Pergamon

A RELATIVE ENTROPY-BASED APPROACH TO IMAGE THRESHOLDING

Chein-I Chang,† Kebo Chen,† Jianwei Wang† and Mark L. G. Althouse‡
† Department of Electrical Engineering, University of Maryland, Baltimore County Campus, Baltimore, MD 21228-5398, U.S.A.
‡ Edgewood Research, Development and Engineering Center, Aberdeen Proving Ground, MD 21010-5432, U.S.A.

Abstract—In this paper, we present a new image thresholding technique which uses the relative entropy (also known as the Kullback–Leiber discrimination distance function) as a criterion of thresholding an image. As a result, a gray level minimizing the relative entropy will be the desired threshold. The proposed relative entropy approach is different from two known entropy-based thresholding techniques, the local entropy and joint entropy methods developed by N. R. Pal and S. K. Pal in the sense that the former is focused on the matching between two images while the latter only emphasized the entropy of the co-occurrence matrix of one image. The experimental results show that these three techniques are image dependent and the local entropy and relative entropy seem to perform better than does the joint entropy. In addition, the relative entropy can complement the local entropy and joint entropy in terms of providing different details which the others cannot. As far as computing saving is concerned, the relative entropy approach also provides the least computational complexity.

Thresholding   Relative entropy   Local entropy   Joint entropy   Co-occurrence matrix

1. INTRODUCTION

Image thresholding often represents a first step in image understanding. In an ideal image where objects are clearly distinguishable from the background, the grey-level histogram of the image is generally bimodal. In this case, a best threshold segmenting objects from the background is one placed right in the valley of two peaks of the histogram. However, in most cases, the grey-level histograms of images to be segmented are always multimodal. Therefore, finding an appropriate threshold for images is not straightforward. Various thresholding techniques have been proposed to resolve this problem.

In recent years, information theoretic approaches based on Shannon's entropy concept have received considerable interest.[1-6] Of particular interest are two methods proposed by N. R. Pal and S. K. Pal[1] which use a co-occurrence matrix to define second-order local and joint entropies. The co-occurrence matrix is a transition matrix generated by changes in pixel intensities. For any two arbitrary grey levels i and j (i, j are not necessarily distinct), the co-occurrence matrix describes all intensity transitions from grey level i to grey level j. Suppose that t is the desired threshold. The t then segments an image into the background which contains pixels with grey levels below or equal to t and the foreground which corresponds to objects having pixels with grey levels above t. This t further divides the co-occurrence matrix into four quadrants which correspond, respectively, to transitions from background to background (BB), background to objects (BO), objects to background (OB) and objects to objects (OO). The local entropy is defined only on two quadrants, BB and OO, whereas the joint entropy is defined only on the other two quadrants, BO and OB. Based on these two definitions, Pal and Pal developed two algorithms, each of which maximizes local entropy and joint entropy, respectively.

In this paper, we present an alternative entropy-based approach which is different from those in references (1–6). Rather than looking into entropies of background or object individually, we introduce the concept of the relative entropy[7] (also known as cross entropy, Kullback–Leiber's discrimination distance and directed divergence), which has been widely used in source coding for the purpose of measuring the mismatching between two sources. Since a source is generally characterized by a probability distribution, the relative entropy can be also interpreted as a distance measure between two sources. This suggests that the relative entropy can be used for a criterion to measure the mismatching between an image and a thresholded bilevel image. One method to apply the relative entropy concept to image thresholding is to calculate the gray-level transition probability distributions of the co-occurrence matrices for an image and a thresholded bilevel image, respectively, then find a threshold which minimizes the discrepancy between these two transition probability distributions, i.e. their relative entropy. The threshold rendering the smallest relative entropy will be selected to segment the image. As a result, the

thresholded bilevel image will be the best approximation to the original image. Since transitions of OB and BO generally represent edge changes in boundaries and transitions of BB and OO indicate local changes in regions, we can anticipate that a thresholded bilevel image produced by the proposed relative entropy approach will best match the co-occurrence matrix of the original image. This observation is demonstrated experimentally by several test images. In general, the performance of all the three methods is image dependent. Although there is no evidence that one is generally better than the others, according to the experiments conducted in this paper, the joint entropy did not work as well as did the local entropy and relative entropy. Interestingly, among all images tested the relative entropy approach seems to be better than the others at finding edges. In addition, our experiments show that the relative entropy seems to be a good complement to the local entropy and joint entropy methods in terms of providing different image details and descriptions from those provided by the local entropy and joint entropy. Finally, an advantage of the relative entropy approach is the computational saving based on arithmetic operations required for calculating entropies compared to the local and joint entropy approaches.

This paper is organized as follows. Section 2 describes previous work on entropy-based thresholding approaches. Section 3 introduces the concept of relative entropy and presents a relative entropy-based thresholding algorithm. In Section 4, experiments are conducted based on various test images in comparison to the local entropy and joint entropy methods described in reference (1). Finally a brief conclusion is given in Section 5.

2. PRELIMINARIES

2.1. Co-occurrence matrix

Given a digitized image of size $M \times N$ with L gray levels $G = \{0, 1, 2, \ldots, L-1\}$, we denote $F = [f(x, y)]_{M \times N}$ to represent an image, where $f(x, y) \in G$ is the grey level of the pixel at the spatial location (x, y). A co-occurrence matrix of an image is an $L \times L$ dimensional matrix, $T = [t_{ij}]_{L \times L}$, which contains information regarding spatial dependency of grey levels in image F as well as the information about the number of transitions between two grey levels specified in a particular way. A widely used co-occurrence matrix is an asymmetric matrix which only considers the grey level transitions between two adjacent pixels, horizontally right and vertically below.[1] More specifically, let t_{ij} be the (i, j)th entry of the co-occurrence matrix T. Following the definition in reference (1),

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

where

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

$$\delta(l, k) = 0, \quad \text{otherwise.}$$

One may like to make the co-occurrence matrix symmetric by considering horizontally right and left, and vertically above and below transitions. It has, however, been found[1] that including horizontally left and vertically above transitions does not provide more information about the matrix or significant improvement. Therefore, it is sufficient to consider adjacent pixels which are horizontally right and vertically below so that the required computation can be reduced.

Normalizing the total number of transitions in the co-occurrence matrix, we obtain the desired transition probability from grey level i to j[1] as follows.

$$p_{ij} = t_{ij} \bigg/ \left(\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} t_{ij} \right). \qquad (2)$$

2.2. Quadrants of the co-occurrence matrix

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

These four quadrants may be separated into two types. If we assume that pixels with grey 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. The probabilities associated with each quadrant are then defined by

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

$$P_B(t) = \sum_{i=0}^{t} \sum_{j=t+1}^{L-1} p_{ij}$$

$$P_C(t) = \sum_{i=t+1}^{L-1} \sum_{j=t+1}^{L-1} p_{ij}$$

$$P_D(t) = \sum_{i=t+1}^{L-1} \sum_{j=0}^{t} p_{ij}. \qquad (3)$$

The probabilities in each quadrant can be further defined by the "cell probabilities" and obtained as
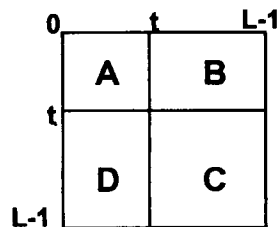

Fig. 1. Quadrants of a co-occurrence matrix.

follows by normalization.

$$p_{ij}^A = p_{ij}/P_A = \frac{t_{ij}\left/\left(\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}t_{ij}\right)\right.}{\sum_{i=0}^{t}\sum_{j=0}^{t}\left(t_{ij}\left/\sum_{i=0}^{L-1}\sum_{j=0}^{L-1}t_{ij}\right.\right)}$$

$$= \frac{t_{ij}}{\sum_{i=0}^{t}\sum_{j=0}^{t}t_{ij}},$$

$$\text{for } 0 \le i \le t, 0 \le j \le t \quad (4)$$

$$p_{ij}^B = p_{ij}/P_B = \frac{t_{ij}}{\sum_{i=0}^{t}\sum_{j=t+1}^{L-1}t_{ij}},$$

$$\text{for } 0 \le i \le t, t+1 \le j \le L-1 \quad (5)$$

$$p_{ij}^C = p_{ij}/P_C = \frac{t_{ij}}{\sum_{i=t+1}^{L-1}\sum_{j=t+1}^{L-1}t_{ij}},$$

$$\text{for } t+1 \le i \le L-1, t+1 \le j \le L-1 \quad (6)$$

$$p_{ij}^D = p_{ij}/P_D = \frac{t_{ij}}{\sum_{i=t+1}^{L-1}\sum_{j=0}^{t}t_{ij}},$$

$$\text{for } t+1 \le i \le L-1, 0 \le j \le t. \quad (7)$$

2.3. Algorithms of Pal and Pal[1]

The algorithms suggested by Pal and Pal[1] attempted to take advantage of spatial correlation in an image. By doing so, they introduced two concepts of second-order entropy based on Equations (4)–(7), which are called local entropy and joint entropy.

Since quadrant A and quadrant C reflect the local transitions from background to background (BB), and object to object (OO), they defined local entropy of background and local entropy of object by $H_B(t)$ and $H_O(t)$, respectively, as follows.

$$H_B^{(2)}(t) = -\frac{1}{2}\sum_{i=0}^{t}\sum_{j=0}^{t}p_{ij}^A\log p_{ij}^A \quad (8)$$

$$H_O^{(2)}(t) = -\frac{1}{2}\sum_{i=t+1}^{L-1}\sum_{j=t+1}^{L-1}p_{ij}^C\log p_{ij}^C. \quad (9)$$

It should be noted that (8) and (9) are determined by the threshold t, thus they are a function of t.

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

$$H_{local}^{(2)}(t) = H_B^{(2)}(t) + H_O^{(2)}(t). \quad (10)$$

The algorithm proposed by Pal and Pal[1] is one to select a threshold which maximizes the $H_{local}^{(2)}$ over t. In this paper, it will be called the local entropy-based algorithm.

Alternatively, quadrant B and quadrant D provide edge information on transitions from background to

object (BO) and object to background (OB). In analogy with the local entropy defined above, another second-order joint entropy of the background and the object was also defined in reference (1) and given as follows by averaging the entropy $H(B; O)$ resulting from quadrant B, and the entropy $H(O; B)$ from quadrant D.

$$H_{joint}^{(2)}(t) = (H(B; O) + H(O; B))/2$$

$$= -\left(\sum_{i=0}^{t}\sum_{j=t+1}^{L-1}p_{ij}^B\log p_{ij}^B\right.$$

$$\left. + \sum_{i=t+1}^{L-1}\sum_{j=0}^{t}p_{ij}^D\log p_{ij}^D\right)\Big/2. \quad (11)$$

The algorithm maximizing (11) is called the joint entropy-based algorithm, which is the second algorithm developed by Pal and Pal.[1]

3. RELATIVE ENTROPY-BASED THRESHOLDING TECHNIQUE

3.1. Definition of relative entropy

Let S be an L-symbol source and p_j and p_j' be two probability distributions defined on S. The relative entropy between p and p' (or equivalently, the entropy of p relative to p') is defined by

$$L(p; p') = \sum_{j=0}^{L-1}p_j\log\frac{p_j}{p_j'}. \quad (12)$$

The definition given by Equation (12) was first introduced by Kullback[7] as a distance measure between two probability distributions, and later was found to be very useful in many applications.[8–12] Since the information contained in an image source can be described by its entropy, which in turn can be completely characterized by source symbol probabilities, the relative entropy basically provides a criterion to measure the discrepancy between two images determined by probability distributions p_j and p_j', respectively. The smaller the relative entropy, the less the discrepancy. It is natural to use relative entropy as a measure of difference between an image and its segmented image; in our case, a bilevel thresholded image. There are several synonyms of relative entropy, e.g. cross entropy, Kullback–Leiber's discrimination distance function and directed divergence.

In order to obtain a bilevel image of good quality, our aim is to find a threshold to segment an image such that the resulting thresholded bilevel image will best match the original image. Using the measure of relative entropy, one can choose the threshold t in such a manner that the grey-level probability distribution p_j' of the thresholded image has minimum relative entropy $L(p; p')$ with respect to that of the original image, p. More specifically, the desired threshold t minimizes the discrepancy between p and p', where p and p' are the grey-level probability distributions of the original image and the resulting thresholded image, respectively.

3.2. Joint relative entropy-based approach

As indicated previously, a thresholding method based on first-order statistics of an image does not consider spatial correlation of an image. Therefore, exploiting the spatial dependency of the pixel values in the image can help to determine a good threshold. It seems reasonable to extend the first-order relative entropy defined in the previous subsection to a second-order joint relative entropy between p_{ij} and p'_{ij}, where p_{ij} and p'_{ij} are the transition probability distributions of the co-occurrence matrices defined by Equation (1) and (2) generated by the original image and the thresholded image, respectively. Since transition probability distributions defined by the co-occurrence matrix contain the spatial information which reflects homogeneity within groups (quadrants A and C in Fig. 1), and changes across boundaries (quadrants B and D in Fig. 1), one can envision that a better result may be obtained if we choose the thresholded bilevel image to be the one which has the best transition match to that of the original image in terms of relative entropy.

Let the joint relative entropy of the probability distributions p_{ij} and p'_{ij} be defined by:

$$L(p;p') = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \log \frac{p_{ij}}{p'_{ij}}, \quad (13)$$

where p_{ij} and p'_{ij} are the transition probabilities from grey level i to grey level j of the original image and the bilevel image, respectively. Minimizing $L(p;p')$ over t generally renders a bilevel 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 to either 0 or 1 which corresponds to background or objects. As a result, there are only two grey levels in the thresholded image. The subscript ij used in the notation of the transition probability p'_{ij} still refers to the grey levels of the original image. In addition, the statistics of pixels not adjacent to one another could also be considered, but the estimation of probabilities for such cases would be very difficult. In this paper we only consider the asymmetric co-occurrence matrix defined in Section 2.1 for joint relative entropy.

3.3. Co-occurrence matrix of a thresholded bilevel image

Let us assume that t is the selected threshold. By assigning 1 to all grey levels above threshold t, $G_1 = \{t+1,\ldots,L-1\}$ and 0 to all grey levels below t, $G_2 = \{0,1,\ldots,t\}$, we obtain a binary image. It should be noted that the grey levels in G_1 will be treated equally likely in probability, as will be the grey levels in G_2. Consequently, the p'_{ij} can be found as follows (see Fig. 1):

$$p'^{(A)}_{ij}(t) = q_A(t) = \frac{P_A(t)}{(t+1) \times (t+1)}$$

$$\text{for } 0 \le i \le t, 0 \le j \le t \quad (14)$$

$$p'^{(B)}_{ij}(t) = q_B(t) = \frac{P_B(t)}{(t+1) \times (L-t-1)}$$

$$\text{for } 0 \le i \le t, t+1 \le j \le L-1 \quad (15)$$

$$p'^{(C)}_{ij}(t) = q_C(t) = \frac{P_C(t)}{(L-t-1) \times (L-t-1)}$$

$$\text{for } t+1 \le i \le L-1, t+1 \le j \le L-1 \quad (16)$$

$$p'^{(D)}_{ij}(t) = q_D(t) = \frac{P_D(t)}{(L-t-1) \times (t+1)}$$

$$\text{for } t+1 \le i \le L-1, 0 \le j \le t. \quad (17)$$

where $P_A(t)$, $P_B(t)$, $P_C(t)$, and $P_D(t)$ are defined by Equation (3). For each selected t, $p'^{(A)}_{ij}(t)$, $p'^{(B)}_{ij}(t)$, $p'^{(C)}_{ij}(t)$, and $p'^{(D)}_{ij}(t)$ are constants in each individual quadrant and only depend upon the quadrant to which they belong. Therefore, we can denote them by $q_A(t)$, $q_B(t)$, $q_C(t)$, and $q_D(t)$, respectively.

3.4. Relative entropy-based algorithm

By expanding Equation (13), we have:

$$L(p;p') = \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \log \frac{p_{ij}}{p'_{ij}}$$

$$= \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \log p_{ij} - \sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \log p'_{ij}. \quad (18)$$

Because the first term in Equation (18) is independent of the threshold t, minimizing the relative entropy described by Equation (13) is equivalent to maximizing the second term of Equation (18).

We can simplify even further the second term of the right side of the Equation (18) as follows:

$$\sum_{i=0}^{L-1} \sum_{j=0}^{L-1} p_{ij} \log p'_{ij} = \sum_A p_{ij} \log q_A(t) + \sum_B p_{ij} \log q_B(t)$$

$$+ \sum_C p_{ij} \log q_C(t) + \sum_D p_{ij} \log q_D(t)$$

$$= P_A(t) \log q_A(t) + P_B(t) \log q_B(t)$$

$$+ P_C(t) \log q_C(t) + P_D(t) \log q_D(t). \quad (19)$$

This implies that in order to obtain a desirable threshold for classifying the object from the background, we need only maximize the last expression in Equation (19) over t. The expression consists of four terms only, each of which is a product of P_i and $\log q_i(t)$ for $i = A$, B, C, D. In comparison with Equations (8) and (9) required for the local entropy and Equation (11) for the joint entropy, the computational load for the relative entropy is significantly reduced. From Equations (8) and (9), $(t+1)^2 + (L-t)^2$ multiplications are required to calculate $p_{ij} \log p_{ij}$ for finding the local entropy, and from Equation (11), $2(t+1)(L-t)$ multiplications for the joint entropy. However, only four multiplications and four divisions are needed in Equations (19) for the

relative entropy. As a result, the computational saving can be tremendous when the size of an image is very large.

4. EXPERIMENTAL RESULTS

In order to see the performance of the relative entropy-based thresholding method we conducted tests for a set of various images. As shown in experiments, the relative entropy approach provides an alternative efficient and effective image thresholding tool. All test images have 256 grey levels. In all experiments, the images labelled (a) are original images; the images labelled (b), (c) and (d) are generated by the local entropy, joint entropy and relative entropy, respectively. All the figures labelled (e) represent the corresponding grey-level histograms of the original images.

Experiment 1: Peppers image, Fig. 2(a)

From Fig. 2(b–d), it is obvious that the joint entropy produced the worst image with threshold $t = 90$, while the local entropy and relative entropy produced an identical image since both generated the same threshold $t = 127$.

Experiment 2: F-16 jet image, Fig. 3(a)

In this image, three methods generated different details, shown in Fig. 3(b–d). For instance, the local entropy with threshold $t = 115$ gave the best description of the lettering "F-16" on the tail, while the relative entropy with threshold $t = 175$ shows more clearly the cockpit, the insignia, and the lettering "US AIR FORCE" on the fuselage. The joint entropy with $t = 137$ produced an image between the quality of the other methods.

Experiment 3: Couple image, Fig. 4(a)

Evidently, the couple image thresholded by the relative entropy produced the best image, Fig. 4(d), compared to those in Figs 4(b) and (c) generated by the local entropy and joint entropy. The threshold used for the relative entropy was 111, whereas both the local entropy and joint entropy used the same threshold 171.

Experiment 4: Building image, Fig. 5(a)

The building image is interesting. Both the local entropy and joint entropy generated close thresholds $t = 166$ and $t = 172$, respectively. As a result, the corresponding thresholded images, Figs 5(b) and (c) are close. However, Fig. 5(d) produced by the relative entropy using threshold $t = 237$ is quite different from Figs 5(b) and (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 are shown in Fig. 5(d). The reason for this is probably that the relative entropy can best match all possible transitions made from one grey level to another. This seems to be justified from the

grey-level histogram of the building image given by Fig. 5(e). It should be noted that the histogram of Fig. 5(e) is very different from that of previous images, Figs 2(e), 3(e) and 4(e).

Experiment 5: Coffee cup image, Fig. 6(a)

Compared to the grey-level histogram of the building image, Fig. 5(e), the coffee cup image has a very similar grey-level histogram distribution, Fig. 6(e). Coincidentally, the relative entropy produced the same threshold $t = 237$ which was used for the building image. The local entropy and joint entropy produced Fig. 6(b) and (c) with $t = 130$ and $t = 156$, respectively. As shown in Fig. 6(b)–(d), Fig. 6(d) picks up the open edge of the cup while Fig. 6(b) and (c) show the side edges of the cup.

Experiment 6: Vase image, Fig. 7(a)

The grey-level histogram of the vase image is very different from that of other images, where its grey levels are distributed more uniformly than others. The image Fig. 7(c) produced by $t = 163$ for the joint entropy does not seem as good as Fig. 7(b) and (d) with $t = 125$ for the local entropy and $t = 132$ for the relative entropy, respectively. In addition, Fig. (d) looks a little better than Fig. 7(b), since there is blurring over the top of the vase in Fig. 7(b).

Experiment 7: Lena image, Fig. 8(a)

Figure 8(b)–(d) shows that the quality of images produced by $t = 159$ for the local entropy, $t = 124$ for the joint entropy and $t = 170$ for the relative entropy is nearly the same except that they pick up different tiny descriptions and details. For instance, relative entropy shows Lena's mouth at the expense of some details of the feather on Lena's hat. In contrast, the local entropy and the joint entropy give a little more detail on the feather while missing Lena's mouth and some details of Lena's hat.

Experiment 8: City image, Fig. 9(a)

Like the Lena image, the city images produced by $t = 123$ for the local entropy, $t = 128$ for the joint entropy and $t = 112$ for the relative entropy are very close. It is interesting to compare the grey-level histograms of these two images. If one of them is flipped over, it is found that their distributions turned out very similar. Therefore, these two experiments should be expected to have similar results.

Based on the experiments conducted above it seems that the grey-level histograms of these eight images can be roughly classified into four categories. The first three experiments (peppers, F-16 jet and couple images) are grouped together into Category 1 since their histograms have many saw-like sharp peaks with short durations. The next two experiments (building and cup images) are in Category 2 because they share the same characteristics of the histograms which are Gaussian-

(a) original image

(b) local entropy (t=127)

(c) joint entropy (t=90)

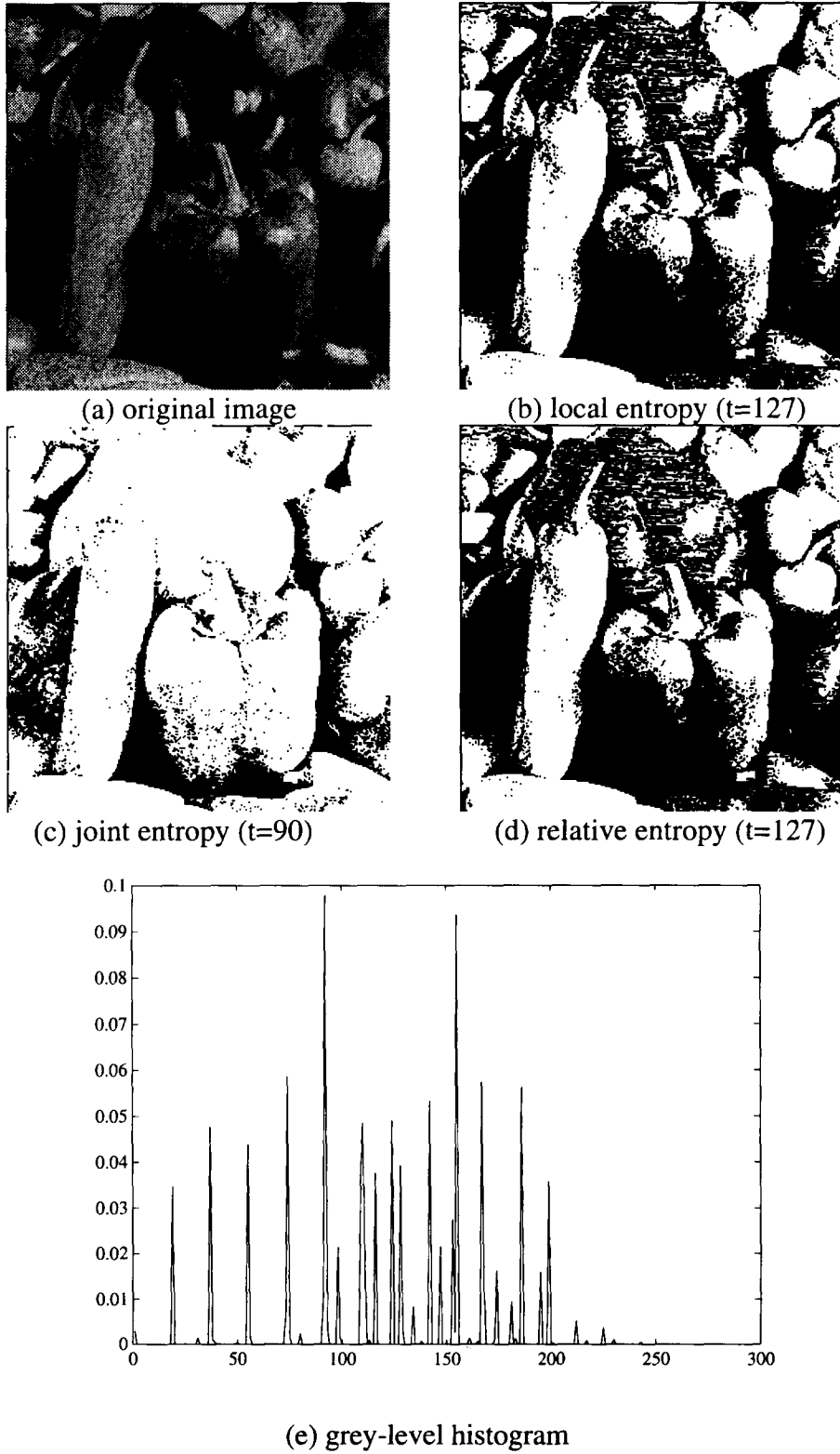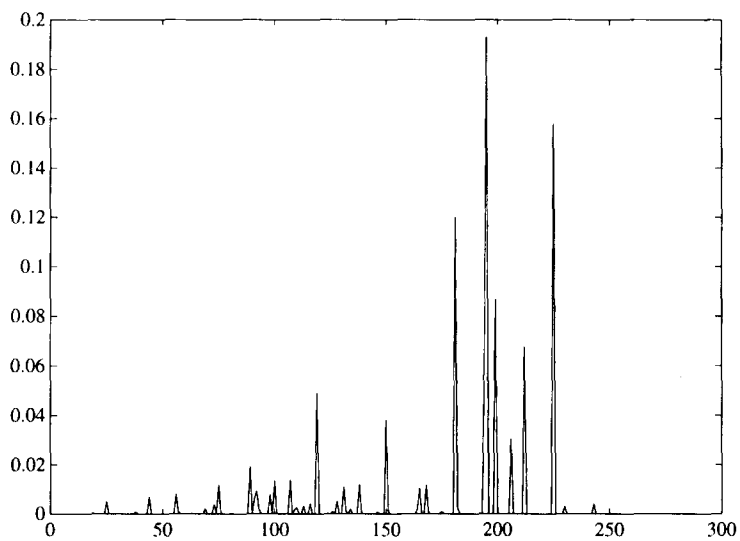(d) relative entropy (t=127)

(e) grey-level histogram

Fig. 2. Peppers image: (a) original image, (b) image obtained by local entropy, $t = 127$, (c) image obtained by joint entropy, $t = 90$, (d) image obtained by relative entropy, $t = 127$, (e) grey-level histogram.
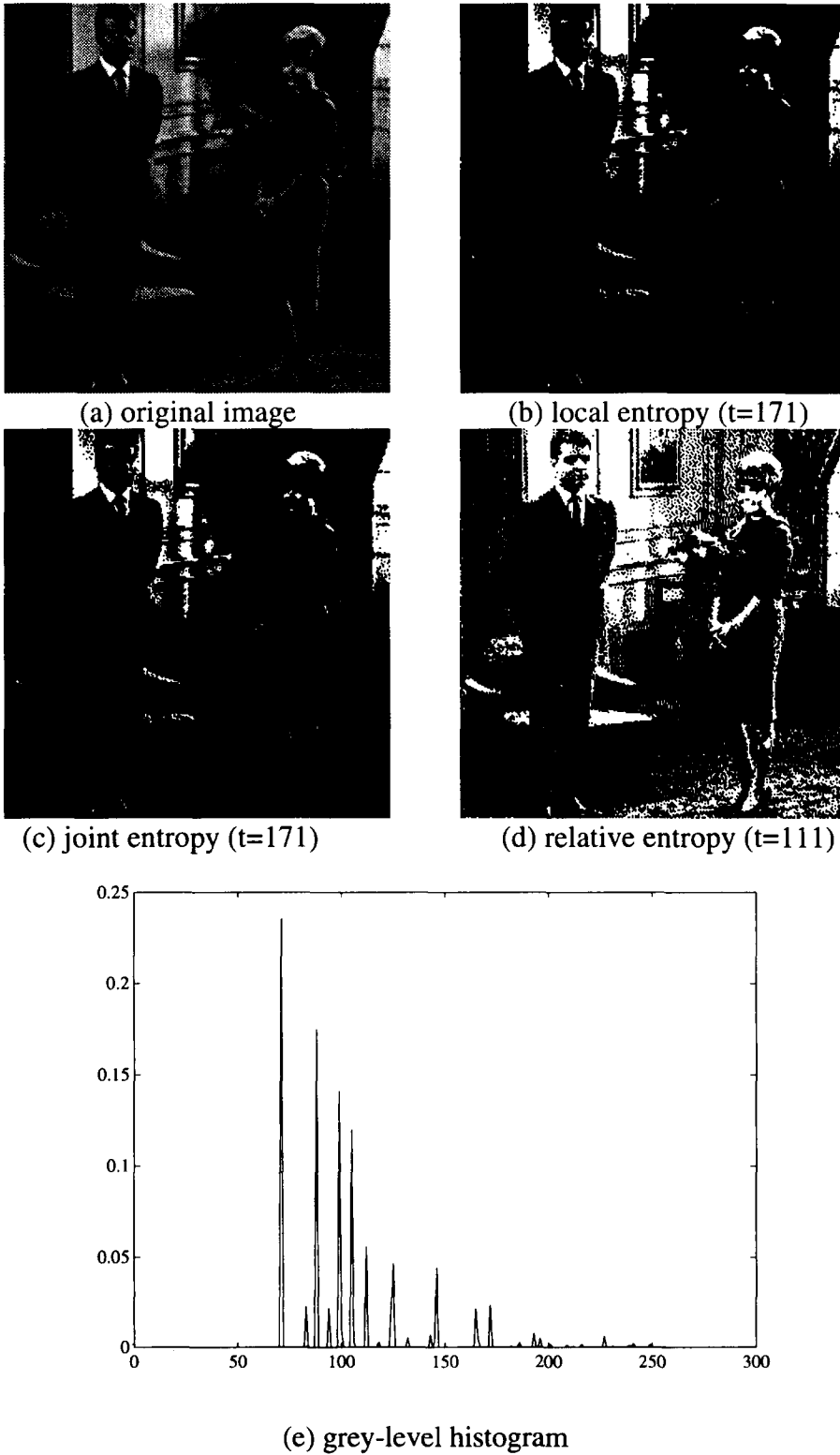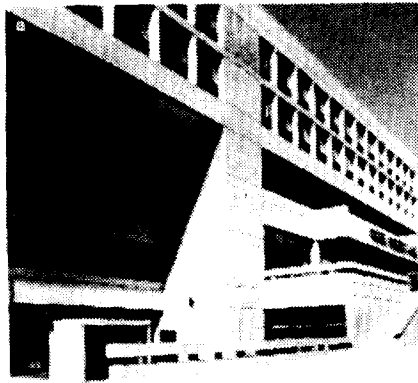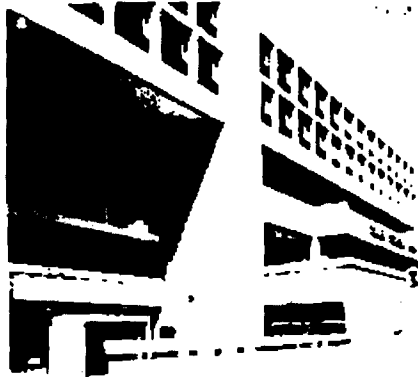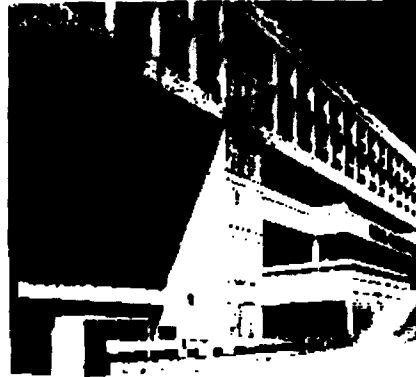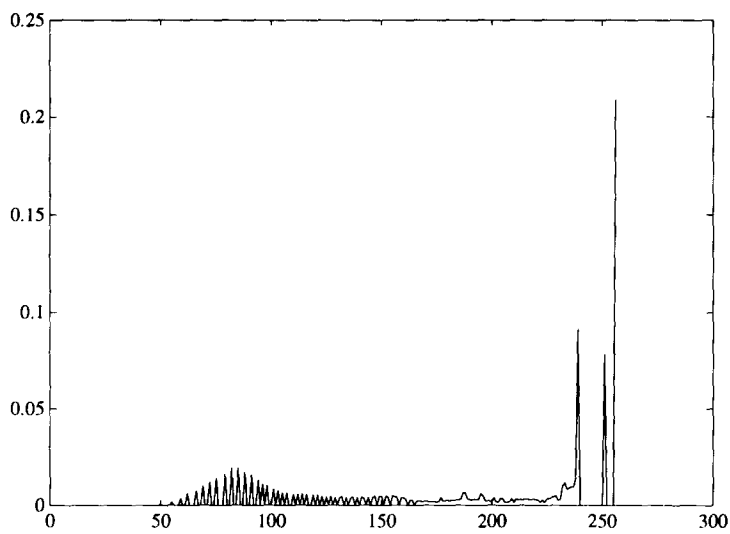
(a) original image

(b) local entropy (t=115)

(c) joint entropy (t=137)

(d) relative entropy (t=175)

(e) grey-level histogram

Fig. 3. F-16 jet image: (a) original image, (b) image obtained by local entropy, $t = 115$, (c) image obtained by joint entropy, $t = 137$, (d) image obtained by relative entropy, $t = 175$, (e) grey-level histogram.
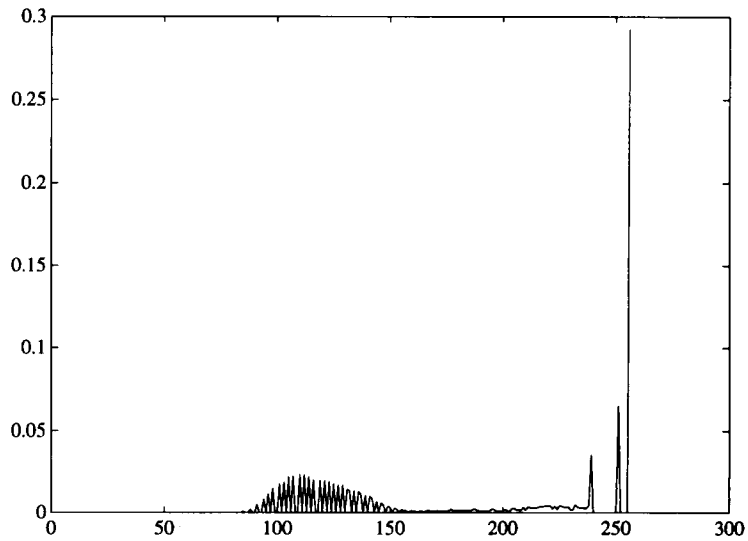
(a) original image

(b) local entropy (t=171)

(c) joint entropy (t=171)

(d) relative entropy (t=111)

(e) grey-level histogram

Fig. 4. Image of couple: (a) original image, (b) image obtained by local entropy, $t = 171$, (c) image obtained by joint entropy, $t = 171$, (d) image obtained by relative entropy, $t = 111$, (e) grey-level histogram.

(a) original image

(b) local entropy (t=166)

(c) joint entropy (t=172)

(d) relative entropy (t=237)

(e) grey-level histogram

Fig. 5. Building image: (a) original image, (b) image obtained by local entropy, $t = 166$, (c) image obtained by joint entropy, $t = 172$, (d) image obtained by relative entropy, $t = 237$, (e) grey-level histogram.
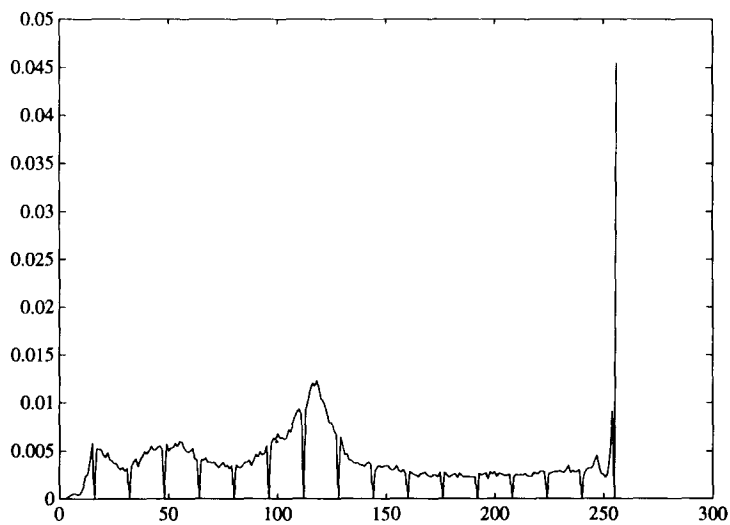
(a) original image   (b) local entropy (t=130)

(c) joint entropy (t=156)   (d) relative entropy (t=237)

(e) Grey-level histogram

Fig. 6. Coffee cup image: (a) original image, (b) image obtained by local entropy, $t = 130$, (c) image obtained by joint entropy, $t = 156$, (d) image obtained by relative entropy, $t = 237$, (e) grey-level histogram.

(a) original image

(b) local entropy (t=125)

(c) joint entropy (t=163)

(d) relative entropy (t=132)

(e) grey-level histogram

Fig. 7. Vase image: (a) original image, (b) image obtained by local entropy, $t = 125$, (c) image obtained by joint entropy, $t = 163$, (d) image obtained by relative entropy, $t = 132$, (e) grey-level histogram.
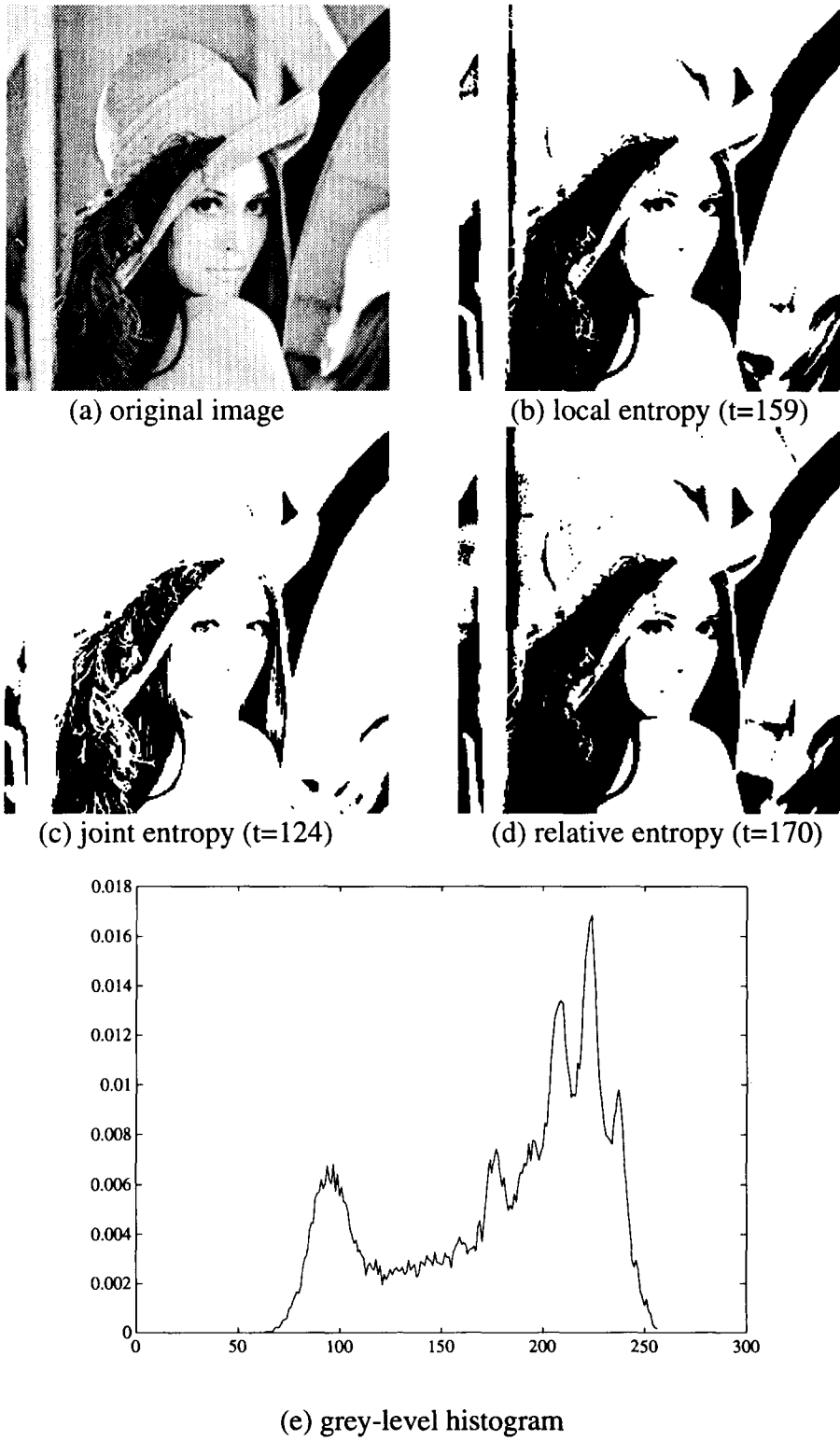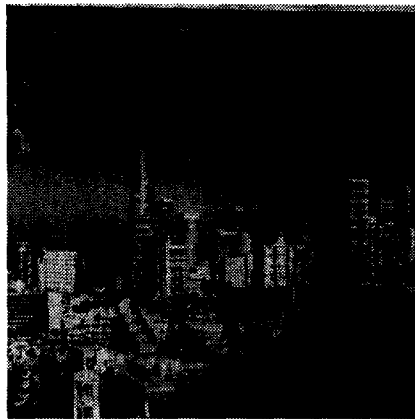
(a) original image

(b) local entropy (t=159)

(c) joint entropy (t=124)

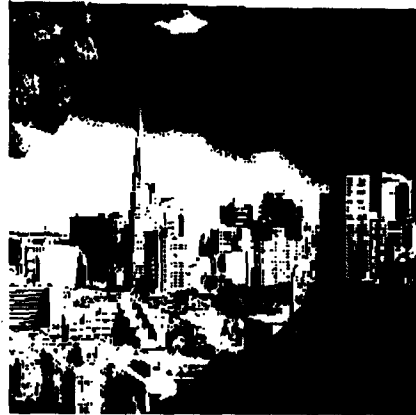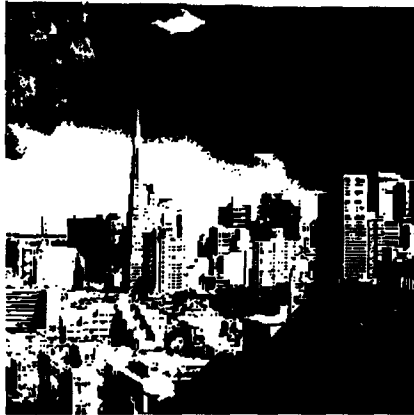(d) relative entropy (t=170)
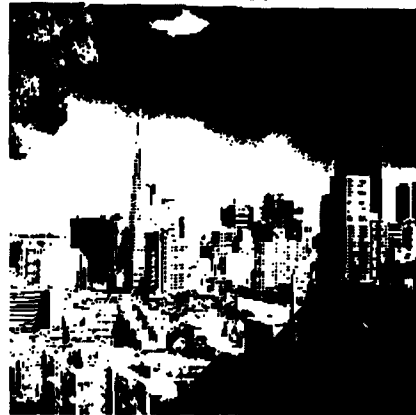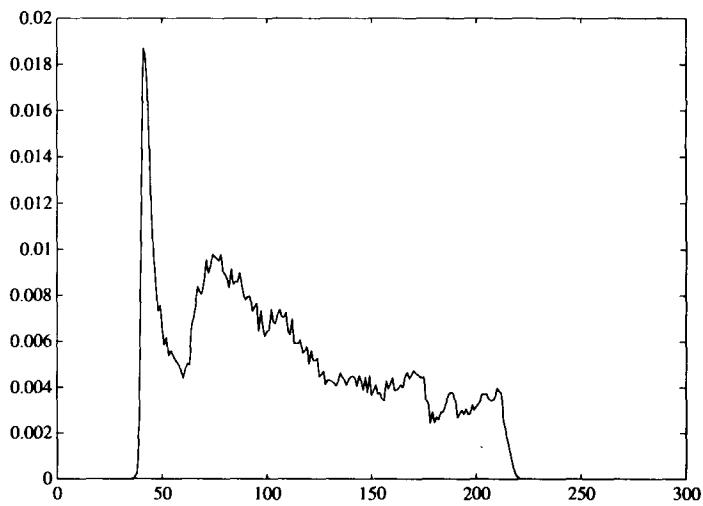
(e) grey-level histogram

Fig. 8. Lena image: (a) original image, (b) image obtained by local entropy, $t = 159$, (c) image obtained by joint entropy, $t = 124$, (d) image obtained by relative entropy, $t = 170$, (e) grey-level histogram.

(a) original image

(b) local entropy (t=123)

(c) joint entropy (t=128)

(d) relative entropy (t=112)

(e) grey-level histogram

Fig. 9. City image: (a) original image, (b) image obtained by local entropy, $t = 123$, (c) image obtained by joint entropy, $t = 128$, (d) image obtained by relative entropy, $t = 112$, (e) grey-level histogram.

Table 1. Images versus thresholds for three methods

Images	Min (grey level)	Max (grey level)	Local entropy	Joint entropy	Relative entropy
1. Peppers	0	242	127	90	127
2. F-16 jet	13	255	115	137	175
3. Couple	70	255	171	171	111
4. Building	40	255	166	172	237
5. Coffee cup	78	255	130	156	237
6. Vase	0	255	125	163	132
7. Lena	57	255	159	124	170
8. City	34	219	123	128	112

like with three sharp peaks. The last two experiments (Lena and city image) are in Category 3 because of their very similar mountain-like histograms. The histogram of the vase image is completely different from those of all previous images and stands itself alone to form Category 4 because its histogram is more or less uniformly distributed. Table 1 summarizes the results for images versus thresholds. As shown in the table, the thresholds are image dependent. Although more experiments need to be performed, our experiments show that the local entropy and relative entropy perform better than does the joint entropy in most cases, and the relative entropy can compete with the local entropy. It is also shown from experiments that the relative entropy approach can complement the local entropy and joint entropy approaches in terms of providing different details which were missed by the local entropy and joint entropy. This is particularly true for the F-16 jet, building and cup images.

5. CONCLUSION

A new thresholding method based on the relative entropy concept is presented in this paper. The idea is to find a threshold which minimizes the mismatching between two transition probability distributions resulting from the co-occurrence matrices of an image and a thresholded image. The proposed approach is different from the local entropy and joint entropy methods suggested by N. R. Pal and S. K. Pal.[1] In order to demonstrate the performance of the relative entropy approach, several images are studied in comparison to the local and joint entropy approaches. The experimental results show that the relative entropy-based method is a good alternative to the local and joint entropy methods. Particularly interesting is that the relative entropy approach can complement the local entropy and joint entropy methods, and it demonstrates a good capability for picking up the edges of objects. In addition, the computational complexity of calculating entropy required for the relative entropy method is far less than required for the local entropy and joint entropy algorithms.

6. SUMMARY

Image thresholding using information theoretic approaches based on Shannon's entropy concept has received considerable interest in recent years. Of particular interest are two methods proposed by N. R. Pal and S. K. Pal which use a co-occurrence matrix to define second-order local and joint entropies. The co-occurrence matrix is a transition matrix generated by changes in pixel intensities. For any two arbitrary grey levels i and j (i, j are not necessarily distinct), the co-occurrence matrix describes all intensity transitions from grey level i to grey level j. Suppose that t is the desired threshold. The t then segments an image into the background which contains pixels with grey levels below or equal to t and the foreground which corresponds to objects having pixels with grey levels above t. This t further divides the co-occurrence matrix into four quadrants which correspond to transitions from background to background (BB), background to objects (BO), objects to background (OB) and objects to objects (OO). The local entropy is defined only on two quadrants, BB and OO, whereas the joint entropy is defined only on the other two quadrants, BO and OB. Based on these two definitions, Pal and Pal developed two algorithms, one which maximizes local entropy, and the other which maximizes joint entropy.

In this paper, we present an alternative entropy-based approach which is different from previous approaches. Rather than looking into entropies of background or object individually, we introduce the concept of the relative entropy (also known as cross entropy, Kullback–Leiber's discrimination distance and directed divergence) that has been widely used in source coding for the purpose of measuring the mismatching between two sources. Since a source is generally characterized by a probability distribution, the relative entropy can be also interpreted as a distance measure between two sources. This suggests that the relative entropy can be used for a criterion to measure the mismatching between an image and a thresholded bilevel image. One method to apply the relative entropy concept to image thresholding is to calculate the grey-level transition probability distributions of the co-occurrence matrices for an image and a thresholded bilevel image, then find a threshold which minimizes the discrepancy between these two transition probability distributions, i.e. their relative entropy. The smaller the discrepancy, the better the matching between the original image and the thresholded image. The threshold rendering the smallest relative entropy will be selected to segment the image. As a result, the thresholded bilevel image will be the

best approximation to the original image. Since transitions of OB and BO generally represent edge changes in boundaries and transitions of BB and OO indicate local changes in regions, we can anticipate that a thresholded bilevel image produced by the proposed relative entropy approach will best match the co-occurrence matrix of the original image. This observation is demonstrated experimentally by several test images. Although there is no evidence of indication that one is generally better than the others, according to the experiments conducted in this paper, the joint entropy did not work as well as did the local entropy and relative entropy. Interestingly, among all images tested the relative entropy approach seems to perform better than the others in finding edges. In addition, our experiments show that the relative entropy seems to be a good complement to the local entropy and joint entropy methods in terms of providing different image details and descriptions from those provided by the local entropy and joint entropy. Finally, an advantage of the relative entropy approach is the computational saving compared to the local and joint entropy approaches based on arithmetic operations required for calculating entropies.

REFERENCES

1. N. R. Pal and S. K. Pal, Entropy thresholding, *Signal Process.* **16**, 97–108 (1989).

2. T. Pun, A new method for grey-level picture thresholding using the entropy of the histogram, *Signal Process.* **2**, 223–237 (1980).
3. T. Pun, Entropic thresholding, a new approach, *Comput. Graphics Image Process.* **16**, 210–239 (1981).
4. J. N. Kapur, P. K. Sahoo and A. K. C. Wong, A new method for gray-level picture thresholding using the entropy of the histogram, *Comput. Graphics Vision Image Process.* **29**, 273–285 (1985).
5. A. S. Abutaleb, Automatic thresholding of gray-level pictures using two-dimensional entropy, *Comput. Vision Graphics Image Process.* **47**, 22–32 (1989).
6. N. R. Pal and S. K. Pal, Entropy: a new definition and its application, *IEEE Trans. Syst. Man, Cybern.* **21**, 1260–1270 (1991).
7. S. Kullback, *Information Theory and Statistics*. Dover, New York (1968).
8. L. D. Davisson and A. Leon-Garcia, A source matching approach to finding minimax codes, *IEEE Trans. Inf. Theory* **26**, 166–174 (1980).
9. C.-I. Chang, S. C. Fan and L. D. Davisson, On numerical methods of calculating the capacity of a continuous-input discrete output memoryless channel, *Inf. Comput.* **86**, 1–13 (1990).
10. C.-I Chang and L. B. Wolfe, Source matching problems revisited, *IEEE Trans. Inf. Theory* **38**, 1391–1395 (1992).
11. L. B. Wolfe and C.-I. Chang, A simple method for calculating the rate distortion function of a source with an unknown parameter, *Signal Process.* **33**, 209–221 (1993).
12. C.-I Chang, L. C. Fan and L. D. Davisson, Computation of the rate-distortion function of a source with uncertain statistics, *Proc. IEEE 1988 International Conference on Communication Systems*, Singapore, pp. 1180–1184, 31 October–3 November (1988).

About the Author—CHEIN-I CHANG received his B.S., M.S., and M.A. degrees from Soochow University, Taipei, Taiwan, in 1973, National Tsing Hua University, Hsinchu, Taiwan, in 1975, and SUNY at Stony Brook in 1977, respectively, all in Mathematics, and M.S.E.E. from the University of Illinois at Urbana-Champaign in 1982, and Ph.D. in Electrical Engineering from the University of Maryland, College Park, in 1987. From 1973 to 1977 he was a teaching assistant at Soochow University, National Tsing Hua University, and SUNY at Stony Brook. From 1977 to 1980 he was an instructor of Mathematics at the University of Illinois. During 1980–1982 he was a research assistant at the University of Illinois and was also a research assistant at the University of Maryland from 1982 to 1986. He was Visiting Assistant Professor from January 1987 to August 1987, and Assistant Professor from 1987 to 1993, and currently is Associate Professor of Electrical Engineering at the University of Maryland, Baltimore County campus. His research interests include information theory and coding, signal detection and estimation, biomedical imaging, multispectral/hyperspectral signal and image processing, sensor/data fusion, neural networks, computer vision and pattern recognition. Dr Chang is a senior member of IEEE and a member of Phi Kappa Phi and Eta Kappa Nu.

About the Author—JIANWEI WANG received her B.S. and M.S. from the Department of Electronic Engineering of Zejing University in 1985 and 1988, respectively. She was a teaching assistant from 1991 to 1993 and research assistant since 1993 in the Department of Electrical Engineering at the University of Maryland, Baltimore County, where she is currently a Ph.D. candidate. Her research interests include image processing, computer vision and pattern recognition.

About the Author—MARK L. G. ALTHOUSE received his B.S. in Physics from the Pennsylvania State University in 1981, and his M.S. in Electrical Engineering from Johns Hopkins University in 1988, and is presently working on his Ph.D. in the Department of Electrical Engineering at the University of Maryland, Baltimore County. Since 1981 he has been employed by the U.S. Army Edgewood Research Development and Engineering Center, Aberdeen Proving Ground, MD, working on the remote detection of chemical and biological agents. From 1985 to 1986 he was an exchange scientist at the German NBC Defence Laboratory in Münster, Germany. Mr Althouse is a member of OSA, SPIE, IEEE, and INNS.