of the building and the outside stairs which was shown in Figure 4(d). The reason for this we believe is that the relative entropy can best match all possible transitions made from one gray level to another in particular regions such quadrants (A,C) or (B,D).

Experiment 4: Vapor cloud-Figure 5(a)

The chemical vapor cloud image shown Figure 5(a) was taken by a forward looking infrared (FLIR) imager and has very narrowly concentrated histogram. The cloud is barely visible in Figure 5(a). As shown in Figures 5(c-i), JRE ($t=199$) generates the best description of the vapor image while JE ($t=206$) producing the worst quality of the image. The images obtained by LE, GE, MET, LRE and GRE seem moderate with thresholds spread from 183 to 199. This example shows that the thresholded images are very sensitive to threshold values. A reckless threshold selection may result in significant degradation in image quality.

Table 1. Images versus thresholds for seven methods

Images	GE	LE	JE	MET	GRE	LRE	JRE
1. Lena	136	159	124	138	170	166	203
2. Coffee cup	156	130	156	238	237	237	121
3.. Building	172	166	172	238	237	213	233
4. Vapor cloud	192	193	206	190	193	192	199

As shown in Table 1, the thresholds are image-dependent. In addition to the above 4 experiments, we have also done more experiments. The results show that LE and JRE perform better than do JE, GE, MET LRE, GRE in most of cases and JRE generally outperforms LE. It is also shown from experiments that relative entropic thresholding can complement entropic thresholding in terms of providing different details which were missed by entropic thresholding. It is particularly true for the above images. Based on conducted experiments, it indicates that relative entropic thresholding tends to generate higher threshold values than does entropic thresholding.

4.2 Comparative Study

In this subsection, seven thresholding methods (LE, JE, GE, Kittler-Illingworth's MET, LRE, JRE and GRE) will be discussed and studied. Their relative performances will be also compared.

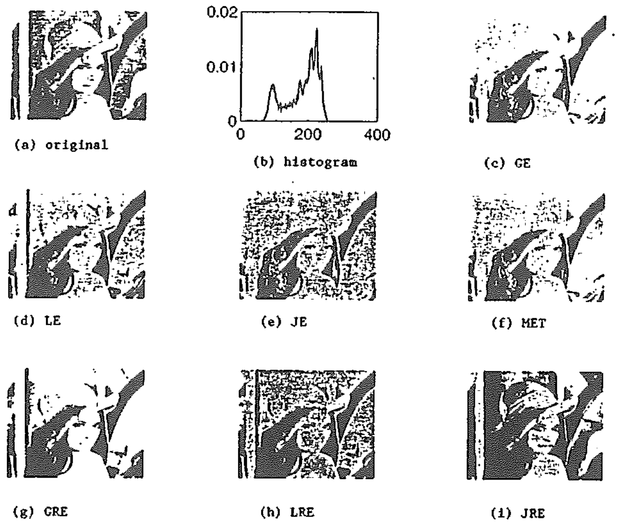

Figure 2: Lena

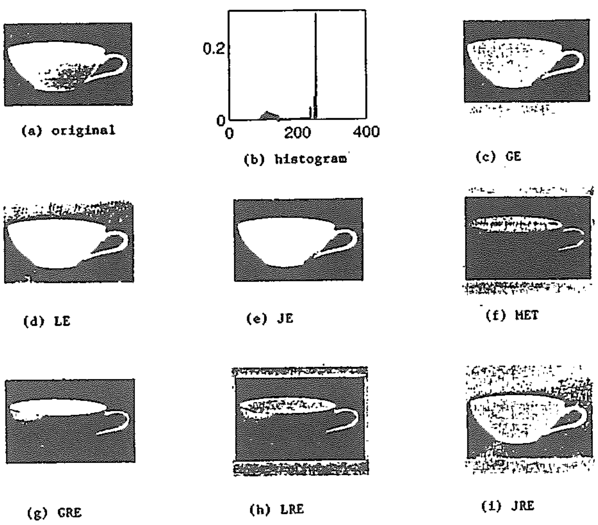

Figure 3: Coffee cup

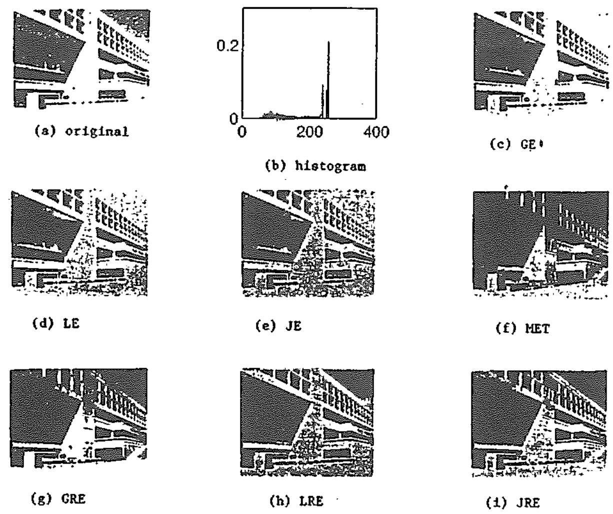

Figure 4: Building

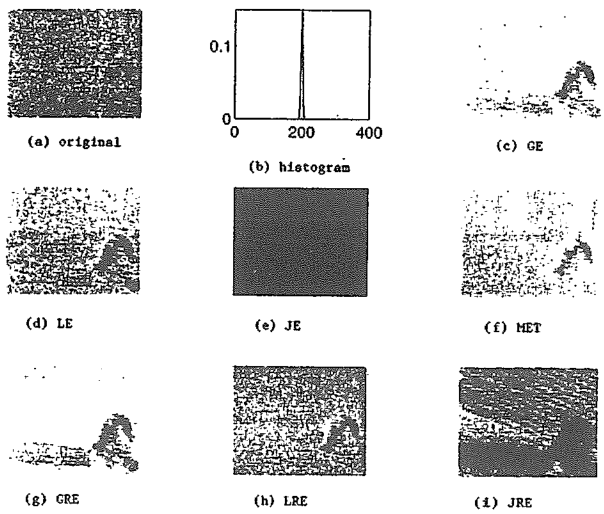

Figure 5:Vapor cloud

4.2.1 Entropic thresholding

The idea of Pal-Pal's methods is to apply the principle of maximum entropy to two classes of quadrants resulting from thresholding the co-occurrence matrix (see Figure 1). One class consists of BB (quadrant A) and FF (quadrant D) which can be considered as *within-class quadrants* and the other is made up of BF (quadrant B) and FB (quadrant C) which can be regarded as *between-class quadrants*. While LE thresholding maximizing the entropy of *within-class quadrants*, JE is to maximize the entropy of *between-class quadrants*. As expected, LE performs better than JE because LE tries to maximize the information contained only in the background and foreground given that no a *priori* knowledge is known. Compared to LE, JE maximizes the information of gray-level transitions between the background and foreground. GE is simply to maximize the overall information in four quadrants. Consequently, the performance of GE is generally moderate between LE and JE. However, if there is the case that one cannot decide LE or JE for thresholding, GE is probably a good candidate for a compromise..

4.2.2 Relative Entropic Thresholding

Relative entropic thresholding is a different concept from entropic thresholding in the sense that the former uses relative entropy as a measure to minimize the mismatch between an image and a thresholded image, whereas the latter applies the maximum entropy principle to thresholding an image. Two relative entropy-based approaches are discussed as follows.

5. Conclusion

In this paper, entropic thresholding and relative entropic thresholding techniques are studied and compared. A total of seven different entropy-based measures are considered. Their performances vary since all the techniques are image-dependent. Nevertheless, experiments have shown that JRE produces the best results. Several studies using entropy/relative entropy-based thresholding are investigated[13]. One application [13,14] is to define a spectral co-occurrence matrix to replace the spatial gray-level co-occurrence matrix for multispectral images. A similar approach to multispectral images for temporal image sequences is also presented in reference [13]. In addition, it is also found [13] that relative entropic thresholding can be used as a criterion for band selection for multispectral/hyperspectral images. An extension to relative entropic thresholding in multiple stages for image segmentation was also considered in [15]. All these applications have demonstrated that relative entropic thresholding may be a promising image processing technique in many areas not limited to threshold selection.

References

[1] Cover and J. Thomas, Elements of Information Theory, John Wiley & Sons, Inc., 1991.

[2] Pun, "A new method for grey-level picture thresholding using the entropy of the histogram," Signal Processing, Vol. 2, 1980, pp. 223-237.

[3] Pun, "Entropic thresholding, a new approach," Comput., Graphics and Image Proc., Vol.16, pp.210-239, 1981.

[4] Kapur, P.K. Sahoo and A.K.C. Wong, "A new method for grey-level picture thresholding using the entropy of the histogram," Comp. Graphics, Vision and Image Proc., vol. 29, pp. 273-285, 1985.

[5] Sklansky, "Image segmentation and feature extraction, "IEEE Trans. Syst. Man Cybern., vol. SMC-8, no. 4, 1978, pp. 237-247.

[6] Haralick, K. Shanmugam and I. Dinstein, "Textural features for image segmentation, "IEEE Trans. Syst. Man Cybern., vol. SMC-3, no. 6, 1973, pp. 610-621.

[7] Conners and C.A. Harlow, "A theoretical comparison of texture algorithms,"IEEE Trans. Pattern Analysis and Machine Intelligence, vol. PAMI-2, no. 3, 1980, pp. 204-222.

[8] Nikhil R. Pal and Sankar K. Pal, "entropic thresholding,"Signal Processing, vol. 16, pp.97-108, 1989.

[9] Kittler and J. Illingworth, "Minimum error thresholding," PatternRecognitoin, vol. 19, no. 1, 1986, pp 41-47. 652-655.

[10] C-I Chang, K. Chen, J. Wang and M.L.G. Althouse, "A relative entropy-based approach to image thresholding," Pattern Recognition, vol. 27, no. 9, Septemberr 1994, pp. 1275-1289.

[11] C-I Chang, J. Wang and M.L.G. Althouse, "Vapor cloud detection using relative entropy thresholding," Signal Processing, Sensor Fusion and Target Recognition III, Volume 2232, SPIE, Orlando, Florida, April, 1994, pp. 276-284.

[12] Kullback, Information Theory and Statistics. New York: Dover, 1968.

[13] Althouse, Vapor Cloud Detection in Multispectral Infrared Image Sequences Using Co-ocurrence Matrix Methods, Ph.D. dissertation, Department of Electrical Engineering, University of Maryland Baltimore County, Baltimore, MD, December 1994.

[14] Althouse and C.-I Chang, "Target detection in multispectral images using spectral co-occurrence matrix and entropy thresholding," accepted for publication in Optical Engineering

[15] C-I Chang, J. Wang and M.L.G. Althouse, "Chemical vapor cloud detection using multistage entropic thresholding," Proc. Scientific Conference on Chemical Defense Research, ERDEC, Aberdeen Proving Ground, MD, November 16-19, 1993, pp. 655-667.