Entropy: A New Definition and its Applications

Nikhil R. Pal and Sankar K. Pal

Abstract—Shannon's definition of entropy is critically examined and a new definition of classical entropy based on the exponential behavior of information-gain is proposed along with its justification. The concept is then extended to gray tone image for defining its global, local and conditional entropy. Based on these definitions four algorithms for object extraction are developed and implemented. One of these algorithms uses a Poisson distribution-based model of an ideal image. Finally, a concept of positional entropy giving an information regarding the location of an object in a scene is introduced.

I. INTRODUCTION

The entropy of a system as defined by Shannon [1], [2] gives a measure of uncertainty about its actual structure. Shannon's function is based on the concept that the information gain from an event is inversely related to its probability of occurrence. Several authors [3]–[11] have used Shannon's concept for image processing and pattern recognition problems. Pun [3]–[4] and Kapur *et al.* [5] used Shannon's concept to define the entropy of an image assuming that an image is entirely represented by its gray level histogram only. As a result segmentation algorithms [3]–[5] using Shannon's function resulted in an unappealing result, viz, same entropy and threshold values for different images with identical histogram.

The present work first of all, introduces a new definition of classical entropy along with its justification. Unlike the logarithmic behavior of Shannon's entropy, the gain function considered here is of exponential nature. Some properties of this definition are also proved. Based on the new concept, three definitions (e.g., global, local and conditional) of entropy of an image are then introduced and applied to formulate four algorithms for image segmentation. One of the algorithms assumes a Poisson distribution to describe the gray level variation within the object and background.

The algorithms are implemented on a set of images and their superiority is established by comparing the results with those of other methods. Finally, a concept of positional entropy to indicate the information on location of an object within a scene is discussed.

II. SHANNON'S ENTROPY AND ITS APPLICATIONS

Shannon [1]–[2] defined the entropy of an n-state system as

$$H = -\sum_{i=1}^{n} p_i \log(p_i) \tag{1}$$

where p_i is the probability of occurrence of the event i and

$$\sum_{i=1}^{n} p_i = 1, \qquad 0 \le p_i \le 1. \tag{2}$$

Intuitively we feel that the gain in information from an event is inversely related to its probability of occurrence. Shannon used

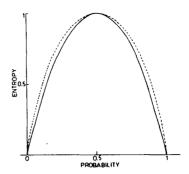$$\Delta I = \log(1/p_i) = -\log(p_i) \tag{3}$$

Fig. 1. Normalized entropy of a two state system (— new entropy, --- Shannon's entropy).
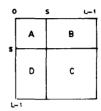

Fig. 2. Quadrants of a co-occurrence matrix.

as the measure for such a gain. The expected value of the gain function is taken as the entropy, i.e.,

$$H = E(\Delta I) = -\sum_{i=1}^{n} p_i \log(p_i). \tag{4}$$

Let $F = [f(x,y)]_{P \times Q}$ be an image of size $P \times Q$, where $f(x,y)$ is the gray value at (x,y); $f(x,y) \in G_L = \{0, 1, \cdots, L-1\}$, the set of gray levels. Let N_i be the frequency of the gray level i. Then

$$\sum_{i=0}^{L-1} N_i = PQ = N(\text{say}). \tag{5}$$

Pun [3], [4] and Kapur *et al.* [5] considered the gray level histogram of F as the outcome of an L-symbol source, independently from the underlying image. In addition to this, they also assumed that these symbols are statistically independent.

III. ENTROPY OF AN IMAGE

Following Shannon's definition of entropy (1), Pun [3] defined the entropy of the image (histogram) as

$$H = -\sum_{i=0}^{L-1} p_i \log p_i; \qquad p_i = N_i / N. \tag{6}$$

Let s be a threshold and $P_s = \sum_{i=0}^{s} p_i$. Then the *a posteriori* entropy of the image is $H'_L(s) = -P_s \log P_s - (1 - P_s) \log (1 - P_s)$. Pun [3] maximized an upper bound of $H'_L(s)$ for threshold selection. On the other hand, Kapur *et al.* [5] considered two probability distributions, one for the object and the other for the background. They maximized the sum of the individual entropy of the object and background.

All these methods [3]–[5] virtually assume that an image is entirely represented only by its histogram. Thus, different images
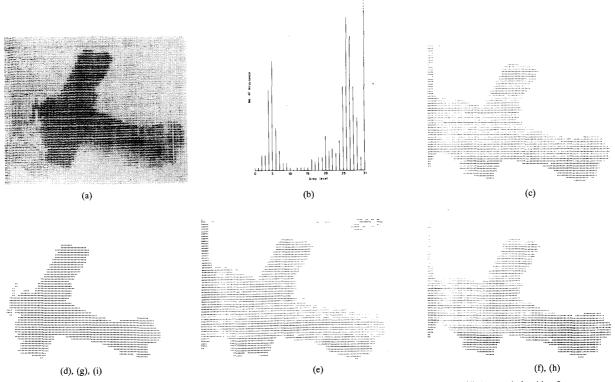
(a)

(b)

(c)

(d), (g), (i)

(e)

(f), (h)

Fig. 3. Input and segmented images of the biplane. (a) Input. (b) Histogram. (c) Proposed algorithm 1. (d) Proposed algorithm 2. (e) Algorithm of pun. (f) Algorithm of Kapur *et al.* (g) Bayes classifier. (h) Proposed algorithm 3. (i) Proposed algorithm 4.

with identical histograms will result in the same entropic value and threshold in spite of their different spatial distributions of gray levels. This is not intuitively appealing. Moreover in the algorithm of Pun [3] the concept of maximization of an upper bound of $H'_L(s)$, for object-background classification is not justified.

IV. NEW DEFINITION OF ENTROPY

A. Justification

Before introducing a new definition of entropy the following points are in order.

1) It is to be noted from the logarithmic entropic measure that as $p_i \rightarrow 0, \Delta I(p_i) \rightarrow \infty$ and $\Delta I(p_i) = -\log(p_i)$ is not defined for $p_i = 0$. On the other hand, as $p_i \rightarrow 1, \Delta I(p_i) \rightarrow 0$ and $\Delta I(p_i = 1) = 0$.

   In practice, the gain in information from an event, whether highly probable or highly unlikely, is expected to lie between two finite limits.

2) In Shannon's theory the measure of ignorance or the gain in information is taken as $\log(1/p_i)$. But statistically ignorance can be better represented by $(1 - p_i)$ than $(1/p_i)$.

   Now if we define the gain in information corresponding to the occurrence of the ith event as

$$\Delta I(p_i) = \log(1 - p_i) \quad \text{or} \quad -\log(1 - p_i) \tag{7}$$

then $\Delta I < 0$ or increases with p_i, which is intuitively unappealing.

   The above problem can be circumvented by considering an

exponential function of $(1 - p_i)$. This is also appropriate while considering the concept of information gain in an image [14], [15].

3) The additive property of the gain function for independent events does not have a strong relevance (impact) in practice. If we consider, the joint occurrence of two independent events, then there is no point in establishing a relationship between the gains of two events as these are the outcomes of two independent systems.

B. Definition [14]

The aforesaid analysis led us to the following desirable properties for the new entropic function.

P1: $\Delta I(p_i)$ is defined at all points in $[0,1]$.
P2: $\text{Lim}_{p_i \rightarrow 0} \Delta I(p_i) = \Delta I(p_i = 0) = k_1, k_1 > 0$ and finite.
P3: $\text{Lim}_{p_i \rightarrow 1} \Delta I(p_i) = \Delta I(p_i = 1) = k_2, k_2 > 0$ and finite.
P4: $k_2 < k_1$
P5: With increase in p_i, $\Delta I(p_i)$ decreases exponentially.
P6: $\Delta I(p)$ and H, the entropy, are continuous for $0 \leq p \leq 1$.
P7: H is maximum when all p_i's are equal.

In other words, $H(p_1, \cdots, p_n) \leq H(1/n, \cdots, 1/n)$.

Under the above frame work let us define the gain in information from an event with probability p_i as

$$\Delta I(p_i) = e^{(1-p_i)} \tag{8}$$

and the entropy

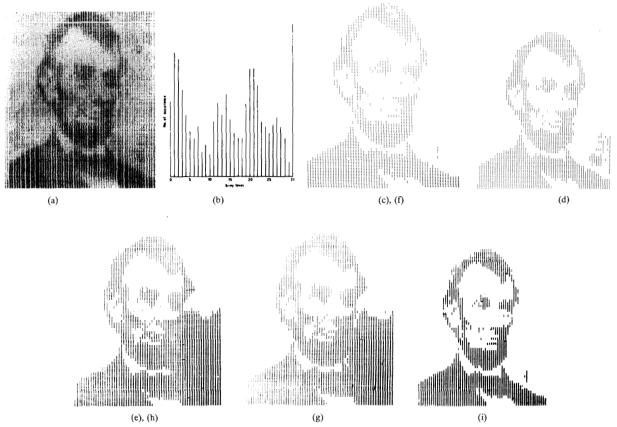$$H = E(\Delta I) = \sum_{i=1}^{n} p_i e^{(1-p_i)}. \tag{9}$$

Fig. 4. Input and segmented images of Abraham Lincoln. (a) Input. (b) Histogram. (c) Proposed algorithm 1. (d) Proposed algorithm 2. (e) Algorithm of pun. (f) Algorithm of Kapur *et al.* (g) Bayes classifier. (h) Proposed Algorithm 3. (i) Proposed algorithm 4.

The normalized entropy H_{nor} can be defined as

$$H_{\text{nor}} = k(H - 1) \tag{10}$$

where $K = 1/\left(e^{1-1/n} - 1\right)$.

It is easy to see that the properties P1 through P6 are satisfied when k_1 and k_2 take the values e and 1, respectively. The proof of the property P7 is given in Appendix A. In this context it may be mentioned that a large number of additional interesting properties are established in [15]. It is also proved in Appendix B that for a binary source H monotonically increases for p in $[0, 0.5)$ and monotonically decreases for p in $(0.5, 1]$ with a maximum at $p = 0.5$.

Fig. 1, which gives the plot of Shannon's entropy and the new entropy (*normalized*) for a two state system, establishes that the basic nature of the information measure remains the same with the new definition.

In the following sections two measures, higher-order entropy and conditional entropy of an image that take into account the spatial dependency of pixel intensities are defined.

A. Higher-Order Entropy

To arrive at the expression of entropy of an image the following theorem can be stated based on the idea of Shannon [1], [2], [12], [15].

Theorem: Let $p(s_i)$ be the probability of a sequence s_i of gray levels of length q. Let us define

$$H^{(q)} = \frac{1}{q} \sum_i p(s_i)e^{1-p(s_i)} \tag{11}$$

where the summation is taken over all gray level sequences of length q. Then for a stationary information source $H = \lim_{q \to \infty} H^{(q)}$ exists. Let us call H as the entropy of the image.

Different values of q gives various orders of entropy.

For $q = 1$ we get

$$H^{(1)} = \sum_{i=0}^{L-1} p_i e^{1-p_i}. \tag{12}$$

Such an entropy is a function of the histogram only and it may be called the "global entropy" of the image.

Similarly, $q = 2$ gives

$$H^{(2)} = 1/2 \sum_i \sum_j p_{ij}e^{1-p_{ij}} \tag{13}$$

where p_{ij} is the probability of co-occurrence of gray levels i and j. $H^{(2)}$ takes into account the spatial distribution of gray levels. Expressions for higher-order entropies $(q > 2)$ can also be deduced in a similar manner. $H^{(i)}$, $i > 2$ may be called "local entropy" of order i of an image.

B. Conditional Entropy

Suppose an image has two distinct portions, the object O and background B with gray levels $\{x_i\}$ and $\{y_i\}$, respectively. Then the conditional entropy of the object O given the background B is

$$H(O/B) = \sum_{x_i \in O} \sum_{y_j \in B} p(x_i/y_j) e^{1-p(x_i/y_j)}. \qquad (14)$$

Similarly, the conditional entropy of the background B given the object O is defined as

$$H(B/O) = \sum_{y_j \in B} \sum_{x_i \in O} p(y_j/x_i) e^{1-p(y_j/x_i)}. \qquad (15)$$

Here the pixel y_j, in general can be an mth order neighbor of the pixel x_i. Since the estimation of such a probability is very difficult, we impose the constraint that x_i and y_j are adjacent pixels. The conditional entropy $H^{(C)}$ of the image can, therefore, be defined as

$$H^{(C)} = (H(O/B) + H(B/O))/2. \qquad (16)$$

V. APPLICATIONS TO IMAGE SEGMENTATION

Based on the new definitions of entropy of an image, the following four algorithms for object-background classification are proposed.

A. Algorithms Based on Global Information

Algorithm 1: Following the concept of Kapur et al. and making use of (12) we can determine the thresholding criterion as follows. Suppose s is an assumed threshold. Then using (12) the global entropy of the black portion of the image can be defined as

$$H_B^{(G)}(s) = \sum_{i=0}^{s} p_i/P_s e^{1-p_i/P_s} \qquad (17)$$

and that of the background or the white portion of the image as

$$H_W^{(G)}(s) = \sum_{i=s+1}^{L-1} p_i/(1-P_s) e^{1-p_i/(1-P_s)}. \qquad (18)$$

Therefore, the total global entropy of the partitioned image is

$$H_T^{(G)}(s) = H_B^{(G)}(s) + H_W^{(G)}(s). \qquad (19)$$

Let $\mathrm{Max}_s\{H_T^{(G)}(s)\} = H_T^{(G)}(k); 0 \leq s, k < L-1$.

The level k can be taken as a threshold for object-background classification of the image.

Algorithm 2: This algorithm assumes a Poisson distribution for describing the variation in gray levels within the object/background. Let us define an ideal imaging process as its understanding is required, for justifying the use of Poisson distribution.

An ideal imaging device can be thought of as a spatial array of individual photon receptors and counters, each having identical properties. It is assumed that each receptor can receive light quanta (photon) independent of its neighboring receptors. The image state of a receptor is completely determined by the number of quanta it receives and records, and each receptor can be in one of a finite number of distinguishable image states. Since the possible number of states of a receptor is finite, after a receptor attains the final state, all other additional incident quanta will remain unrecorded. In other words, the receptor gets saturated.

When we feel the exposure level over the entire imaging device as uniform, the number of incident quanta is found to follow a Poisson distribution with a sufficient degree of validity [13]. In other words,

if the uniform exposure level is such that, each receptor has, on the average q quanta then

$$p = \text{probability of a counter receiving } r \text{ quanta}$$
$$= q^r e^{-q}/r!, \qquad \text{for } r = 0, 1, \cdots, q, \cdots. \qquad (20)$$

In the case of a digital image, each pixel can be viewed as a receptor. Like the ideal imaging system, the spatial resolution here depends also on the spatial size of the pixel and each pixel can have only a finite number of states with a saturation level. The observed gray level of a pixel is nothing but the effect of the received quanta by the corresponding receptor. Larger the number of recorded quanta, the higher is the gray value.

Let us now consider a scene consisting of an ideal object and an ideal background. An ideal object means that the entire object surface has uniform characteristics (i.e., constant coefficient of reflection, constant temperature distribution, made up of same material and so on). Similarly, the ideal background also has uniform characteristics, but obviously different from those of the object. When we take the photograph of such an ideal scene illuminated by a uniform light, the scene itself acts as the source of light for the imaging system. Though the illumination of the entire scene is uniform, the object and background exposure levels for the imaging system will be of two different natures since they have different characteristics. Ignoring the interaction between the quanta emitted by the object with that emitted by the background, we can say that the recorded uniform image will have two uniform illuminations, one corresponding to the object and the other due to background.

We assume that the pixel value of a cell is equivalent to the recorded quanta by the corresponding receptor cell of the imaging system. In other words, it implies that the pixel values for uniform illumination follow a Poisson distribution. Thus the gray level histogram of a digital image will be a combination of two Poisson distributions characterized by two different parameters λ_O and λ_B.

In the object extraction problem, our aim is to find the gray level l, $\lambda_O < l < \lambda_B$, such that the gray levels ranging from zero (0) to l represent the object (with uniform illumination λ_O) and the gray levels in the range $l+1$ to $L-1$ form the background (with uniform illumination λ_B) or vice versa.

In order to select such a level (threshold) we maximize the sum of entropy of the two Poisson distributions. For each assumed threshold l, $0 \leq l \leq L-1$, the maximum likelihood (ML) estimate of the parameters (λ_O and λ_B) for both of the distributions are found out and used to compute the entropy as follows.

The estimated probability p_r^O of an object cell to have a gray value

$$r = \frac{\lambda_O^r e^{-\lambda_O}}{r!}, \qquad 0 \leq r \leq l \qquad (21)$$

and $p_r^B = $ Poisson probability of a background cell to have a gray value

$$r = \frac{\lambda_B^r e^{-\lambda_B}}{r!}, \qquad l+1 \leq r \leq L-1. \qquad (22)$$

Thus the sum of entropy of the object and background is

$$H_l = \sum_{i=O}^{l} p_i^O e^{1-p_i^O} + \sum_{i=l+1}^{L-1} p_i^B e^{1-p_i^B}. \qquad (23)$$

The threshold level for object-background classification, is selected by maximizing H_l with respect to l.
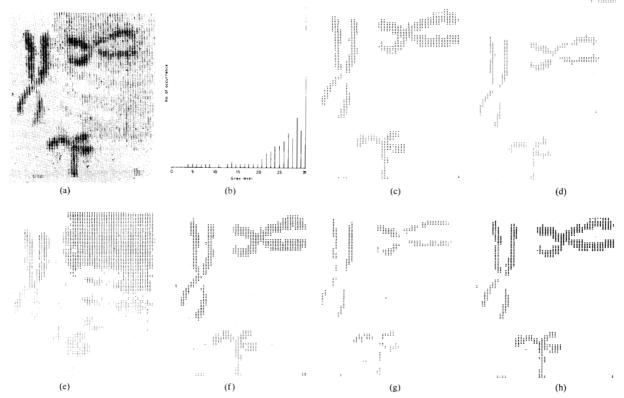
Fig. 5. Input and segmented images of three chromosomes. (a) Input. (b) Histogram. (c) Proposed algorithm 1. (d) Proposed algorithm 2. (e) Algorithm of pun. (f) Algorithm of Kapur *et al.* (g) Proposed algorithm 3. (h) Proposed Algorithm 4.

B. Algorithms Based on Local Information

Algorithm 3: Let us now describe another algorithm based on (13), which takes into account the spatial details of an image. Such a method is dependent on the probability of co-occurrence p_{ij} of pixel intensities i and j.

If s, $0 \leq s < L - 1$, is a threshold, then s partitions the co-occurrence matrix into four quadrants, namely A, B, C, D (Fig. 2).

Let us define the following quantities.

$$P_A = \sum_{i=0}^{s} \sum_{j=0}^{s} p_{ij}. \qquad P_B = \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} p_{ij}$$

$$P_C = \sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} p_{ij} \qquad P_D = \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} p_{ij}. \qquad (24)$$

Normalizing the probabilities within individual quadrant, we get the following cell probabilities for different quadrants:

$$p_{ij}^A = \frac{p_{ij}}{P_A} \quad \text{for } 0 \leq i. \qquad j \leq s \qquad (25)$$

$$p_{ij}^B = \frac{p_{ij}}{P_B} \quad \text{for } 0 \leq i \leq s \quad \text{and} \quad s+1 \leq j \leq L-1 \qquad (26)$$

$$p_{ij}^C = \frac{p_{ij}}{P_C} \quad \text{for } s+1 \leq i. \qquad j \leq L-1 \qquad (27)$$

and

$$p_{ij}^D = \frac{p_{ij}}{P_D} \quad \text{for } s+1 \leq i \leq L-1 \quad \text{and} \quad 0 \leq j \leq s. \qquad (28)$$

Now with the help of (13) and (25) the second-order local entropy of the object can be defined as

$$H_A^{(2)}(s) = \frac{1}{2} \sum_{i=0}^{s} \sum_{j=0}^{s} p_{ij}^A e^{1 - p_{ij}^A}. \qquad (29)$$

Similarly, the second-order entropy of the background can be written as

$$H_C^{(2)}(s) = \frac{1}{2} \sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} p_{ij}^C e^{1 - p_{ij}^C}. \qquad (30)$$

Hence the total second-order entropy of the partitioned image is

$$H_T^{(2)}(s) = H_A^{(2)}(s) + H_C^{(2)}(s). \qquad (31)$$

The gray level corresponding to the maximum of $H_T^{(2)}(s)$ gives the threshold for object-background classification.

Algorithm 4: This algorithm is based on the concept of the conditional entropy (14)–(16). Let t_{ij} be an entry of the quadrant B (Fig. 2), then t_{ij} gives the number of transitions, such that i belongs to the object and j belongs to the background, and i and j are adjacent. Therefore, p_{ij}^B's of (26) give the probabilities required by (14). Similarly, p_{ij}^D's of (28) correspond to the probabilities of (15). Thus,

$$H(O/B) = \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} p_{ij}^B e^{1 - p_{ij}^B} \qquad (32)$$
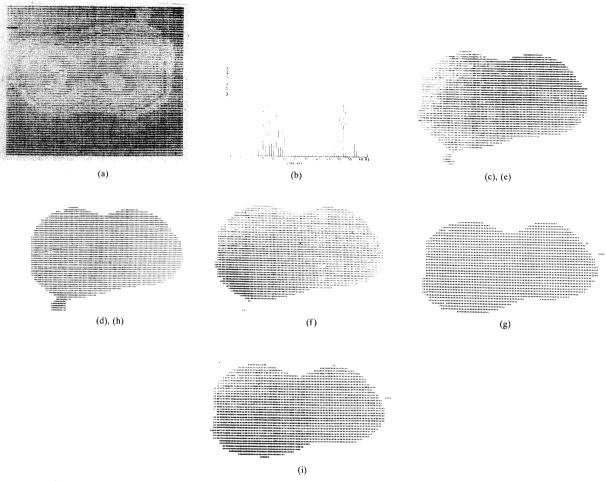
(a)

(b)

(c), (e)

(d), (h)

(f)

(g)

(i)

Fig. 6. Input and segmented images of blurred chromosome. (a) Input. (b) Histogram. (c) Proposed algorithm 1. (d) Proposed algorithm 2. (e) Algorithm of pun. (f) Algorithm of Kapur *et al.* (g) Bayes classifier. (h) Proposed Algorithm 3. (i) Proposed algorithm 4.

and

$$H(B/O) = \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} p_{ij}^{D} e^{1-p_{ij}^{D}}. \tag{33}$$

Now the conditional entropy of the image is

$$H_T^{(C)}(s) = (H(O/B) + H(B/O))/2. \tag{34}$$

To get the threshold, we minimized $H_T^{(C)}(s)$ with respect to s.

VI. IMPLEMENTATION AND RESULTS

The performance of the algorithms proposed in Section V has been demonstrated on a set of five images. We shall first of all concentrate on Algorithms 1 and 2 to show their superiority over other algorithms based on global information only. Then the results of Algorithms 3 and 4 will be presented.

A. Results of Methods Based on Global Information

Figs. 3(a)–7(a) represent different input images used, while Figs. 3(b)–7(b) depict the corresponding gray level histograms. In order to compare the results of the proposed Algorithms 1 and 2

with some of the existing methods, algorithm of Pun [3] and Kapur *et al.* [5] have been implemented. In addition to these, another algorithm based on Bayes misclassification error [12] (will be called as Algorithm *) has been considered here.

The threshold levels produced by different global methods are displayed in Table I. For the image of a biplane Fig. 3(a) the outputs produced by different methods are shown in Fig. 3(c)–(g). From the results one can see that except (23) and Algorithm * the propeller in front of the biplane is lost. For the other methods some portion of the background got mixed up with the object.

The thresholded outputs for the Lincoln image (Fig. 4(a)) are shown in Fig. 4(c)–(g). Though the histogram has a number of deep valleys, the proposed Algorithm 2 has been able to extract correctly the object and in fact, produced the best result. The result produced by Algorithm * is close to that obtained by Algorithm 2 but the other methods have failed to extract the object correctly.

In order to show the effectiveness of the algorithms for images with unimodal histogram (Fig. 5(b)), an image of three chromosomes (Fig. 5(a)) has been considered. In this case it is found that Algorithm 1 and the method of Kapur *et al.* [5] have resulted in good object-background classification (Fig. 5(c) and (f)) keeping the legs of a chromosome connected.
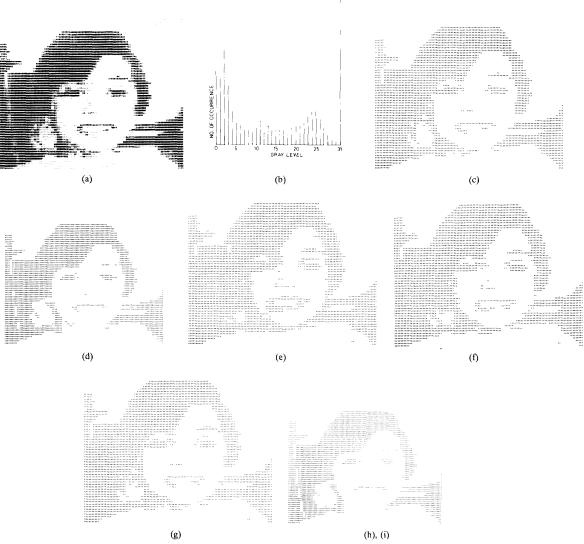
Fig. 7. Input and segmented images of boy. (a) Input. (b) Histogram. (c) Proposed algorithm 1. (d) Proposed algorithm 2.
(e) Algorithm of pun. (f) Algorithm of Kapur *et al.* (g) Bayes classifier. (h) Proposed Algorithm 3. (i) Proposed algorithm 4.

In case of the image of the blurred chromosome (Fig. 6(a)) the segmented results produced by all the methods are found to be comparable.

For the image of the boy (Fig. 7(a)), Algorithm 1 is seen to be able to extract the face preserving its shape and also to separate the ring in the right ear of the boy (Fig. 7(c)), which is not the case for others (Figs. 7(d)–(g)).

From the above discussion, it appears that the overall performance of Algorithm 2 is consistently satisfactory compared to other global information-based algorithms.

B. Results of Methods Based on Local Information

All the five images described earlier have also been used as input to the proposed Algorithms 3 and 4. The threshold levels obtained by these methods are displayed in Table II. In case of the

biplane (Fig. 3(a)), Algorithm 4 has produced much better result than Algorithm 3 (Figs. 3(h)–(i)). Results produced by Algorithm 3 is comparable to those obtained by [5] and (19) (Table I). Similar is the case for Lincoln image, where Algorithm 4 created much better result than that of Algorithm 3 (Figs. 4(h) and (i)). The threshold obtained by Algorithm 4 is close to those given by (23) and Algorithm *. Whereas, the result of Algorithm 3 is almost same with that of others shown in Table I. For the image of chromosomes (Fig. 5(a)) Algorithm 4 has again resulted in a more appealing segmented version (Fig. 5(h)) than others.

Finally, for the image of the boy (Fig. 7(a)) both Algorithms 3 and 4 have produced same result (Fig. 7(h)–(i)).

VII. Positional Entropy

The new entropy introduced in Section III or Shannon's entropy,

TABLE I
THRESHOLDS FOR OBJECT BACKGROUND CLASSIFICATION (BASED ON GLOBAL INFORMATION)

	THRESHOLDS				
Images	Proposed Algo. −1 (19)	Proposed Algo. −2 (23)	Algorithm of Pun [3]	Algorithm Kapur et al. [5]	Algorithm * [12]
BIPLANE (Fig. 3)	22	12	24	21	12
LINCOLN (Fig. 4)	15	11	16	15	08
CHROMOSOMES (Fig. 5)	19	12	27	20	—
BLURRED CHROMOSOME (Fig. 6)	32	31	32	33	43
BOY (Fig. 7)	11	09	15	13	12

TABLE II
THRESOLDS FOR OBJECT-BACKGROUND CLASSIFICATION
(BASED ON LOCAL INFORMATION)

	THRESHOLDS	
Images	Proposed algorithm −3 (31)	Proposed algorithm −4 (34)
BIPLANE (Fig. 3)	21	12
LINCOLN (Fig. 4)	16	09
CHROMOSOMES (Fig. 5)	10	17
BLURRED CHROMOSOME (Fig. 6)	31	41
BOY (Fig. 7)	08	08

```
A  .   .   . G  .   .   . B
.                         .
.                         .
.                         .
E                         F
.                         .
.                         .
.                         .
D  .   .   . H  .   .   . C
```
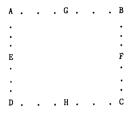
Fig. 8. Different positions of object for correlation.

and

$$\sigma_j = \left[\frac{\sum (j - \bar{J})^2}{O_s} \right]^{1/2}. \tag{37}$$

Such a correlation can have either a positive, negative or zero value depending on the position of the object in the scene. The "Object correlation" will be positive, if major portion of the object is along the diagonal AC (Fig. 8). On the other hand, r will be negative when the object is concentrated along the diagonal BD (Fig. 8) and r takes the value zero when the object lies symmetrically on any one of the axes EF and GH (Fig. 8). Let us now define the positional entropy (H^P) as follows

$$H^P = \sum_i \sum_j w_{i,j} p_{i,j} e^{1-p_{i,j}} \tag{38}$$

where $p_{i,j}$ is the probability of occurrence of the gray level that occupies the (i,j)th cell:

$$w_{i,j} = 1/d_A(i,j), \quad \text{when } r > 0 \tag{39a}$$
$$= 1/d_D(i,j), \quad \text{when } r < 0 \tag{39b}$$
$$= 1/d_G(i,j), \quad \text{when } r = 0 \quad \text{and}$$
$$\sum_{(i,j)\in\text{Object}} (j - \bar{J}) = 0 \tag{39c}$$
$$= 1/d_E(i,j), \quad \text{when } r = 0 \quad \text{and}$$
$$\sum_{(i,j)\in\text{Object}} (j - \bar{J}) \neq 0. \tag{39d}$$

Here, $d_x(i,j)$ denotes the distance from the point x (x can be any of the points A, D, G, and E (Fig. 8)).
Equation (38) for H^P can be written as

$$H^P = \sum \sum w_{i,j} p_{i,j} e^{1-p_{i,j}} \tag{40}$$

gives a measure of our ignorance about the system. For an image, the global entropy gives a measure of grayness ambiguity. The local and conditional entropies, on the other hand, give information about the spatial ambiguity by measuring intraset homogeneity and interset contrast. In addition to this one may be interested in an information about the position of an object on a scene.

The proposed entropic measures are able to extract (segment) the object irrespective of its position on the scene and if the object is moved from one place to another, on the scene, keeping the gray level distribution same then the entropy does not change. This observation led us to define another entropic measure that will give an idea about the location of the object. Such a measure may be named as "positional entropy."

A. Definition

Before introducing the measure let us define the "object correlation" (r) as

$$r = \frac{\sum_{(i,j)\in\text{Object}} (i - \bar{I})(j - \bar{J})}{\sigma_i \sigma_j O_s} \tag{35}$$

where $\bar{I} = P/2 + 0.5$, $\bar{J} = Q/2 + 0.5$, O_s is the size of the object, and $P \times Q$ is the size of the image:

$$\sigma_i = \left[\frac{\sum (i - \bar{I})^2}{O_s} \right]^{1/2} \tag{36}$$

Fig. 9. Illustration of positional entropy.

$$= \sum_{l=0}^{s} \left\{ p_l e^{1-p_l} \sum_{f(i,j)=l} w_{i,j} \right\}$$

$$= \sum_{l=0}^{s} w_l p_l e^{1-p_l} \tag{40}$$

where $w_l = \sum_{f(i,j)=l} w_{i,j}$ and the gray levels ranging from 0 to s constitute the object. Similar definition using the logarithmic entropy can also be written.

Equation (41) can be viewed as an weighted entropy. Such a measure takes into account the position of the object both on the image plane and on the gray scale. Thus if the same object (same gray level distribution) is placed at different position on the scene, the positional entropy will change. Again keeping the position of the object fixed, if the gray level distribution is changed, then also the positional entropy changes. This characteristic of H^P, may be used to find the approximate location of the object on the scene.

Case -1: $r = 0$: If the object lies symmetrically any where on the axis EF or GH (Fig. 8), the correlation (r) will be zero. Note that $\sum (j - \overline{J}) = 0$ for an object lying on the axis GH. Thus a high (low) value of H^P will indicate the position of the object near $G(H)$. Otherwise, (i.e., if $(j - \overline{J}) \neq 0$) high and low values of H^P will correspond to the vicinity of E and F, respectively.

Case -2: $r > 0$: In this case the object is expected to lie along the diagonal AC (Fig. 8). Hence if H^P is high, the major part of the object will be in the first quadrant, otherwise in the third quadrant.

Case -3: $r < 0$: Under this situation a high value of H^P will indicate the position of the major part of the object in the fourth quadrant (Fig. 8). Similarly, a low value of H^P will correspond to the second quadrant.

The previous discussion merely introduces a concept for providing information regarding the location of an object in a scene. Further experimental investigation is required to assign the value of H^P to the fuzzy hedges "high" and "low."

As an illustration, Fig. 9(a)–(j) demonstrate the aforementioned feature of H^P for different positions of an object of size 4×8 in an 8 level 16×16 image. The corresponding H^P and r values conform to what was discussed earlier.

VIII. CONCLUSION

A new definition of probabilistic entropy based on the exponential behavior of information gain is proposed along with its justification. Its properties are also found to include those of Shannon's entropy. Based on this concept various definitions of image-entropy (namely, global, local, and conditional) are introduced.

Four algorithms for object-background classification are proposed. Two of them are based on the global entropy while the remaining two take the local and conditional entropy into account. One of the global entropic algorithms assumes a Poisson distribution for describing the gray level variation within the object/background. The analysis of the actual image formation process led us to the assumption of Poisson distribution. The performance of the proposed algorithms is compared with some existing entropic thresholding methods and with a Bayes classifier using normal distribution. The proposed global entropic method using Poisson distribution and the conditional entropic method are found to be consistently superior to other methods for a wide range of input images.

Besides these, a concept of positional entropy together with its formulation has also been attempted. This may be used to find the approximate location of an object in a scene.

APPENDIX A

Proof of P7 (Section III-B):

$$H = \sum_{i=1}^{n} p_i e^{(1-p_i)}$$

$$0 \leq p_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1 \tag{41}$$

or

$$H = p_1 e^{(1-p_1)} + p_2 e^{(1-p_2)} + \cdots + p_{n-1} e^{(1-p_{n-1})}$$
$$+ (1 - p_1 - p_2 - \cdots - p_{n-1}) e^{(p_1+p_2+\cdots+p_{n-1})}$$

as

$$p_n = 1 - p_1 - p_2 - \cdots p_{n-1}. \tag{44}$$

Now taking the partial derivative of H with respect to p_i, $1 \leq i < n$ and equating that to zero we get,

$$\frac{\partial H}{\partial p_i} = 0, \qquad i = 1, 2, \cdots, n-1 \tag{45}$$

or

$$e^{(1-p_i)} - p_i e^{(1-p_i)} - e^{(p_1+p_2+\cdots+p_{n-1})}$$
$$+ (1 - p_1 - p_2 - \cdots - p_{n-1}) e^{(p_1+p_2+\cdots+p_{n-1})} = 0 \tag{46}$$

or

$$(1 - p_i) e^{(1-p_i)} = (p_1 + p_2 + \cdots + p_{n-1}) e^{(p_1+p_2+\cdots+p_{n-1})} \tag{47}$$

Now writing

$$1 - p_i = x_i \tag{48}$$

and

$$p_1 + p_2 + \cdots + p_{n-1} = y \tag{49}$$

we get

$$x_i e^{x_i} = y e^y. \tag{50}$$

Now define a function $f(x) = x e^x \; 0 \leq x \leq 1$

Claim: $f(x)$ is a bijection, i.e., $f(x)$ maps x uniquely.

Proof: Let x_1 and x_2 be two points in $[0, 1]$, i.e., $0 \leq x_1, x_2 \leq 1$ Then

$$f(x_1) = x_1 e^{x_1}$$

and

$$f(x_2) = x_2 e^{x_2}$$

If

$$x_1 > x_2 \quad \text{then } e^{x_1} > e^{x_2} \to f(x_1) > f(x_2)$$

and

$$\text{if } x_1 < x_2 \quad \text{then } e^{x_1} < e^{x_2} \to f(x_1) < f(x_2).$$

Thus if x_1 and x_2 are different then $f(x_1)$ and $f(x_2)$ will also be different. Therefore, $f(x_1) = f(x_2)$ if and only if $x_1 = x_2$. Hence, $f(x) = x e^x$ is a bijection.

Using this fact and (50) one can write

$$x_i = y \quad \text{for } i = 1, 2, \cdots, n-1$$
$$\text{or} \quad 1 - p_i = p_1 + p_2 + \cdots + p_{n-1} \quad \text{for } i = 1, 2, \cdots, n-1.$$

Now taking the summation on both sides over $i = 1, 2, \cdots, n - 1$ we get

$$\sum_{i=1}^{n-1} (1 - p_i) = \sum_{i-1}^{n-1} (p_1 + p_2 + \cdots + p_{n-1}) \tag{51}$$

or

$$(n - 1) - (p_1 + p_2 + \cdots + p_{n-1}) = (n - 1)$$
$$\cdot (p_1 + p_2 + \cdots + p_n) \tag{52}$$

or

$$(n - 1) = n(p_1 + p_2 + \cdots + p_{n-1}) \tag{53}$$

or

$$n - 1 = n(1 - p_n) \tag{54}$$

or

$$np_n = 1 \tag{55}$$

or

$$p_n - 1/n. \tag{56}$$

Similarly, expressing other p_i's, $i = 1, 2, \cdots n - 1$, in terms of the remaining probabilities one can prove the condition of maximality (as H is concave) of H as $p_i = 1/n$ for all $i = 1, 2, \cdots, n$.

(The proof can also be done using Jensen inequality.)

APPENDIX B

For a binary source with probabilities p and $(1 - p)$ the entropy H of the system takes the following form.

$$H = pe^{1-p} + (1 - p)e^p, \qquad 0 \le p \le 1.$$

We need to show that H monotonically increases in $[0, 0.5)$ and monotonically decreases in $(0.5, 1]$ with a maximum at $p = 0.5$.

Differentiating H with respect to p we get

$$\frac{dH}{dp} = e^{1-p} - pe^{1-p} - e^p + (1 - p)e^p \tag{57}$$

$$= (1 - p)e^{(1-p)} - pe^p. \tag{58}$$

Now

$$\text{if } p \in [0, 0.5) \quad \text{then } (1 - p)e^{(1-p)} > pe^p$$

and

$$\text{if } p \in (0.5, 1] \quad \text{then } (1 - p)e^{(1-p)} < pe^p.$$

Therefore,

$$\frac{dH}{dp} > 0 \quad \text{if } p \in [0, 0.5) \tag{59}$$

$$\frac{dH}{dp} < 0 \quad \text{if } p \in (0.5, 1] \tag{60}$$

and

$$\frac{dH}{dp} = 0 \quad \text{if } p = 0.5. \tag{61}$$

Hence the proof.

ACKNOWLEDGMENT

REFERENCES

[1] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.,* vol. 27, pp. 379–423, July, 1948.

[2] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication,* Urbana, The Univ. Ill. Press, 1949.

[3] T. Pun, "A new method for gray-level picture thresholding using the entropy of the histogram," *Signal Processing,* vol. 2, pp. 223–237, 1980.

[4] T. Pun, "Entropic thresholding: A new approach," *Signal Processing,* vol. 2, pp. 210–239, 1981.

[5] 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," *Computer Graphics, Vision and Image Processing,* vol. 29, pp. 273–285, 1985.

[6] S. K. Pal, "A note on the quantitative measure of image-enhancement through fuzziness," *IEEE Trans. Pattern Anal. Machine Intell.,* vol. PAMI-4, pp. 204–206, 1982.

[7] S. K. Pal, R. A. King and A. A. Hashim, "Automatic gray-level thresholding through index of fuzziness and entropy," *Pattern Recog. Lett.,* vol. 1, pp. 141–146, 1983.

[8] S. K. Pal, "A measure of edge ambiguity using fuzzy sets," *Pattern Recog. Lett.,* vol. 4, pp. 51–56, 1986.

[9] S. K. Pal and P. K. Pramanik, "Fuzzy measures in determining seed points in clustering," *Pattern Recog. Lett.,* vol. 4, pp. 159–164, 1986.

[10] S. K. Pal and B. Chakraborty, "Fuzzy set theoretic measure for automatic feature evaluation," *IEEE Trans. Syst., Man, Cybern.,* vol. 16, no. 5, pp. 754–760, 1986.

[11] S. K. Pal and D. Duttamajumdar, *Fuzzy Mathematical Approach to Pattern Recognition.* New York: Wiley, 1985.

[12] E. L. Hall, *Computer Image Processing and Recognition.* New York: Academic, 1979, pp. 381–382.

[13] J. C. Dainty and R. Shaw, *Image Sciences.* New York: Academic, 1974, pp. 1–31.

[14] S. K. Pal and N. R. Pal, "Object-background classification using a new definition of entropy," *Proc. Int. Conf. Syst., Man, Cybern.,* Beijing, China, vol. 2, 1988, pp. 773–776.

[15] N. R. Pal, "On image information measures and object extraction," Ph.D. thesis, the Indian Statistical Institute, Calcutta, Mar. 1990.