ON IMAGE INFORMATION MEASURES AND OBJECT EXTRACTION

by

NIKHIL RANJAN PAL

Computer Science Unit
Indian Statistical Institute
Calcutta

A thesis submitted in partial fulfilment of
the requirements for the degree of
Doctor of Philosophy
of the Indian Statistical Institute
February 1990

To my Late father

ACKNOWLEDGEMENT

The author takes this opportunity to express his profound respect, deep sense of gratitude, and indebtedness to Dr. Sankar K. Pal , professor, Electronics and communication sciences unit, Indian Statistical Institute, Calcutta, not only for his valuable guidance but also for his constant inspiration and encouragement to complete the work.

Invaluable discussions with Prof. D. Dutta Majumdar, Indian Statistical Institute, and his important comments helped significantly to complete the research work, and are gratefully acknowledeged. Fruitful discussions with Prof. A. C. Mukhopadhyay, Dr. B. Roy , Sri S. C. Nandy and Mr. B. Mitra, computer science unit, are thankfully acknowledged. Constant inspiration, co-operation and encouragement provided by Prof. S. K. Pal, Prof. S. C. Kundu and Sri R. K. Naha, Computer science unit, are gratefully acknowledged.

Heartfelt thanks are due to the author's friends and colleagues, especially Mr. A. Ghosh, Mr. D. Bhandari, Mr. R. Das and Mr. A. Das for their sincere co-operation and help to complete the thesis.

The author's thanks are also due to the Computer science unit, and Electronics and communication sciences unit for allowing the author to use their laboratories. Acknowledgement is also due to the members of the reprography unit for doing the photographic work.

At last, but not the least, the author expresses his gratefulness to his mother, brother and sister in law for their hearty inspiration and help in various respects.

CONTENTS Page

LIST OF FIGURES

LIST OF TABLES

Chapter - 1

INTRODUCTION AND SCOPE OF THE WORK

1.1 INTRODUCTION

The field of image processing deals with the manipulation of data which are inherently two-dimensional in nature. The techniques of image processing stem from two principal application areas, namely, improvement of pictorial information for human interpretation and processing of scene data for automatic machine perception. These areas together have experienced a vigorous growth in recent years because they have offered a number of important applications in solving scientific and engineering problems. In biological and medical sciences, we are interested in automatic analysis and interpretation of radiographs, cell images and tissue micrographs. In metallurgical, geological and environmental sciences, we are concerned with the automatic analysis of fractographic pictures, electron-microscopic images, and particle morphology. In law enforcement, we need machines which can interpret fingerprints, handwriting and speech spectrograms. In earth and space sciences, we are interested in automatic interpretation of satellite pictures and aerophotos. In computer and information sciences, we need algorithms for automatic recognition and reading of alphanumeric characters, drawing, geometrical configurations and pictures. In industrial application, we are interested in robots for product assembly, testing of semifinished products and quality control by automated

inspection of pictures generated by infrared , visible radiation, X-rays or ultrasounds.

Let us now give a brief introduction of the various image processing operations.

1.2 IMAGE PROCESSING PRELIMINARIES

An image is a two dimensional light intensity function $f'(x,y)$, where (x,y) denotes the spatial co-ordinate and $f'(x,y)$ denotes the brightness (intensity) value at (x,y). A digital image, on the other hand, is a two dimensional discrete function $f(x,y)$ which has been digitized both in spatial co-ordinates and brightness. The brightness value $f(x,y)$ is called the gray value (gray level) at (x,y). Thus a digital image can be viewed as a two dimensional matrix whose row and column indices identify a point, called a pixel, in the image and the corresponding matrix element value identifies the intensity level, called the gray level at that point. Throughout this report a digital image will be represented by

$$F = [f(x,y)]_{PxQ} \qquad (1.1)$$

where PxQ is the size of the image and $f(x,y) \in G_L = \{0,1,..,L-1\}$, the set of gray levels.

Image processing or picture processing refers to operations that transform a digital image into other digital images, which are, in some sense, improved versions of the input image, while image analysis/vision deals with shapes, positions, color, and

sizes of various objects in the scene. The task of processing includes digitization, compression, enhancement, segmentation, restoration and construction. It also includes description of the scene in terms of relationships among the parts which involve matching, representation and description of the image.

Digitization is a process which converts the continuous brightness and spatial co-ordinates into discrete components. The quality of the processed image, obviously, depends on the resolution of digitization both in gray levels and spatial co-ordinates. The higher the resolution in either scale, the better is the quality of the processed image. However, nothing can be gained by increasing the resolutions beyond the spatial and gray scale resolution capabilities of the receiver. On the other hand, the objective of compression is to represent the image by as few bits as possible for the purpose of reducing the transmission and / or storage overhead. A good coverage of the various aspects of digitization and compression can be found in [1,3].

Often, the available image may not be of desired quality for some specific purpose. The objective of enhancement techniques is to transform the degraded image in such a way that the resultant image is more suitable for that specific purpose. The frequency domain methods of enhancement are based on the modification of the Fourior transform of the image, while the spatial domain methods work directly on the image plane. A large number of methods are

available in the literature [1-11]. The objective of restoration [1,2,12-20] is the same as that of the enhancement. In this case either the degradation function is known or can be estimated. Here we try to model the degradation process and given a degraded image we try to make as good an estimate as possible of the original picture. Normally, a restoration process will require some form of knowledge about the degradation function.

In many occasions we need to match two pictures with one another or to search for a given pattern within a scene, e.g, searching a target in the terrian. These involve problem of matching [1,21-27].

Segmentation is a process of partitioning the image into nonoverlapping meaningful regions, depending on the problem being considered. This is an essential step in image analysis/vision. The success of an industrial inspection system will depend on the quality of segmentation. A large variety of methods based on various approaches is available [1,3,28-107] in the literature.

Once the meaningful regions of the image have been extracted, the relationship among various components can be established. There are different types of picture models namely, declarative , procedural etc., which are used in image description. Once the regions of the image have been segmented we try to characterise the image by a set of region descriptors which are resonably

insensitive to changes in size, rotation, translation . This will, not only reduce the raw data, but also, bring out features which will help to differentiate between regions with different attributes. In the next step we try to organise these regions into a meaningful relational structure. Details of several methods, in this regard, may be obtained in [1,2,137-149].

The present thesis confines itself in providing a set of new image information measures and in formulating a number of image segmentation algorithms for object extraction. The information measures involve the co-occurrence matrix of an image and the concept of entropy. Before presenting the scope of the thesis in detail, we shall present a review of the various image segmentation techniques along with the entropy definitions as a measure of information.

1.3 REVIEW ON IMAGE SEGMENTATION

Segmentation is the first essential step of low level vision. For example, in a vision guided car-assembly system, the robot needs to pick up the appropriate components from the bin. For this, segmentation followed by recognition is required. Segmentation is a process of partitioning the image into some non-intersecting regions such that each region is homogeneous and union of no two adjacent regions is homogeneous. Formally, it can be defined [28] as follows. If F is the set of all pixels and P() is a uniformity (homogeneity) predicate defined on groups of

connected pixels, then segmentation is a partitioning of the set F into a set of connected subsets or regions (S_1, S_2,..,S_n) such that

$$\bigcup_{i=1}^{n} S_i = F \quad \text{with } S_i \cap S_j = \phi \ , \ i \neq j \ . \qquad (1.2)$$

The uniformity predicate $P(S_i)$ = true for all regions (S_i) and $P(S_i \cup S_j)$ = false, when S_i is adjacent to S_j.

There is no single method which can be considered good for all images, nor all methods are equally good for a particular type of image. There are various ways in which one can attempt to extract the segments. Broadly, there are two approaches namely, classical approach and fuzzy mathematical approach. Under the classical approach we have segmentation based on histogram thresholding, edge detection, relaxation, and semantic and syntactic approach. In addition to these,there are certain other methods which do not fall clearly to any one of the above classes. Similarly, fuzzy mathematical approach also has methods based on edge detection, thresholding and relaxation. These techniques are described below.

1.3.1 Gray Level Thresholding

Histogram thresholding is one of the old, simple and popular techniques for image segmentation, when the image is composed of regions with different gray level ranges. If these regions are distinct, the histogram of the image usually shows different

peaks, each corresponding to one region and adjacent peaks are likely to be separated by a valley. For example, if the image has a distinct object on the background, the gray level histogram is likely to be bimodal with a deep valley. In this case, the bottom of the valley (T) is taken as the threshold for object background separation. In other words, gray levels less than or equal to T are taken as parts of the object (background) and others are considered to be its background (object).

Therefore, when the histogram has a deep (or a set of) valley(s), selection of threshold(s) becomes easy because , it becomes a problem of detecting a valley. However, normally the situation is not like this and threshold selection is not a trivial job. There are various methods [1,2,28-107] available for this. For example, Otsu [36] maximised a measure of class separability. He maximised the ratio of the between class variance to the total variance to obtain thresholds. Nagawa and Rosenfeld [76] assumed that the object and background populations are distributed normally with distinct means and standard deviations. Under this assumption they selected the threshold by minimising the total misclassification error. This method is computationally involved. Kittler and Illinworth [55], under the same assumption of normal mixture, suggested a computationally less involved method. They proposed a method which optimises a criterion function related to average pixel classification rate that finds out an approximate minimum error threshold.

Pun [59] assumed that an image is the outcome of an L-symbol source. He maximised an upper bound of the total a-posteriori entropy of the partitioned image for the purpose of selecting the threshold. Kapur et. al. [33], on the other hand, assumed two probability distributions, one for the object area and the other for the background area. They then, maximised the total entropy of the partitioned image in order to arrive at the threshold level.

All these methods have certain drawbacks, because they take into account only the histogram information (ignoring the spatial details). As a result, the algorithm may fail to detect thresholds if these are not properly reflected in the histogram by valley, which is normally the case.

There are many thresholding schemes which use spatial information, insteads of histogram information. For example, the busyness measure of Weszka and Rosenfeld [73] is dependent on the co-occurrence of the adjacent pixels in an image. They minimised the busyness measure in order to arrive at the threshold for segmentation. Deravi and Pal [74] minimised the conditional probability of transition across the boundary between two regions. This method also uses the local information contained in the co-occurrence matrix of the image. These methods also threshold the histogram, but since they make use of the spatial details, they result in more meaningful segmentation than the methods which use only histogram information. Based on the co-occurrence matrix, Chanda et. al. [75] have given an average-contrast measure

for segmentation. Haralick et. al. [131] have also defined several co-occurrence based measures for texture classification.

The philosophy behind gray level thresholding, "pixels with gray level $\leq$ T fall into one region and the remaining pixels belong to another region ", may not be true on many occasions. This leads one to go for the methods of pixel classification, instead of gray level thresholding.

It may be mentioned here that the algorithms proposed in this thesis are based on gray level thresholding. Methods of Pun [59] , Kapur et. al. [33], Wezska and Rosenfeld [73], Deravi and Pal [74], Chanda et. al. [75], and a Bayes classifier [3] have been explained in detail in the respective chapters while comparing their performence with those of the proposed methods.

1.3.2 Relaxation

Relaxation [1] is an iterative approach to segmentation in which 'classification' decision about each pixel can be taken in parallel. Decisions made at neighbouring points in the current iteration are then combined to make decision in the next iteration.

There are two types of relaxation : probabilistic and fuzzy. Let us discuss here the probabilistic relaxation. Suppose a set of pixels $\{f_1, f_2, \ldots, f_n\}$ are to be classified into m classes $\{C_1, C_2,$

..,C_m). We assume that for each pair of class assignments $f_i \in C_j$ and $f_h \in C_k$, there exists a quantitative measure of compatibility C(i,j;h,k) of this pair, i.e., the class assignments of pixels are interdependent. It is reasonable to assume that a positive value of C(i,j;h,k) indicates the compatibility of $f_i \in C_j$ and $f_h \in C_k$, while a negative value represents incompatibility and a zero "don't care" situation. The function C need not be symmetric.

Let p_{ij} represent the probability that $f_i \in C_j$, $1 \le i \le n$ and $1 \le j \le m$, with $0 \le p_{ij} \le 1$, $\sum_j p_{ij} = 1$. Intuitively, if p_{hk} is high and C(i,j;h,k) is positive, We should try to increase p_{ij} since it is compatible with the high-probability event $f_h \in C_k$. Similarly, if p_{hk} is high and C(i,j;h,k) is negative, we try to reduce p_{ij} as it is incompatible with $f_h \in C_k$. On the other hand, if p_{hk} is low or C(i,j;h,k) is nearly zero, we do not like to change p_{ij} as either $f_h \in C_k$ has a low probability or is irrelevent to $f_i \in C_j$. Keeping the above analysis in mind, we need a method of deriving p_{ij}^{r+1} from p_{ij}^{r}, given an estimate of p_{ij}^{0} ; p_{ij}^{r} is the probability of $f_i \in C_j$ at the *rth* iteration. One of the methods for combining results of the *rth* step to get the estimate of (r+1)*th* step can be described as follows.

$$p_{ij}^{r+1} = \frac{p_{ij}^{r} \cdot (1+q_{ij}^{r})}{\sum_{j=1}^{m} p_{ij}^{r} (1+q_{ij}^{r})} \qquad (1.3)$$

where $$q_{ij}^{r} = \frac{1}{n-1} \sum_{\substack{h=1 \\ h \neq i}}^{n} \left(\sum_{k=1}^{m} (C(i,j;h,k)\ p_{hk}^{r}) \right) \qquad (1.4)$$

Various types of compatibility functions can be used depending on the problem at hand. However, a general method of defining C can be done in terms of the a-priori and conditional probabilities of $f_i \in C_j$ and $f_h \in C_k$ as follows.

$$C(i,j;h,k) = log(R(i,j;h,k)) \qquad (1.5)$$

where $R(i,j;h,k) = p(f_i \in C_j \,/\, f_h \in C_k) \,/\, p_i(f_i \in C_j)$

Now, if $f_h \in C_k$ is compatible to $f_i \in C_j$ then $p(f_i \in C_j \,/\, f_h \in C_k)$ would be greater than $p(f_i \in C_j)$ making R greater than 1. On the other hand, when they are incompatible R will be less than 1; and R becomes 1 when $f_i \in C_j$, is independent of $f_h \in C_k$. Therefore, $log(R(i,j;h,k))$ satisfies the desired properties of a compatibility function.

1.3.3 Surface Based Segmentation

Besl and Jain [49] have developed an imge segmentation algorithm based on the assumption that the image data exhibits surface coherence, i.e., image data may be interpreted as noisy samples from a piecewise-smooth surface function. Though, this method is probably most useful for range images (in a range image, each pixel represents distance to the physical surface from a known reference surface), the method can be used to segment any type of image that can be modelled as a noisy sampled version of a piecewise-smooth graph surface. Their method is based on the concept that any arbitrary smooth surface can be decomposed into

one of the eight possible surface types : peak, pit, ridge, valley, saddle valley, flat (planar) and minimal. These simple surfaces can be well approximated for the purpose of segmentation by bivariate polynomials of order ≤ 4.

The first stage of the algorithm creates a surface type label image based on the local information (using mean curvature and Gaussian curvature images). The second stage takes the original image and the surface type image as input and performs iterative region growing using the variable order surface fitting. In the variable order surface fitting , first they try to represent the point in a seed region by a planar surface. If this simple hypothesis of planar surface is found to be true then the seed region is grown on the planar surface fit. If the simple hypothesis fails, then the next more complicated hypothesis of biquadratic surface fit is tried. If this is satisfied, the region is grown based on that form otherwise, the next complicated form is tried. The process is terminated when either the region growing has converged (same region obtained twice) or when all preselected hypotheses fail. In the later case, possibly a higher order surface should be tried.

1.3.4 Clustering

Hoffman and Jain [51] have developed a method of segmentation and classification of range images. They have used a clustering algorithm to segment the image into surface patches. Different

types of clustering algorithms including methods based on minimal spanning tree, mutual nearest neighbour, hierarchical clustering and square error clustering have been attempted. The square error clustering has been found to be the most successful method for range images. The feature set used contains the co-ordinate position (x,y), the depth value $f(x,y)$ and the estimated unit surface normal vector. The unit surface normal vector is normal to the tangent plane at a point which is obtained by finding the best (in the least square sense) fitting plane over a 5x5 neighbourhood. In the second phase of the method this patches are classified as planar, convex or cocave. In order to make the method of classification more effictive they have combined three different methods, namely 'non-parametric trend test for planarity', 'curvature planarity test' and the eigen value planarity test.

In the final stage, boundaries between adjacent surface patches are classified as crease or noncrease edge, and this information is then used to merge adjacent compatible patches to result in reasonable faces of the object.

1.3.5 **Gibbsian Model**

There are many iamge segmentation methods [64-67] that use the spatial interaction model like Markov Random Field [MRF] or Gibbs Random Field [GRF] to model images. Here the scene is assumed to be a GRF and the observed digital image is a noise

(normally independent and identically distributed additive noise is assumed) corrupted realization of the scene. Various statistical criteria are then used to estimate the scene random field from the observed image.

1.3.6 Edge Detection

Segmentation can also be obtained through detection of edges of various regions, which normally tries to locate the points of abrupt changes in gray level intensity values. Since edges are local feature, they are determined based on local information. A large variety of methods are available in the literature [1-3,86-100] for edge finding. There are different types of differential operators such as Robert gradient, Sobel gradient, Prewitt gradient and the Laplacian operator [1-2]. These difference operators respond to changes in gray level or average gray level. The magnitude of the edge value at (x,y) under different gradient operators can be defined as

$$e(x,y) = \{|G_1|+|G_2|\} \qquad (1.6)$$

$$\text{or} \qquad \sqrt{G_1 + G_2} \qquad (1.7)$$

where G_1 and G_2 under different gradient operators are defined as follows.

For the Robert's operator G_i's are defined as

$$G_1 = f(x+1,y+1) - f(x,y) \qquad (1.8)$$

$$G_2 = f(x,y+1) - f(x+1,y) \qquad (1.9)$$

This gradient operator, not only responds to edges but also to isolated points. Prewitt's and Sobel's operators are obtained as follows.

$$G_1 = \frac{1}{2+w}[\ (f(x+1,y+1) + w.f(x+1,y) + f(x+1,y-1)) - (f(x-1,y+1)+w.f(x-1,y)+f(x-1,y-1))] \quad (1.10)$$

and

$$G_2 = \frac{1}{2+w}[\ (f(x-1,y+1) + w.f(x,y+1) + f(x+1,y+1)) - (f(x+1,y-1)+w.f(x,y-1)+f(x-1,y-1))] \quad (1.11)$$

In the equations 1.10 and 1.11 when w = 1 we get Prewitt's operator and with w = 2 we get Sobel's operator.

For Prewitt's operator the response to the diagonal edge is weak, while for Sobel's operator the response to diagonal edge is not that weak as it gives a greater weight to points lying close to (x,y). However, both Prewitt's and Sobel's operators possess greater noise immunity.

The operators considered so far are called the first difference operator. Laplacian, on the other hand, is a second difference operator. The Laplacian operator

$$\frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \quad (1.12)$$

has the digital equivalent as

$$\nabla^2 f(x,y) = f(x+1,y)+f(x-1,y)+f(x,y+1)+f(x,y-1)-4\ f(x,y) \quad (1.13)$$

The digital Laplacian being a second difference operator, has a zero response to linear ramps. It responds strongly to corners, lines, and isolated points.

Thus for a noisy picture, unless it has a low contrast, the noise will produce higher Laplacian values than the edges. Moreover, the digital Laplacian is not orientation invariant.

1.3.7 Laplacian of a Gaussian

A good edge detector, should be a filter with the following two features. First, it should be a differential operator, taking either a first or second spatial derivative of the image. Second, it should be capable of being tuned to act at any desired scale, so that a large filters can be used to detect blurry shadow edges, and small ones to detect sharply focussed fine details. The second requirement is very useful as intensity changes occur at different scales in an image.

According to Marr and Hildreth [94] the most satisfactory operator fulfilling these conditions is the Laplacian of Gaussian (L-G) operator. It is normally denoted by $\nabla^2 G$; where the Laplacian is given by

$$\nabla^2 = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2} \qquad (1.14)$$

and

$$G = e^{-\frac{x^2+y^2}{2\pi\sigma^2}} \qquad (1.15)$$

is a two dimensional Gaussian distribution, with standard deviation σ^2. The Gaussian part of the L-G operator blurs the image, wiping out all structures at scales much smaller than σ of the Gaussian.

The Gaussian blurring function is preferred over others because it has the desirable property of being smooth and localised in both the spatial and the frequency domain. In order to find the intensity change at a given scale, Marr and Hildreth, first filtered the image with the $\nabla^2 G$ filter and then found the zero crossings in the filtered image.

The space described by the scale parameter σ and the zero-crossing curves is called the scale-space. The behaviour of edges in the scale-space produced by the L-G operator has been recently studied by Lu and Jain [97]. In order to formulate rules for reasoning in the scale space they studied dislocation of edges, false edges, and merging of edges with nice mathematical frames.

1.3.8 An Average Edge Detector

In the case of a noise free image, the edge angle can be measured accurately, but in real life images, noise can not be avoided and it makes difficult to estimate the true edge angles. Kittler et. al. [87] suggested three methods to improve the edge angle estimate obtained from Sobel's operator.

All the three methods involve averaging of the outputs of the Sobel's operator over a 3x3 window. One of the methods, which ignores the effect of the central pixel, at which the angle estimate is wished, is found to produce best result. They have justified this counter intuitive view also.

1.3.9 Facet Model

Haralick [96] attacked the problem of edge and region detection from a new angle. He assumed that the observed image is an ideal image with noise added. Each region in the image is a sloped plane. In order to determine the edge between two pixels, best fitted sloped planes over a neighbourhood of each pixel are found. Edges are declared at locations having significant different planes on either side of them. The least square procedure has been used to estimate the parameters of a sloped surface for a given neighbourhood. Appropriate F-statistic has been used to test the significance of the difference of the estimated slope from a zero slope or the significance of the difference of the estimated slopes of adjacent neighbours.

1.3.10 Edge Thresholding

Various operators have been discussed so far to get edge values. All the edges produced by those operators are, normally, not significant edges when viwed by human beings. Therefore, one needs to find out prominent (valid) edges from the output of the edge operators. Kundu and Pal [95] have suggested a method of

thresholding to extract the prominent edges based on the psychovisual phenomena. Haddon [99] developed a technique to derive a threshold for any edge operator , based on the noise statistics of the image.

1.3.11 Methods Based on Fuzzy Set Theory

Zadeh (1965) introduced the concept of fuzzy sets in which imprecise knowledge can be used to define an event. A fuzzy set A is represented as

$$A = \{x_i \ / \ \mu_A(x_i),\ i=1,2,\ldots,n\} \qquad (1.16)$$

where $\mu_A(x_i)$ gives the degree of belongingness of the element x_i to the set A.

In the case of an ordinary set an element either belongs to or does not belong to a set. Thus an ordinary set A is defined by the characteristic function $C_A : R^n \dashrightarrow \{0,1\}$. On the other hand, a fuzzy set A is defined by a characteristic function

$$\mu_A : R^n \dashrightarrow [0,1].$$

The function μ_A associates with each $x \in R^n$ a grade of membership to the set A and is known as the membership function . Thus a fuzzy set (A) in a space of points X={x} is a class of events with a continuum of grades of membership and is characterised by a membership function $\mu_A(x)$ which associates with each point in X a real number in the interval [0 1] with the value of $\mu_A(x)$ at x representing the grade of memnbership of x in A.

Similarly, a property P defined on an event x_i is a function $P(x_i)$ which can have values only in the interval [0,1]. A set of these functions which assigns the degree of possessing some property P by the event x_i constitutes what is called a property set.

This theory has also been extensively used to solve various image processing problems [100-107]. We shall consider here a few methods of segmentation and edge detection based on this concept.

1.3.11.1 Fuzzy Edge Detection

Pal and King [100] used a nonsymmetrical membership function G_s to get the fuzzy property plane from the intensity plane. The G_s is defined as

$$G(F(x,y)) = (1+|\hat{f} - f(x,y)|/ F_d)^{-F_e} \qquad (1.17)$$

where $\hat{f}$ is a reference level, F_e and F_d are the exponential and denominational fuzzifiers respectively. If $\hat{f} = f_{max}$ the maximum gray level, G approximates the standard S function [105] of Zadeh and when $\hat{f}$ = some other level, $0 < \hat{f} < f_{max}$, it approximates the standard π function [105] of Zadeh. The G function under the above cases will be denoted by G_s and G_π respectively. They used this G_s and G_π function in conjunction with an intensification operator INT to intensify the contrast in the image. Finally, an inverse transformation is also applied to get the spatial domain enhanced

image. Edges of this enhanced image can, then be easily found with any spatial domain techniques. Details of the method can be obtained in [102].

1.3.11.2 Fuzzy Thresholding

Pal et. al. [101] have also used ambiguity measures like, Index of fuzziness and entropy of a fuzzy set [102,116] to detect the valley in a histogram for the purpose of gray level thresholding. Their method is based on the fact that the ambiguity in a fuzzy set is maximum when all the supporting points have membership value 0.5 and ambiguity decreases as more and more supporting points move away from 0.5. Recently, Pal and Rosenfeld [106] used the concept of fuzzy geometry in detecting thresholds by optimising "compactness" of an image which takes into account the spatial information together with gray level information.

1.3.11.3 Fuzzy Clustering

The fuzzy ISODATA or fuzzy C-means clustering algorithm [108] has also been used to image segmentation [107] . The fuzzy C-means algorithm uses an iterative optimisation of an objective function based on a weighted similarity measure between the pixels in the image and each of C cluster centers. Local extrema of this objective function indicate an optimal clustering of the input data. The objective function that is minimised is given by

$$W_m (U,V) = \sum_{k=1}^{n} \sum_{i=1}^{c} (\mu_{ik})^m (d_{ik})^2 \qquad (1.18)$$

where μ_{ik} is the fuzzy membership value of the k *th* pixel in the i-th cluster ; d_{ik} is any inner product induced norm metric; m controls the nature of clustering with hard clustering at m=1 and increasingly fuzzier clustering at higher values of m; V is the set of C-cluster centers and U is the fuzzy C partition of the image.

Backer [118] developed a very general clustering strategy which has been applied to different types of data including images. The set of samples C is first partitioned into m (number of clases) disjoint sets as an initial guess of the desired partition. Then a membership function is assigned to those initialised clusters according to some "point to point subset affinity" mechanism for all points in C. In fact, he suggested a number of affinity mechanisms based on the distance concept, the neighbourhood concept, and the probabilistic concept. Updating of the partitions (repartition, reclassification) is then done under the guidance of some criterion function which characterises the partition. He considered three different types of criterion functions, based on measures of fuzziness, inter fuzzy set distance, and measure of fuzzy similarity. If changes occur in the earlier step, the process of assigning membership function and updation is repeated otherwise, the algorithm terminates.

The above mentioned discussions made so far are related only to the segmentation techniques. Before discussing the scope of the thesis, let us now explain the various entropy definitions used as a measure in image processing problems.

1.4 Entropy Definitions

The term entropy was first used by Clausius in 1864. His concept was a nonprobabilistic one. In statistical thermodynamics L. Boltzmann(1896) defined entropy as a probabilistic function of positions and velocities of all particles included in a physical system. Though Boltzmann made no explicit reference to information, his entropy gives a measure of disorder in a physical system. This measure of disorder also indicates the uncertainty concerning the states of the individual particles. Therefore, Boltzmann may be considered the pioneer of information theory. By anlogy with Boltzmann's expression for the probabilistic entropy, Shannon [109,110] in 1948, introduced entropy as a measure of information of a probabilistic experiment.

The entropy of a system as defined by Shannon[109,110] gives a measure of our ignorance about its actual structure. In the context of information theory, Shannon's function is based on the concept that information gain from an event is inversely related to its probability of occurrence. The logarithmic behavior of entropy is considered to incorporate the additive property of gain in information. But, there are some other probabilistic definitions of entropy [111-115] using logarithmic function, which do not satisfy such an additive property.

There is another type of entropy measure [116-125] defined on a fuzzy set. This reflects the amount of difficulty (ambiguity) in

deciding whether an element can be considered a member of the set or not.

1.4.1 Classical Entropy

Shannon's Definition [109-110]

Shannon defined the entropy of an n-state system as

$$H = -\sum_{i=1}^{n} p_i \log (p_i) \qquad (1.19)$$

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

$$\sum_{i=1}^{n} p_i = 1 \quad , \; 0 \leq p_i \leq 1.$$

In case of a binary system, the entropy becomes

$$H = - p \log (p) - (1-p) \log(1-p) \qquad (1.20)$$

The entropy H is claimed to express a measure of ignorance about the actual structure of the system.

Other Definitions

Though Shannon's definition of entropy is the oldest and most widely used definition of entropy, there are other definitions too, of which some are probabilistic and some are non probabilistic. We shall, consider here some probabilistic definitions only.

Renyi [111,115] extended Shannon's entropy to an incomplete probability distribution by defining the entropy of order α as

$$H_n = \frac{1}{(1-\alpha)} \log \frac{\sum_{k=1}^{n} p_k^{\alpha}}{\sum_{k=1}^{n} p_k} \qquad (1.21)$$

where $\alpha > 0$, $\alpha \neq 1$, $\sum_{k=1}^{n} p_k \leq 1$.

Verma [112,115] proposed two expressions, namely

$$H_n^{A\,\alpha\,m} = \frac{1}{m-\alpha} \log \frac{\sum_{k=1}^{n} p_k^{\alpha-m+1}}{\sum_{k=1}^{n} p_k} \qquad (1.22)$$

where $m-1 < \alpha < m$, $m \geq 1$ and $\sum p_k \leq 1$ and

$$H_n^{B\,\alpha\,m} = \frac{1}{m(m-\alpha)} \log \frac{\sum_{k=1}^{n} p_k^{\alpha/m}}{\sum_{k=1}^{n} p_k} \qquad (1.23)$$

with $0 < \alpha < m$, $m \geq 1$, $\sum p_k \leq 1$.

Note that both of these expressions give Renyi's α - entropy for $m = 1$. The definition introduced by Kapur[113,115] is also a logarithmic function of some combination of probabilities. He

defined entropy of order α and type β as

$$H_n^{\alpha\ \beta} = \frac{1}{1-\alpha}\ log\ \frac{\sum_{k=1}^{n} p_k^{\alpha+\beta-1}}{\sum_{k=1}^{n} p_k^{\beta}} \qquad (1.24)$$

where $\alpha > 0$, $\alpha \neq 1$, $\beta \geq 1$ and $\sum p_k \leq 1$.

A more general type of definition of entropy of order α and type β_k , $1 \leq \beta_k \leq n$ is introduced by Rathie [114,115] as

$$H_n^{\alpha\ \beta_1 \ldots \beta_k} = \frac{1}{1-\alpha}\ log\ \frac{\sum_{k=1}^{n} p_k^{\alpha+\beta_k-1}}{\sum_{k=1}^{n} p_k^{\beta_k}} \qquad (1.25)$$

with $\alpha > 0$, $\alpha \neq 1$, $\beta_k > 1$ and $\sum p_k \leq 1$.

Equation (1.25) reduces to equation (1.24) when $\beta_k = \beta$ for $k=1,\ldots,n$. It is to be noticed that all the above mentioned measures are logarithmic functions of the probabilities. It is interesting to note that all these definitions involve logarithm of some combination of probabilities.

Weighted Entropy

In all the definitions discussed so far, every state of the system has been given equal importance to arrive at the expression for entropy. But there are many areas dealing with random events in which consideration of only probabilities is not enough for entropy. It is necessary to take into account the probabilities and some qualitative characteristics of the event. In order to

account for the subjective aspects related to the goal set by the experimenter it is necessary to attach a qualitative weight to every elementary event. Thus the gain in the information from an event with probability p_i can be defined as

$$\Delta I(p_i , w_i) = - w_i \log (p_i) \qquad (1.26)$$

and the weighted entropy as

$$H^{W} = \sum_{i=1}^{n} p_i \Delta I(p_i , w_i) \qquad (1.27)$$

1.4.2 Entropy of Fuzzy Sets

In the areas of pattern recognition and image processing, it is often required to get some idea about the degree of ambiguity (fuzziness) present in a fuzzy set. A measure of fuzziness is expected to give the average amount of difficulty that is involved in taking a decision whether an element belongs to the set or not [116-125]. It is to be noted that this is not a property of the individual element of the set but, a property of the set as a whole. Therefore, if every element of the set has a membership value of 0.5, then the fuzziness in the set should be maximum. There have been different definitions of entropy for a fuzzy set to provide measures of fuzziness. Here we shall consider only definition provided by Deluca and Termini [120] as that is a straight forword application of Shannon's function. Instead of the probability, the membership function has been used here to give a measure of fuzziness(ambiguity) present in a set. They defined the entropy of a fuzzy set **A** as

$$H(A) = k \sum_{i=1}^{n} S_n(\mu_A(x_i)) \qquad (1.28)$$

where S_n is the Shannon's function having the form

$$S_n(x) = -x \log(x) - (1-x) \log(1-x) \qquad (1.29)$$

and k is a normalising constant.

The entropy H(A) has the following properties.

P1 : H(A) is minimum if and only if $\mu_A(x_i) = 0$ or 1 for all i.

P2 : H(A) is maximum if and only if $\mu_A(x_i) = 0.5$ for all i.

P3 : $H(A) \geq H(A^*)$, where A^* is any sharpened version of A.

A sharpened version of A is defined as

$$\mu_{A^*}(x_i) \geq \mu_A(x_i) \quad \text{if} \quad \mu_A(x_i) \geq 0.5$$

and

$$\mu_{A^*}(x_i) \leq \mu_A(x_i) \quad \text{if} \quad \mu_A(x_i) \leq 0.5.$$

P4 : $H(A) = H(\bar{A})$, where $\bar{A}$ is the complement set of A.

It is very easy to see that with proper choice of k, properties P1 to P4 are satisfied by H(A) of equation 1.28.

H(A) is thus seen to use Shannon's function but its meaning is quite different from the classical entropy (equation 1.19); because no probabilistic concept is needed to define it. H(A) provides the degree of fuzziness which expresses on a global level, the average amount of difficulty (or ambiguity) in deciding whether an element would be considered to be a member of A or not. In the subsequent discussion, the term "entropy of a fuzzy set"

has often been referred as "fuzzy entropy". We shall now explain the scope of the work presented in the thesis.

1.5 SCOPE OF THE WORK

The thesis introduces a number of measures of image information and demonstrates their applications to image segmentation problem. Some of the measures directly contribute towards information theory and hence are applicable to any information source including images while, the remaining measures are developed exclusively for images. Justification for developing each measure has been provided. The performance of the proposed methods has been compared with those of the several related methods.

1.5.1 Measures of Homogeneity and contrast [29,61,84]

In chapter 2, some of the co-occurrence based methods of segmentation as described in section 1.3.1 have been critically analysed and measures like "contrast" and "homogeneity" of regions have been defined using the co-occurrence matrix of an image. The measure of "contrast" involves the concept of logarithmic response (adaptibility with background intensity) of the human visual system. Based on these measures, two algorithms for image segmentation have been formulated. Provisions are also kept in two different ways to remove the undesirable segments. The incorporation of the psychovisual facts and the region size in the

objective function made the methods superior to many other co-occurrence based algorithms. The performance of the proposed methods is compared with those of some existing ones.

1.5.2 **Higher order and Conditional Entropy** [31,63]

Limitations arising out of the improper interpretation of the conept of entropy in the thresholding algorithms of Pun [59] and Kapur et. al [33] have been discussed in chapter 3. Definitions of the higher order probabilistic entropy (entropy of order q), giving a measure of homogeneity within a region and the conditional entropy, providing a measure of contrast, have then been introduced based on Shannon's theory of information. The existing definition of global entropy is found to be its special case when $q = 1$.

Based on these measures various algorithms for object extraction from an image have been developed. Since the proposed algorithms use the spatial information present in an image they produce improved results as compared to those in [33,59].

1.5.3 **Measures Based on a New Classical Entropy** [30,68,69,85]

Chapter 4 justifies the need for an alternative definition of the classical (probabilistic) entropy and introduces a new definition which uses an exponential gain function instead of the logarithmic function of Shannon. This new definition has been established based on a large number of desirable properties along with their proofs. The proposed definition has also been found

extendable to define the entropy of a fuzzy set. It has then been used to formulate algorithms for image segmentation using the concept of higher order entropy and conditional entropy as discussed in chapter 3.

An attempt has then been made to provide a definition of positional entropy which gives a measure of information about the location of an object in a scene.

1.5.4 Poisson Distribution Based Image Model and Parametric Methods [32]

A Poisson distribution based digital image model derived from the physics of image formation has been introduced in chapter 5 of the thesis. The digital image model leads us to accept Poisson distribution for the gray levels instead of the widely used normal distribution. Under this model, some parametric segmentation algorithms have been formulated based on the maximum entropy principle and the minimum χ^2-statistic. Superiority of the Poisson distribution over the normal distribution has been established considering some global information based parametric algorithms. Results indicate that image information is more appropriately represented by the exponential entropy than by the logarithmic entropy .

1.5.5. Higher order Fuzzy Entropy and Hybrid Entropy [62,126,127]

Chapter 6 discusses the inadequacy of the existing measures of fuzziness to handle various real life problems and establishes

the need for a new concept of "higher order entropy" of a fuzzy set (higher order fuzzy entropy) .The higher order fuzzy entropy introduced here is then used to define an index of similarity of supports in a fuzzy set.

The second part, indicates the drawbacks of various attempts to bridge between the probabilistic and possibilistic uncertainties and introduces the definition of "hybrid entropy" of a set. The hybrid entropy provides a measure of uncertainty arising out of both probability distribution and possibility distribution. When the fuzziness in a set is removed it boils down to the classical entropy of a two state system. Different useful properties of both definitions along with their proofs have also been given.

Finally, the higher order fuzzy entropy has been used to develop algorithms for segmentation and edge detection. The results obtained here establish the superiority of the exponential entropy over the Shannon's entropy , in the case of an image. Possible use of the hybrid entropy as an objective criterion in channel noise removal and in image enhancement is also highlighted.

1.5.6 Conclusion and Further Scope of Work

The concluding remarks with further scope of work are presented in chapter 7.

In each chapter the results of the proposed methods have been compared with some similar existing methods. Sufficient details of

those methods have been discussed in the respective chapters (which are not covered in the review done in section 1.2) for the convenience of the reader. Attempts have also been made to make the chapters stand-alone so that the back and forth references can be avoided.

Chapter 2

MEASURES OF HOMOGENEITY AND CONTRAST

2.1 INTRODUCTION

We have discussed in chapter 1, several techniques of image segmentation based on global and local information of an image. One of the techniques based on global information is histogram thresholding which selects the vally points as threshold levels. For images where the histogram does not have sharp valleys (i.e., having flat valley or local minima) the histogram is usually sharpened [1,70-72,101] by a sutiable transformation so that the task of selecting valleys becomes easier. These transformations usually require some parameters, the choice of which has significant impact in determining the number of thresholds. The co-occurrence matrix, on the other hand, uses local spatial information of an image and provides information regarding the number of transitions between any two gray levels in the image. This information has been used by various authors namely, Weszka and Rosenfeld [73], Deravi and pal [74] and Chanda et. al [75] for segmentation.

The measures on co-occurrence matrix reported by the above mentioned authors did not consider the fact of logarithmic response of human visual system (HVS) [2,95,128-130] in measuring 'Contrast' between regions in an image. The work reported in this

chapter attempts to bring this factor into consideration while defining a measure of 'Contrast' in addition to defining another measure called 'homogeneity' within a region. The combination of these two measures made the algorithms effective in determining threshold levels. The 'Contrast' measure ensures three things. Firstly, it assures the fact that a transition (change in the gray level) from i to $i+l$ and a transition from i to $i+k$, $l \neq k$, create different impression. Secondly, a constant change in gray level at different positions on the gray scale results in different impression and finally, if the object and background intensities are interchanged the contrast value remains unaltered.

Furthermore, provisions are also kept in two different ways for eliminating the undesirable segments. In the first approach, noninformative single gray level regions are eliminated by a separate merging algorithm. The second approach on the other hand, inherently attempts to eliminate such smaller regions irrespective of their intensity-width.

The effectiveness of the algorithms alongwith its comparison with three other methods [73-75] has been demonstrated on a set of images.

Since the proposed methods [29,61,84] and the methods in [73-75] , use the co-occurrence matrix, let us first of all define it before explaining the methods.

2.2. DEFINITION OF CO-OCCURRENCE MATRIX [29]

The co-occurrence matrix (or the transition matrix) of the image F is an LxL dimensional matrix that gives an idea about the transition of intensity between adjacent pixels. In other words, the (i,j)th entry of the matrix gives the number of times the gray level j follows the gray level i (i.e., the gray level j is an adjacent neighbour of the level i) in a specific fashion.

Let a denote the (i,j)th pixel in F, a_8 be the set of eight neighbouring pixels of a, and $b \in a_8$; i.e.,

$$b \in a_8 = \{(i,j-1),\ (i,j+1),\ (i+1,j),\ (i-1,j),\ (i-1,j-1),\ (i-1,j+1),\ (i+1,j-1),\ (i+1,j+1)\}.$$

Define $t_{lk} = \sum_{\substack{a \in F \\ b \in a_8}} \delta$

where $\delta = 1$, if the gray level value of a is l and that of b is k,

$\delta = 0$, otherwise.

Obviously, t_{lk} gives the number of times the gray level k follows gray level l in any one of the eight directions. The matrix $T = [t_{lk}]_{LxL}$ is, therefore, the co-occurrence matrix of the image F.

One may get different definitions of the co-occurrence matrix by considering different subsets of a_8, i.e.,considering $b \in a_8'$, where $a_8' \subseteq a_8$ .The co-occurrence matrices may be either

non-symmetric or symmetric. One of the non-symmetrical forms can be defined considering,

$$t_{lk} = \sum_{i=1}^{P} \sum_{j=1}^{Q} \delta \qquad (2.1)$$

with

$\delta = 1$, if $f(i,j) = l$ and $f(i,j+1) = k$

or $f(i,j) = l$ and $f(i+1,j) = k$;

and $\delta = 0$, otherwise.

On the other hand, the following definition of t_{lk} gives a symmetric co-occurrence matrix,

$$t_{lk} = \sum_{i=1}^{P} \sum_{j=1}^{Q} \delta \qquad (2.2)$$

where $\delta = 1$, if $f(i,j) = l$ and $f(i,j+1) = k$

or

$f(i,j) = l$ and $f(i,j-1) = k$

or

$f(i,j) = l$ and $f(i+1,j) = k$

or

$f(i,j) = l$ and $f(i-1,j) = k$;

$\delta = 0$, otherwise.

2.3 MEASURES FOR THRESHOLDING

Since the co-occurrence matrix contains information regarding the spatial distribution of gray levels in the image, several

workers [73-75] have used them for segmentation. Three of these measures have been considered here in details as they have been used for comparison.

2.3.1 Busyness Measure [73]

For thresholding at gray level s, Weszka and Rosenfeld defined the busyness measure as follows

$$\text{Busy}(s) = \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij} + \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij} \qquad (2.3)$$

The co-occurrence matrix used in equation 2.3 is symmetric (using equation 2.2).

The sum of the entries of the shaded portion in Fig. 2.1, represents the Busyness measure for the level s. For an image with only two regions, say, object and background, the value of s for which the minimum of Busy(s) occurs, gives the threshold. Similarly, for an image having more than two regions the Busyness measure provides a set of minima corresponding to different thresholds.

2.3.2 Conditional Probability Measure [74]

Deravi and Pal have given a measure for the conditional probability of transition from one region to another using a non-symmetric co-occurrence matrix:

$$t_{lk} = \sum_{i=1}^{P} \sum_{j=1}^{Q} \delta \qquad (2.4)$$

with

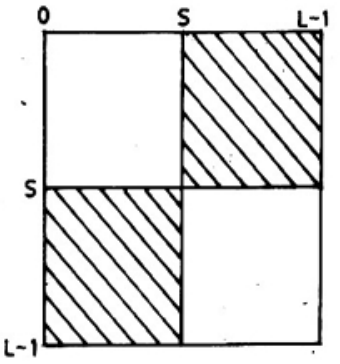


Fig. 2.1 Pictorial representation of Busyness measure

$\delta = 1$, if $f(i,j) = l$, $f(i,j+1) = k$, and $f(i+1,j) = k$;

$\delta = 0$, otherwise.

If the threshold is at s, the conditional probability of the intensity transition from the region $[0,s]$ to $[s+1, L-1]$, i.e., the probability of any intensity level from the class $[0,s]$ being followed by a level from the class $[s+1, L-1]$ in the fashion given by equation 2.4, is

$$P_1 = \frac{\sum_{i=0}^{s} \sum_{J=s+1}^{L-1} t_{ij}}{\sum_{i=0}^{s} \sum_{j=0}^{s} t_{ij} + \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij}} \qquad (2.5)$$

Similarly, the conditional probability of intensity transition from the region $[s+1, L-1]$ to $[0,s]$, is

$$P_2 = \frac{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}}{\sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} t_{ij} + \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}} \qquad (2.6)$$

The conditional probability, $P_c(s)$ of transition across the boundary is defined as

$$P_c(s) = (P_1 + P_2)/2 \qquad (2.7)$$

The lower the value of $P_c(s)$, the lower is the probability of transition across the boundary between two classes $[0,s]$ and $[s+1, L-1]$. That means, a minimum of $P_c(s)$ will correspond to a threshold such that most of the transitions are within the class and few are across the boundary. Therefore, a set of minima of $P_c(s)$ would be obtained corresponding to different thresholds in F.

2.3.3 **Average Contrast Measure** [75]

Chanda et.al have also used the co-occurrence matrix for thresholding. They defined an average contrast measure as

$$AVC(s) = \frac{\sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij} * (i-j)^2}{\sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij}} + \frac{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij} * (i-j)^2}{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}} \qquad (2.8)$$

AVC(s) shows a set of maxima corresponding to the thresholds among various regions in F. In the computation of t_{ij} they considered only vertical transitions in the downward direction.

Note here that all of the three measures described before are basically based on some weighted combinations of the number of entries in the shaded and blank regions of Fig. 2.1.

2.4 PROPOSED MEASURES OF CONTRAST AND HOMOGENEITY [29,31,84]

In this section, we are presenting an algorithm for segmentation on the basis of the information of contrast and homogeneity (between/within the regions in F) as obtained from the co-occurrence matrix (equation 2.1). The concept of the Human visual system has been incorporated in the contrast measure to make the method of segmentation closer to the way, a human being makes use of the intensity distribution for segmentation (of course, human being uses some higher level knowledges too). Such a segmented image when used for enhancement will result in a better enhanced image as the method of segmentation tries to simulate the way, a human being perceives the intensity variation.

Before describing the algorithm let us first explain some facts of the human visual system.

2.4.1 Human Psychovisual Facts

In psychophysiology, contrast C refers to the ratio of difference in luminance of an object B_o and its immediate surrounding B [2] i.e.,

$$C = |B_o - B| \; / \; B \; = \frac{\Delta B}{B} \qquad (2.9)$$

The perceived grayness of a surface depends on its local background and the perceived contrast remains constant if the measure C of the contrast between object and local background remains constant.

The visual increment threshold (or the just noticeable difference) is defined as the amount of light ΔB_T necessary to add to a visual field of intensity B such that it can be discriminated from a reference field of same intensity B. It therefore, gives a limit for a perceivable change in luminance or intensity.

At low intensity near absolute visual threshold (mere presence or absence of light intensity detectable under dark adapted condition), the visual increment threshold ΔB_T is constant. With increasing B, ΔB_T can be approximated by a linear function of B, i.e., $\Delta B_T = \alpha.B$ (the factor α is called the weber ratio). When this linear relationship is valid one speakes of **'Weber behaviour'**.

Fig. 2.2 presents such a characteristic response in the log ΔB_T - log B plane. The Weber behaviour is characterised by unit slope of the curve. The preceding region with slope 1/2 is known as the De Vries - Rose region characterised by $\Delta B_T = \alpha_1\sqrt{B}$, which is a valid approximation, only followed in a small restricted region. Under saturation, the visual increment threshold ΔB_T does not follow the Weber behaviour and this deviation from weber behaviour is shown by the dashed line [128].

Therefore, if the brightness value of an object is higher (lower) than its surrounding or background or a reference intensity B by such an amount that it corresponds to a point on or above the curve (Fig. 2.2), the object will only then appear brighter (darker) i.e., discriminable to the human visual system

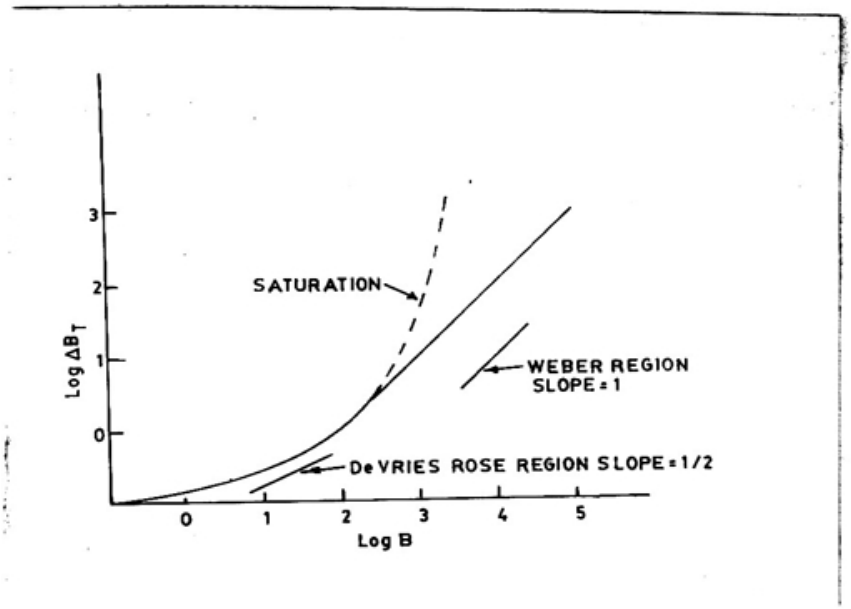


Fig. 2.2 Variation of log ΔB_T with log B

(HVS). Furthermore, an equal amount of ΔB value created at different background intensity (B-value) does not result in equal perceivable change to HVS. For example, the discrimination ability in the De Vries - Rose region is greater than that in the Weber region, and this ability decreases with increase in the value of B. The possible reason for this deterioration in discrimination ability can be attributed to nonlinearities inherent in the visual system.

2.4.2 **Measures of Contrast and Homogeneity** [29,61,84]

It has already been discussed that the problem of segmentation is to partition the set G_L of gray levels into some non-intersecting subsets such that each segment is as homogeneous as possible while the contrast between any segment and its neighbouring segments is as high as possible. Two such measures, namely the contrast of a segment with its neighbouring segments and the homogeneity of a segment are defined in the following. A composite measure of the two is then used to select the threshold levels in the image F.

Let the gray levels ranging from K to M form one of the segments, say, R_1 of the image F, i.e., $R_1 = [K,M]$, $K \leq M$.

Define $C^B_{K,M}$, the contrast of the segment R_1 with respect to other segments as folows:

$$C^{B}_{K,M} = \frac{\sum_{i\in R_1}\sum_{j\notin R_1} t_{ij} * W_{ij}}{C_1 * \sum_{i\in R_1}\sum_{j\notin R_1} t_{ij}} = \frac{\sum_{i=k}^{M}\sum_{j<K \text{ or } j>M} t_{ij} * W_{ij}}{C_1 * \sum_{i=K}^{M}\sum_{j<K \text{ or } j>M} t_{ij}} \qquad (2.10)$$

From section 2.4.1 and Fig. 2.2 it is seen that because of the logarithmic behaviour of the HVS, the discrimination ability decreases with an increase in the value of background intensity B. This is what is also reflected by the contrast measure C defined in equation 2.9. In order to incorporate this property, we have introduced the weighting factor W_{ij} in the expression 2.10.

However, one may note that the direct use of $\Delta B/B$ for W_{ij} will obviously change the contrast value if the object and background intensities are interchanged. Since this is not intuitively appealing, W_{ij} may be defined as

$$\text{either} \quad W_{ij} = |i-j|/(i+j) \qquad (2.11.1)$$

$$\text{or} \quad W_{ij} = |i-j|/\max\{i,j\} \qquad (2.11.2)$$

$$\text{or} \quad W_{ij} = |i-j|/\min\{i,j\} \qquad (2.11.3)$$

It is, therefore, seen that W_{ij} ensures equal contribution to equation 2.10 when the object and background intensities are interchanged in addition to the property of deterioration of discrimination ability with increase in the background intensity.

In case of a complex image, the visual system does not adapt to a single intensity level; instead, it adapts to an average

level which depends on the nature of the image [3]. It is therefore, more logical to choose equation 2.11.1 for W_{ij}. In the denominator of equation 2.10 , the term $\Sigma\Sigma t_{ij}$ is used to make the measure independent of the size of regions while the constant C_1 is introduced to make $0 \leq C^B_{K,M} \leq 1$. C_1 obviously depends on the choice of W_{ij} and is equal to the maximum possible value of W_{ij}.

Therefore,

$$C_1 = (L-1)/(L+1) \quad \text{for equation 2.11.1}$$
$$= (L-1)/L \quad \text{for equation 2.11.2}$$
$$= (L-1) \quad \text{for equation 2.11.3}$$

where L is the maximum level in F.

Therefore, it appears from equation 2.10 that if $i=j$, then $C^B_{K,M} = 0$ i.e., contrast in $[K,M]$ is minumum. On the other hand, if $i=1$ and $j=L$ then contrast is maximum (=1) since for all t_{ij} that are considered in $C^B_{K,M}$ the values of W_{ij} become equal to C_1, thereby making the numerator and denominator same.

The t_{ij} s considered in the equation 2.10 are shown by the shaded portion in Fig. 2.3, which gives the total number of transitions across the boundary of the segment R_1, i.e., from the region $[K,M]$ to its outside.

Again, define $C^W_{K,M}$, the homogeneity of the region $[K,M]$ as

$$C^W_{K,M} = 1 - \frac{\sum_{i\in R_1}\sum_{j\in R_1} t_{ij} * |i-j|}{(L-1) * \sum_{i\in R_1}\sum_{j\in R_1} t_{ij}} = 1 - \frac{\sum_{i=k}^{M}\sum_{j=K}^{M} t_{ij} * |i-j|}{(L-1) * \sum_{i=K}^{M}\sum_{j=K}^{M} t_{ij}} \qquad (2.12)$$

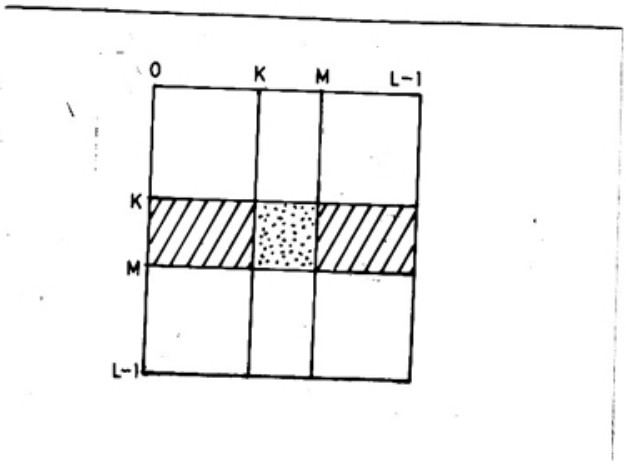


Fig. 2.3 Pictorial representation of contrast-homogeneity measure

The t_{ij}s considered in equation 2.12 are shown by the dotted portion in Fig.2.3. $C^W_{K,M}$ lies in the interval [0,1] i.e., $0 \leq C^W_{K,M} \leq 1$.

If R_1 is perfectly homogeneous, $K = M$ (i.e., the region contains only one gray level) then $C^W_{K,M} = 1$ as $|i-j| = 0$ for all i and j.

With the decrease of the homogeneity of R_1, $|i-j|$ increases and approaches to $(L-1)$. As a result, the ratio in equation 2.12 approaches to unity and $C^W_{K,M}$ tends to zero. Thus $C^B_{K,M}$ and $C^W_{K,M}$ are found to increase with increase in contrast and homogeneity, respectively, in a region $[K,M]$. On the basis of these two measures we define a composite measure

$$g_{K,M}(C^W_{K,M}, C^B_{K,M}) = C^W_{K,M} * C^B_{K,M} \qquad (2.13)$$

Since $0 \leq C^W_{K,M} \leq 1$ and $0 \leq C^B_{K,M} \leq 1$, g also lies between 0 and 1.

The level at which g attains a maximum value can, therefore, be considered a boundary (or threshold) between regions.

2.5 APPLICATION OF THE PROPOSED MEASURE FOR SEGMENTATION

2.5.1 Extraction of Starting Regions - Algorithm 1

In order to extract the thresholds in an image F, we start with $R_1 = [0,0]$ and increase the size of R_1 one by one to the right side of the gray scale till we get a local maximum of $g_{K,M}$

i.e., the process is started with $K = 0$, $M = 0$ and M is incremented one by one until $g_{K,M}$ attains a maximum value. If the maximum occurs at the gray level K_1, then K_1 corresponds to a threshold and gray levels ranging from 0 to K_1 represent a region of the image F. Then we start with $K = M = K_1 + 1$, and the process is repeated as described above till we get the next maximum, say at K_2. The gray levels ranging from K_1+1 to K_2 thus constitute another segment. In this way the process is carried on until the entire gray scale is exhausted.

2.5.1.1 **Merging of Single Valued Regions.**

The composite measure $g_{K,M}$ is found to be very much sensitive to highly uniform regions. In other words, a region containing only one gray level is likely to be detected as a separate segment, although one does not desire to have such a segment. To avoid this we have suggested here a merging algorithm whereby such a single-valued segment is accepted, if the transitions within the segment is higher than those across it. This criterion enables one to retain only those regions which have significant size in the spatial domain.

Therefore, the algorithm first of all finds out the regions of single gray level and determines whether such a region should be merged or not. If it decides to merge a region, then the next task is to determine the adjacent region (left or right) to which it is to be merged.

Let $R_i = [M,M]$ be the region under consideration.

Let T_W be total number of transitions within the region R_i, then

$$T_W = t_{M,M} \cdot \tag{2.14}$$

If T_O = total number of transitions from R_i to all other outside regions, then

$$T_O = \sum_{i=M}^{M} \sum_{j \neq M} t_{ij} \tag{2.15}$$

Decision Rules.

If $T_W > T_O$, then the size of the region can be taken as reasonably big. Therefore, accept the region, do not merge it; otherwise, merge the region to either of its adjacent regions. Now the problem is to select the adjacent region to which it is to be merged.

Case-1. If $M = 0$, merge R_i to the right adjacent region.

Case-2. If $M = L\text{-}1$, merge R_i to the left adjacent region.

Case-3. If $0 < M < L\text{-}1$, we proceed as follows.

Let T_L be the total number of transitions from R_i to its left adjacent region. If the left adjacent region contains gray levels ranging from L_1 to L_2 then

$$T_L = \sum_{i=M}^{M} \sum_{j=L_1}^{L_2} t_{ij} \tag{2.16}$$

Suppose T_R is the total number of transitions from R_i to the right adjacent region and the right adjacent region contains gray levels ranging from R_1 to R_2 then

$$T_R = \sum_{i=M}^{M} \sum_{j=R_1}^{R_2} t_{ij} \qquad (2.17)$$

If $T_L > T_R$, then merge R_i to the left adjacent region; otherwise, merge it to the right adjacent region.

The foregoing decision rule can also be formulated in a more general way as follows.

Let T_t be the total number of transitions from R_i to all regions including itself,

$$\text{i.e.,} \qquad T_t = \sum_{i=M}^{M} \sum_{j=0}^{L-1} t_{ij} \qquad (2.18)$$

$$\text{and} \qquad \theta = T_W / T_t, \quad \theta < 1.$$

Now if $\theta > \theta_H$, accept the region, otherwise, merge it, where θ_H is some preassigned threshold value.

$\theta_H = 0.5$ obviously gives the original decision rule.

The advantage of defining the decision rule in this manner is that in an interactive environment one can change θ_H, if necessary, and compare the results to pick up the most appropriate value of θ_H for a particular type of image.

2.5.2 Incorporating Size of Regions - Algorithm 2 [29]

In Algorithm 1 we have incorporated a separate merging phase in order to eliminate the undesirable small (noninformative) regions. The following algorithm introduces the concept of size of regions into the composite measure $g_{K,m}$ and thereby attempts to eliminate the need of a separate merging phase.

Let us define $g'_{K,M} = g_{K,M} * \left(\frac{A_R}{A} \right) * \left(\frac{1}{d_R} \right)^{\alpha}$ (2.19)

where

A_R = size of the region $[K,M]$, i.e., number of pixel intensities in F which are in the range $[K,M]$,

A = size of the Image = $P * Q$,

α = a parameter, with $0 \leq \alpha \leq 1$,

d_R= Length of the region $[K,M]$ = M-K+1.

The first factor (A_R/A) gives more weight to the regions of greater size, resulting in an increase in $g'_{K,M}$ value. Thus if the histogram of an image has a sharp valley, $g'_{K,M}$ value is very likely to decrease when M is set at the valley point. On the other hand, for a small region with a weak valley, such a decrease in g' value (for detecting the region) is less likely to occur. In such a situation it is further to be noted that the next following region (valley) becomes much less likely to be detected because of the cummulative increase of region size d_R and hence A_R .

In order to circumvent this situation, a second factor $(1/d_R)^{\alpha}$ is introduced which compensates the above effect by reducing the value of g' with increase in d_R . α controls the rate of decrease in g' value. As α increases, the number of regions is likely to increase.

To extract the thresholds under this algorithm we proceed exactly in the same way as we have mentioned earlier for Algorithm 1, considering $g'_{K,M}$ for $g_{K,M}$.

2.6 Computational Steps

Algorithm 1 : Extraction of initial thresholds

S0 : $K = 0$, $M = 0$, *no-of-thresholds* = 0

S1 : *Previous-value* = $g_{K,M}$

S2 : $M = M + 1$

If M is greater than L-1 , then go to S5; Otherwise,

Current-value = $g_{K,M}$.

S3 : If *Current-value* is not less than *Previous-value*, then

Previous-value = *Current-value* and go to S2.

S4 : *no-of-thresholds* = *no-of-thresholds* + 1,

THRESHOLD(*no-of-thresholds*) = $M - 1$ (*THRESHOLD* is an array that stores the threshold levels),

$K = M$, go to S1.

S5 : End of extraction of thresholds

Merging Algorithm (output of Algorithm 1 is input to this algorithm)

S0 : Find the next region of single gray level. If there is none then STOP.

S1 : (Suppose the region under consideration is $R_i = [M,M]$)
Compute T_w = The number of transitions within R_i.
Compute T_t = the total number of transitions from R_i to all the regions including itself. Compute $\theta = T_w / T_t$.

S2 : If θ is greater than θ_H (a predetermined value), then go to S0

S3 : If M is the first gray level with nonzero frequency, then merge the region R_i to its right adjacent region and go to S0.

S4 : If M is the last gray level with non zero frequency, then merge the region R_i to its left adjacent region and go to S0.

S5 : Compute T_R and T_L, the number of transitions from R to its right and left adjacent regions, respectively.

S6 : If T_L is greater than T_R, then merge R_i to its left adjacent region. Otherwise, merge R_i to its right adjacent region.

S5 : Go to S0.

2.7 IMPLEMENTATION AND RESULTS

The segmentation algorithms described in the previous section are implemented on a set of four different images [3] having dimension 64 x 64 with 32 gray levels. Figs. 2.4.1, 2.5.1, 2.6.1 and 2.7.1 represent the original input images, while Figs. 2.4.2, 2.5.2, 2.6.2 and 2.7.2 represent the corresponding gray level histograms. These images are produced on a line printer by over printing different character combinations for different gray levels.

Fig. 2.4.1 represents an image of Mona Lisa. It is to be noted that the gray level histogram [Fig. 2.4.2] is almost unimodal (having some local minima). When porposed Algorithm 1 (without merging) is applied to it, four thresholds namely, 0,1,6 and 17 are produced. The corresponding segmented image is shown in Fig. 2.4.7, where different segments are represented by different textures. When the merging algorithm is applied to it, the segment [1,1] (Table 2.1) is merged to its right adjacent segment. The segmented image so obtained after merging is shown in Fig. 2.4.3. Comparing Fig 2.4.3 and 2.4.7 we find that there is an undesirable region inside the hair of Mona Lisa, which after being merged results in a more meaningful segmentation [Fig 2.4.3]. Fig. 2.4.7 is shown, as an illustration, only to demonstrate the effect of the merging algorithm in selecting final thresholds.

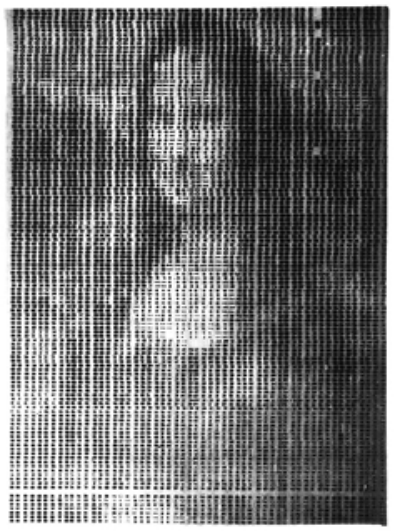


Fig. 2.4.1 Input

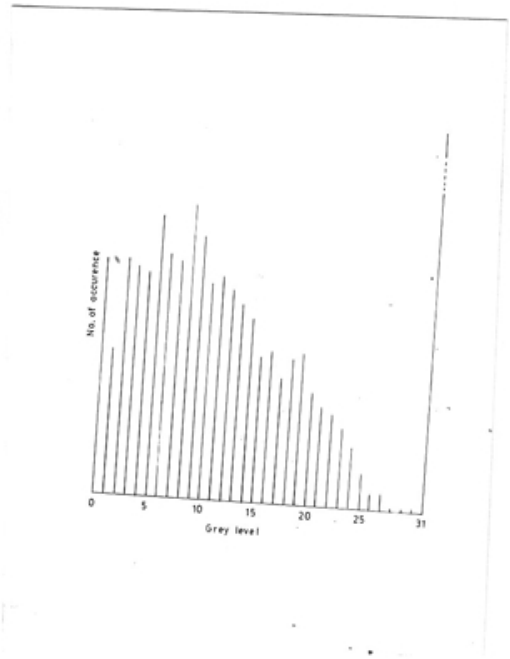


Fig. 2.4.2 Histogram

Fig. 2.4.3 Segmented image by Eqn . 2.13(after merging)

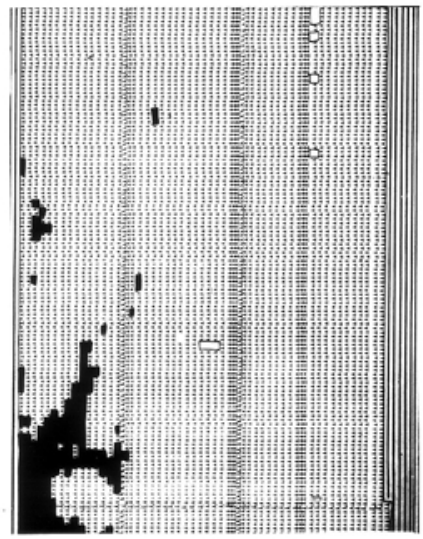


Fig. 2.4.4 Segmented image by Eqn. 2.3

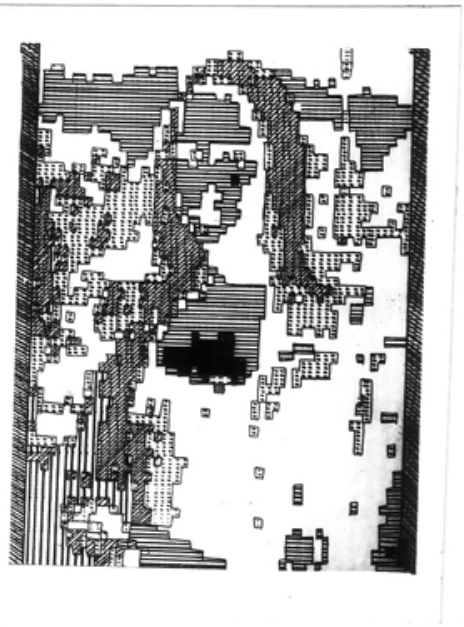


Fig. 2.4.5 Segmented image by Eqn. 2.7

Fig. 2.4.6 Segmented Image by Eqn. 2.8

Fig. 2.4.7 Segmented image by Eqn . 2.13(before merging)

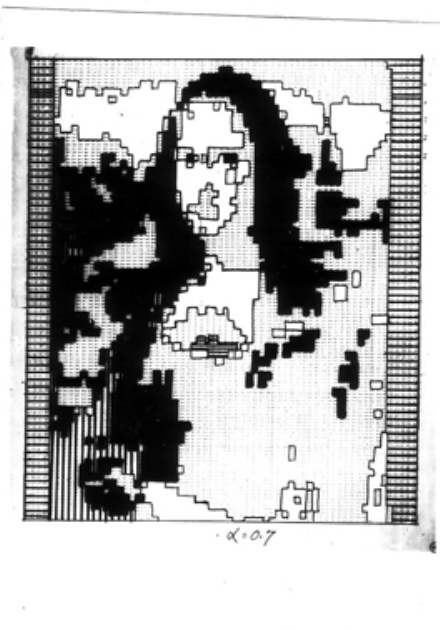


Fig. 2.4.8 Segmented image by Eqn. 2.19 (α = .7)

Fig. 2.4 Input and segmented images of Mona Lisa

Table 2.1

Thresholds for Various Methods

IMAGE	Proposed method		Method of Wezska and Rosenfeld	Method of Deravi and Pal	Method of Chanda et.al.
	Before merging	After merging			
MONALISA Fig.2.4	0,1,6,17	0,6,17	0,30	0,3,6,17, 25,30	0,3,6,17, 28,30
LINCOLN Fig.2.5	0,4,9,13, 17,18,24	0,4,9,13, 17,24	0,4,9,12, 17,24,30	4,9,12,17, 23,30	5,10,12,17, 22,27,30
JET Fig.2.6	0,3,7,23, 24,26	0,3,7,24, 26	0,3,5,7, 18,29	3,7,11, 18,29	7,11,13,18, 21,23
BIPLANE Fig.2.7	0,1,2,10, 17,18,23	1,10,17, 23	0,10,17, 22,30	0,2,10,17, 22,30	0,10,14,17, 21,29

Fig. 2.5.1 is an image of Abraham Lincoln, and the corresponding gray level histogram [Fig. 2.5.2] is found to have a number of deep valleys. The thresholds (before and after merging) generated by the proposed method 1 are shown in Table 2.1. The output segmented image is shown in Fig. 2.5.3.

To demonstrate the validity of the method 1 for images with flat and wide valleys in their histogram, the algorithm is applied to the image of a jet [Fig. 2.6.1]. One can see in Fig. 2.6.1 that the right wing of the jet has vanished inside the cloud in such a way that apparently it is very difficult to trace the boundary of the right wing. The present algorithm 1 is found to be successful in separating out that wing from the cloud. Fig. 2.6.3 represents the segmented image, while the thresholds are shown in Table 2.1. As a typical illustration, Fig. 2.6.9 shows the plot of $g_{K,M}$ value for the image of the Jet.

Fig. 2.7.1 represents the image of a biplane having two dominant modes in its histogram [Fig. 2.7.2]. From Fig. 2.7.3 the object is found to be clearly separated from the background.

The above results were obtained by using equation 2.11.1 while computing $W_{i,j}$ of the contrast measure. Experiments were also carried out with equations 2.11.2 and 2.11.3, and the corresponding performances were found to be almost similar.

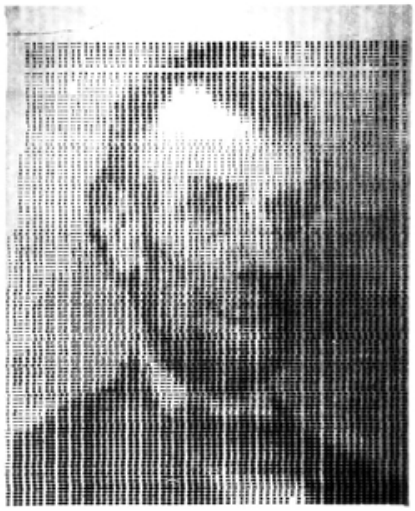


Fig. 2.5.1 Input

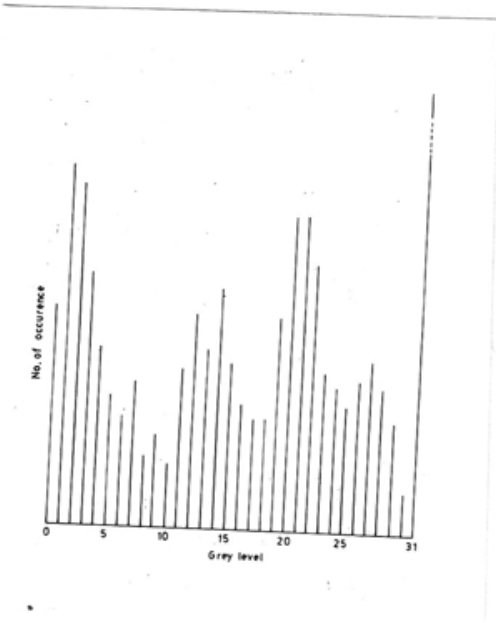


Fig. 2.5.2 Histogram

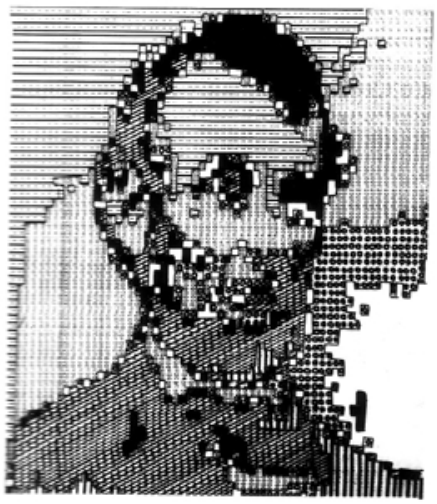


Fig. 2.5.3 Segmented image by Eqn . 2.13(after merging)

Fig. 2.5.4 Segmented image by Eqn. 2.3

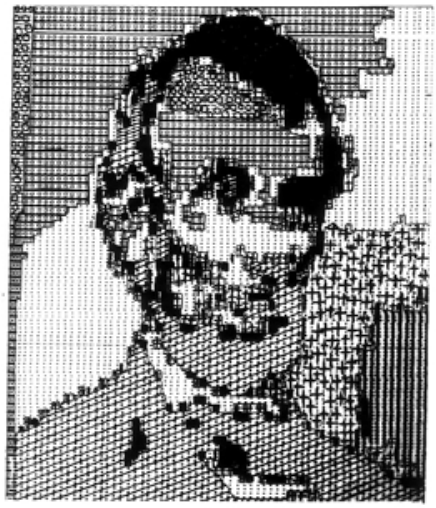


Fig. 2.5.5 Segmented image by Eqn. 2.7

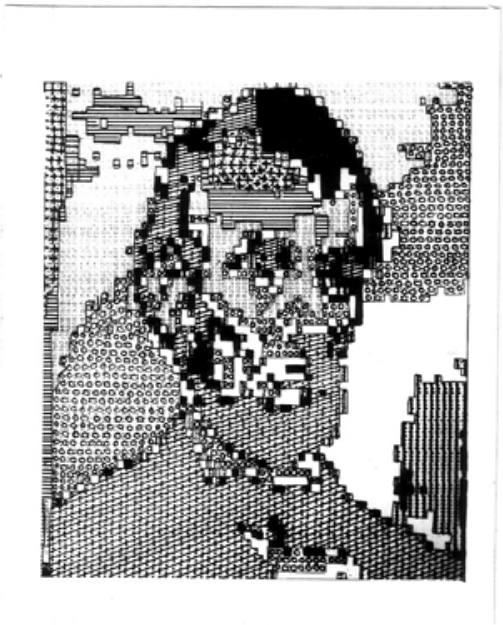


Fig. 2.5.6 Segmented Image by Eqn. 2.8

Fig. 2.5.7 Segmented image Eqn. 2.19 (α = .7)

Fig. 2.5 Input and segmented images of Abraham Lincoln

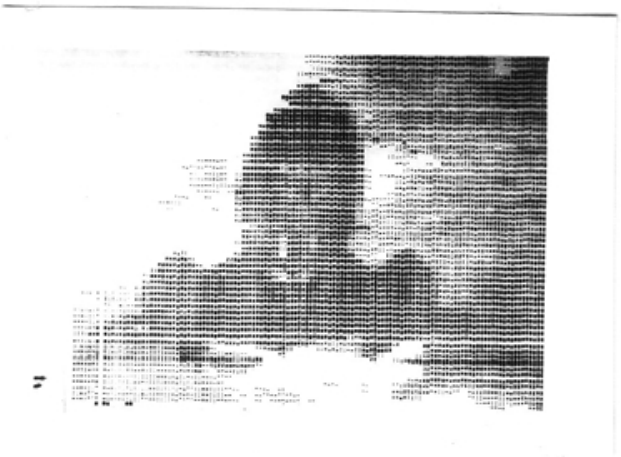


Fig. 2.6.1 Input

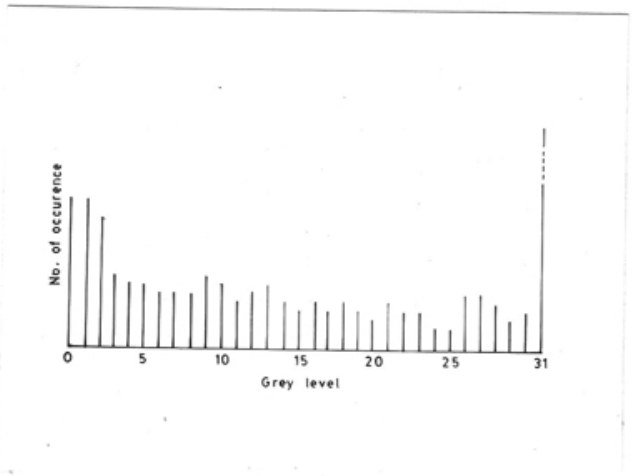


Fig. 2.6.2 Histogram

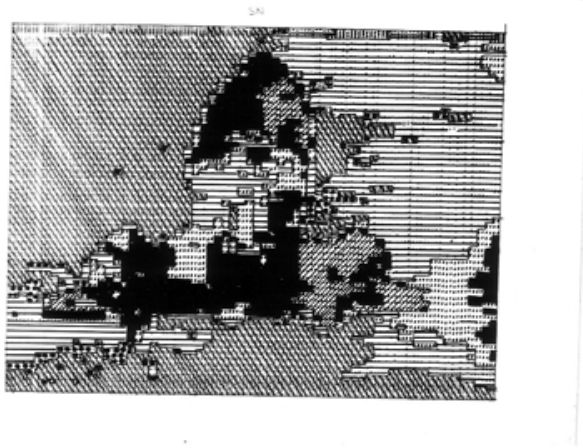


Fig. 2.6.3 Segmented image by Eqn . 2.13(after merging)

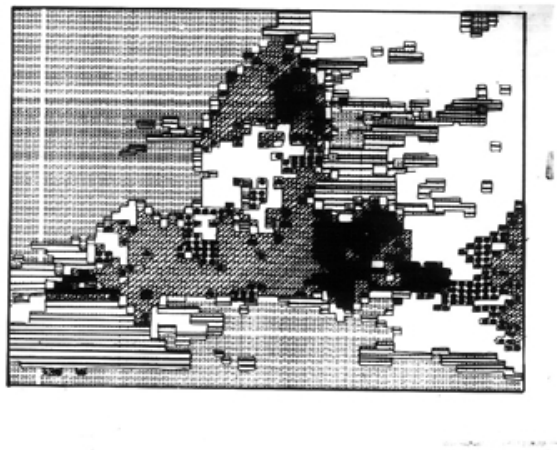


Fig. 2.6.4 Segmented image by Eqn. 2.3

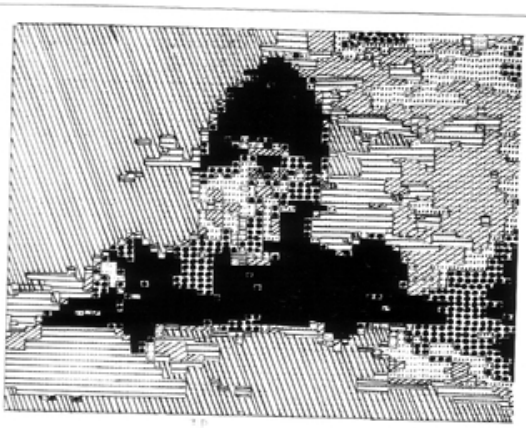


Fig. 2.6.5 Segmented image by Eqn. 2.7

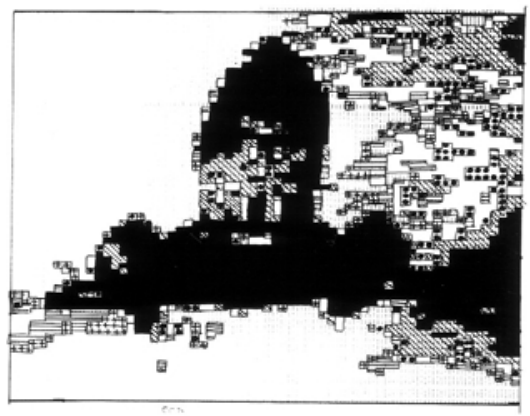


Fig. 2.6.6 Segmented image by Eqn. 2.8

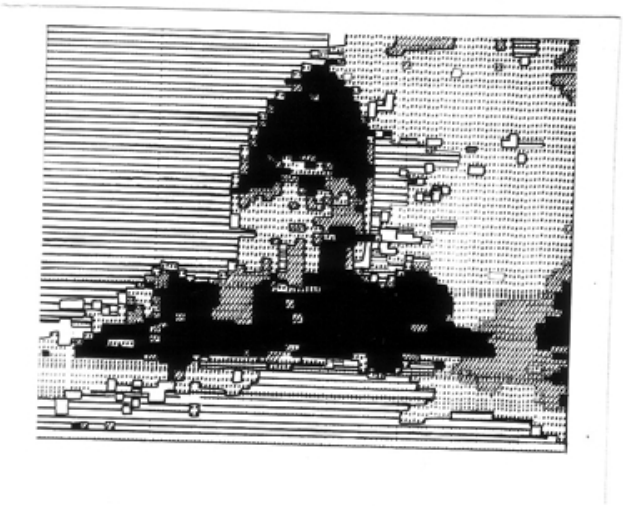


Fig. 2.6.7 Segmented image by Eqn. 2.13 (θ_H=.75)

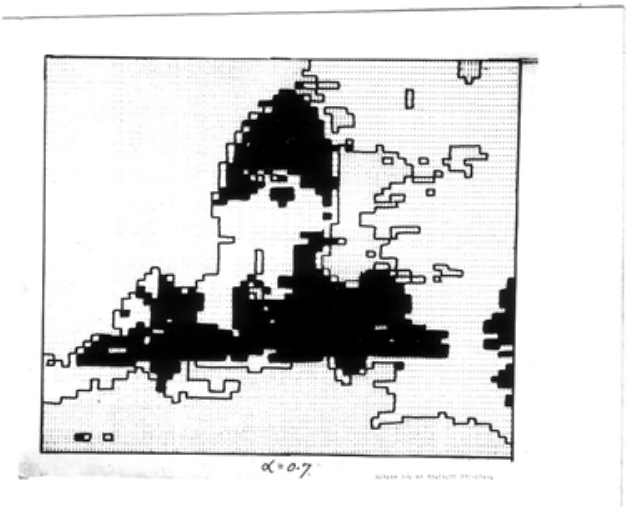


Fig. 2.6.8 Segmented Image by Eqn. 2.19(α=.7)

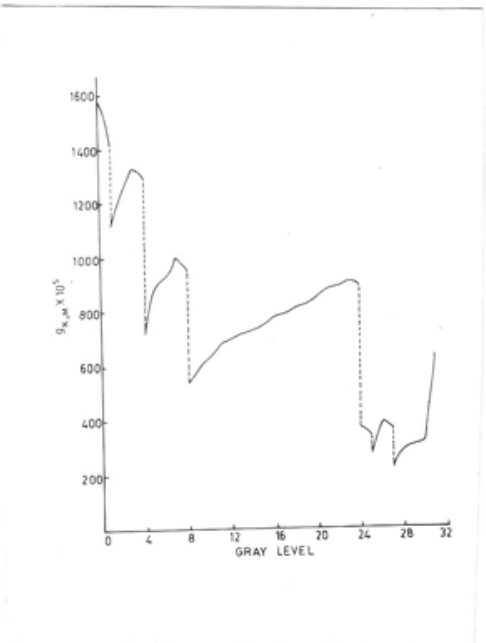


Fig. 2.6.9 Plot of $g(K,M)$

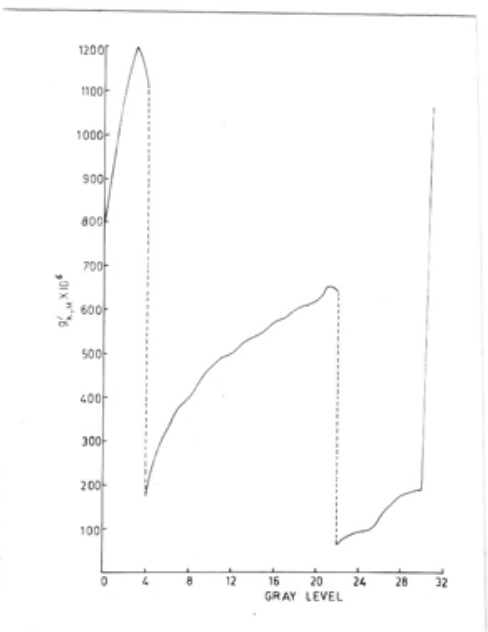


Fig. 2.6.10 Plot of $g'(K,M)$ (α=0.6)

Fig. 2.6 Input and segmented image of Jet

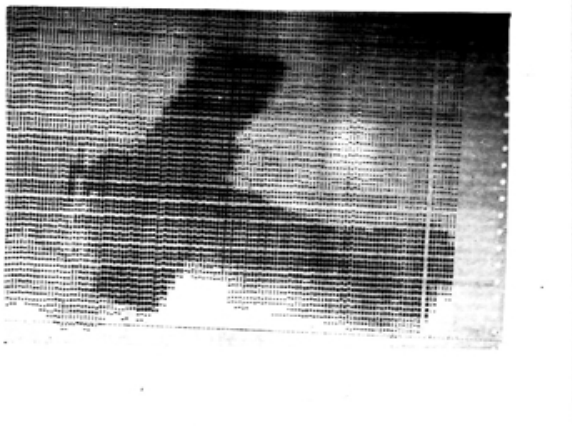


Fig. 2.7.1 Input

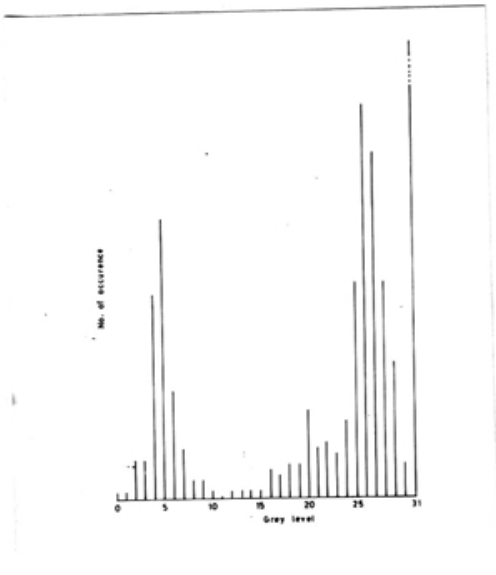


Fig. 2.7.2 Histogram

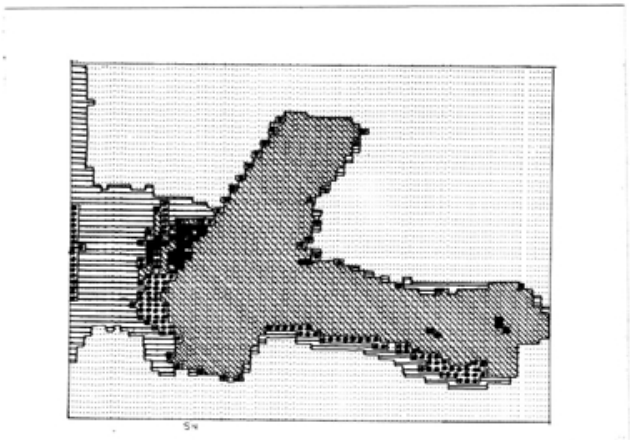


Fig. 2.7.3 Segmented image by Eqn . 2.13(after merging)

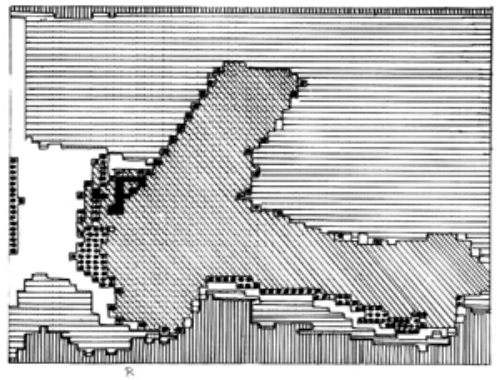


Fig. 2.7.4 Segmented image Eqn. 2.3

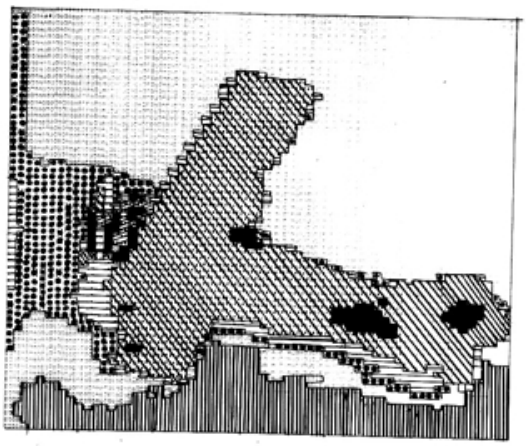


Fig. 2.7.5 Segmented image by Eqn. 2.7

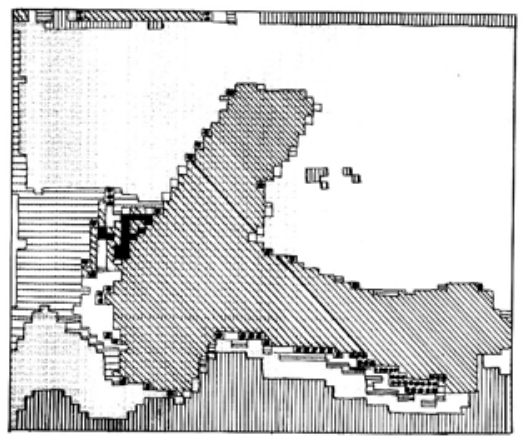


Fig. 2.7.6 Segmented image by Eqn. 2.8

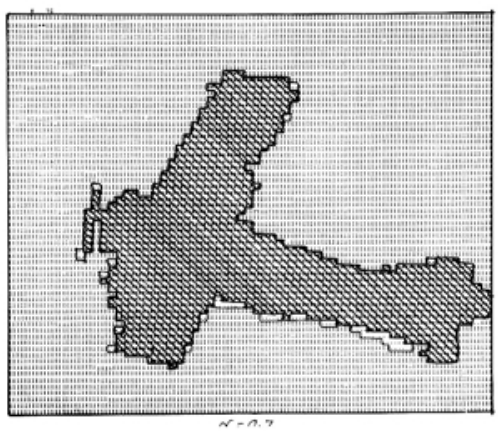


Fig. 2.7.7 Segmented Image by Eqn. 2.19(α=.7)

Fig. 2.7 Input and segmented images of biplane

Table 2.2 shows the thresholds for these images obtained by the proposed Algorithm 2 for different values of α. As discussed in Section 2.5.2 the number of segments is found to increase with increases in the value of α. Note, too, from Table 2.2 that the level 17 for Mona Lisa (unimodal histogram) which was detected at $\alpha = 0.2$ was found to be missing for values of $\alpha = 0.4$ and 0.6. The threshold at 17 or its around again appeared for $\alpha \geq 0.7$. Similar is the case for Jet having almost flat and wide valleyed histogram where the level 7 which was detected at α=0.2, was lost at α = 0.4, 0.6 and 0.7 and reappeared for higher α values. Such a situation did not arise for Lincoln and biplane which have deep valleys. This fact can be explained as follows.

Suppose the first threshold is detected at a gray level $l = l_1$ for $\alpha = \alpha_1$, which means that g' kept on increasing up to gray level l_1 and at $(l_1 + 1)$, g' started falling. Now if for a higher value of $\alpha = \alpha_2$, $\alpha_2 > \alpha_1$, the first threshold is detected earlier at a gray level $l = l_2$, $l_2 < l_1$, then we can say that l_2 is detected, as $(1/d_R)^{\alpha}$ has reduced the value of g' at (l_2+1) to a value which is less than the value of g' at l_2. The threshold at $l = l_1$ which was detected with $\alpha=\alpha_1$, may not now be detected with an α not very high from α_1, because there will not be a sharp fall in g' value as the histogram has weak valleys. Moreover, the value of $(1/d_R)^{\alpha}$ with $d_R = l_1+1$ and $\alpha = \alpha_1$ may be smaller than $(1/d_R)^{\alpha}$ with $d_R = l_1-l_2$, and $\alpha = \alpha_2$ as d_R in the second case is reduced to l_1-l_2 from l_1+1, unless α_2 is very high. Therefore, unless α_2 is

Table 2.2

Thresholds for Different 'α'

IMAGE	α=.2	α=.4	α =.6	α=.7	α=.8	α=1.0
MONALISA Fig.2.4	17,28	6,25	0,6,25, 28	0,6,17, 24,28	0,6,16,24, 25,28	0,3,6,9,14, 22,24,25,28
LINCOLN Fig.2.5	5	4,17	4,9,17	4,9,17, 29	4,9,17,24, 29	2,4,7,9,15, 17,23,29
JET Fig.2.6	7	3,23	3,21	3,21	3,7,16,23, 28	0,2,3,7,11, 13,16,19, 23,24,28
BIPLANE Fig.2.7	10	10	8,10	8,10	0,7,10	2,7,9,23, 29

sufficiently high, the threshold at l_1 will not be detected. However, if the histogram has a deep valley at l_1, it will be detected as the fall in g value will be sharp.

Fig. 2.4.8, 2.5.7, 2.6.8 and 2.7.7 show the segmented images for Algorithm 2 when α is considered to be 0.7, as an illustration. Fig 2.6.10 shows the plot of the $g'_{K,M}$ value for the image of the Jet.

2.7.1 Comparison With The Existing Algorithms

In order to compare the performance of the algorithms with those of some of the existing algorithms based on co-occurrence matrix, we have considered algorithms of Weszka and Rosenfeld [73], Deravi and Pal [74], and Chanda et. al [75]. The thresholds obtained by these methods are also shown in Table 2.1.

Equation 2.3 is found to fail to extract all the meaningful regions of the image of Mona Lisa. It has selected only three segments [Fig. 2.4.4] with thresholds at 0 and 30, as a result, most of the important information is lost.

Equations 2.4-2.7, on the other hand, has detected two extra segments in the chest and one extra region in the hair [Fig. 2.4.5], while equation 2.8 has also produced two similar segments one of them is of smaller size than the corresponding one produced by equations 2.4-2.7 [Fig. 2.4.6]). From Fig 2.4.3, it is seen that the regions generated by the proposed method 1 where

these additional regions are absent, create a better impression to the eye. However, for Algorithm 2, (with $\alpha = 0.7$) the hair of Mona Lisa did not split, and it has two extra regions in her chest [Fig. 2.4.8].

For the image of Lincoln (Fig. 2.5) all the methods except the present ones have divided the fore-head into two regions. Furthermore, Algorithm 1 and the method by Weszka and Rosenfeld have divided the beard into two regions, which is not the case for the other three methods. Finally, Algorithm 2 created the smallest number of segmented regions [Fig. 2.5.7] and created a better impression to the eye.

In the case of the jet, equation 2.8 has failed to discriminate between the cloud and right wing [Fig. 2.6.6] while the other two methods like ours are successful in doing so [Figs.2.6.3, 2.6.4 and 2.6.5]. Algorithm 1 and equation 2.3 produced comparable results, while the result produced by equations. 2.4-2.7 [Fig.2.6.5] seems to create a better impression to the eye, but, if we alter the value of θ_H from 0.5 to 0.75 in the merging algorithm, the first two regions (Table 2.1) are merged and the resulting image [Fig.2.6.7] is found to be much improved. However, Fig.2.6.8 (by Algorithm 2) which has least number of segments created best result.

In the case of the biplane, all the methods were able to detect its contour (Fig.2.7). However, equations 2.4-2.7 are found

to generate two additional regions inside the tail of the biplane which are absent for other cases. The background is found to be clustered in two parts by all but our methods. Furthermore, all the methods except ours have divided the shade of the biplane into three or more regions, which is two and one for Algorithm 1 and 2, respectively.

From Table 2.1 it apears that the equations 2.3, 2.4-2.7, and 2.8 detected, except for the Jet, a threshold at the end of the gray scale (e.g., 30 for Mona Lisa, Lincoln and biplane). These thresholds correspond to some undesirable regions at the frame of the images. The incorporation of the factor W_{ij} in equation 2.10, which accounts for the nonlinear behaviour of HVS, has been found to be able to eliminate such occurrences. However, Algorithm 2 has failed to eliminate such segments for higher value of α.

2.8 CONCLUSION

Two algorithms for image segmentation are described using the measures of "homogeneity" and "contrast" within/between regions of an image. The first algorithm has a separate merging phase to eliminate the undesirable segments. The second algorithm, on the other hand, eliminated the need of such a separate merging phase by considering the region size while extracting the thresholds. Both of the algorithms simulate the way in which human beings perceive the brightness variation in an image.

Note here that when a human being attempts to segment an image, in addition to his/her logarithmic response to contrast he / she makes use of certain other higher level knowledge which is not considered in the proposed algorithms. Hence it is not expected that the proposed algorithms would segment an image exactly in the same way a human observer does. However, the incorporation of the weighting function $w_{i,j}$ to simulate only the contrast response of the human visual system has resulted in consistently better segments for images with different types of histograms.

It is found that none of the existing methods could generate consistently good segments as that of the proposed algorithms for all images. Some of the existing methods have generated extra segments for some images, while some methods have failed to detect meaningful segments. Both of the proposed algorithms are suitable for an interactive environment.

It is seen that the choice of α in Algorithm 2 is critical. As α increases, the number of regions also increases, resulting in an acceptable segmentation around $\alpha = 0.5$. However, for a good image having sharp valleys in the gray level histogram, the typical value of α may be less than 0.5. On the other hand, an α grater than 0.5 can be taken for images having weak valleys in the histogram or a unimodal histogram.

The proposed methods are expected to be less sensitive to noise as the psychovisual facts have been considered. No information theoretic concept has been used here to define the proposed co-occurrence based measures. In the next chapter some more measures of contrast and homogeneity will be defined, alongwith their applications, based on Shannon's theory of information.

Chapter 3

HIGHER ORDER AND CONDITIONAL ENTROPY

3.1. INTRODUCTION

This chapter consists of three parts. In the first part, the various applications [33,58,59] of Shannon's entropy in image segmentation have been criticised. The second part introduces two new definitions of image entropy. Their applications to image segmentation problem have then been demonstrated in the third part of the chapter.

Pun [58,59] used Shannon's concept to define the entropy of an image assuming that an image is entirely represented by its gray level histogram only. He used this concept to derive an expression for an upper bound of the a-posteriori entropy and finally used it to segment an image into object and background. Kapur et al.[33] have found some flaws in Pun's derivation and also used a similar concept to partition the image into object and background. They, instead of considering one probability distribution for the entire histogram, used two separate probability distributions: one for the object and the other for the background. The total entropy of the image is then maximised to arrive at the threshold for segmentation.

Both these methods were developed without highlighting the adequateness of the concept of Shannon's entropy in the case of an image. For example, the dependency of pixel intensities in an

image and hence its spatial distribution are not taken into account in defining its entropy. As a result, different images with identical histograms will always result in the same entropic value and same threshold. This is, of course, not intuitively acceptable. Moreover, in the algorithm of Pun [59] the maximisation of an upper bound of the a-posteriori entropy, to avoid the trivial result with the a-posteriori entropy, is not justified.

In order to circumvent this, two other definitions of entropy for images, namely entropy of order q, $q = 1,2,\ldots,$ and the conditional entropy of an image have been defined. Entropy of order unity ($q = 1$) corresponds to the global entropy as used by Pun [58,59] and Kapur et al. [33]. Higher order ($q > 1$) entropies take into account the information contained in a sequence of gray levels of length q and thus denote the various local entropies of an image. The conditional entropy, on the other hand, gives an average amount of information that may be obtained from a region, when one has viewed the remaining portion of the image.

These new concepts are then introduced to develop two algorithms for the object-background image classification (segmentation) problem. The effectiveness of these algorithms is demonstrated for a set of images with different types of histograms and their superiority in performance over those of Pun [59] and Kapur et al. [33] is also established.

3.2. ENTROPIC MEASURES FOR IMAGE PROCESSING

Shannon [109,110] defined the entropy of a system as a function of the probability of occurence of different states of the system. If a system has n different states with probability of occurence

p_i, $i = 1,2,\ldots,n$, $\sum_{i=1}^{n} p_i = 1$, then the entropy, H of the system is defined as

$$H = -\sum_{i=1}^{n} p_i \log_2 p_i. \qquad (3.1)$$

Based on the concept of Shannon's entropy, different authors have defined the entropy for an image. Let us discuss here those measures and the associated problems when applied to image processing.

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

Pun [58,59] and Kapur et al. [33] cosidered 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.

Following Shannon's definition of entropy (equqtion 3.1), Pun [59] defined the entropy of the image (histogram) as

$$H = -\sum_{i=0}^{L-1} p_i \log_2 p_i; \qquad p_i = N_i/N, \tag{3.2}$$

for the image segmentation problem.

3.2.1. Evaluation Function of Pun [59]

Let s be the threshold which classifies the image into object and background. Let N_B and N_W be the numbers of pixels in the black and white portions of the image. Then the a-posteriori probability of a black pixel is $P_B = N_B/N$ and that of a white pixel is $P_W = N_W / N$. Thus, the a posteriori entropy of the image is:

$$H_L'(s) = -P_B \log_2 P_B - P_W \log_2 P_W$$

$$= -P_S \log_2 P_S - (1 - P_S) \log_2 (1 - P_S) \tag{3.3}$$

as

$$P_S = \sum_{i=0}^{s} p_i \quad \text{and} \quad P_W = 1 - P_S \; . \tag{3.4}$$

Since the maximisation of H_L' gives the trivial result of $P_S = 1/2$, Pun maximised an upper bound $g(s)$ of $H_L'(s)$, where

$$g(s) = \frac{H_B^S \; . \; \log_2 P_S}{H_L \; . \; \log_2 [\max(p_0, p_1, \ldots, p_S)]}$$

$$+ \frac{(H_L - H_B^S) \; . \; \log_2 (1 - P_S)}{H_L \; . \; \log_2 [\max(p_{S+1}, p_{S+2}, \ldots, p_{L-1})]} \tag{3.5}$$

where

$$H_L = -\sum_{i=0}^{L-1} p_i \log_2 p_i \qquad (3.6)$$

and

$$H_B^s = -\sum_{i=0}^{s} p_i \log_2 p_i \qquad (3.7)$$

The value of s which maximises $g(s)$ can be taken as the threshold for object and background classification.

3.2.2. Method of Kapur, Sahoo, And Wong [33]

Kapur et al. have also used Shannon's concept of entropy but from a different point of view. They, instead of considering one probability distribution for the entire image, considered two probability distributions, one for the object and the other for the background. The sum of the individual entropy of the object and background is then maximised. In other words, this will result in equiprobable gray levels in each region and thus maximises the sum of homogeneities in gray levels within object and background by making the gray levels equiprobable in either region.

If s is an assumed threshold, then the probability distribution of the gray levels over the black portion of the image is :

$$\frac{p_0}{Ps}, \frac{p_1}{Ps}, \ldots, \frac{p_s}{Ps},$$

and that of the white portion is

$$\frac{p_{s+1}}{1-Ps}, \frac{p_{s+2}}{1-Ps}, \ldots, \frac{p_{L-1}}{1-Ps}.$$

The entropy of the black portion (object) of the image is :

$$H_B^{(s)} = - \sum_{i=0}^{s} \frac{p_i}{Ps} \log_2(p_i/P_s) \qquad (3.8)$$

and that of the white portion is

$$H_W^{(s)} = - \sum_{i=s+1}^{L-1} \frac{p_i}{1-P_s} \log_2(p_i/(1-P_s)). \qquad (3.9)$$

The total entropy of the image is then defined as

$$H_T^{(s)} = H_B^{(s)} + H_W^{(s)}. \qquad (3.10)$$

In order to select the threshold they maximised $H_T^{(s)}$. In other words, the value of s which maximises $H_T^{(s)}$ gives the threshold for object and background classification.

3.2.3. Some Remarks

All the methods [33,58,59] discussed so far virtually assume that an image is entirely represented only by its histogram. Thus, different images with identical histograms will result in same entropic value in spite of their different spatial distributions of gray levels. This is, of course, not intuitively appealing. For example, consider Figs. 3.1 and 3.2 where "dotted","shaded" and "blank" pixels represent the gray levels l_1, l_2 and $l_3 (l_1 < l_2 < l_3)$ respectively. Both Fig.3.1 and Fig.3.2 have identical histograms but different spatial distributions of gray levels. As a result, the entropy(information content) of Fig. 3.1 and Fig. 3.2 is expected to be different.

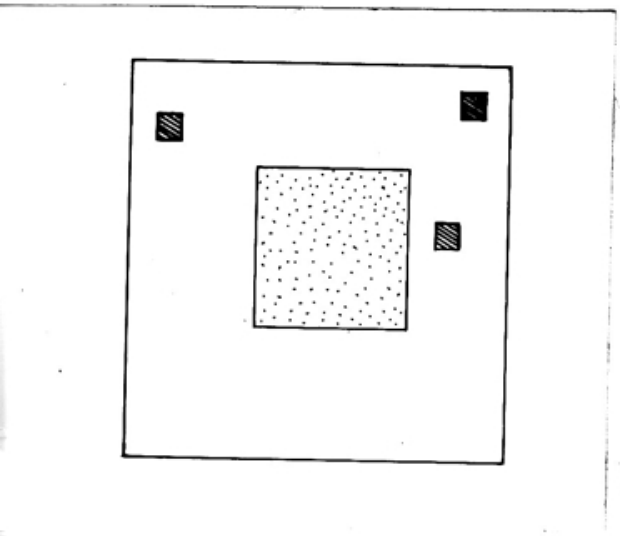


Fig. 3.1 Illustration of image entropy

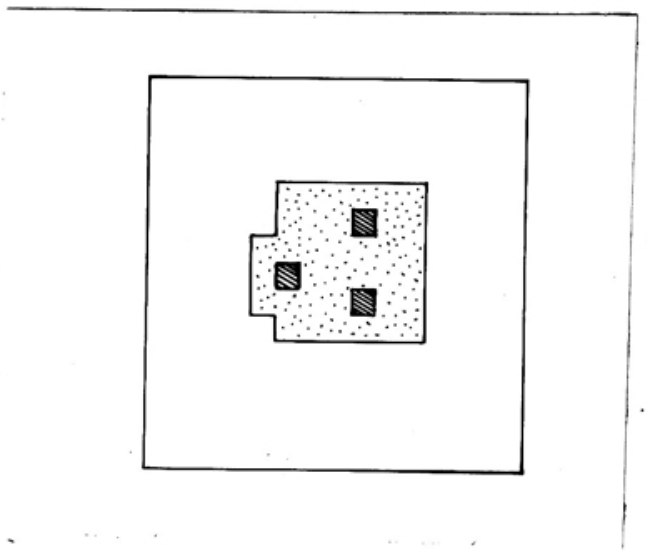


Fig 3.2 Image with same histogram, different entropy

Under those definitions all images with identical histograms but different spatial distributions of gray levels will, therefore, give rise to the same threshold value. Our experience and intuition also do not support this. For example, l_1 should be the threshold in Fig. 3.1, whereas, it is l_2 in Fig. 3.2 for object-background classification.

In the algorithm of Pun [59], the concept of maximisation of the evaluation function $g(s)$ (equation 3.5) , an upper bound of $H_L'(s)$, for object-background classification is not justified. For example, the maximum value of equation 3.5 may even correspond to a minimum value of the a-posteriori entropy (equation 3.3).

Moreover, all these methods have used Shannon's concept of entropy in image processing without highlighting its adequateness in the case of an image.

3.3 ENTROPY OF AN IMAGE [31,63]

3.3.1. Global And Local Entropy

We know that in an image pixel intensities are not independent of each other. This dependency of pixel intensities can be incorporated by considering sequences of pixels to estimate the entropy. In order to arrive at the expression of entropy of an image the following theorem due to Shannon (2,109) can be stated.

Theorem. Let $p(s_i)$ be the probability of a sequence s_i of gray levels of length q, where a sequence s_i of length q is defined as a permutation of q gray levels. Let us define

$$H^{(q)} = -\frac{1}{q} \sum_i p(s_i) \log_2 p(s_i), \tag{3.11}$$

where the summation is taken over all gray level sequences of length q. Then $H^{(q)}$ is a monotonic decreasing function of (q) and $\lim_{q \to \infty} H^{(q)} = H$, the entropy of the image.

For different values of q we get various orders of entropy.

Case-1:

$q = 1$, i.e., sequence of length one. If $q = 1$, we get

$$H^{(1)} = -\sum_{i=0}^{L-1} p_i \log_2 p_i, \tag{3.12}$$

where p_i is the probability of occurrence of the gray level i

Such an entropy is a function of the histogram only and it may be called the "global entropy" of the image. Therefore, different images with identical histograms would have same $H^{(1)}$ value irrespective of their contents. The definitions used by Pun and Kapur et al., in fact belong to *Case*-1.

Case-2:

$q = 2$, i.e. sequences of length two, Hence

$$H^{(2)} = -\frac{1}{2} \sum_i p(s_i) \log_2 p(s_i),$$

where s_i is a sequence of gray level of length two,

$$= -\frac{1}{2}\sum_i \sum_j p_{ij} \log_2 p_{ij}, \qquad (3.13)$$

where p_{ij} is the probability of co-occurrence of the gray levels i and j. Therefore, $H^{(2)}$ can be obtained from the co-occurrence matrix.

$H^{(2)}$ takes into account the spatial distribution of gray levels. Therefore, two images, with identical histograms but different spatial distributions will result in different entropy, $H^{(2)}$ values. Expressions for higher order entropies ($q > 2$) can also be deduced in a similar manner. $H^{(i)}, i \geq 2$, may be called the "local entropy" of order i of an image. In this context mention must be made of the work of Haralick et. al, who gave some entropy measures for texture classification [131]. Their measures are conceptually different from those defined above and are dependent on different types of spatial-dependence matrix which consider only pairs of pixels.

3.3.2. Conditional Entropy

Suppose an image has two distinct portions, the object X and the background Y. Suppose the object consists of the gray levels $\{x_i\}$ and the background contains the gray levels $\{y_i\}$. The conditional entropy of the object X given the background Y, i.e. the average amount of information that may be obtained from X given that one has viewed the background Y, can be defined as

$$H(X/Y) = -\sum_{x_i \in X} \sum_{y_j \in Y} p(x_i/y_j) \log_2 p(x_i/y_j). \qquad (3.14)$$

Similarly, the conditional entropy of the background Y given the object X is defined as

$$H(Y/X) = - \sum_{y_j \in Y} \sum_{x_i \in X} p(y_j/x_i) \log_2 p(y_j/x_i) . \qquad (3.15)$$

The pixel y_j, in general, can be an m th order neighbour of the pixel x_i, i.e., y_j can be the m th pixel after x_i. Since the estimation of such a probability is very difficult, we impose another constraint on x_i and y_j of equations 3.14 and 3.15. In addition to $x_i \in X$ and $y_j \in Y$, we also impose the restriction that x_i and y_j must be adjacent pixels. Thus, equations 3.14 and 3.15 can be rewritten as

$$H(X/Y) = - \sum_{\substack{x_i \in X \\ (x_i, y_j)\,adjacent}} \sum_{y_j \in Y} p(x_i/y_j) \log_2 p(x_i/y_j) \qquad (3.16)$$

$$H(Y/X) = - \sum_{\substack{y_j \in Y \\ (y_j, x_i)\,adjacent}} \sum_{x_i \in X} p(y_j/x_i) \log_2 p(y_j/x_i) \qquad (3.17)$$

The conditional entropy of the image can, therefore, be defined as

$$H^{(C)} = (H(X/Y) + H(Y/X)) / 2 \qquad (3.18)$$

where X and Y represent the object and background of the image, respectively.

3.4 APPLICATION TO IMAGE SEGMENTATION [31,63]

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

3.4.1 Algorithm 1 (Based on Second order Entropy)

We are now going to describe an algorithm based on equation 3.13, which takes into account the spatial details of an image. Such a method is dependent on the probability of co-occurrences of pixel intensities, that can be estimated from the co-occurrence matrix defined in chapter - 2 (section 2.2).

Here, we have used an asymmetric co-occurrence matrix which considers the horizontally right and vertically lower transitions only. It has been found experimentally [29,74] that the consideration of a symmetric co-occurrene matrix by taking into account both right-left and upper-lower transitions needs some additional computations without changing much the information content in it. Furthermore, consideration of both horizontal and vertical transitions allows all the edges to participate in threshold selection.

Thus, t_{ij} is defined as follows :

$$t_{ij} = \sum_{l=1}^{P} \sum_{k=1}^{Q} \delta \ , \qquad (3.19.1)$$

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

$\delta = 0$, otherwise.

The probability of co-occurrence of p_{ij} of gray levels i and j can, therefore, be written as

$$p_{ij} = t_{ij} \Big/ \left(\sum_{i} \sum_{j} t_{ij} \right), \qquad (3.19.2)$$

where obviously $0 \leq p_{ij} \leq 1$.

If s, $0 \leq s \leq L-1$, is a threshold, then s partitions the co-occurrence matrix into four quadrants, namely A, B, C and D (Fig. 3.3).

Let us define the following quantities:

$$P_A = \sum_{i=0}^{s} \sum_{j=0}^{s} p_{ij} , \qquad (3.20.1)$$

$$P_B = \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} p_{ij} , \qquad (3.20.2)$$

$$P_C = \sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} p_{ij} , \qquad (3.20.3)$$

$$P_D = \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} p_{ij} , \qquad (3.20.4)$$

Normalising the probabilities within each individual quadrant, such that the sum of the probabilities of each quadrant is equal to one(1), we get the following cell probabilities for different quadrants :

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

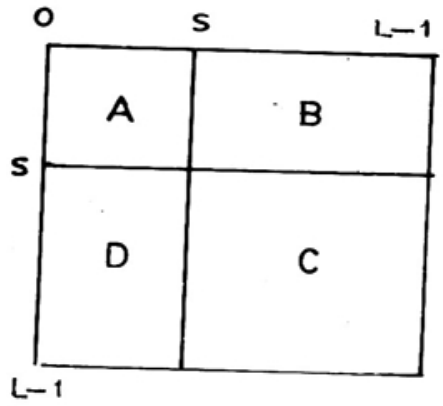


Fig. 3.3 Quadrants of co-occurrence matrix

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

for $0 \le i, j \le s$. (3.21)

Similarly,

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

for $0 \le i \le s$, $s+1 \le j \le L-1$; (3.22)

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

for $s+1 \le i, j \le L-1$; (3.23)

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

for $s+1 \le i \le L-1$, $0 \le j \le s$. (3.24)

Now with the help of equations 3.13 and 3.21, the second order local entropy of the object can be defined as

$$H_A^{(2)}(s) = -\frac{1}{2} \sum_{i=0}^{s}\sum_{j=0}^{s} p_{ij}^{A} \log_2 p_{ij}^{A} . \quad (3.25)$$

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

$$H_C^{(2)}(s) = -\frac{1}{2} \sum_{i=s+1}^{L-1}\sum_{j=s+1}^{L-1} p_{ij}^{C} \log_2 p_{ij}^{C} . \quad (3.26)$$

Hence the total second order local entropy of the object and the background can be written as

$$H_T^{(2)}(s) = H_A^{(2)}(s) + H_C^{(2)}(s). \qquad (3.27)$$

The gray level corresponding to the maximum of $H_T^{(2)}(s)$ gives the threshold for object-backgrond classification. The copmputational steps involved are given in section 3.5.1.

3.4.2. Algorithm 2 (Based on Conditional Entropy)

This algorithm is based on the concept of conditional entropy (equations 3.16-3.18). Suppose s is an assumed threshold. Then pixels with gray level values ranging from 0 to s constitute the object, while the remaining pixels with gray values lying between $s+1$ and L-1 correspond to the background. Let t_{ij} be an entry of the quadrant B (Fig 3.3), 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 as defined in equation 3.22 gives the probability that gray levels i and j belong to the object and background, respectively, and they are adjacent. Thus p_{ij}^B s of equation 3.22 give the probabilities required by equation 3.16. Similarly, p_{ij}^D s of equation 3.24 correspond to the probabilities of equation 3.17.

Therefore,

$$\text{H(Object/Background)} = \text{H(O/B)} = -\sum_{i=0}^{s}\sum_{j=s+1}^{L-1} p_{ij}^B \log_2 p_{ij}^B \qquad (3.28)$$

and

$$H(\text{Background/Object}) = H(B/O) = -\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} p_{ij}^{D} \log_2 p_{ij}^{D} \quad (3.29)$$

Now the average conditional entropy of the partitioned image is

$$H_T^{(C)}(s) = (H(O/B) + H(B/O)) / 2. \quad (3.30)$$

In order to get the threshold for object-background classification $H_T^{(C)}(s)$ is maximised with respect to s. The computational steps for this method is presented in section 3.5.2.

3.5. COMPUTATIONAL STEPS.

3.5.1 Algorithm 1

S1. Compute the co-occurrence matrix.

S2. $s = 0$, $Max = 0$, $Th = 0$

S3. For $i = 0$ to s
For $j = 0$ to s
Compute $p_{ij}^{\wedge}$ (using equation 3.21)

S4. For $i = s + 1$ to L - 1
For $j = s + 1$ to L - 1
Compute p_{ij}^{C} (using equation 3.23)

S5. Compute $H_T^{(2)}(s)$ (using equation 3.27)

S6. If $Max < H_T^{(2)}(s)$ then

$Max = H_T^{(2)}(s)$

$Th = s$

S7. $s = s + 1$

If $s < L - 1$ then
go to S3.

S8. Th is the desired threshold.

STOP.

3.5.2 Algorithm 2

S1. Compute the co-occurrence matrix.

S2. $s = 0$, $Max = 0$, $Th = 0$

S3. For $i = 0$ to s
For $j = s+1$ to L-1

Compute p^{B}_{ij} (using equation 3.22)

S4. For $i = s + 1$ to $L - 1$
For $j = 0$ to s

Compute p^{D}_{ij} (using equation 3.24)

S5. Compute $H_T^{(C)}(s)$ (using equation 3.30)

S6. If $Max < H_T^{(C)}(s)$ then

$Max = H_T^{(C)}(s)$

$Th = s$

S7. $s = s + 1$
If $s < L - 1$ then
go to S3.

S8. Th is the desired threshold.

STOP.

3.6. IMPLEMENTATION AND RESULTS [31,63]

The segmentation (object-background classification) algorithms described in Section 3.5 are implemented on a set of four images . Figs. 3.4.1, 3.5.1, 3.6.1 and 3.7.1 represent the input images while Figs. 3.4.2, 3.5.2, 3.6.2 and 3.7.2 represent the corresponding gray level histograms. The threshold levels produced by different methods are presented in Table 3.1.

Fig. 3.4.1 represents the image of a biplane with two dominant modes in its gray level histogram (Fig. 3.4.2). The segmented images produced by Algorithm 1, Algorithm 2, the method of Pun, and the method of Kapur et al. are shown in Figs. 3.4.3-6, respectively. From the results one can see that, except for the conditional entropic method (Algorithm 2), the propeller in front of the biplane is lost. In all but Algorithm 2, some portion of the background got mixed up with the object though the histogram of the image has two dominant modes. The methods of Pun (Fig. 3.4.5) and that of Kapur et al. (Fig. 3.4.6) have produced comparable results to that of Algorithm 1.

Figs. 3.5.1 and 3.5.2 represent the input image of Abraham Lincoln and its gray level histogram, respectively. The histogram has a number of deep valleys. The thresholds produced by different methods are shown in Table 3.1 and the corresponding segmented images are shown in Figs 3.5.3 - 6 . In this case too, all the methods except the conditional entropic method (Algorithm 2) have

Table - 3.1

Thresholds for Object-Background classification

Images	Thresholds			
	Proposed Algorithm 1 (Equn. 3.27)	Proposed Algorithm 2 (Equn. 3.30)	Algorithm of Pun (Equn. 3.5)	Algorithm of Kapur et al. (Equn. 3.10)
Biplane (Fig. 3.4)	22	12	24	21
Lincoln (Fig. 3.5)	17	10	16	15
Saturn (Fig. 3.6)	0	17	13	05
Blurred Chromosome (Fig. 3.7)	33	41	32	33

See Fig. 2.7.1
page - 85

See Fig. 2.7.2
page - 85

Fig. 3.4.1 Input

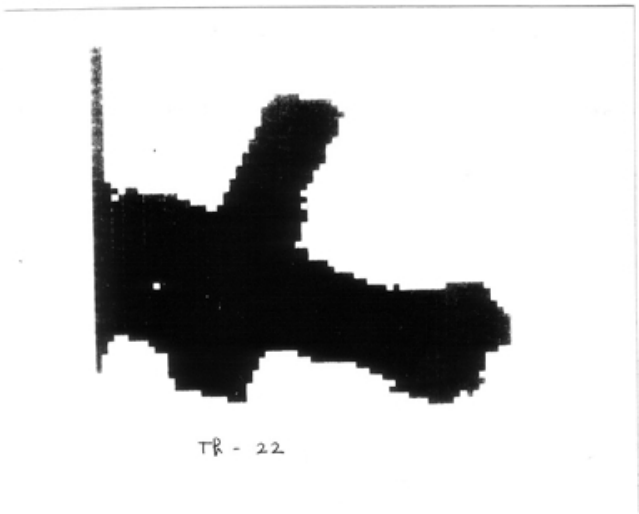


Fig. 3.4.2 Histogram

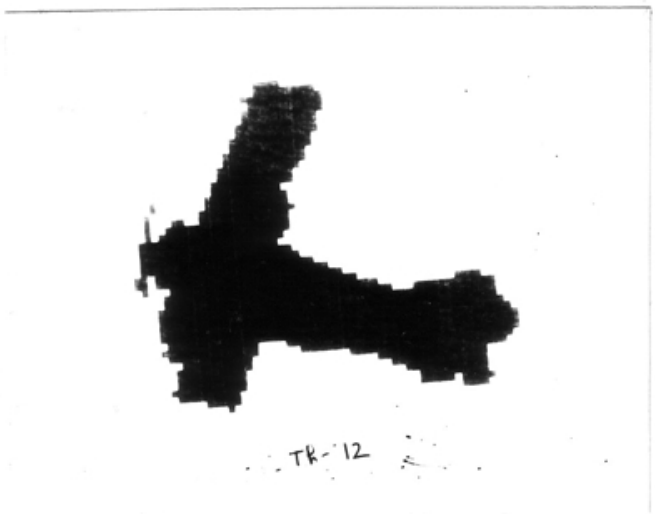


Fig. 3.4.3 Segmented image by Eqn. 3.27

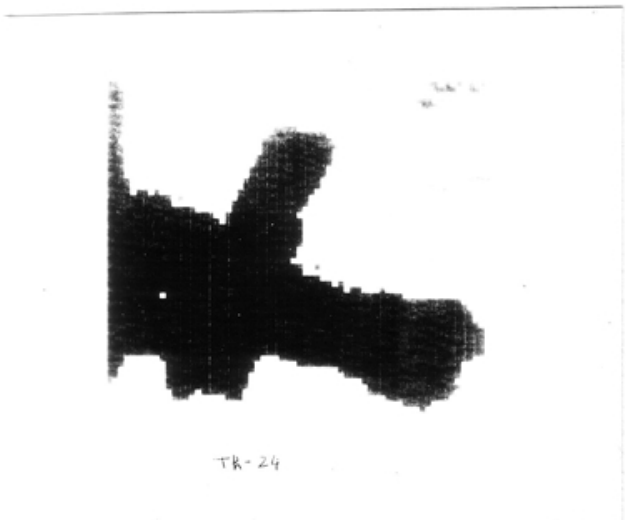


Fig. 3.4.4 Segmented image by Eqn 3.30

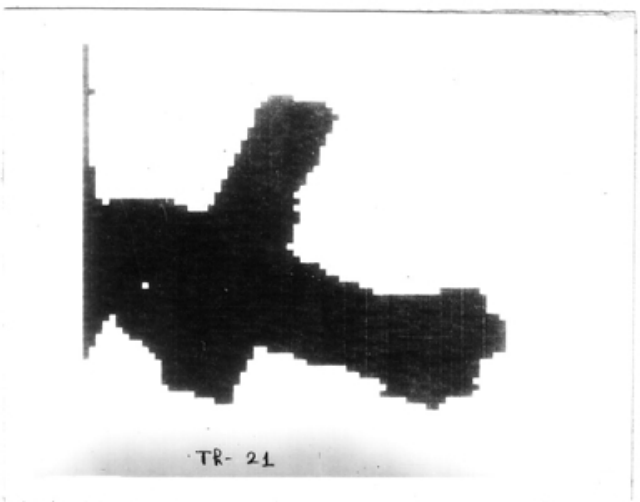


Fig. 3.4.5 Segmented image by Eqn. 3.5

Fig. 3.4.6 Segmented image by Eqn. 3.10

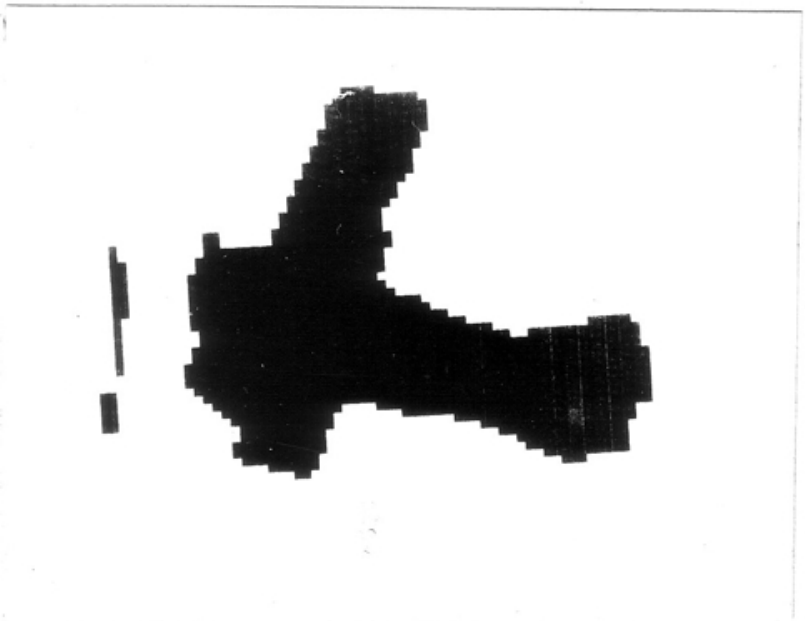


Fig. 3.4.7 Segmented image by [82]

Fig. 3.4 Input and segmented images of biplane

See Fig. 2.5.1
page - 81

Fig. 3.5.1 Input

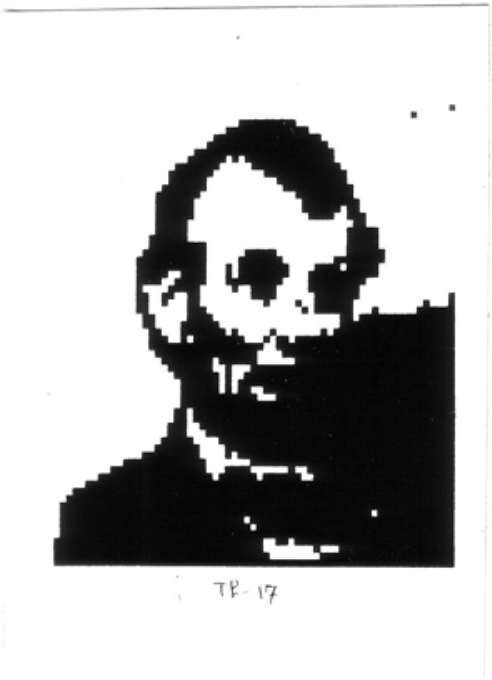


Fig. 3.5.3 Segmented image by Eqn. 3.27

See Fig. 2.5.2
page - 81

Fig. 3.5.2 Histogram

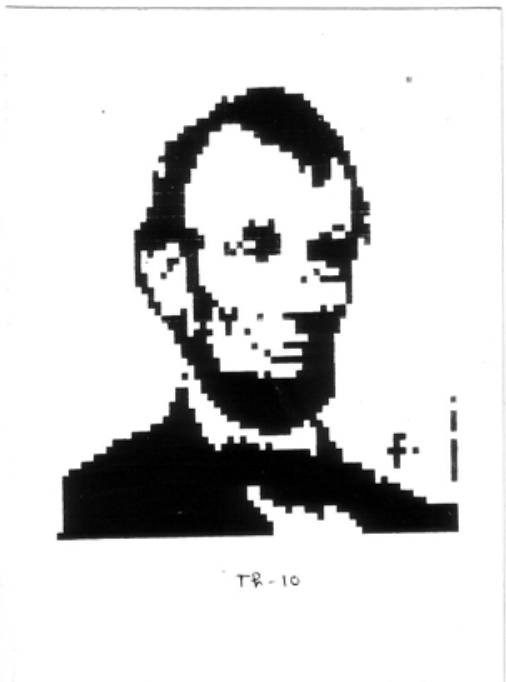


Fig. 3.5.4 Segmented image by Eqn 3.30

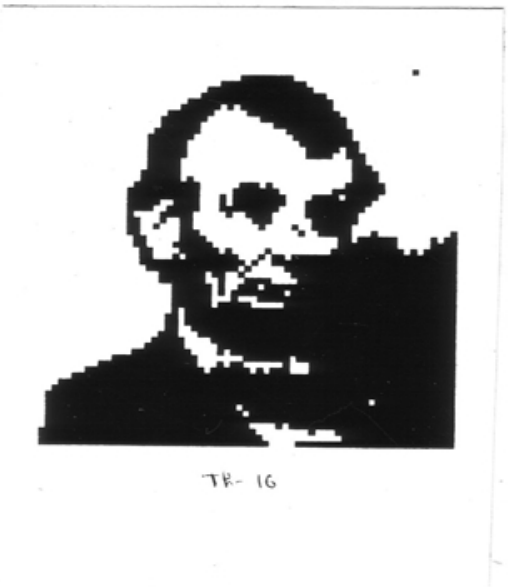


Fig. 3.5.5 Segmented image by Eqn. 3.5

Fig. 3.5.6 Segmented image by Eqn. 3.10

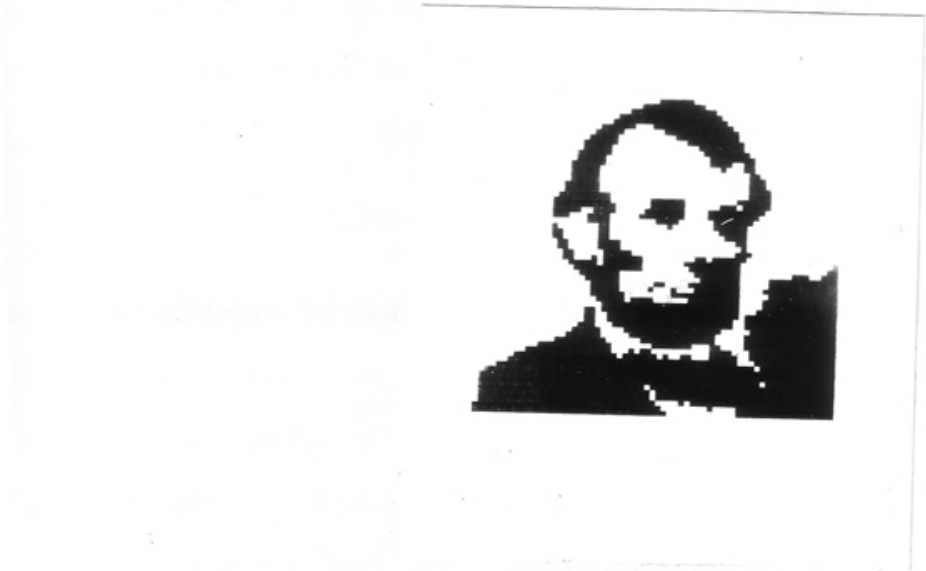


Fig. 3.5.7 Segmented image by [82]

Fig. 3.5 Input and segmented images of Abraham Lincoln

produced comparable result. The best result is produced by Algorithm 2 (equation 3.30) which has clearly separated the object from the background. All other methods failed to discriminate between the beard and the background at the south-east corner of the image.

Let us now consider an image of Saturn (Fig. 3.6.1). The characteristics of its histogram are depicted in Fig. 3.6.2. More or less, all the algorithms are found here to partition well the Saturn from the background. For Algorithm 1 and the method of Kapur et. al. a small portion of the background got mixed up with the right side of the ring. On the other hand, the thin shade on the left side of the ring is made visible by Algorithm 2 and the method of Pun . Of these two, the boundaries of Saturn produced by Algorithm 2 are more smooth and thus closely resemble that of the input.

The algorithms are also tested on an image of blurred chromosome (Fig. 3.7.1) having a bimodal histogram (Fig 3.7.2). Here too, all the methods except the conditional entropic method (Algorithm 2) have produced similar results. However, the best classification is done by Algorithm 2. This also conforms well to the recent work of Pal and Rosenfeld [106].

Recently, Abutaleb [82] suggested a thresholding algorithm based on 2-dimensional entropy. He regarded a pixel value and the

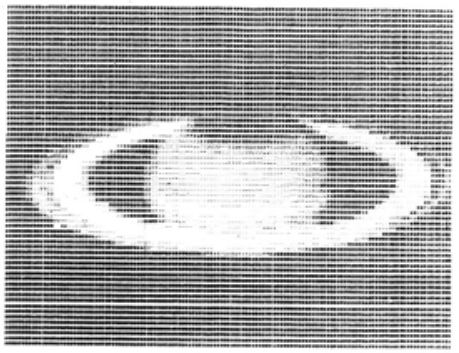


Fig. 3.6.1 Input

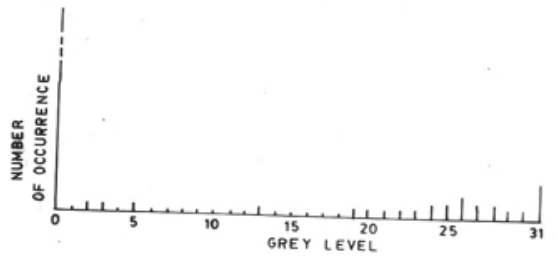
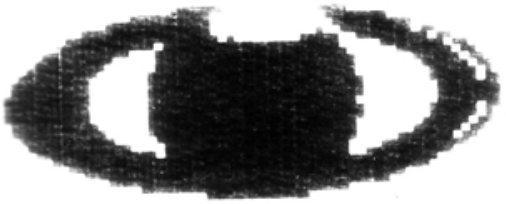
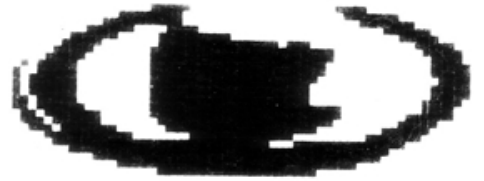


Fig. 3.6.2 Histogram



TR-0

Fig. 3.6.3 Segmented image by Eqn. 3.27

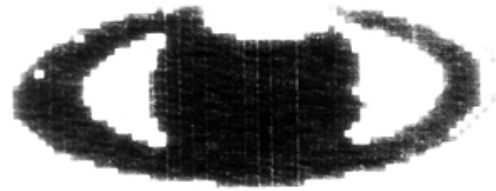


TR-17

Fig. 3.6.4 Segmented image by Eqn 3.30

TR-13

Fig. 3.6.5 Segmented image by Eqn. 3.5



TR-5

Fig. 3.6.6 Segmented image by Eqn. 3.10

Fig. 3.6 Input and segmented images of saturn

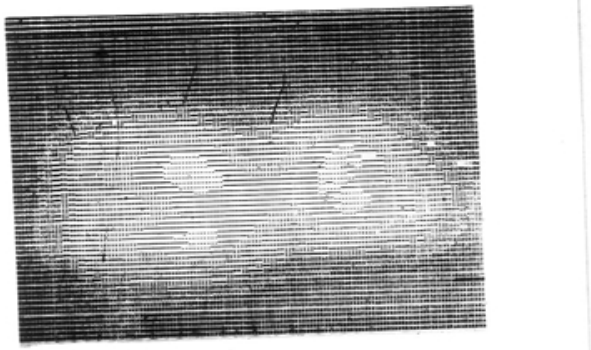


Fig. 3.7.1 Input

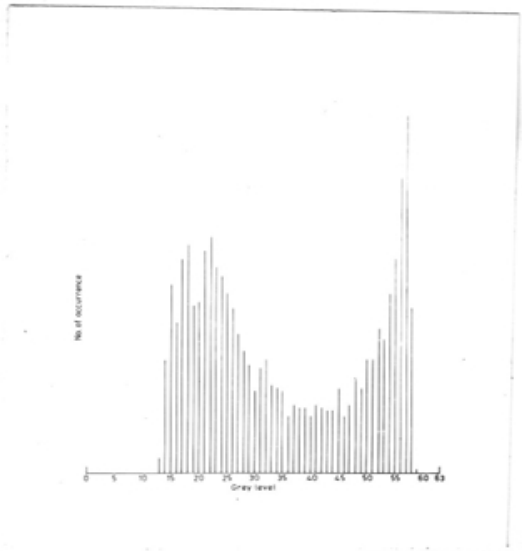


Fig. 3.7.2 Histogram

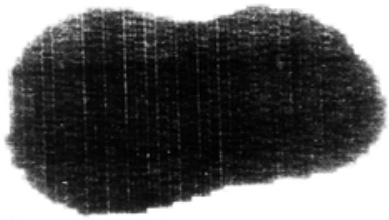


Fig. 3.7.3 Segmented image by Eqn. 3.27 and Fig. 3.7.6 by Eqn. 3.10

Fig. 3.7.4 Segmented image by Eqn 3.30

Fig. 3.7.5 Segmented image by Eqn. 3.5

Fif. 3.7 Input and segmented images of blurred chromosome

average gray level value of its neighbourhood as a pair, for the purpose of computation of the two dimensional histogram of the image. The 2-dimensional probability distribution is obtained by finding frequencies of such pairs. A threshold vector (s,t) , s for the pixel value and t for the average value is obtained by maximising the sum of entropies of object and background. Regarding this approach, the following points must be mentioned here.

a) Consideration of the threshold vector like (s,t), instead of (s,s) as we have used, increases the computational overhead extensively.

b) Computation of the average is also time consuming and may result in improper classification of the boundary pixels.

c) To reduce some computation, he assumed that the sum of entries in the quadrants representing transitions across the boundary is zero. This seems to be a very unrealistic assumption. Moreover, those entries contain important information as established by our conditional entropic method.

The method [82] is also implemented, in a part of the experiment, on the images of the biplane and Lincoln for comparison. The results [Figs. 3.4.7 , 3.5.7] conform to the point (b) mentioned above. It is further to be noted that, the algorithm in [82] considers L^2 threshold vectors whereas, it is L in our case.

3.7 EXTENSION TO MULTITHRESHOLDING [31]

The proposed algorithm can be extended to multithresholding problems using the concept of conditional entropy for more than two classes. For example, if $S_R = \{s_1, s_2, \ldots, s_R\}$ denotes R thresholds in an image with (R+1) regions, then S_R partitions the co-occurrence matrix of the image into (R+1) × (R+1) blocks. Define p_{ij}^{lm} ; $l,m = 1,2,\ldots,R+1$, the normalised probabilities for the block (l,m) as

$$p_{ij}^{lm} = \frac{t_{ij}}{\sum_{i=s_{l-1}}^{s_l} \sum_{j=s_{m-1}}^{s_m} t_{ij}} \quad ,$$

$$s_{l-1} \leq i \leq s_l \quad , \; s_{m-1} \leq j \leq s_m \qquad (3.31)$$

with $s_0 = 0$ and $s_{R+1} = L-1$. The total conditional entropy of the partioned image (into R+1 regions) will then be

$$H_T^{(C)}(S_R) = - \sum_{l=1}^{R+1} \sum_{\substack{m=1 \\ m \neq l}}^{R+1} \sum_{i=s_{l-1}}^{s_l} \sum_{j=s_{m-1}}^{s_m} p_{ij}^{lm} \log_2 p_{ij}^{lm} \; . \qquad (3.32)$$

The vector $S_R = \{ s_1, s_2, \ldots, s_R \}$ which maximises $H_T^{(C)}(S_R)$ can be taken as the set of thresholds for multiple segmentation of the image.

3.8. CONCLUSION

The concept of different order local entropy and conditional entropy of an image are introduced. The first order local entropy corresponds to the global entropy, as used by Pun[58,59] and by Kapur et al. [33].

Two algorithms for object-background classification are proposed whereby it is possible to segment/extract the object from the background. The results are compared with those of the existing entropic thresholding methods and are found to be superior for a wide class of images. Extension of the conditional entropic method to multithresholding has also been done.

The effect of noise on the performance of the algorithms, defining quantitative measures for objective evaluation of the resulting thresholded images, and establishment of a link between the entropies presented here and some of the various Markovian models [132], may constitute a part of further investigation.

Chapter 4

MEASURES BASED ON A NEW CLASSICAL ENTROPY

4.1 INTRODUCTION

In the previous chapter, we have exploited the entropy definition of Shannon in providing various image entropy measures. The present chapter provides an alternative definition of classical (probabilistic) entropy, in general, based on an exponential behavior of gain function and it has then been used to define various image entropy measures.

In the first part of this chapter Shannon's definition is critically analysed and the need for a new definition, in general, is justified. Unlike the logarithmic behavior of Shannon's entropy, the gain function considered in the new definition is of exponential nature so that the gain in information from an event i with probability of occurrence p_i is defined at all points with bounds at both ends. All other properties except the additive property for independent events(which has been shown to be not much relevant in practice) of Shannon's entropy are also proved.

An extension of the proposed definition is made to fuzzy sets for defining a nonprobabilistic entropy in the second part of the chapter. This nonprobabilistic entropy is found to satisfy all the desired properties stated by De luca and Termini [120], and Pal [102]. A large number of interesting properties of the aforesaid definitions has been stated and proved.

In this context it should be mentioned that a more informative definition of entropy of a fuzzy set together with its application has been be introduced in chapter 6.

Based on the new concept, three definitions (e.g., global, local and conditional) of entropy of an image, as done in chapter 3, are then introduced. As an application of these definitions three algorithms (similar to those in chapter 3) are developed for image segmentation; one of which considers only the global information and the remaining algorithms use local information.

The algorithms are finally implemented on a set of images with widely different types of histogram. The superiority of the proposed methods is then established by comparing the results with those of Pun [59] and Kapur et. al. [33]. It is to be noted that in this chapter only the exponential probabilistic entropy has been used for segmentation. The fuzzy entropy defined here will be used in chapter 6 where a more general and informative definition of fuzzy entropy has been introduced.

The various image entropy measures defined so far, represent either the gray level ambiguity or the spatial ambiguity but, they do not say anything about the location of an object within a scene. In order to account for this fact another important measure, "positional entropy", has finally been introduced. The

positional entropy which gives a measure of information about the location of an object within a scene, may be used in scene analysis / computer vision.

4.2 JUSTIFICATION BEHIND SHANNON'S ENTROPY

In section 1.4.1, we have seen the expression for Shannon entropy of an n-state system as

$$H = -\sum_{i=1}^{n} p_i \log (p_i) \qquad (4.1)$$

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

$$\sum_{i=1}^{n} p_i = 1 \quad , \text{ o} \leq p_i \leq 1 \qquad (4.2)$$

In order to explain why such an expression is taken as a measure of ignorance, let us critically examine the philosophy behind Shannon's entropic measure with an example described below.

Suppose a six faced die, covered with a box is placed on a table and someone is asked to guess about the top most face of the die. Since the exact state of the die is not known, he/she can describe the state of the die by the probability distribution of occurrence of different faces on the top. In other words, the state of the die can be expressed by specifying p_i , i=1,....,6; where p_i is the probability that the i th face is the top most one. Obviously,

$$0 \leq p_i \leq 1 \quad \text{and} \quad \sum_{i=1}^{6} p_i = 1. \qquad (4.3)$$

When the box is opened, the state of the die becomes known to us and we gain some information. A very natural question arises, 'How much information did we gain ?'

Let $p_k = \max_i \{ p_i \}$: the most probable event and

$p_m = \min_i \{ p_i \}$: the least probable event.

Now if the k th face appears on the top, the gain in information would be minimum, whereas the occurrence of the m th face on the top would result in the maximum gain.

Thus we see that the gain in information from an event is inversely related to its probability of occurrence. This, of course, intuitively seems all right. For example, if some body says, " The sun rises in the east", the information content of the statement is practically nil. On the other hand, if one says, "he is ten feet in hight", the information content of the statement is very high as it is an unlikely event. A very often used measure of such a gain is

$$\Delta I = \log (1/p_i) = - \log (p_i) \qquad (4.4)$$

In order to justify the logarithmic function the following points can be stated.

a) It gives additive property of information. To make it more clear, suppose two independent events m and n, with

probabilities of occurrence p_m and p_n, have occurred jointly, then the additive property says

$$\Delta I (p_m \cdot p_n) = \Delta I (p_m) + \Delta I (p_n) \qquad (4.5)$$

where $(p_m \cdot p_n)$ is the probability of joint occurrence of the events m and n. Thus the additive property can be stated as follows. The information gain from the joint occurrence of more than one independent events is equal to the sum of information gain from their individual occurrence.

b) The gain in information from an absolutely certain event is zero, i.e., $\Delta I(p_i = 1) = 0$.

c) As p_i increases, $\Delta I(p_i)$ decreases.

Referring back to our experiment of die, the expected gain in information from the experiment can be written as

$$H = E (\Delta I) = -\sum_{i=1}^{6} p_i \log(p_i) \; . \qquad (4.6)$$

The value of H denotes the entropy(Shannon's entropy) of the system. Thus for an n-state source, the entropy may be defined as in equation (4.1).

4.3 NEW DEFINITION OF ENTROPY [30,68,69]

4.3.1 Justification

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

a) It is to be noted from the logarithmic entropic measure

that as p_i ——> 0, $\Delta I(p_i)$ ——> ∞ and $\Delta I(p_i) = -\log(p_i)$ is not defined for $p_i = 0$. On the other hand, as p_i ——> 1, $\Delta I(p_i)$ ——> 0 and $\Delta I(p_i = 1) = 0$.

Thus we see that information gain from an event is neither bounded at both ends nor defined at all points. In practice, the gain in information from an event, whether highly probable or highly unlikely, is expected to lie between two finite limits. For example, as more and more pixels in an image are analysed, the gain in information increases and when all the pixels are inspected the gain attains its maximum value, irrespective of the content of the image.

b) In Shannon's theory the measure of ignorance or the gain in information is taken as $\log(1/p_i)$, i.e., ignorance is inversely related to p_i. But mathematically, a more sound expression is possible to arrive at. If u_i is the uncertainty of the i th event then using the knowledge of probability one can write $u_i = 1-p_i$. Since u_i is the unlikeliness (i.e., the probability of non occurrence), statistically ignorance can be better representd by $(1-p_i)$ than $(1/p_i)$.

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

$$\Delta I(p_i) = \log(1-p_i) \qquad (4.7)$$

then $\Delta I < 0$ which is intuitively unappealing. Furthermore, consideration of $-\log(1-p_i)$ as the gain in information leads to

the fact that $\Delta I(p_i)$ increases with p_i, which is again not desirable.

The above problem can be circumvented by considering an exponential function of $(1-p_i)$ instead of the logarithmic behavior. This is also appropriate while considering the concept of information gain in an image.

For example consider Fig. 4.1 Suppose the images [Figs. 4.1.1-4.1.5] have only two gray levels: one corresponding to the lines (black portions) and the other corresponding to the white portion. In the case of the first image we have analysed only few black pixels and from this image we can't say firmly about the content of the image. At this stage we see that it can be either a curtain or the hair of a face or something else. From the image 4.1.2 we can say that it is not a curtain (i.e., some gain in knowledge) while from the image 4.1.3 one can realise that it is a face. The image 4.1.4 says that it is a face with the mouth. However, the image 4.1.5 does not say any thing more than what is described by image 4.1.4, though the number of black pixels (hence probability) has increased.

Let $\Delta I(1)$, $\Delta I(2)$, $\Delta I(3)$, $\Delta I(4)$ and $\Delta I(5)$ be the information content of the images 4.1.1 thru 4.1.5 respectively. Now define the following quantities, representing the changes in gain;

$$G1 = \Delta I(2) - \Delta I(1) \qquad (4.8)$$

$$G2 = \Delta I(3) - \Delta I(2) \qquad (4.9)$$

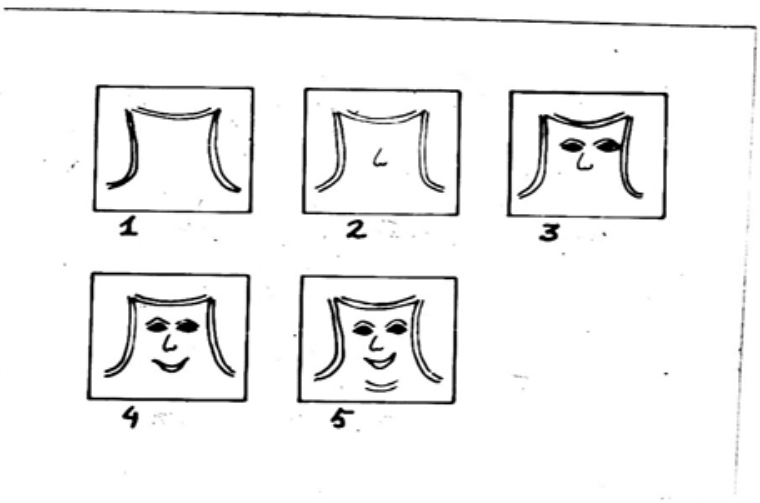


Fig. 4.1 Variation of gain in information in images

$$G3 = \Delta I(4) - \Delta I(3) \qquad (4.10)$$

$$G4 = \Delta I(5) - \Delta I(4) \qquad (4.11)$$

Obviously G1 > G2 > G3 > G4 $\cong$ 0.

The above observation and the fact that information gain approaches a finite limit when more and more pixels (increase in N_i and hence p_i) are analysed, strengthen the assertion that the gain in information (i.e., increase in knowledge or decrease in ignorance) is exponential in nature.

c) The additive property of the gain function for independent events does not have a strong relevance(impact) in practice. It can be justified as follows. The gain function ΔI is used to define the entropy of a system. If the system has n different states, only one of the states can occur at a time. The question of the additive property of the gain function then does not arise for defining the entropy of only one system. If we consider, on the other hand, the joint occurrence of two events (outcome of two independent systems), then there is no point in establishing a relationship between the gains of two events as these are the outcomes of two independent systems. For example, consider the cases of two independent systems viz., throwing a die and shooting a target. For the first system, the possible outcomes are 1,2,...,6; whereas, there are only two (success or failure) outcomes for the second experiment. Suppose the results of the two experiments are '5' and 'failure', respectively. Then, is there any justification in establishing a relationship between the gains associated with the two events? The answer is possibly 'no'.

4.3.2 **Definition**

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 : $\underset{p_i \text{ ---> } 0}{Lim} \Delta I(p_i) = \Delta I(p_i=0) = k_1$, $k_1 > 0$ and finite.

P3 : $\underset{p_i \text{---> } 1}{Lim} \Delta I(p_i) = \Delta I(p_i = 1) = k_2$, $k_2 > 0$ and finite.

P4 : $k_2 < k_1$

P5 : With increase in p_i, $\Delta I(p_i)$ decreases exponentially. In other words, with increase in the uncertainty (u_i) the gain in information increases 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, $H(p_1,\ldots,p_n) \leq H(1/n,\ldots,1/n)$.

P8 : In case of 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 $p = 0.5$.

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

$$\Delta I(p_i) = e^{u_i} = e^{(1 - p_i)} \qquad (4.12)$$

and

the entropy $H = E(\Delta I) = \sum_{i=1}^{n} p_i \, e^{(1 - p_i)}$ (4.13)

The normalised entropy H_{nor} can be defined as follows,

$$H_{nor} = k (H - 1) \text{ where } K = 1 / (e^{1-1/n} - 1) \quad (4.14)$$

It is easy to see that the properties p1 thru P6 are satisfied when k_1 and k_2 take the values *e* and 1 respectively. The proofs of properties P7 and P8 are given in the next section.

Figure 4.2 gives the plot of Shannon's entropy and the new entropy (normalised) for a two state system. It is quite interesting to note that the two curves are almost identical indicating that the basic nature of the information measure remains the same with the new definition.

4.3.3 Proofs

Proof of P7.

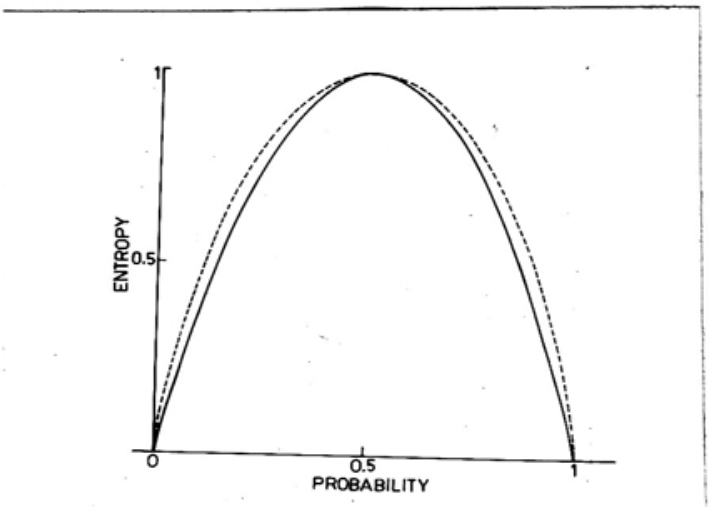
$$H = \sum_{i=1}^{n} p_i \, e^{(1-p_i)} \quad (4.15)$$

$$0 \le p_i \le 1 \quad \text{and} \quad \sum_{i=1}^{n} p_i = 1 \quad (4.16)$$

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

$$(1 - p_1 - p_2 - \ldots\ldots - p_{n-1}) \, e^{(p_1+p_2+\ldots+p_{n-1})} \quad (4.17)$$

as $p_n = 1 - p_1 - p_2 - \ldots p_{n-1}$.



..... Logarithmic
—— exponential
(normalised)

Fig. 4.2 Plot of Shannon's entropy and exponential entropy for a two state system

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

$$\frac{\partial H}{\partial p_i} = 0 \ , \ i = 1,2,....,n-1 \qquad (4.18)$$

or $e^{(1-p_i)} - p_i e^{(1-p_i)} - e^{(p_1+p_2+....+p_{n-1})} +$

$$(1-p_1-p_2-...-p_{n-1}) e^{(p_1+p_2+....+p_{n-1})} = 0 \qquad (4.19)$$

$$\text{or } (1-p_i) e^{(1-p_i)} = (p_1+p_2+....+p_{n-1}) e^{(p_1+p_2+....+p_{n-1})} \qquad (4.20)$$

Now writing $1 - p_i = x_i$ (4.21)

and

$$p_1 + p_2 +....+ p_{n-1} = y \qquad (4.22)$$

we get $x_i e^{x_i} = y e^{y}$ (4.23)

Now define a function $f(x) = x e^{x}$ $0 \leq x \leq 1$ (4.24)

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 $x_1 > x_2$ then $e^{x_1} > e^{x_2}$ ===> $f(x_1) > f(x_2)$

and

if $x_1 < x_2$ then $e^{x_1} < e^{x_2}$ ===> $f(x_1) < f(x_2)$

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 equation 4.23 one can write

$$x_i = y \quad \text{for } i = 1,2,\ldots,n-1$$

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

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

$$\sum_{i=1}^{n-1} (1-p_i) = \sum_{i=1}^{n-1} (p_1 + p_2 + \ldots + p_{n-1}) \qquad (4.25)$$

or $$(n-1) - (p_1 + p_2 + \ldots + p_{n-1})$$

$$= (n-1)(p_1 + p_2 + \ldots + p_{n-1}) \qquad (4.26)$$

or $$(n-1) = n\,(p_1 + p_2 + \ldots + p_{n-1}) \qquad (4.27)$$

or $$n-1 = n\,(1-p_n) \qquad (4.28)$$

or $$n\,p_n = 1 \qquad (4.29)$$

or $$p_n = 1/n \qquad (4.30)$$

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

Proof of P8.

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

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

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

$$\frac{dH}{dp} = e^{1-p} - p\, e^{1-p} - e^{p} + (1-p)e^{p} \qquad (4.31)$$

$$= (1-p)\, e^{(1-p)} - p\, e^{p} \qquad (4.32)$$

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

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

Therefore, $\frac{dH}{dp} > 0$ if $p \in [0,0.5)$; (4.33.1)

$\frac{dH}{dp} < 0$ if $p \in (0.5,1]$ (4.33.2)

and $\frac{dH}{dp} = 0$ if $p = 0.5$. (4.33.3)

Hence the proof.

4.3.4 Some Additional Properties (Theorems) on The Proposed H.

Theorem 4.3.4.1

$$H = \sum_{i=1}^{n} p_i\, e^{1-p_i}\ , \ 0 \le p_i \le 1\ ;\ \sum_{i=1}^{n} p_i = 1$$

is a concave function.

Proof.

Let $f(p) = p\, e^{1-p}$ (4.34)

then $f'(p) = (1-p)\, e^{1-p}$ (4.35)

and $f''(p) = (p - 2)e^{1-p}$ (4.36)

i.e., $f''(p) < 0$ for $0 \leq p \leq 1$.

We know that a function $f(x)$ is concave in the interval (a,b), if at all points in the interval (a,b) the second derivative of $f(x)$ is negative , i.e., $f''(x) < 0$, hence $f(p)$, defined above is a concave function. It is also known that a sum of concave functions is also a concave function. Hence H is a concave function.

Theorem 4.3.4.2

Any change towards equalization of the probabilities $p_1, p_2, .., p_n$ increases H. Thus if $p_1 < p_2$ and we increase p_1, decreasing p_2 by an equal amount so that p_1 and p_2 become more nearly equal, then H increases.

Proof.

Suppose $p_1 < p_2$. Now p_1 is increased by an amount δ ($\delta > 0$) and p_2 is reduced by the same amount, such that $p_1 + \delta < p_2 - \delta$. With this change p_1 and p_2 has become more equal. In order to prove the theorem, we need to show that,

$$H(p_1 + \delta, p_2 - \delta, p_3, \ldots, p_n) - H(p_1, p_2, p_3, \ldots, p_n) > 0 \quad (4.37)$$

or $$(p_1+\delta)\, e^{1-p_1-\delta} + (p_2-\delta)\, e^{1-p_2+\delta} + \sum_{i=3}^{n} p_i\, e^{1-p_i} - \sum_{i=1}^{n} p_i\, e^{1-p_i} > 0 \quad (4.38)$$

or $(p_1+\delta)\, e^{1-p_1-\delta} + (p_2-\delta)\, e^{1-p_2+\delta} - p_1\, e^{1-p_1} - p_2\, e^{1-p_2} > 0$

or $p_1\, e^{1-p_1}\, (e^{-\delta} - 1) + p_2\, e^{1-p_2}\, (e^{\delta} - 1) +$

$$\delta\, (e^{1-p_1-\delta} - e^{1-p_2+\delta}) > 0 \qquad (4.40)$$

Let $\delta\, (e^{1-p_1-\delta} - e^{1-p_2+\delta}) = M \qquad (4.41)$

Now $p_1 + \delta < p_2 - \delta$

$\Rightarrow 1 - p_1 - \delta > 1 - p_2 + \delta$

$\Rightarrow \delta\, (e^{1-p_1-\delta} - e^{1-p_2+\delta}) > 0$

i.e., $M > 0$.

Let $p_1\, e^{1-p_1}\, (e^{-\delta} - 1) + p_2\, e^{1-p_2}\, (e^{\delta} - 1) = K \qquad (4.42)$

or $K = p_2\, e^{1-p_2}\, (e^{\delta} - 1) - p_1\, e^{1-p_1}\, (1 - e^{-\delta}) \qquad (4.43)$

or $K = p_2\, e^{1-p_2}\, (e^{\delta} - 1) - p_1\, e^{1-p_1}\, \left(\frac{e^{\delta} - 1}{e^{\delta}}\right) \qquad (4.44)$

Since $p\, e^{1-p}$ is a monotonically increasing function of p, for $p \in [0,1]$, $p_1\, e^{1-p_1} < p_2\, e^{1-p_2}$. In addition to this $\{(e^{\delta}-1)/e^{\delta}\} < (e^{\delta} - 1)$ as $e^{\delta} > 1$ for $\delta > 0$.

Hence $K > 0$.

Thus

$$H(p_1+\delta, p_2-\delta, p_3, \ldots, p_n) - H(p_1, p_2, \ldots, p_n) = M + K > 0 \qquad (4.45)$$

Hence the proof.

Theorem 4.3.4.3

If any zero probability is changed to a nonzero probability by reducing some other probability then the entropy increases. i.e.,

$$H(\delta, p_2-\delta, p_3, \ldots, p_n) - H(0, p_2, \ldots, p_n) > 0.$$

Proof

$$H(\delta, p_2-\delta, p_3, \ldots, p_n) - H(0, p_2, \ldots, p_n)$$

$$= \delta\, e^{1-\delta} + (p_2-\delta)\, e^{1-p_2+\delta} + \sum_{i=3}^{n} p_i\, e^{1-p_i} - \sum_{i=2}^{n} p_i\, e^{1-p_i}$$

$$= \delta\, e^{1-\delta} + (p_2-\delta)\, e^{1-p_2+\delta} - p_2\, e^{1-p_2} \qquad (4.46)$$

$$= \delta\, e^{1-\delta} + (p_2-\delta)\, e^{1-p_2+\delta} - \delta\, e^{1-p_2} - (p_2-\delta)\, e^{1-p_2} \qquad (4.47)$$

$$= \delta\,(\, e^{1-\delta} - e^{1-p_2}\,) + (\, p_2 - \delta\,) e^{1-p_2} (e^{\delta} - 1)$$

Now $\delta < p_2$

$\Rightarrow 1 - \delta > 1 - p_2$

$\Rightarrow e^{1-\delta} > e^{1-p_2}$

$\Rightarrow \delta(\, e^{1-\delta} - e^{1-p_2}\,) > 0$

Again $\delta > 0$

$\Rightarrow e^{\delta} > 1$

$\Rightarrow (e^{\delta} - 1) > 0$

$\Rightarrow (\, p_2 - \delta\,) e^{1-p_2} (e^{\delta} - 1) > 0$

Hence the proof.

Theorem 4.3.4.4

H is minimum, if and only if all the p_i but one are zero, this one having the value of unity.

Proof.

$p_i \geq 0$ and $\sum p_i = 1$ define a convex region and H is a concave function defined over it. We have already seen in theorem 4.3.4.3 that if any non-zero probability is splitted into two non-zero components, the entropy of the system increases. Suppose in the system two states i and j , have probabilities p_i and p_j . Now if the probability of the states i and j are changed to 0 and $(p_i + p_j)$, the entropy of the system will be reduced. This reduction process of entropy can be continued, till $(n - 1)$ states assume zero probabilities and one state assume the probability value of unity(1) . Further reduction is not possible as $\Sigma p_i = 1$. Moreover, this is true for any probability vector. In addition to this H is a symmetric function (i.e., permutation of p_i does not change the entropy) . Thus, H will be minimum when all p_i s are zero except one , which takes a value of unity.

Hence,

$$H(1,0,0,\ldots,0) = H(0,1,0,\ldots,0) = \ldots\ldots = H(0,0,\ldots,1) = H_{min}$$

4.3.5 Properties of Compound Probabilistic Experiment

Consider two probabilistic experiment A and B with possible outcomes $(a_1, a_2, \ldots\ldots, a_n)$ and $(b_1, b_2, \ldots, b_m)$ respectively.

Define (A * B) as a compound probabilistic experiment with outcomes of the form (a_k, b_l); $k=1,2,\ldots,n$; $l=1,2,\ldots,m$.

Let p_{kl} be the probability of the event (a_k, b_l) . Now, if p_k be the probability of the outcome a_k in the first experiment regardless of the second then,

$$p_k = \sum_{l=1}^{m} p_{kl} \tag{4.48}$$

Similarly, q_l , the probability of b_l in the second experiment regardless of the first is

$$q_l = \sum_{k=1}^{n} p_{kl} \tag{4.49}$$

Now, $p(a_k/b_l)$, the conditional probability of a_k given that b_l has occured is given by

$$p(a_k/b_l) = \frac{p_{kl}}{q_l} = p_{k/l} \quad \text{(say)} \tag{4.50}$$

Similarly, $$p(b_l/a_k) = \frac{p_{kl}}{p_k} = q_{l/k} \tag{4.51}$$

Therefore, the entropy of the system A, given that b_l has occured in the experiment B is given by

$$H_n(A/b_l) = \sum_{k=1}^{n} p_{k/l} \; e^{1-p_{k/l}} \tag{4.52}$$

and

$$H_n(A/B) = \sum_{l=1}^{m} \sum_{k=1}^{n} q_l \; p_{k/l} \; e^{1-p_{k/l}} \tag{4.53}$$

Similarly,

$$H_m(B/A) = \sum_{k=1}^{n} \sum_{l=1}^{m} p_k \, q_{l/k} \, e^{1-q_{l/k}} \qquad (4.54)$$

Again, the entropy of the compound experiment (A * B) is defined as

$$H_{nm}(A * B) = \sum_{k=1}^{n} \sum_{l=1}^{m} p_{kl} \, e^{1-p_{kl}} \qquad (4.55)$$

Under the above framework, the following theorems can be stated .

Theorem 4.3.5.1.

$$H_m(B/A) \leq H_m(B)$$

and

$$H_n(A/B) \leq H_n(A)$$

This theorem says that the average amount of gain in information from the realisation of the experiment A (or B) can decrease if another experiment B (or A) has already been realised.

Proof.

We know that for any real valued continuous concave function $f(x)$ over [0,1] and for any set of $(x_1, x_2, \ldots, x_n)$ over [0,1] the following inequality holds good.

$$\sum_{k=1}^{n} \lambda_k \, f(x_k) \leq f\left(\sum_{k=1}^{n} \lambda_k x_k\right) \quad \text{(Jensen inequality)} \qquad (4.56)$$

where λ_i s are non-negative real numbers such that $\sum_{i=1}^{n} \lambda_i = 1$.

It has been already proved that $f(x) = x\, e^{1-x}$ is a concave function over [0,1]. Now taking $\lambda_k = p_k$ and $x_k = q_{l/k}$, we have

$$\left(\sum \lambda_k = 1 \text{ and } 0 \le x_k \le 1 \right)$$

$$\sum_{k=1}^{n} p_k\, q_{l/k}\, e^{1-q_{l/k}} \le \left(\sum_{k=1}^{n} p_k\, q_{l/k} \right) e^{1-\sum p_k q_{l/k}} \qquad (4.57)$$

$$\text{or} \quad \sum_{k=1}^{n} p_k\, q_{l/k}\, e^{1-q_{l/k}} \le q_l\, e^{1-q_l} \qquad (4.58)$$

Now summing over l, we get

$$H_m(B/A) \le H_m(B) \qquad (4.59)$$

Similarly, it can also be proved that

$$H_n(A/B) \le H_n(A) \qquad (4.60)$$

Theorem 4.3.5.2

$$H_{nm}(A * B) \le H_n(A) + H_m(B/A)$$

Proof.

$$H_{nm}(A * B) = \sum_{k=1}^{n} \sum_{l=1}^{m} p_{kl}\, e^{1-p_{kl}} \qquad (4.61)$$

$$= \sum_{k=1}^{n} \sum_{l=1}^{m} p_k\, q_{l/k}\, e^{1-p_k\, q_{l/k}} \qquad (4.62)$$

$$\le \sum_{k=1}^{n} \sum_{l=1}^{m} p_k\, q_{l/k} \left(e^{1-p_k} + e^{1-q_{l/k}} \right) \qquad (4.63)$$

(as $e^{1-x} + e^{1-y} \ge e^{1-xy}$; $0 \le x, y \le 1$; will be shown latter)

$$= \sum_{k=1}^{n} \sum_{l=1}^{m} p_k\, q_{l/k}\, e^{1-p_k} + \sum_{k=1}^{n} \sum_{l=1}^{m} p_k\, q_{l/k}\, e^{1-q_{l/k}} \qquad (4.64)$$

$$= \sum_{k=1}^{n} \left(\sum_{l=1}^{m} p_{kl} \right) e^{1-p_k} + \sum_{k=1}^{n} \sum_{l=1}^{m} p_k \, q_{l/k} \, e^{1-q_{l/k}} \qquad (4.65)$$

$$= \sum_{k=1}^{n} p_k \, e^{1-p_k} + \sum_{k=1}^{n} \sum_{l=1}^{m} p_k \, q_{l/k} \, e^{1-q_{l/k}} \qquad (4.66)$$

$$= H_n(A) + H_m(B/A) \qquad (4.67)$$

Hence,

$$H_{nm}(A * B) \leq H_n(A) + H_m(B/A) \qquad (4.68)$$

Proof of $e^{1-x} + e^{1-y} \geq e^{1-xy}; \quad 0 \leq x, y \leq 1.$

we need to show

$$e^{1-x} + e^{1-y} \geq e^{1-xy}; \quad 0 \leq x, y \leq 1$$

or $e^{-x(1-y)} + e^{-y(1-x)} \geq 1$

We know that $log(1-x) = -x - \frac{x^2}{2} - \frac{x^3}{3} - .....$

or $log\ (1-x) \leq -x$

or $1-x \leq e^{-x}$

Using the above result, we can write that

$e^{-x(1-y)} + e^{-y(1-x)} \geq 1 - x(1-y) + 1 - y(1-x)$

or $e^{-x(1-y)} + e^{-y(1-x)} \geq 2 - x - y + 2xy$

or $e^{-x(1-y)} + e^{-y(1-x)} \geq 2 - x - y + xy$

or $e^{-x(1-y)} + e^{-y(1-x)} \geq 1 + (1-y)\ (1-x)$

or $e^{-x(1-y)} + e^{-y(1-x)} \geq 1$ as $0 \leq x, y \leq 1.$

or $e^{1-x} + e^{1-y} \geq e^{1-xy}$

Hence the proof.

Theorem 4.3.5.3

$$H_{nm}(A * B) \leq H_n(A) + H_m(B)$$

This result follows from theorems 4.3.5.1 and 4.3.5.2.

Theorem 4.3.5.4

Let $p(s_i)$ be the probability of a sequence of symbols s_i of length q, then $H^{(q)}$, the entropy of order q of the system is defined as

$$H^{(q)} = \frac{1}{q} \sum_i p(s_i)\, e^{1-p(s_i)} \tag{4.69}$$

where the summation is taken over all possible sequence of length q. If the information source is stationary, then the entropy of the source H is given by

$$H = \lim_{q \to \infty} H^{(q)} = \lim_{q \to \infty} \frac{1}{q} H_q \quad , \tag{4.70}$$

denoting

$$\sum_i p(s_i)\, e^{1-p(s_i)} = H_q \tag{4.71}$$

Proof.

Suppose the source can generate symbols

$$(g_1, g_2, \ldots, g_L) = G_L \quad \text{(say)} \tag{4.72}$$

Let us now consider a sequence of symbols $(x_1, x_2, .., x_n)$, $x_i \in G_L$, such that x_i occurs at time t_i. Let us denote such a sequence by $(x_1/t_1, \ldots, x_n/t_n)$. Suppose $p(x_1/t_1, \ldots, x_n/t_n)$ represents the probability of the sequence $(x_1/t_1, \ldots, x_n/t_n)$. The information source is said to be stationary, if for any integer r the following is true.

$$p(x_1/t_1+r,\ldots,x_n/t_n+r) = p(x_1/t_1,\ldots,x_n/t_n) \qquad (4.73)$$

Let $s_q^t = (x_1/t, x_2/t+1, \ldots, x_q/t+q-1)$ and S_q^t is the finite probability space whose elementery events are sequences of the form s_q^t. It is to be noted that there are L^q such sequences. Therefore, the entropy of such a finite probability space S_q^t can be written as

$$H(S_q^t) = \sum_{S_q^t} p(x_1/t,\ldots,x_q/t+q-1)\, e^{1-p(x_1/t,\ldots,x_q/t+q-1)} \qquad (4.74)$$

Since the source is stationary, we can write that, $s_q^t = s_q^0$.

Hence

$$H(S_q^t) = H(S_q^0) = H^{(q)} \qquad (4.75)$$

i.e., the entropy $H(S_q^t)$ does not depend on t, but on the probabilistic structure of the source and q. Now consider a sequence of length $(q+k)$ as,

$$s_{q+k}^0 = (x_1/0, x_2/1, \ldots, x_q/q-1, x_{q+1}/q, \ldots, x_{q+k}/q+k-1) \qquad (4.76)$$

and the associated finite probability space S_{q+k}^0 .

Now,

$$s_{q+k}^0 = (x_1/0, x_2/1, \ldots, x_q/q-1) \cap (x_{q+1}/q, \ldots, x_{q+k}/q+k-1) \qquad (4.77)$$

$$= s_q^0 \cap s_k^q \quad ; \text{ as this implies a sequence of length } (q+k)$$

in which s_k^q follows s_q^0 .

Thus we have,

$$S^0_{q+k} = S^0_q * S^q_k \qquad (4.78)$$

i.e., the probability space S^0_{q+k} is equal to the compound (product) probabilistic space $S^0_q * S^q_k$.

We have already seen that (theorem 4.3.5.2)

$$H(A * B) \leq H(A) + H(B/A)$$

So
$$H(S^0_q * S^q_k) \leq H(S^0_q) + H(S^q_k / S^0_q) \qquad (4.79)$$

$$\leq H(S^0_q) + H(S^q_k) \qquad (4.80)$$

(by theorem 4.3.5.1)

Since the source is stationary, $S^q_k = S^0_k$

$$H(S^0_q * S^q_k) \leq H(S^0_q) + H(S^0_k) \qquad (4.81)$$

or
$$H(S^0_{q+k}) \leq H(S^0_q) + H(S^0_k) \qquad (4.82)$$

For the sake of notational simplicity, we can write,

$$H^{(q+k)} \leq H^{(q)} + H^{(k)} \qquad (4.83)$$

From equation 4.83 one can write

$$H^{(nk)} \leq n. H^{(k)} \qquad (4.84)$$

Taking $k = 1$, $H^{(n)} \leq n H^{(1)}$ (4.85)

or
$$\frac{H^{(n)}}{n} \leq H^{(1)} \qquad (4.86)$$

Where $H^{(1)}$ is the entropy of the finite probability space S^0_1 (i.e, only sequence of length one).

Now when the source can geneate L distinct symbols, $H^{(1)} \leq e^{1-1/L}$, (already proved). Now since $H^{(n)}/n \leq H^{(1)}$ we can write,

$$\inf_n \frac{H^{(n)}}{n} \leq H^{(1)} \tag{4.87}$$

Let ws write $$H = \inf_n \frac{H^{(n)}}{n} \tag{4.88}$$

Thus for every $\varepsilon > 0$, there exists a positive integer m, such that

$$\frac{H^{(m)}}{m} - H < \varepsilon \tag{4.89}$$

or $$\frac{H^{(m)}}{m} < H + \varepsilon \tag{4.90}$$

Now, we can always find a pair of integers n and r such that $n > m$ and $(r-1)m < n \leq rm$. From equations 4.83 and 4.84 we can write

$$H^{(n)} \leq H^{(rm)} \leq r\, H^{(m)} \tag{4.91}$$

or $$H^{(n)} \leq r\, H^{(m)} \tag{4.92}$$

or $$\frac{H^{(n)}}{n} \leq \frac{r}{n} H^{(m)} \tag{4.93}$$

or $$\frac{H^{(n)}}{n} \leq \frac{rm}{n} \frac{H^{(m)}}{m} \tag{4.94}$$

or $$\frac{H^{(n)}}{n} \leq \frac{r}{r-1} \frac{H^{(m)}}{m} \tag{4.95}$$

(as $r/(r-1) > (rm)/n$ will be shown latter)

or $$\frac{H^{(n)}}{n} < \frac{r}{r-1} (H + \varepsilon) \tag{4.96}$$

(Using equation 4.90)

Therefore,

$$Lim_{n}\ Sup\ \frac{H^{(n)}}{n} \leq H + \varepsilon \qquad (4.97)$$

(as $n \to \infty$, $r \to \infty$ and $\lim_{r\to\infty} (r/(r-1)) \to 1$)

Since ε is a positive number, we can write

$$Lim_{n}\ Sup\ \frac{H^{(n)}}{n} \leq H \qquad (4.98)$$

Again, we have seen that (equation 4.88)

$$\frac{H^{(n)}}{n} \geq H$$

Therefore, $$Lim_{n}\ Inf\ \frac{H^{(n)}}{n} \geq H \qquad (4.99)$$

Now combining equations 4.98 and 4.99, we get

$$\lim_{n\to\infty} \frac{H^{(n)}}{n} = H \quad ; \text{ Hence the proof.}$$

Proof of ($r/(r-1) > (rm)/n$)

$n/(rm) = (rm - x)/(rm)$ where $0 \leq x \leq m-1$

$= 1 - x/(rm)$

Since $0 \leq x \leq m-1$, $x/(rm) < 1/r$

Now $(r-1)/r = 1 - 1/r < 1 - x/(rm)$

$= (rm-x)/(rm)$

$= n/(rm)$

Thus $(r-1)/r < n/(rm)$

$\Rightarrow r/(r-1) > (rm)/n$

An Alternative Proof of P7.

We have already proved that H is a real valued concave function defined on the interval [0,1]. In order to prove **P7**, we

shall make use of the well known Jensen inequality for a real valued continuous function.

Let $f(x)$ be a real valued continuous concave function defined on the interval $[a,b]$. Then for any $x_1, x_2, \ldots, x_n \in [a,b]$ and for any set of non negative real numbers $a_1, a_2, \ldots, a_n$ such that $\sum_{i=1}^{n} a_i = 1$, we have

$$\sum_{i=1}^{n} a_i f(x_i) \leq f\left(\sum_{i=1}^{n} a_i x_i\right) \tag{4.100}$$

We have already seen that $f(x) = x\, e^{1-x}$ is a real valued continuous concave function over $[0,1]$.

Therefore, setting

$$a = 0,\ b = 1,\ x_i = p_i,\ a_i = 1/n$$

and $f(x) = x\, e^{1-x}$ we get,

$$\sum_{i=1}^{n} \frac{1}{n} p_i\, e^{1-p_i} \leq \left(\sum_{i=1}^{n} \frac{1}{n} p_i\right) e^{\left(1 - \sum \frac{1}{n} p_i\right)} \tag{4.101}$$

$$\text{or} \quad \frac{1}{n} \sum_{i=1}^{n} p_i\, e^{1-p_i} \leq \frac{1}{n}\, e^{1-1/n} \tag{4.102}$$

$$\text{or} \quad \sum_{i=1}^{n} p_i\, e^{1-p_i} \leq e^{1-1/n} \tag{4.103}$$

$$\text{Thus} \quad H(p_1, p_2, \ldots, p_n) \leq H\left(\frac{1}{n}, \frac{1}{n}, \ldots, \frac{1}{n}\right) \tag{4.104}$$

Hence the proof.

4.3.6 Extension of the New Concept to Fuzzy Sets

Based on the exponential entropy, let us define a new expression for the entropy of a fuzzy set A as

$$H'(A) = \frac{1}{n} \sum_{i=1}^{n} [\ \mu_A(x_i)\ e^{1-\mu_A(x_i)} + (1-\mu_A(x_i))\ e^{\mu_A(x_i)}\ ] \qquad (4.105)$$

Like the equation 1.28, H'(A) also satisfies all the properties P1 to P4 of section 1.2.5.2. Proofs of those properties for equation 4.105 are given in the next section. Therefore, equation 4.105 can be regarded as a measure of fuzziness in a set which gives the average amount of difficulty (ambiguity) in deciding whether an element would be considered to be a member of a set or not.

It is to be noted that this definition has been given here as an application of the exponential entropy. A more general and informative entropy definition of fuzzy sets will be introduced in chapter 6 along with its application. Furthermore, an attempt is also made there to combine the probabilistic and fuzzy uncertainties by defining a measure, "hybrid entropy" of a set. This entropy measure reduces to the classical entropy of a two state system when the fuzziness is removed.

4.3.6.1 Proofs for the Entropy of Fuzzy Sets

For the sake of notational simplicity $\mu_A(x_i)$ will be written as μ_i.

Proof of P3.

(We shall prove P3 first and then P1 and P2 will be proved , this will make the proof of P1 and P2 easier)

$$H'(A) = \frac{1}{n} \sum_{i=1}^{n} \{ \mu_i e^{(1-\mu_i)} + (1-\mu_i) e^{\mu_i} \} \quad (4.106)$$

$$= \frac{1}{n} \sum_{i=1}^{n} f(\mu_i) \quad (4.107)$$

where $f(\mu_i) = \mu_i e^{(1-\mu_i)} + (1-\mu_i) e^{\mu_i}$ (4.108)

In order to prove P3 it is enough to show that $f(\mu_i)$ monotonically increases for $\mu_i \in [0,0.5)$, monotonically decreases for $\mu_i \in (0.5,1]$ and attains the maximum value at $\mu_i = 0.5$.

Now differentiating $f(\mu_i)$ with respect to μ_i we get,

$$\frac{df}{d\mu_i} = (1-\mu_i) e^{1-\mu_i} - \mu_i e^{\mu_i} \quad (4.109)$$

Now if $\mu_i \in [0,0.5)$, then $(1-\mu_i) e^{1-\mu_i} > \mu_i e^{\mu_i}$;

if $\mu_i \in (0.5,1]$ then $(1-\mu_i) e^{1-\mu_i} < \mu_i e^{\mu_i}$

and if $\mu_i = 0.5$ then $(1-\mu_i) e^{1-\mu_i} = \mu_i e^{\mu_i}$.

Therefore, we see that

$$\frac{df}{d\mu_i} > 0 \quad \text{if } \mu_i \in [0,0.5) \quad (4.110)$$

$$< 0 \quad \text{if } \mu_i \in (0.5,1] \quad (4.111)$$

$$= 0 \quad \text{if } \mu_i = 0.5. \quad (4.112)$$

Hence the proof.

Proof of **p1**.

If $\mu_i = 0$ or 1 then

$$\mu_i e^{(1-\mu_i)} + (1 - \mu_i) e^{\mu_i} = 1 \quad (4.113)$$

It has been already shown that such a funtion monotonically increases with μ over $\mu \in [0,0.5)$ and monotonically decreases with μ for $\mu \in (0.5,1]$. Therefore, each term of H'(A) assumes the minimum value of 1, when $\mu_i = 0$ or 1, as shown in equation 4.113. In other words, $\underset{\mu_i \in [0,1]}{minimum} \{\mu_i\ e^{(1-\mu_i)} + (1-\mu_i)\ e^{\mu_i}\} = 1$ and it is attained when $\mu_i = 0$ or 1. Therefore, H'(A) will attain the minimum value when $\mu_i = 0$ or 1 for all $i = 1,2,\ldots,n$.

The minimum value of $H'(A) = \frac{1}{n} \sum 1 = 1$.

Proof of P2

It has already been shown that,

$f(x) = x\ e^{(1-x)} + (1-x)\ e^{x}$ is a concave function of x over [0,1] and it assumes the maximum when x=0.5. Now, since H'(A) is a summation of n such functions. Therefore, H'(A) attains the global maximum value when each of $f(\mu_i)$ attains the maximum value. Thus H'(A) attains the global maximum value when each $\mu_i = 1/2$; for $i=1,\ldots n$.

Proof of P4.

$$H'(A) = \frac{1}{n} \sum_{i=1}^{n} \{ \mu_i\ e^{1-\mu_i} + (1-\mu_i)\ e^{\mu_i} \} \qquad (4.114)$$

$$= \frac{1}{n} \sum_{i=1}^{n} \{ (1-\mu_i)\ e^{\mu_i} + \mu_i\ e^{1-\mu_i} \qquad (4.115)$$

$$= H'(\bar{A}) \quad .$$

Hence the proof.

4.3.7. Additional Theorem on H'(A).

Theorem 4.3.7.1. Let A and B are two fuzzy subsets in a space of points $X = \{x_i\}$ with membership functions $\mu_A(x_i)$ and $\mu_B(x_i)$ respectively. Then

$$H'(A \cup B) + H'(A \cap B) = H'(A) + H'(B) \qquad (4.116)$$

Proof

Let us denote

$$D = \{ x_i \ / \ \mu_A(x_i) \geq \mu_B(x_i) \} \qquad (4.117)$$

and

$$D' = \{ x_i \ / \ \mu_A(x_i) < \mu_B(x_i) \qquad (4.118)$$

Then,

$$D \cup D' = X \qquad (4.119)$$

and

$$D \cap D' = \varnothing \qquad (4.120)$$

Let us define

$$f(\mu) = \mu \, e^{1-\mu} + (1-\mu) \, e^{\mu} \qquad (4.121)$$

then $$H'(A) = \frac{1}{n} \sum_{x_i \in X} f(\mu_A(x_i)) \qquad (4.123)$$

and

$$H'(B) = \frac{1}{n} \sum_{x_i \in X} f(\mu_B(x_i)) \qquad (4.124)$$

$$H'(A \cup B) = \frac{1}{n} \sum_{x_i \in X} f(\mu_{A \cup B}(x_i)) \qquad (4.125)$$

and

$$H'(A \cap B) = \frac{1}{n} \sum_{x_i \in X} f(\mu_{A \cap B}(x_i)) \qquad (4.126)$$

We know that,

$$\mu_{A\cup B}(x_i) = max \{ \mu_A(x_i) , \mu_B(x_i) \}, \quad \forall x_i \in X \tag{4.127}$$

and $$\mu_{A\cap B}(x_i) = min \{ \mu_A(x_i) , \mu_B(x_i) \}, \quad \forall x_i \in X \tag{4.128}$$

Therefore, we can write,

$$\sum_{x_i \in X} f(\mu_{A\cup B}(x_i)) = \sum_{x_i \in D} f(\mu_A(x_i)) + \sum_{x_i \in D'} f(\mu_B(x_i)) \tag{4.129}$$

Similarly,

$$\sum_{x_i \in X} f(\mu_{A\cap B}(x_i)) = \sum_{x_i \in D'} f(\mu_A(x_i)) + \sum_{x_i \in D} f(\mu_B(x_i)) \tag{4.130}$$

Now taking the sum of equations 4.129 and 4.130 we obtain the following.

$$\sum_{x_i \in X} f(\mu_{A\cup B}(x_i)) + \sum_{x_i \in X} f(\mu_{A\cap B}(x_i)) =$$

$$\sum_{x_i \in X} f(\mu_A(x_i)) + \sum_{x_i \in X} f(\mu_B(x_i)) \tag{4.131}$$

as $D \cup D' = X$ and $D \cap D' = \emptyset$.

Now dividing equation 4.131 by n, we get

$$H'(A \cup B) + H'(A \cap B) = H'(A) + H'(B) \tag{4.132}$$

and this completes the proof.

4.4. ENTROPY OF AN IMAGE [30].

4.4.1 Higher Order Entropy (Exponential).

We have already explained in chapter 3 that pixel intensities in an image are not independent of each other. This

dependency of pixel intensities can be incorporated by considering sequences of pixels to estimate the entropy of an image. Based on theorem 4.3.5.4 the following definition for the higher order entropy of an image can be given.

Let $p(s_i)$ be the probability of a sequence s_i of gray levels of length q. Then the Entropy of order q is defined as follows.

$$H^{(q)} = \frac{1}{q} \sum_i p(s_i)\, e^{1-p(s_i)} \qquad (4.133)$$

Different values of q will generate various orders of entropy. It is to be noted that this expression for entropy of order q is different from that described in chapter-3.

As in the case of equations 3.12 and 3.13, the H^1 and H^2 , considering the exponential gain function can be written as

$$H^{(1)} = \sum_{i=0}^{L-1} p_i\, e^{1-p_i} \qquad (4.134)$$

and

$$H^{(2)} = 1/2 \sum_i \sum_j p_{ij}\, e^{1-p_{ij}}. \qquad (4.135)$$

Here H^1 is dependent on the gray level histogram and H^2 is dependent on the co-occurrence matrix of an image.

4.4.2 Conditional Entropy (Exponential).

Under the new frame work, the expression for H(X/Y) and H(Y/X) are given as

$$H(X/Y) = \sum_{x_i \in X} \sum_{y_j \in Y} p(x_i / y_j)\, e^{1-p(x_i/y_j)} \qquad (4.136)$$

and

$$H(Y/X) = \sum_{y_j \in Y} \sum_{x_i \in Y} p(y_j / x_i)\, e^{1-p(y_j/x_i)} \qquad (4.137)$$

where the pixel y_j , in general can be an m th order neighbour of the pixel x_i. When we impose the constraint that x_i and y_j are adjacent pixels, equations 4.136 and 4.137 can be rewritten as.

$$H(X/Y) = \sum_{\substack{x_i \in X \\ (x_i , y_j) \text{ adjacent}}} \sum_{y_j \in Y} p(x_i / y_j)\, e^{1-p(x_i/y_j)} \qquad (4.138)$$

and

$$H(Y/X) = \sum_{\substack{y_j \in Y \\ (y_j , x_i) \text{ adjacent}}} \sum_{x_i \in X} p(y_j / x_i)\, e^{1-p(y_j/x_i)} \qquad (4.139)$$

The conditional entropy $H^{(C)}$ of the image can, therefore, be defined as the average of equations 4.138 and 4.139.

Thus,

$$H^{(C)} = (H(X/Y) + H(Y/X)) / 2 \qquad (4.140)$$

4.5. APPLICATION TO IMAGE SEGMENTATION [30,68]

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

4.5.1 Algorithm Based on Global Information (Algorithm 1)

Following the concept of Kapur et. al.(as mentioned in chapter 3, section 3.2.2, and making use of the equation 4.134 we can determine the thresholding criterion as follows. Suppose s is an assumed threshold which partitions the image into object(black) and background(white). Using equation 4.134 the global entropy of the object (or the black portion) of the image can be defined as

$$H_B^{(G)}(s) = \sum_{i=0}^{s} p_i/P_s \; e^{1-p_i/P_s} \qquad (4.141)$$

where $P_s = \sum_{i=0}^{s} p_i$ and $\sum_{i=0}^{L-1} p_i = 1$;

and the global entropy of the background (or white portion) of the image as

$$H_W^{(G)}(s) = \sum_{i=s+1}^{L-1} p_i/(1-P_s) \; e^{1-p_i/(1-P_s)} \qquad (4.142)$$

Therefore, the total global entropy of the partitioned image is

$$H_T^{(G)}(s) = H_B^{(G)}(s) + H_W^{(G)}(s) \qquad (4.143)$$

Let $\underset{s}{\text{Maximum}} \{ H_T^{(G)}(s) \} = H_T^{(G)}(k)$; $0 \le s,k \le L-1$.

Then the level k can be taken as a threshold for object-background classification of the image. The threshold so obtained will classify the object and background in such a way that the sum of

information in the background and object is maximised; i.e., the resulting distribution of gray levels in the partioned image would be uniform in the best possible way.

However, like the entropic methods of Pun and Kapur et. al. , this algorithm (equations 4.141-4.143) is also seen to be a function of the gray level histogram of the image only. In other words, different images with identical histogram would result in the same threshold value irrespective of the contents of the image.

4.5.2 Algorithms Based on Local Information

4.5.2.1 Algorithm 2.

Let us now describe another algorithm based on equation 4.135 , which takes into account the spatial details of an image. As mentioned in section 3.4.1, an assumed threshold s, $0 \le s \le L-1$, partitions the co-occurrence matrix into four quadrants, namely A, B, C, D [Fig. 3.3], and the corresponding cell probabilities are as follows:

$$p_{ij}^{A} = \frac{t_{ij}}{\sum_{i=0}^{s} \sum_{j=0}^{s} t_{ij}} \qquad (4.144)$$

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

$$p_{ij}^{B} = \frac{t_{ij}}{\sum_{i=0}^{s} \sum_{j=s+1}^{L-1} t_{ij}} \qquad (4.145)$$

for $0 \le i \le s$ and $s+1 \le j \le L-1$.

$$p_{ij}^{C} = \frac{t_{ij}}{\sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} t_{ij}} \qquad (4.146)$$

for $s+1 \leq i, j \leq L-1$

and

$$p_{ij}^{D} = \frac{t_{ij}}{\sum_{i=s+1}^{L-1} \sum_{j=0}^{s} t_{ij}} \qquad (4.147)$$

for $s+1 \leq i \leq s$ and $0 \leq j \leq s$.

The co-occurrence matrix $[t_{ij}]$ used here is defined in section 3.4.1 . Now in terms of the exponential entropy (equation 4.135) the second order local entropy of the object and the background can be defined as follows

$$H_A^{(2)}(s) = \frac{1}{2} \sum_{i=0}^{s} \sum_{j=0}^{s} p_{ij}^{A} \; e^{1-p_{ij}^{A}} \qquad (4.148)$$

and

$$H_C^{(2)}(s) = \frac{1}{2} \sum_{i=s+1}^{L-1} \sum_{j=s+1}^{L-1} p_{ij}^{C} \; e^{1-p_{ij}^{C}} \qquad (4.149)$$

Hence the total second order entropy of the partitioned image can be written as

$$H_T^{(2)}(s) = H_A^{(2)}(s) + H_C^{(2)}(s) \qquad (4.150)$$

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

4.5.2.2 **Algorithm 3.**

Like Algorithm 2 of chapter 3, this algorithm is based on the concept of the conditional (exponential) entropy (equations 4.138-4.140). Let t_{ij} be an entry of the quadrant B [Fig. 3.3].Then, p_{ij}^{B} s of equation 4.145 give the probabilities required by equation 4.138. Similarly, p_{ij}^{D} s of equation 4.147 correspond to the probabilities of equation 4.139.

Therefore,

$$H(\text{Object/Background}) = H(O/B)$$

$$= \sum_{i=0}^{s} \sum_{j=s+1}^{L-1} p_{ij}^{B} \; e^{1-p_{ij}^{B}} \qquad (4.151)$$

and $H(\text{Background/Object}) = H(B/O)$

$$= \sum_{i=s+1}^{L-1} \sum_{j=0}^{s} p_{ij}^{D} \; e^{1-p_{ij}^{D}} \qquad (4.152)$$

Now the conditional entropy of the image is

$$H_T^{(C)}(s) = (H(O/B) + H(B/O)) / 2 \qquad (4.153)$$

In order to get the threshold for object-background classification $H_T^{(C)}(s)$ is maximised with respect to s.

4.6. IMPLEMENTATION AND RESULTS

The performance of the algorithms proposed in section-4.5 has been demonstrated on a set of five images having widely different types of histograms. We shall first of all concentrate on Algorithm 1 and show its superiority over other algorithms based

on global information only. Then the results of algorithms 2 and 3 which are based on local information will be presented.

4.6.1 Results of Methods Based on Global Information

Figures 4.3.1 - 4.7.1 represent the different input images used, while figures 4.3.2 - 4.7.2 depict the corresponding gray level histograms. In order to compare the results of the proposed Algorithm 1 with some of the existing methods, algorithm of Pun[59] and kapur et. al. [33] have been implemented. In addition to these , another algorithm based on Bayes misclassification error [2] has been considered here. Though the details of this method is available in [2], for the convenience of the reader, it is given in brief below.

Suppose the gray level distribution for the object follows $N(\mu_1,\sigma_1^2)$ and that of the background follows $N(\mu_2,\sigma_2^2)$. Then

$$p_i(x) = \frac{1}{\sigma_i \sqrt{2\pi}} \; e^{-\frac{1}{2}\left(\frac{x-\mu_i}{\sigma_i}\right)^2} \qquad (4.154)$$

where x denotes the brightness value and i can take the value 1 or 2 corresponding to probability distribution over the object or background.

Let l be a threshold level for object-background classification. Then the probability of classification of a background pixel as an object pixel is equal to

$$v_1 = \int_{-\infty}^{l} p_2(x)\, dx \qquad (4.155)$$

while the misclassification probability of an object pixel is

$$v_2 = \int_{l}^{\infty} p_1(x)\, dx \qquad (4.156)$$

If P_1 and P_2 are the proportion of pixels in the object and the background respectively (i.e., $P_1 + P_2 = 1$) then the expected value of the misclassification error is given by

$$E_l = P_1 v_2 + P_2 v_1 \qquad (4.157)$$

Therefore, minimising E_l , with respect to l , the threshold level for object-background classification is obtained.

Implementation of such a method requires the estimation of (μ_i, σ_i), $i=1,2$; the parameters of the distributions. For a gray level l (assuming $l+\varepsilon$ as the threshold, ε is a small positive quantity), $(\mu_i, \sigma_i, v_i, P_i)$, $i=1,2$ and E_l are estimated as follows.

$$\hat{\mu}_1 = \left(\sum_{i=0}^{l} i \cdot N_i\right) / \sum_{i=0}^{l} N_i \qquad (4.158)$$

$$\hat{\mu}_2 = \left(\sum_{i=l+1}^{L-1} i \cdot N_i\right) / \sum_{i=l+1}^{L-1} N_i \qquad (4.159)$$

$$\hat{\sigma}_1^2 = \left(\sum_{i=0}^{l} (i-\hat{\mu}_1)^2 \cdot N_i\right) / \sum_{i=0}^{l} N_i \qquad (4.160)$$

$$\hat{\sigma}_2^2 = \left(\sum_{i=l+1}^{L-1} (i-\hat{\mu}_2)^2 \cdot N_i\right) / \sum_{i=l+1}^{L-1} N_i \qquad (4.161)$$

$$\hat{v}_1 = \int_{-\alpha}^{l+\varepsilon} p_2(x)\, dx \qquad (4.162)$$

$$\hat{v}_2 = \int_{l+\varepsilon}^{\alpha} p_1(x)\, dx \qquad (4.163)$$

$$\hat{P}_1 = \sum_{i=0}^{l} N_i \;/\; \sum_{i=0}^{L-1} N_i \qquad (4.164)$$

$$\hat{P}_2 = 1 - \hat{P}_1 \qquad (4.165)$$

and

$$E_l = \hat{P}_1 \cdot \hat{v}_2 + \hat{P}_2 \cdot \hat{v}_1 \qquad (4.166)$$

where N_i is the number of pixels with gray level i and the ε has been introduced to make the range of integration continuous.

The threshold levels produced by different global methods are displayed in Table 4.1. Fig. 4.3.1 represents the image of the biplane with two dominant modes in its gray level histogram [Fig. 4.3.2]. The segmented images produced by different methods are shown in Figs. 4.3.3 - 4.3.6. From the results one can see that except equation 4.166 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 methods of Pun and Kapur et. al. have produced results comparable to that of Algorithm 1.

Figs. 4.4.1-4.4.2 depict the input image of Abraham Lincoln and its gray level histogram respectively. The thresholded images are shown in Figs. 4.4.3-4.4.6. Here also the results produced by equation 4.166 is better than that by the other methods.

TABLE - 4.1

Thresholds for object background classification

(Based on global information)

Images	THRESHOLDS			
	Proposed algorithm -1 (Eqn. 4.143)	Algorithm of Pun [59]	Algorithm Kapur et.al [33]	Bayes classifir
BIPLANE (Fig.4.3)	22	24	21	12
LINCOLN (Fig.4.4)	15	16	15	08
CHROMOSOMES (Fig.4.5)	19	27	20	--
BLURRED CHROMOSOME (Fig. 4.7)	32	32	33	43
BOY (Fig. 4.7)	11	15	13	12

See Fig. 2.7.1
page- 85

See Fig. 2.7.2
page- 85

Fig. 4.3.1 Input

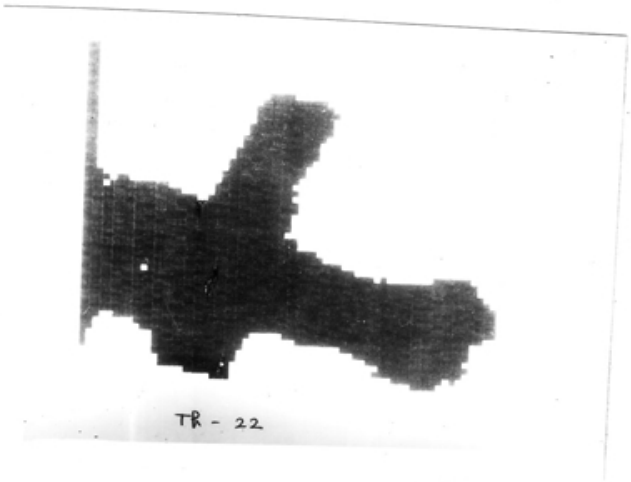


Fig. 4.3.3 Segmented image by Eqn. 4.143

Fig. 4.3.2 Histogram

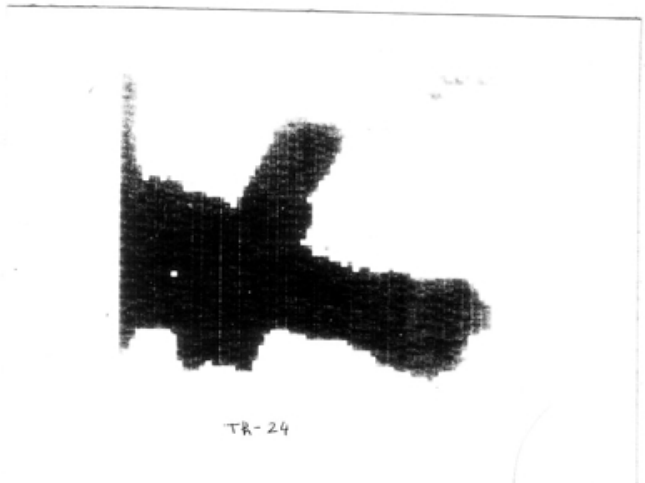


Fig. 4.3.4 Segmented image by Eqn 3.5

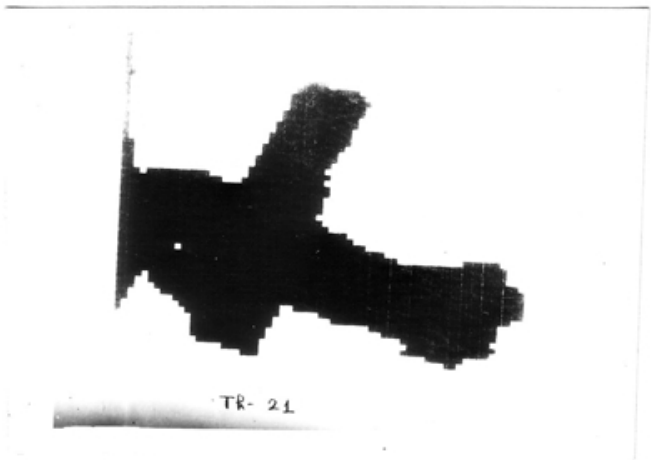


Fig. 4.3.5 & 4.3.7 segmented image by Eqns. 3.10 & 4.150

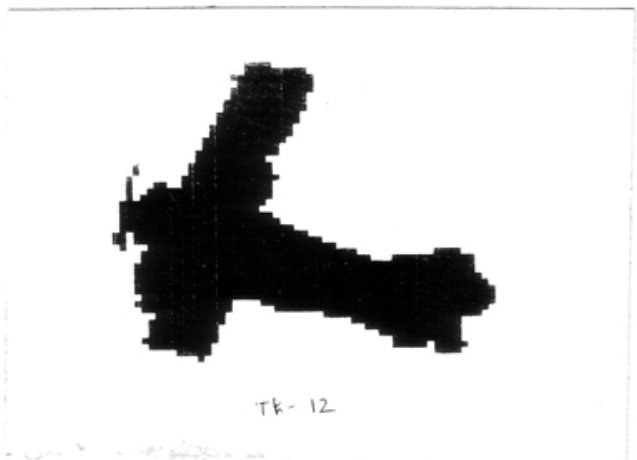


Fig. 4.3.6 & 4.3.8 Segmented by Eqns. 4.166 & 4.153

Fig. 4.3 Input and segmented images of biplane

See Fig. 2.5.1
page- 81

See Fig. 2.5.2
page- 81

Fig. 4.4.1 Input

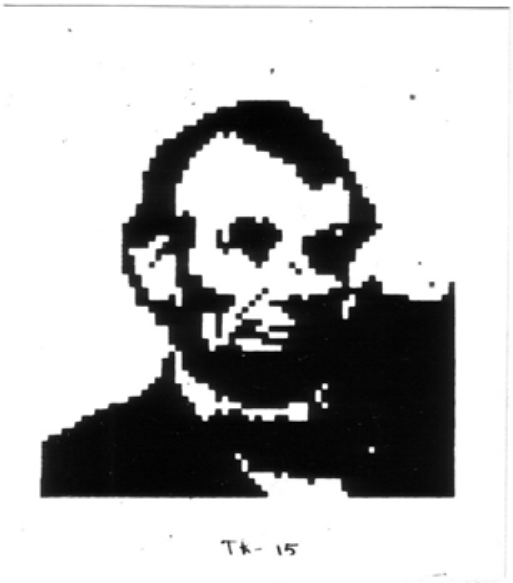


Fig. 4.4.2 Histogram

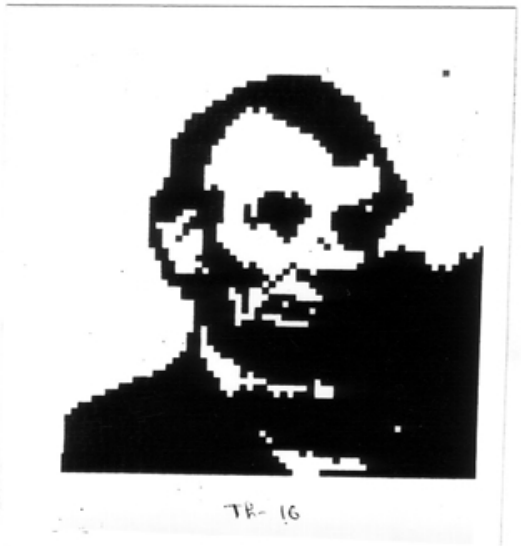


Fig. 4.4.3 & 4.4.5 Segmented image by Eqns. 4.143 & 3.10

Fig. 4.4.4 & 4.4.7 Segmented image by Eqns 3.5 & 4.150

Fig. 4.4.6 segmented image by eqn. 4.166

Fig. 4.4.8 segmented image by eqn. 4.153

Fig. 4.4 Input and segmented image of Abraham Lincoln

In order to show the effectiveness of the algorithms for images with unimodal histogram [Fig. 4.5.2] , an image of three chromosomes [Fig. 4.5.1] has been considered. In this case it is found that equation 4.143 (proposed Algorithm 1) and the method of Kapur et. al. have resulted in good object-background classification [Fig. 4.5.3 and 4.5.5] keeping the legs of a chromosome connected. The method of Pun has totally failed as shown in Fig. 4.5.4 to extract the chromosomes. It is to be noticed that no threshold is detected by the equation 4.166. This is because the assumption of the mixture of normal distributions is possibly not valid here.

In case of the image of the blurred chromosome [Fig. 4.6.1] with a bimodal histogram [Fig. 4.6.2], the segmented results produced by all the methods are found to be comparable, although the threshold detected by equation 4.166 is markedly different from others.

For the image of the boy [Fig. 4.7.1], the proposed algorithm (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.4.7.3]. The other methods have either disturbed the shape of the face due to the selection of a lower threshold value or have failed to separate the ring because of the detection of a threshold at a higher value [Figs. 4.7.4-6].

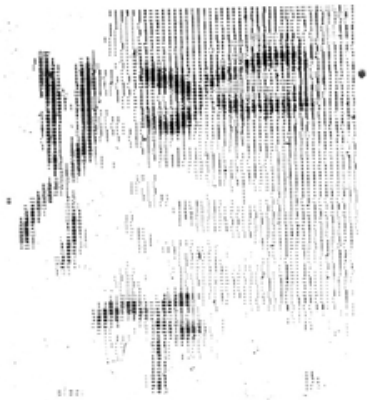


Fig. 4.5.1 Input

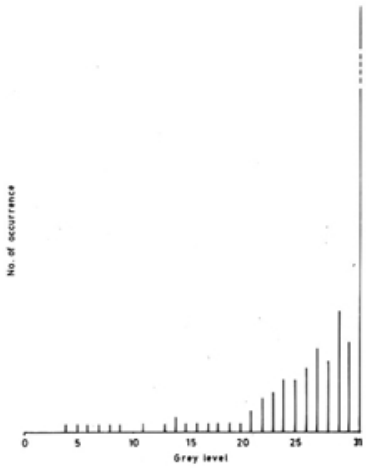


Fig. 4.5.2 Histogram

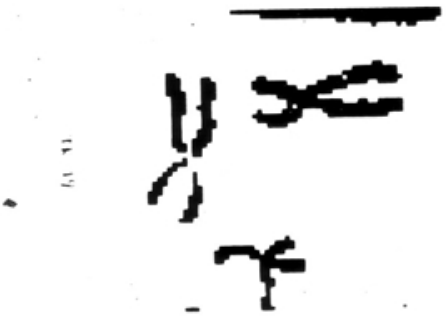


Fig. 4.5.3 Segmented image by eqn. 4.143

Fig. 4.5.4 Segmented image by eqn. 3.5

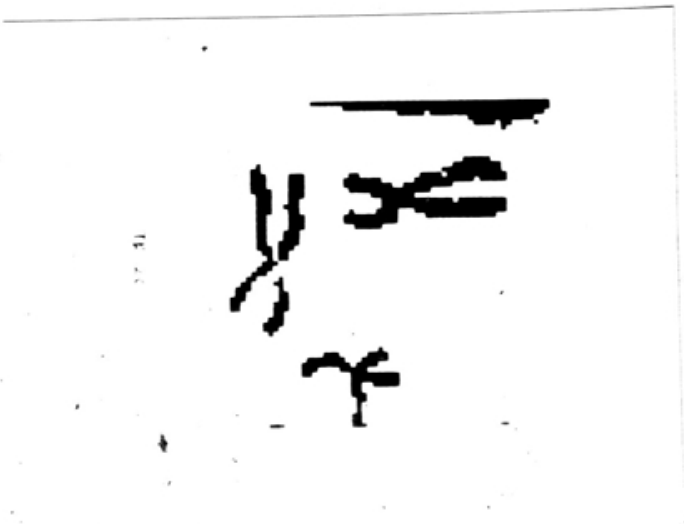


Fig. 4.5.5 segmented image by eqn. 3.10

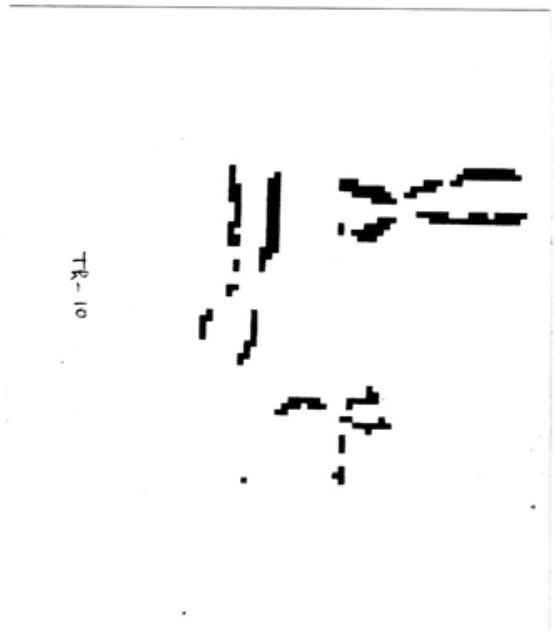


Fig. 4.5.6 segmented image by eqn. 4.150

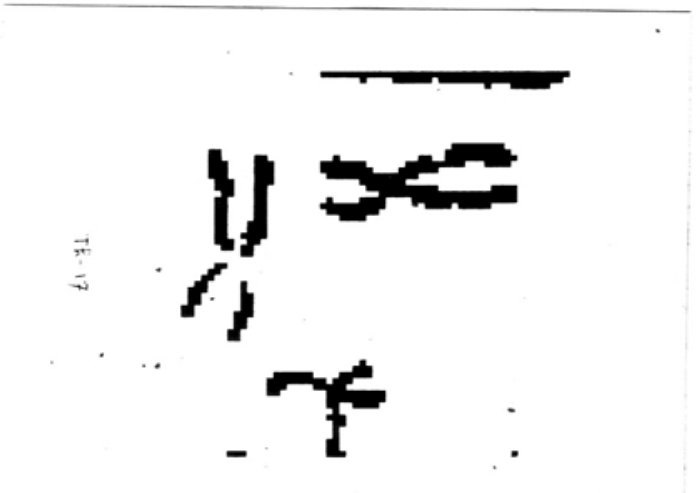


Fig. 4.5.7 segmented image by Eqn. 4.153

Fig. 4.5 Input and segmented image of three chromosomes

See Fig. 3.7.1
page- 120

See Fig. 3.7.2
page- 120

Fig. 4.6.1 Input

Fig. 4.6.2 Histogram

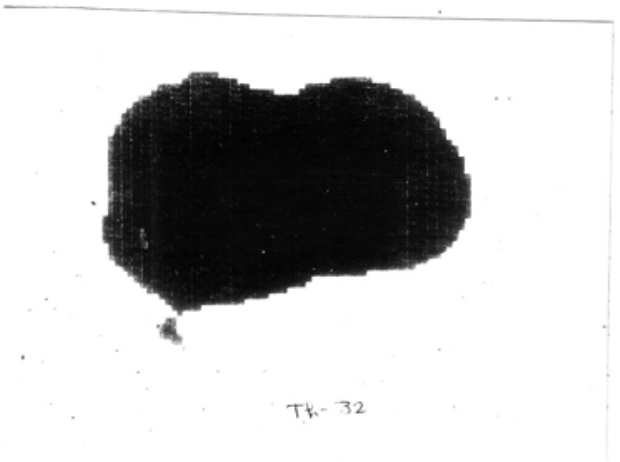


Fig. 4.6.3 and 4.6.4 segmented image by Eqns. 4.143 & 3.5

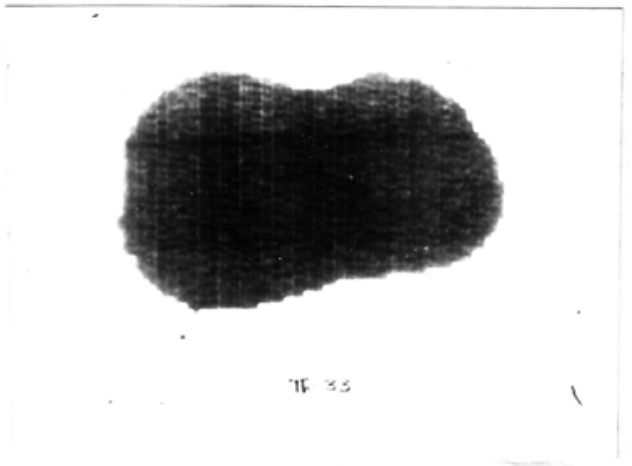


Fig. 4.6.5 Segmented image by Eqn 3.10

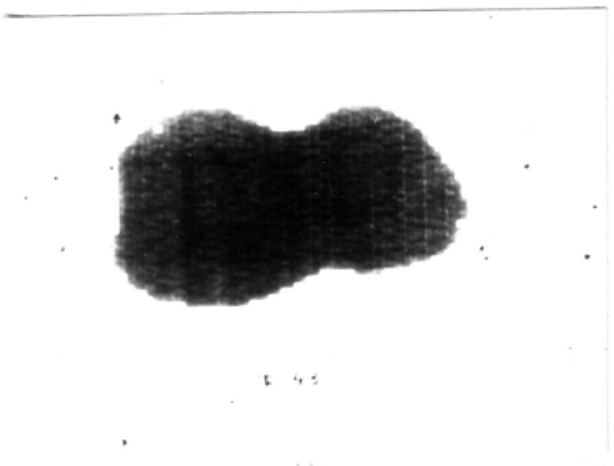


Fig.4.6.6 segmented image by Eqn. 4.166

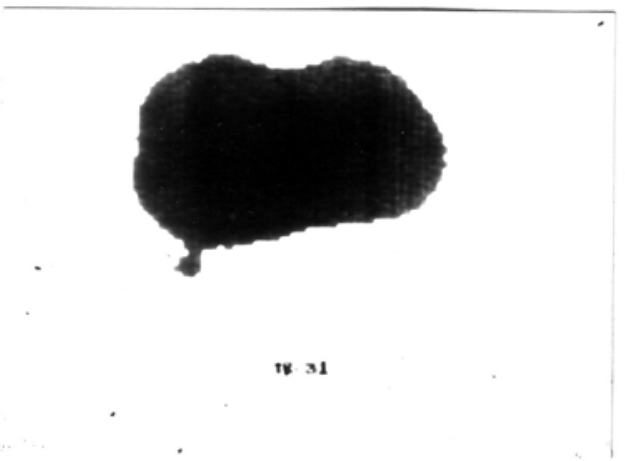


Fig. 4.6.7 Segmented image by Eqns. 4.150

TR-41

Fig. 4.6.8 SEgmented image by Eqn. 4.153

Fig. 4.6 Input and segmented images of blurred chromosome

Fig. 4.7.1 Input

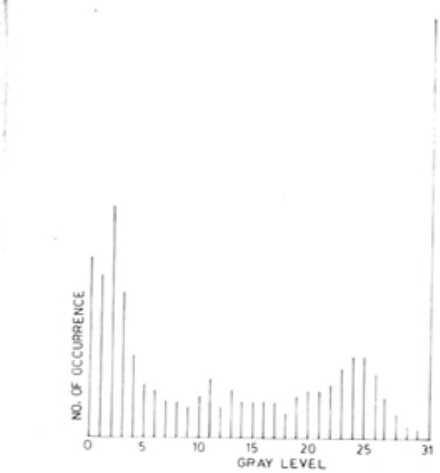


Fig. 4.7.2 Histogram

Fig. 4.7.3 Segmented image by Eqn. 4.143

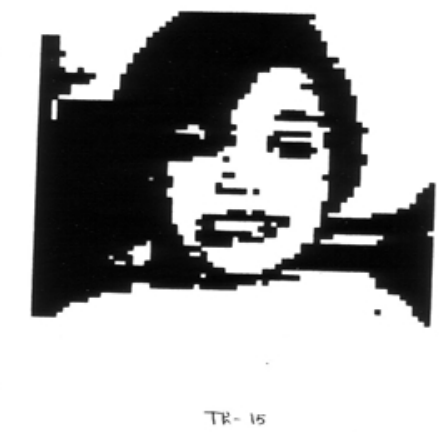


Fig. 4.7.4 Segmented image by Eqns 3.5

Fig. 4.7.5 segmented image by eqn. 3.10

Fig. 4.7.6 segmented image by eqn. 4.166

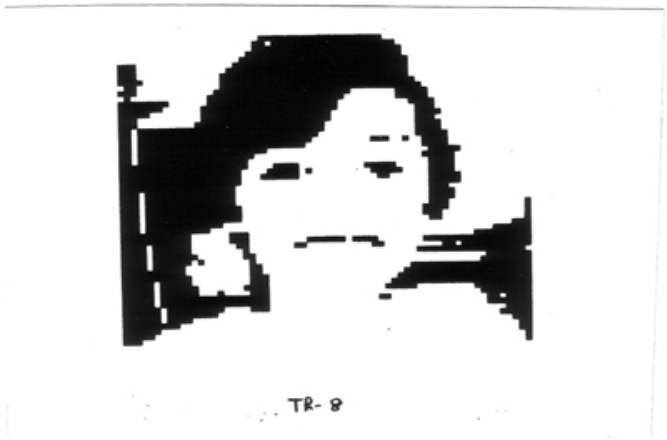


Fig. 4.7.7 and 4.7.8 Segmented image by Eqns. 4.150 and 4.153

Fig. 4.7 Input and segmented image of boy

4.6.2 Results of Methods Based on Local Information .

All the five images described earlier have also been used as input to the proposed algorithms 2 and 3. The threshold levels obtained by these methods for different imput images are displayed in Table - 4.2. In case of the biplane [Fig. 4.3.1], Algorithm 3 has produced much better result than Algorithm 2 [Figs. 4.3.7-8]. Results produced by Algorithm 2 is comparable to those obtained by the method of Kapur et. al. and equation 4.143 [Table 4.1]. Similar is the case for Lincoln image, where Algorithm 3 created much better result than that of Algorithm 2 [Figs. 4.4.7 and 4.4.8]. The threshold obtained by Algorithm 3 is close to those given by equation 4.166 [Fig. 4.6.6, 4.6.8] whereas, the result of Algorithm 2 is almost same with that of others shown in Table 4.1.

Finally, for the image of the boy [Fig. 4.7.1] both algorithms 2 and 3 have produced same threshold. The thresholded images are shown in Fig. 4.7.7. Considering Figs. 4.7.3-8, it appears that for this image some global methods have produced better results than these methods.

It is to be noted that, the results obtained by the exponential entropy are not much different from those of the corresponding algorithms of chapter 3 which use the logarithmic Shannon's entropy . However, the merit of the exponential entropy will be visualised in chapters 5 and 6.

The new entropy introduced in section 4.4 or Shannon's entropy, gives a measure of our ignorance about the system.

TABLE - 4.2

Thresholds for object-background classification
(Based on local information)

Images	THRESHOLDS	
	Proposed algorithm - 2 (Eqn. 4.150)	Proposed algorithm - 3 (Eqn. 4.153)
BIPLANE (Fig.4.3)	21	12
LINCOLN (Fig. 4.4)	16	09
CHROMOSOMES (Fig. 4.5)	10	17
BLURRED CHROMOSOME (Fig. 4.7)	31	41
BOY (Fig.4.7)	08	08

Referring to the example of die, the entropy provides with the expected value of our ignorance about the top most face of the die. For an image, the global entropy gives a measure of grayness ambiguity. The local and conditional entropy, on the other hand, give information about the spatial ambiguity by measuring intraset homogeneity and interset contrast. Although it has been established that these measures of information are quite useful, there is another kind of information, one may be interested in. For example, in case of the die one may be interested in the information about the position of the die on the plane, i.e., where the die lands after it is thrown.

Similarly, for an iamge, information about the location of the object is very important. For example, in scene analysis problem and in the field of robotics and computer vision it becomes highly desirable to have the information about the position of the object within a scene. The proposed entropic measures are able 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 idea led us to define another entropic measure which will give an idea about the location of the object. Such a measure may be named as 'Positional Entropy'. In the following section there is an attempt to introduce such an entropic measure.

4.7 Positional Entropy [85]

Before introducing the measure let us define the 'Object Correlation' as

$$r = \frac{\sum_{(i,j) \in \text{Object}} (i-\bar{I}).(j-\bar{J})}{\sigma_i \, \sigma_j \, O_s} \qquad (4.167)$$

where $\bar{I} = P/2 + 0.5$; $\bar{J} = Q/2 + 0.5$;

O_s is the size of the object ;

P * Q is the size of the image ;

$$\sigma_i = \left[\frac{\sum (i-\bar{I})^2}{O_s} \right]^{1/2} \qquad (4.168)$$

and

$$\sigma_j = \left[\frac{\sum (j-\bar{J})^2}{O_s} \right]^{1/2} \qquad (4.169)$$

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. 4.8]. On the other hand r will be negative when the object is concentrated along the diagonal BD [Fig. 4.8] and r takes the value zero when the object lies symmetrically on any one of the axes EF and GH [Fig. 4.8].

Let us now define the positional entropy (H^P) as follows.

$$H^P = \sum_i \sum_j w_{i,j} \cdot p_{i,j} \, e^{1-p_{i,j}} \qquad (4.170)$$

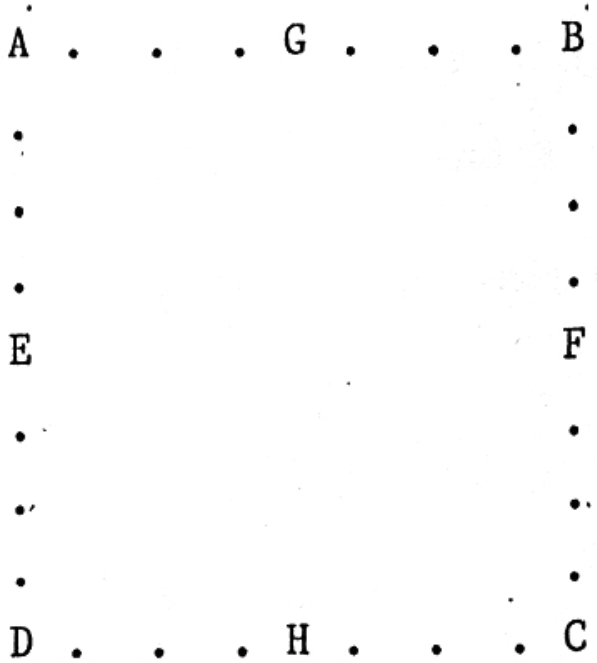


Fig.4.8 Different positions of object for correlation.

Where $p_{i,j}$ is the probability of occurrence of the gray level that occupies the (i,j) th cell (note that this $p_{i,j}$ is different from the p_{ij} introduced earlier as the probability of co-occurrence). w_{ij} is defined as

$$w_{i,j} = 1/d_A(i,j) \quad ; \text{ when } r > 0 \tag{4.171.1}$$

$$= 1/d_D(i.j) \quad ; \text{ when } r < 0 \tag{4.171.2}$$

$$= 1/d_G(i,j) \quad ; \text{ when } r = 0 \text{ and } \sum_{(i,j) \in \text{Object}} (j-\bar{J}) = 0 \tag{4.171.3}$$

$$= 1/d_E(i,j) \quad ; \text{ when } r = 0 \text{ and } \sum_{(i,j) \in \text{Object}} (j-\bar{J}) \neq 0 \tag{4.171.4}$$

Where $d_x(i,j)$ denotes the distance from the point x (x can be any of the points A, D, G, and E [Fig. 4.8].

The equation 4.170 for H^P can be written as

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

$$= \sum_{l=0}^{s} p_l \; e^{1-p_l} \sum_{f(i,j)=l} w_{i,j} \tag{4.172}$$

$$= \sum_{l=0}^{s} w_l \; p_l \; e^{1-p_l} \tag{4.173}$$

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 4.173 can be viewed as an weighted entropy introduced in section 1.4. 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. Depending on the value of r , three possible cases may arise.

Case - 1 $r = 0$

If the object lies symmetrically any where on the axis EF or GH [Fig. 4.8] , the correlation (r) will be zero. Note that $\Sigma(j-\bar{J}) = 0$ 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). Similarly, if the object is on the axis EF, then 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. 4.8]. 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. 4.8]. Similarly, a low value of H^P will correspond to the second quadrant.

The above 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'.

The global and local/conditional entropy provides information necessary for extracting/classifying an object. After it is extracted, its location can be estimated with its positional entropy measure.

As an illustration, Figs. 4.9.1-4.9.10 demonstrate the above mentioned feature of H^P for different positions of an object of size 4 x 8 in an 8 level 16 x 16 image. The corresponding H^P and r values conform to what been discussed earlier.

4.8 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. The idea is also found to be extendable for defining the non-probabilistic entropy of a fuzzy set. A number of interesting properties of the new probabilistic and fuzzy entropy have been proved.

Three algorithms for object-background classification are proposed. One of them is based on the global entropy while the

1	3	2	2	1	1	4	3	8	8	8	8	8	8	8	8
3	2	1	2	1	2	2	1	8	8	8	8	8	8	8	8
3	2	1	3	2	3	1	2	8	8	8	8	8	8	8	8
4	4	3	3	2	1	3	1	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Correlation = .853039

Pos. Entropy = .672327

Fig. 4.9.1

8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	1	3	2	2	1	1	4	3
8	8	8	8	8	8	8	8	3	2	1	2	1	2	2	1
8	8	8	8	8	8	8	8	3	2	1	3	2	3	1	2
8	8	8	8	8	8	8	8	4	4	3	3	2	1	3	1

Correlation = .853039
Pos. Entropy = .155415

Fig. 4.9.2

8	8	8	8	8	8	8	8	1	3	2	2	1	1	4	3
8	8	8	8	8	8	8	8	3	2	1	2	1	2	2	1
8	8	8	8	8	8	8	8	3	2	1	3	2	3	1	2
8	8	8	8	8	8	8	8	4	4	3	3	2	1	3	1
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Correlation = -.853039
Pos. entropy = .160799

Fig. 4.9.3

8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
1	3	2	2	1	1	4	3	8	8	8	8	8	8	8	8
3	2	1	2	1	2	2	1	8	8	8	8	8	8	8	8
3	2	1	3	2	3	1	2	8	8	8	8	8	8	8	8
4	4	3	3	2	1	3	1	8	8	8	8	8	8	8	8

Correlation = -.853039
Pos. Entropy = .732388

Fig. 4.9.4

8	8	8	8	1	3	2	2	1	1	4	3	8	8	8	8
8	8	8	8	3	2	1	2	1	2	2	1	8	8	8	8
8	8	8	8	3	2	1	3	2	3	1	2	8	8	8	8
8	8	8	8	4	4	3	3	2	1	3	1	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Correlation = 0.0
Pos. Entropy = 1.0495

Fig. 4.9.5

8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	1	3	2	2	1	1	4	3	8	8	8	8
8	8	8	8	3	2	1	2	1	2	2	1	8	8	8	8
8	8	8	8	3	2	1	3	2	3	1	2	8	8	8	8
8	8	8	8	4	4	3	3	2	1	3	1	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Correlation = 0.0
pos. Entropy = .344425

Fig. 4.9.6

8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	1	3	2	2	1	1	4	3	8	8	8	8
8	8	8	8	3	2	1	2	1	2	2	1	8	8	8	8
8	8	8	8	3	2	1	3	2	3	1	2	8	8	8	8
8	8	8	8	4	4	3	3	2	1	3	1	8	8	8	8

Correlation = 0.0
Pos. Entropy = .203937

Fig. 4.9.7

8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
1	3	2	2	1	1	4	3	8	8	8	8	8	8	8	8
3	2	1	2	1	2	2	1	8	8	8	8	8	8	8	8
3	2	1	3	2	3	1	2	8	8	8	8	8	8	8	8
4	4	3	3	2	1	3	1	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Correlation = 0.0
Pos. Entropy = .847361

Fig. 4.9.8

8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	1	3	2	2	1	1	4	3
8	8	8	8	8	8	8	8	3	2	1	2	1	2	2	1
8	8	8	8	8	8	8	8	3	2	1	3	2	3	1	2
8	8	8	8	8	8	8	8	4	4	3	3	2	1	3	1
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Correlation = 0.0
Pos. Entropy = .242418

Fig. 4.9.9

8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	1	3	2	2	1	1	4	3	8	8
8	8	8	8	8	8	3	2	1	2	1	2	2	1	8	8
8	8	8	8	8	8	3	2	1	3	2	3	1	2	8	8
8	8	8	8	8	8	4	4	3	3	2	1	3	1	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8
8	8	8	8	8	8	8	8	8	8	8	8	8	8	8	8

Correlation = - .633322
Pos. Entropy = .189947

Fig. 4.9.10

Fig. 4.9 Illustartions of positional entropy

remaining two take the local and conditional entropy into account. The performance of the proposed algorithms is compared to some existing entropic thresholding methods and with a Bayes classifier using normal distribution. The proposed conditional entropic method is 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.

Note that no application of the fuzzy entropy (equation 4.105), introduced in this chapter, has been shown here. However, its application along with a more general definition of entropy (called higher order entropy) of a fuzzy set has been given in chapter 6.

Chapter 5

POISSON DISTRIBUTION BASED IMAGE MODEL AND PARAMETRIC METHODS

5.1 INTRODUCTION

In the previous chapters we have developed several methods for object extraction. These methods are of non parametric nature which involve estimation of the probability of occurrence p_i of the ith level in a region without making use of an appropriate distribution which the gray level variation may follow. There are some parametric methods [2] for object-background classification which assume the well known normal distribution for the gray level without justfication. All these methods do not consider the theory of formation of an image while formulating an algorithm.

The work presented here is an attempt to formulate object extraction algorithms based on the theory of formation of an ideal image. An ideal imaging process has been described which shows that the gray level distributions within the object and background can be approximated with Poisson distributions characterised by two different parameters. In order to establish the validity of the proposed concept, two approaches for object-background classification have been adopted. The first approach is dependent on the maximum entropy principle whereas, the other one is based on the minimum χ^2 statistic. The χ^2 statistic is often used as a criterion for testing goodness of fit of a distribution. Therefore, if an appropriate distribution is assumed, the minimum χ^2 is likely to produce a good segmentation.

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 [133]. In other words, if the uniform exposure level is such that, each receptor has, on the average 'q' quanta then

$$p_r = \text{probability of a counter receiving r quanta}$$
$$= q^r e^{-q}/r! \qquad \text{for} \quad r=0,1,....q,\cdots \tag{5.1}$$
$$= \text{probability of counters receiving r quanta.}$$

If the number of received quanta exceeds the saturation level(s), the excess quanta will not be reflected in the recorded value. In fact in any photographic process, in addition to this upper limit, there is also a lower limit, i.e, threshold (T) to the number of recorded quanta. In other words, as long as the number of incident quanta is less than T, no quanta will be recorded and the T th incident quanta will be recorded as one. The behaviour of the recorded quanta is shown in Fig. 5.1.

Though the number of incident quanta follow a Poisson distribution with average number λ, the average value of the recorded number of quanta will be different from λ because of the two limits mentioned above. The average value, *a*, of the recorded number of quanta is given by

$$a = \sum_{x=0}^{\infty} x' . \frac{e^{-\lambda} . \lambda^x}{x!} \tag{5.2}$$

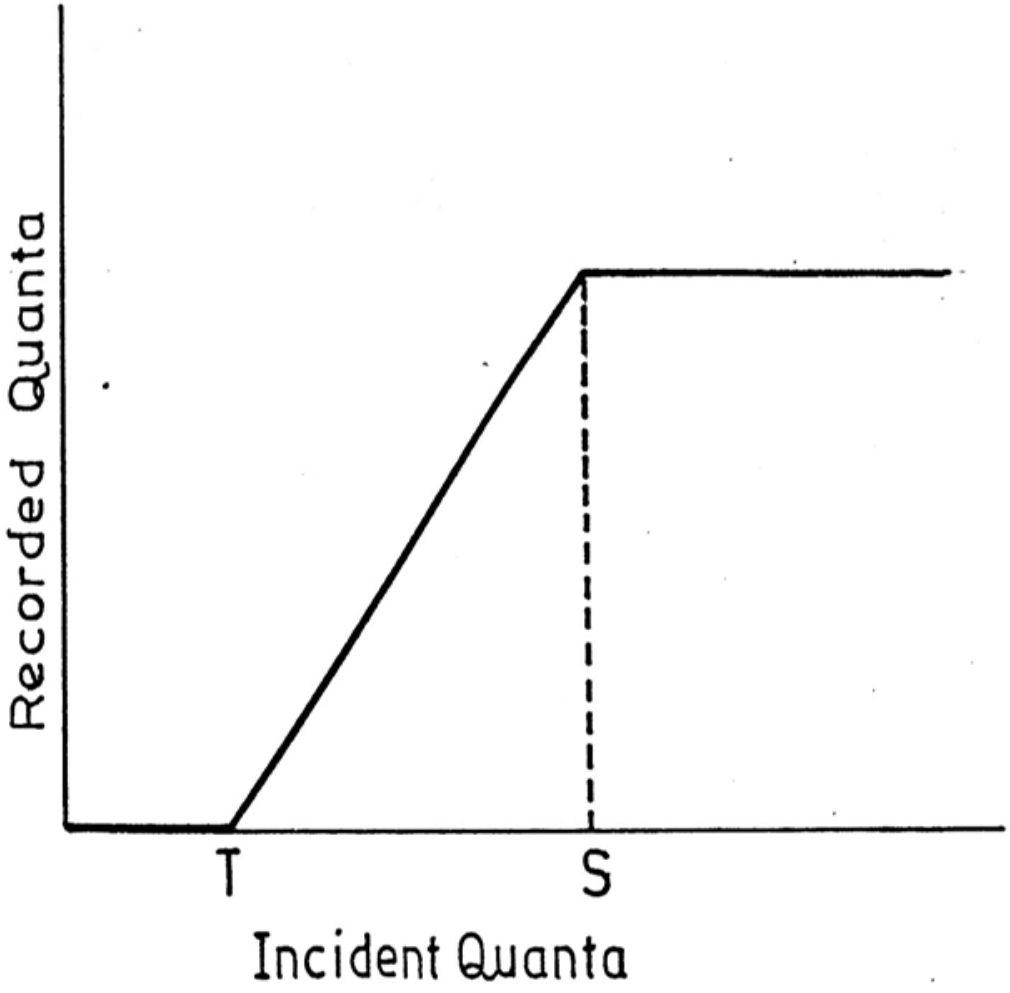


Fig. 5.1 Plot of received vs recorded quanta

where $x' = 0$ for $x < T$

$= x - T + 1$ for $T \leq x < S$

$= S - T + 1$ for $x \geq S$.

On the other hand, $$\lambda = \sum_{x=0}^{\infty} x \frac{e^{-\lambda} . \lambda^x}{x!} \qquad (5.3)$$

obviously, $a \neq \lambda$.

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. For the sake of simplicity we assume here that the probability of the number of incident quanta less than T or greater than S is very small. Thus, under the above assumption, the average number of recorded quanta 'a' (equation 5.2) will be, for all practical purposes, equal to λ. Therefore, the number of recorded quanta will approximately follow a Poisson distribution with average value of λ.

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 that of the object. When we take the photograph of such an ideal scene illuminated by 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 characterised by two different parameters λ_O and λ_B .

To establish the validity of the above model, the idea of Poisson distribution has been used to develop various image segmentation algorithms. Two approaches namely, entropy maximisation and χ^2 statistic minimisation are adopted in the aforesaid context.

The superiority of Poisson distribution over the commonly used normal distribuion for gray level in a digital image has been demonstrated by considering the same algorithms with normal distribution.

5.3 MAXIMUM ENTROPIC THRESHOLDING (MAXET)

We know that the entropy of an *n*-state system as defined by Shannon is given by

$$H = -\sum_{i=1}^{n} p_i \log p_i \quad , \sum_{i=1}^{n} p_i = 1, \quad 0 \le p_i \le 1 \qquad (5.4)$$

where p_i is the probability of the *i*th state of the system. The drawback of this measure of information has been mentioned in chapter 4 and the entropy of an n-state system as suggested there is

$$H = \sum_{i=1}^{n} p_i \, e^{1-p_i} \quad , \quad \sum_{i=1}^{n} p_i = 1, \quad 0 \le p_i \le 1 \qquad (5.5)$$

The term $\{-\log(p_i)\}$ i.e., $\log(1/p_i)$ in equation 5.4 or e^{1-p_i} in 5.5 is called the gain in information from the occurrence of the *i*th event. Let us denote it by $\Delta I(p_i)$, i.e, $\Delta I(p_i)$ is the gain in information from the occurrence of the *i*th state of the n-state system. Thus, in general we can write the expression of entropy as

$$H = \sum_{i=1}^{n} p_i \, \Delta I(p_i),$$

where $\Delta I(p_i)$ can either be $\log(1/p_i)$ or e^{1-p_i} depending on the definition used.

Suppose $g_O(x)$ is the probability function (or probability density function) of the gray level x over the object and $g_B(x)$ is the same over the background. The maximum entropic thresholding principle may now be stated as follows. Partition the image into two non intersecting regions (say, object and background) such that the total entropy of $g_O(x)$ and $g_B(x)$ is maximised. In other words, when $g_O(x)$ and $g_B(x)$ are discrete, maximise

$$\sum_{x \in \text{object}} g_O(x) . \Delta I(g_O(x)) + \sum_{x \in \text{background}} g_B(x) . \Delta I(g_B(x)) \qquad (5.6)$$

On the other hand, when $g_O(x)$ and $g_B(x)$ are continuous, one needs to maximise,

$$\int_{x \in \text{object}} g_O(x) . \Delta I(g_O(x)) \, dx + \int_{x \in \text{background}} g_B(x) . \Delta I(g_B(x)) \, dx \qquad (5.7)$$

In the previous section we have justified that the gray levels within the object and the background follow Poisson distribution with two different parameters λ_O and λ_B .

Thus $$g_O(x) = \frac{e^{-\lambda_O} \lambda_O^x}{x!} \qquad (5.8)$$

and $$g_B(x) = \frac{e^{-\lambda_B} \lambda_B^x}{x!} \qquad (5.9)$$

Therefore, the problem is to find a gray level l, $0 < l < L-1$, $\lambda_O < l < \lambda_B$, such that the gray levels ranging from 0 to l represent the object with expected uniform illumination λ_O and the gray levels in the range $l+1$ to $L-1$ constitute the background with average illumination λ_B. In order to acheive this we maximise equation 5.6. Maximisation of equation 5.6 requires the evaluation of $g_O(x)$ and $g_B(x)$ and hence requires the values of λ_O and λ_B which are to be estimated from the input digital image. There are various methods of estimation of the parameter λ of a Poisson distribution. We use here the maximum likelyhood (ML) estimate of λ. For some hypothetical boundary l, the ML estimate of λ_O and λ_B are given as follows

$$\hat{\lambda}_O = \left(\sum_{i=0}^{l} i.N_i \right) / \sum_{i=0}^{l} N_i \qquad (5.10)$$

$$\hat{\lambda}_B = \left(\sum_{i=l+1}^{L-1} i.N_i \right) / \sum_{i=l+1}^{L-1} N_i \qquad (5.11)$$

(^ is used to indicate the estimated values)

Thus the estimated probability of a gray level x, $x \in$ object, is given by

$$\hat{p}_x^O = \frac{e^{-\hat{\lambda}_O} \hat{\lambda}_O^x}{x\,!} \qquad (5.12)$$

and that for an $x \in$ background is

$$\hat{p}_x^B = \frac{e^{-\hat{\lambda}_B} \hat{\lambda}_B^x}{x\,!} \qquad (5.13)$$

Therefore, for an assumed boundary l, $0 < l < L-1$, the total entropy of the partioned image can be written as

$$H_l = \sum_{x=0}^{l} \hat{p}_x^O . \Delta I(\hat{p}_x^O) + \sum_{x=l+1}^{L-1} \hat{p}_x^B . \Delta I(\hat{p}_x^B) \qquad (5.14)$$

Based on H_l (equation 5.14), the following two algorithms can be formulated.

5.3.1 Algorithm 1

In this method we take $\Delta I(p_x) = e^{1-p_x}$ and maximise equation 5.14 with respect to l.

5.3.2 Algorithm 2

This algorithm uses Shannon's formula of entropy, i.e., it takes $\Delta I(p_x) = \log(1/p_x)$.

In order to strengthen the concept of Poisson distribution we have also experimented with MAXET using the normal distribution to describe the probability distribution of gray levels within a homogeneous region. Thus, we assume that the gray level x is continuous and

$$g_O(x) \approx N(\mu_O, \sigma_O^2) \qquad (5.15)$$

and $$g_B(x) \approx N(\mu_B, \sigma_B^2), \quad \text{i.e.,} \qquad (5.16)$$

$$g_O(x) = (1/(\sigma_O \sqrt{2\pi})) \exp\{-\tfrac{1}{2}((x-\mu_O)/\sigma_O)^2\} \qquad (5.17)$$

and $$g_B(x) = (1/(\sigma_B \sqrt{2\pi})) \exp\{-\tfrac{1}{2}((x-\mu_B)/\sigma_B)^2\} \qquad (5.18)$$

where μ_O, μ_B are means of the two normal distributions and σ_O, σ_B are the standard deviations of the same. With the above forms of probability density functions, we maximise the total entropy of the partitioned image as expressed by equation 5.7 . However, the use of equation (5.17 and 5.18) demands the knowledge of μ_O, μ_B, σ_O and σ_B . Like the previous method, we use here the ML estimators to estimate the μ and σ of the distributions. The ML estimates of the parameters are given by

$$\hat{\mu}_O = \left(\sum_{i \equiv 0}^{l} i.N_i \right) / \sum_{i \equiv 0}^{l} N_i \qquad (5.19)$$

$$\hat{\mu}_B = \left(\sum_{i \equiv l+1}^{L-1} i.N_i \right) / \sum_{i \equiv l+1}^{L-1} N_i \qquad (5.20)$$

$$\hat{\sigma}_O^2 = \left(\sum_{i \equiv 0}^{l} (i-\hat{\mu}_O)^2 . N_i \right) / \sum_{i \equiv 0}^{l} N_i \qquad (5.21)$$

$$\hat{\sigma}_B^2 = \left(\sum_{i \equiv l+1}^{L-1} (i-\hat{\mu}_B)^2 . N_i \right) / \sum_{i=I+1}^{L-1} N_i \qquad (5.22)$$

with the above parameter values, $g_O(x)$ and $g_B(x)$ can be determined and the following algorithms can be formulated.

5.3.3 algorithm 3

In this algorithm, we use $\Delta I(g(x)) = e^{1-g(x)}$ and maximise, for an assumed threshold l,

$$H_l^e = \int_{x \in \text{object}} g_O(x).\Delta I(g_O(x))\, dx + \int_{x \in \text{background}} g_B(x).\ \Delta I(g_B(x))\, dx \qquad (5.23)$$

In this method x (here l) is assumed to be continuous. While selecting the optimum threshold, we compute H_1^e for different l at unit interval over the entire range of the gray levels. Since an image can possess only discrete levels, one dimensional grid search with unit interval sufficies.

5.3.4 Algorithm 4

This algorithm is same as algorithm 3, except that it uses Shannon's entropy, i.e., uses $\Delta I(g(x)) = \log(1/g(x))$.

5.4 MINIMUM CHI-SQUARE THRESHOLDING (MINCST)

Let us describe here another set of thresholding algorithms based on minimun Chi-square (χ^2) statistic to demonstrate the appropriateness of Poisson distribution for gray levels. Let N_1, $N_2,\ldots,N_k$ be the observed frequencies of k different classes and p_1, $p_2,\ldots,p_k$ be the hypothetical probabilities of a normal distribution. In order to test the goodness of fit of the hypothetical probabilities to the observed ones, K. Pearson [134] suggested the critrion,

$$\chi^2 = \sum_{i\equiv 1}^{k} \frac{(N_i - N.p_i)^2}{N.p_i} = \sum_{i\equiv 1}^{k} \frac{(\text{observed} - \text{expected})^2}{\text{expected}} \qquad (5.24)$$

where N is the total number of observations and the asymptotic distribution (a.d) of χ^2 as defined in equation 5.24 is a $\chi^2(k-1)$, i.e, a chi-square with (k-1) degrees of freedom.

The more general problem may be to test whether the class probabilities are some specific functions of a fewer parameters which may be known. If the class probabilities are some specified functions, $p_1(\theta)$, $p_2(\theta),\ldots,p_k(\theta)$, where θ has q (say) components i.e, $\theta=(\theta_1,\theta_2,\ldots,\theta_q)$ and $\hat{\theta}$ is an efficient estimator of θ, then under suitable conditions [134] it can be shown that

$$\chi^2 = \sum_{i=1}^{k} \frac{(N_i - N.p_i(\hat{\theta}))^2}{N.p_i(\hat{\theta})}$$

is a $\chi^2(k-1-q)$.

This statistic is known as χ^2 statistic,and is very often used to measure the goodness of fit of some hypothetical distribution to an observed data. Lower the value of χ^2, better is the fit. The expression for χ^2 may also be viwed as a weighted sum of squared deviation of the expected frequencies from the observed ones.

For the image segmentation problem, the minimum χ^2 thresholding (MINCST) principle may thus be stated as follows. 'Given the distributions, that the gray levels may follow, partition the image into two nonintersecting regions, such that the sum of χ^2 (over the object and the background) is minimised'. Based on this principle two algorithms, one using Poisson distribution and the other using normal distribution, are suggested.

5.4.1 **Algorithm 5**

Let $g_O(x)$ and $g_B(x)$ (equations 5.8 and 5.9) be the probability functions of the gray levels in the object and background regions respectively. For an arbitrary threshold l, the observed frequencies of gray levels over the object are given by

$$N_i^O = N_i, \quad i=0,1,2,..l \tag{5.25}$$

and those over the background are

$$N_i^B = N_i, \quad i=l+1, l+2,...,L-1 \tag{5.26}$$

On the other hand, the estimated probability of a gray level i, $0 \leq i \leq l$, i.e., i is in the object, is given by $g_O(i)$, $i=0,1,2,...l$, and that over the background is given by $g_B(i)$, $i=l+1,l+2,..L-1$. Normalising $g_O(i)$ and $g_B(i)$, over the object and the background regions respectively, we get the expected probabilities as follows.

$$\hat{p}_i^O = g_O(i) \; / \sum_{j=0}^{l} g_O(j), \quad i=0,1,2...l \tag{5.27}$$

$$\hat{p}_i^B = g_B(i) \; / \sum_{j=l+1}^{L-1} g_B(j), \quad i=l+1,l+2,..,L-1 \tag{5.28}$$

Thus the expected frequencies of the gray levels over the object and background are given by the following equations.

$$\hat{N}_i^O = \hat{p}_i^O \cdot N_O, \; i=0,1,2,..,l \tag{5.29}$$

$$\hat{N}_i^B = \hat{p}_i^B \cdot N_B, \; i=l+1,l+2,...,L-1 \tag{5.30}$$

where N_O and N_B are the total number of observations within the object and the background respectively. The parameters λ_O and λ_B which are required for the evaluation of $g_O(x)$ and $g_B(x)$ are estimated using the ML estimator (described in section 5.3). Therefore, for a threshold l, the total Chi-square is given by

$$\chi_l^2 = \sum_{i=0}^{l} \frac{(N_i^O - \hat{N}_i^O)^2}{\hat{N}_i^O} + \sum_{i=l+1}^{L-1} \frac{(N_i^B - \hat{N}_i^B)^2}{\hat{N}_i^B} \qquad (5.31)$$

The optimum threshold is obtained by minimising χ_l^2 with respect to l.

5.4.2 Algorithm 6

This algorithm is also based on the principle of MINCST, but it assumes normal distributions (equations 5.17 and 5.18) for the gray levels. In order to compute the χ^2, for an arbitrary threshold l, we proceed as follows.

For each gray level y, y = 1,2,..., consider a band (y-0.5, y+0.5) and compute I_y as follows.

$$I_y = \int_{y-0.5}^{y+0.5} g_O(x)\, dx \qquad \text{for } 0 < y < l \qquad (5.32)$$

and

$$I_y = \int_{y-0.5}^{y+0.5} g_B(x)\, dx \qquad \text{for } l + 1 < y < L - 1 \qquad (5.33)$$

I_y gives the probability of the gray levels lying in the range (y-0.5, y+0.5). The expected frequency in the band (y-0.5,y+0.5) will then be

$$\hat{N}_y^O = N_O \cdot I_y \qquad \text{for } 0 < y < l \tag{5.34}$$

or $$\hat{N}_y^B = N_B \cdot I_y \qquad \text{for } l + 1 < y < L - 1 \tag{5.35}$$

while the observed frequency of the band around y will be N_y, $y = 1,2,\ldots$.

Note that, we have not considered above the gray levels $y < l$ and $y \geq l$ for the object; similarly, gray levels $y \leq l + 1$ and $y > L - 2$ for the background. In order to account for this, in case of object, we compute

$$I_0 = \int_{-\infty}^{0.5} g_O(x)\, dx \tag{5.36}$$

and $$I_l = \int_{l-0.5}^{\infty} g_O(x)\, dx \tag{5.37}$$

so that

$$\hat{N}_0^O = N_O \cdot I_0 \tag{5.38}$$

and $$\hat{N}_l^O = N_O \cdot I_l \tag{5.39}$$

It is to be noted here that the sum of all the band integrals for $g_O(x)$ will be equal to one, i.e., $\sum_{y=0}^{l} I_y = 1$. Similar computation can also be done for the background probability distribution. Once we know the observed and expected frequencies in different classes, the χ^2 (equation 5.31) can be computed and minimised with respect to l for selecting the threshold.

5.5 RESULTS

All the six algorithms discussed in section 5.3 and 5.4 have been implemented on a set of four images with widely different types of histograms. In addition to this, the global entropy based method of Kapur et. al., as discussed in section 3.2.2 has also been implemented for comparison of results with those of the new methods suggested in this chapter. It has been discussed in chapter 3 that Kapur et. al. maximised the total entropy of the partitioned image without considering any appropriate probability distribution that the gray levels may follow.

Table 5.1 , displays the thresholds produced by different methods for various images. Figs. 5.2.1 and 5.2.2 represent the input image of a biplane and its gray level histogram, respectively. Figs. 5.2.3 - 5.2.9 give the segmented images produced by different methods.

From the results, one can observe that algorithms 1,5 and 6 have produced same result and infact it is the best segmentation of the biplane image. Algorithm 2 has produced almost identical results to that of the above three methods. On the other hand, Algorithms 4 and the method of Kapur et. al. could not extract the propeller infront of the biplane. It is to be noticed that for this image all the algorithms that used Poisson distribution resulted in good and comparable segments. Though the biplane has a

Table 5.1

Thresholds Produced by different methods

IMAGES	MAXENT				MINCST		
	Alg. 1	Alg. 2	Alg. 3	Alg. 4	Alg. 5	Alg. 6	Alg. of Kapur etal
PLANE (Fig.5.2)	12	13	16	19	12	12	21
LINCOLN (Fig.5.3)	11	12	9	15	9	6	15
BOY (Fig.5.4)	9	11	9	14	12	5	13
CHROMO. (Fig.5.5)	12	16	19	19	16	-	20

See Fig. 2.7.1
Page - 85

See Fig. 2.7.2
Page - 85

Fig. 5.2.1 Input

Fig. 5.2.2 Histogram

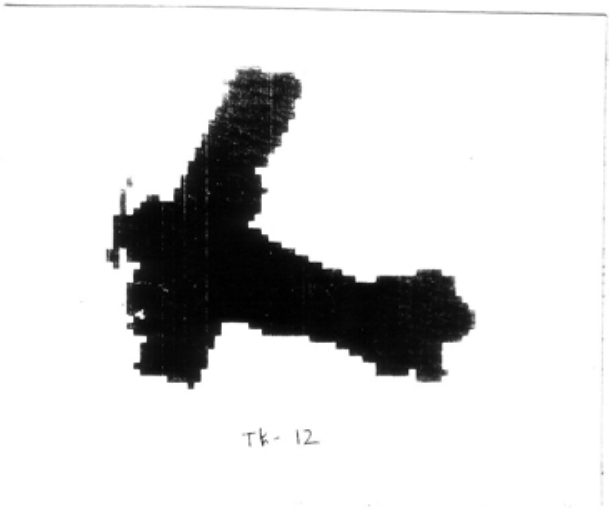


Fig. 5.2.3, 5.2.7 & 5.2.8 Segmented image by Eqns. 5.14(exp. ent.), 5.31 (Pos. dist.)and 5.31 (Normal dist.)

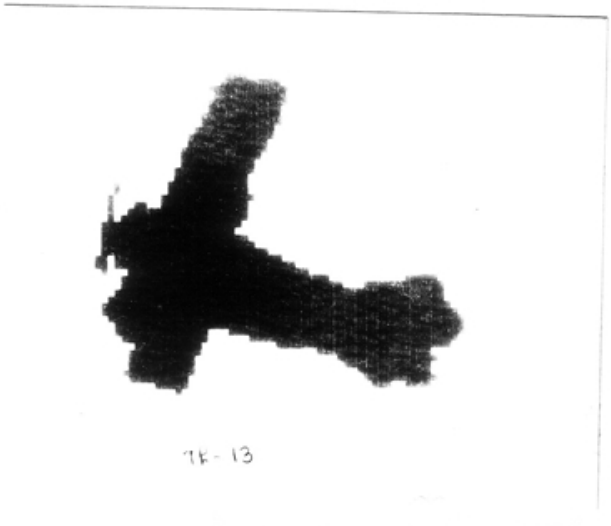


Fig. 5.2.4 Segmented image by Eqn. 5.14 (log.ent.)

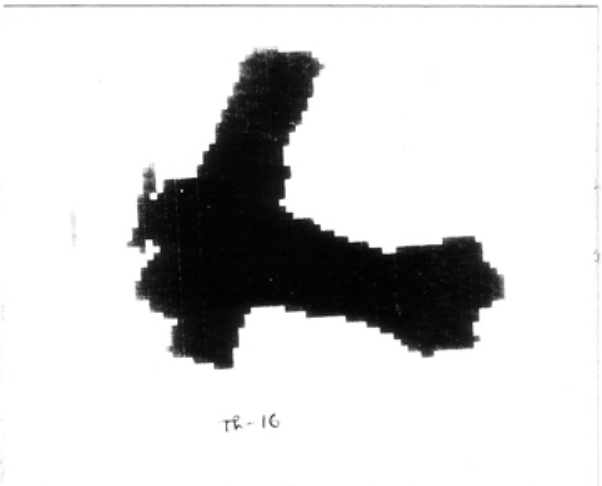


Fig. 5.2.5 Segmented image by eqn. 5.23 (exp. ent.)

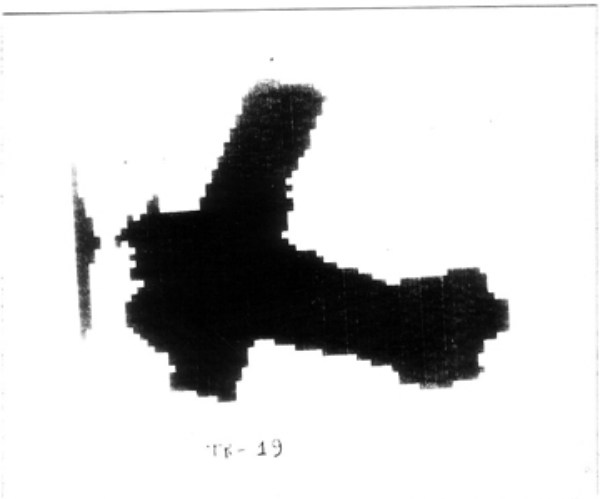


Fig. 5.2.6 Segmented image by Eqn. 5.23 (log. ent.)

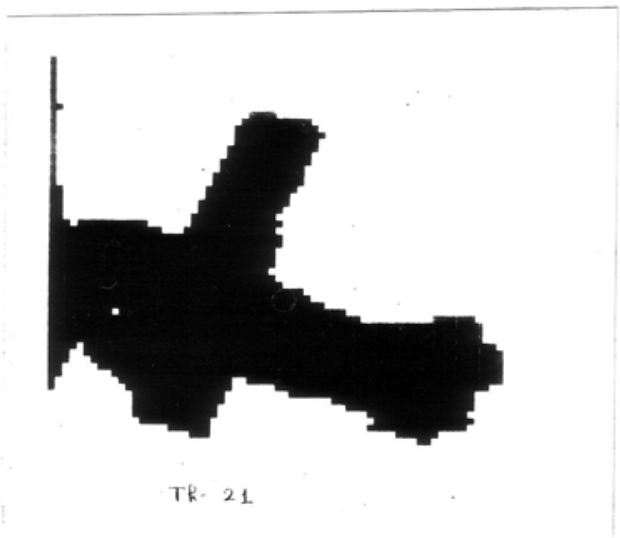


Fig. 5.2.9 Segmented image by Eqn. 3.10

Fig. 5.2 Input and segmented image of biplane

bimodal histogram with a deep valley, Algorithm 3 and 4 (which use normal distribution) and algorithm of Kapur et. al. could not make appropriate segmentation.

The Lincoln image [Fig. 5.3.1] has a multimodal histogram [Fig. 5.3.2]. In this case, the best segmentation is produced by Algorithm 1 [Fig. 5.3.3]. Algorithms 2,3 and 5 have more or less correctly extracted the object [Figs. 5.3.4, 5.3.5 and 5.3.7]. On the other hand, Algorithm 4 and the algorithm of Kapur et. al. , which have resulted in same threshold value, could not separate the beard of Lincoln at the lower right hand side corner of the image [Figs. 5.3.6 and 5.3.9]. Algorithm 6 has produced a very low threshold. Here also all the algorithms which use Poisson distribution give better results than the normal distribution based algorithms. Another point to be observed is that the exponential entropy gives a better threshold than the logarithmic entropy.

Figs. 5.4.1 and 5.4.2 depict the input image and the gray level histogram of a boy. Here, except Algorithm 6, which has resulted in a low threshold, all other methods have produced acceptable results [Figs. 5.4.3 - 9]. However, a critical anlysis of the results [Figs. 5.4.3 - 9] shows that the best result is produced by Algorithm 2 and 5 [Fig. 5.4.4 and 5.4.7], as they have preserved the features of the face and also extracted the ring in the left ear of the image.

See Fig. 2.5.1
page- 81

See Fig. 2.5.2
page- 81

Fig. 5.3.1 Input

Fig. 5.3.2 Histogram

Fig. 5.3.3 Segmented image by Eqn. 5.14(exp. ent.)

Fig. 5.3.4 Segmented image by Eqn 5.14(log. ent.)

Fig. 5.3.5 & 5.3.7 segmented image by eqns. 5.23 (exp. ent.) and 5.31 (Pos. dist.)

Fig. 5.3.6 & 5.3.9 segmented image by eqns. 5.23(log ent.) and 3.10.

Fig. 5.3.8 Segmented image by Eqn. 5.31 (normal dist.)

Fig. 5.3 Input and segmented image of Abraham Lincoln

See Fig. 4.7.1
page- 176

See Fig. 4.7.2
page- 176

Fig. 5.4.1 Input

Fig. 5.4.2 Histogram

Fig. 5.4.3 & 5.4.5 Segmented image by Eqns. 5.14(exp. ent.) and 5.23 (exp. ent)

Fig. 5.4.4 Segmented image by Eqn 5.14(log. ent)

Fig. 5.4.6 segmented image by eqns. 5.23 (log. ent.)

Fig. 5.4.7 segmented image by eqns. 5.31(Pos. dist.)

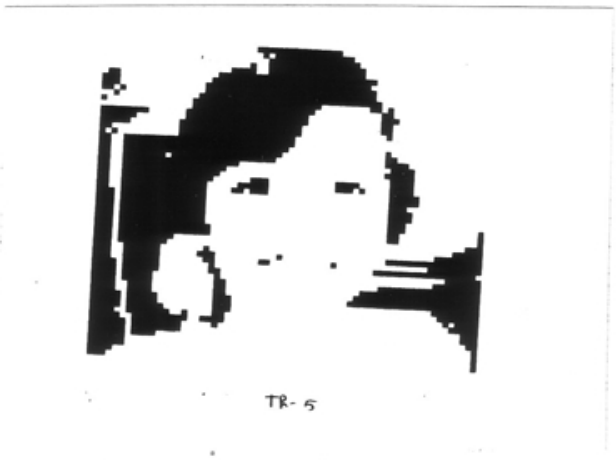


Fig. 5.4.8 Segmented image by Eqn. 5.31 (normal dist.)

Fig. 5.4.9 Segmented image by Eqn. 3.10

Fig. 5.4 Input and segmented image of boy

The proposed algorithms are also tested on an image of a set of three chromosomes [Fig. 5.5.1] with unimodal histogram [Fig. 5.5.2]. In this case, except algorithms 1 and 6 all others have produced comparable results[Figs. 5.5.4 - 8]. Algorithm 1 has produced a thinned version of the chromosomes [Fig. 5.5.3] while, Algorithm 6 has failed to make any segmentation. Although the algorithms described here are all based on global information of an image, the thresholds obtained by Poisson distribution based methods conform well to those obtained by local information based methods (e.g., conditional entropic segmentation technique described in chapters 3 and 4). Based on the above observations, the following points can be mentioned.

All the algorithms which use Poisson distribution (Algorithms 1,2 and 5) consistently result in a better segmentation than the others. This observation establishes the validity of our digital image model based on Poisson distribution. It reveals that Poisson distribution is much more appropriate than normal distribution for the gray levels in a digital image.

The exponential entropic methods (Algorithms 1 and 3) are found to be much better than the logarithmic entropic methods (Algorithms 2,4 and method of Kapur et. al.). This shows that for a digital image, the exponential entropy is possibly a more appropriate measure of information than the Shannon's entropy.

The minimum chi-square thresholdig principle is an excellent method provided an appropriate distribution is assumed.

See Fig. 4.5.1
Page - 172

See Fig. 4.5.2
Page - 172

Fig. 5.5.1 Input

Fig. 5.5.2 Histogram

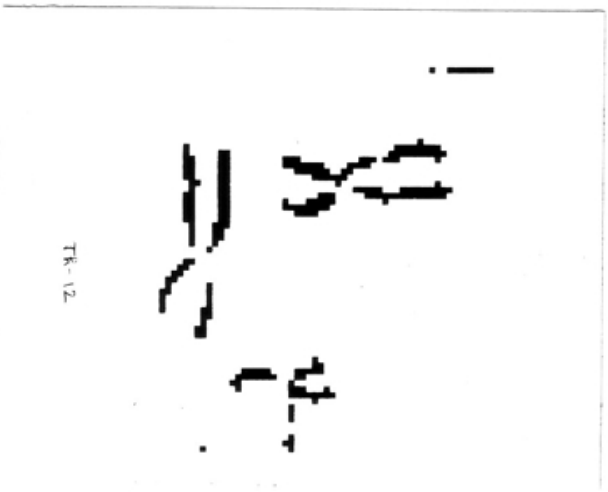


Fig. 5.5.3, Segmented image image by Eqn. 5.14(exp. ent.),

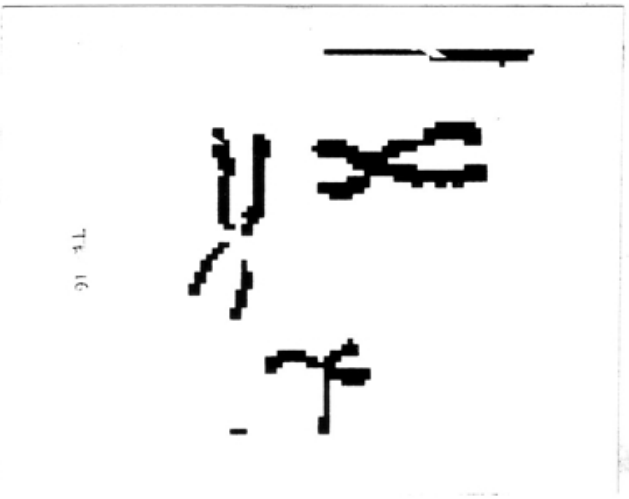


Fig. 5.5.4 & 5.5.7 Segmented image by Eqn. 5.14(log. ent.) and 5.31 (Pos. dist.)

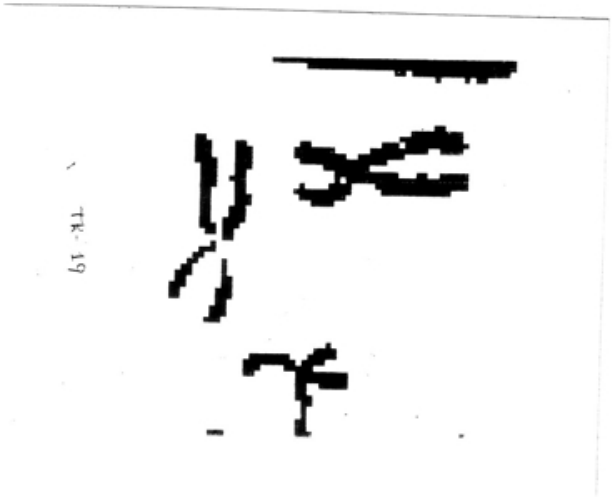


Fig. 5.5.5 & 5.5.6 Segmented image by eqns. 5.23 (exp. ent.)& 5.23 (log. ent.)

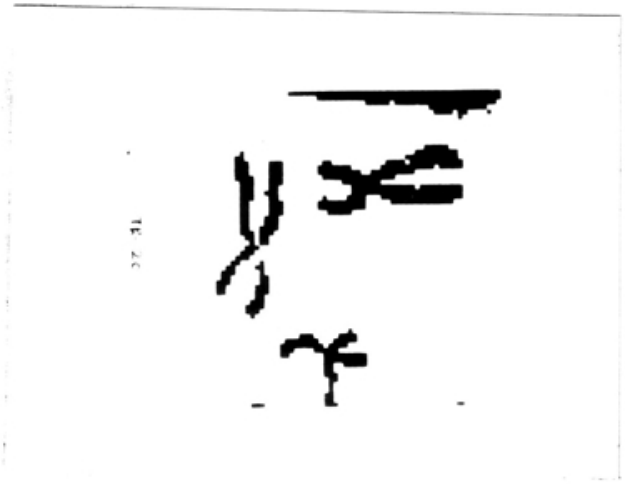


Fig. 5.5.8 Segmented image by Eqn. 3.10

Fig. 5.5 Input and segmented image of three chromosomes

5.6 CONCLUSION

An ideal-image model for a gray tone digital image has been described. It has been found that the gray levels in an image follow a Poisson distribution. Based on this concept, various parametric algorithms for object background classification have been formulated. Two principles namely, entropy maximisation and χ^2 minimisation have been adopted here while formulating the algorithms. Thresholds obtained are found to be highly satisfactory for a wide range of input images.

In order to strengthen the appropriateness of Poisson distribution, the same algorithms have been implemented with normal distribution. In a part of the experiment, the results have also been compared with those of an entropic thresholding algorithm which does not assume any parametric distribution. In all cases, the algorithms based on Poisson distribution consistently resulted in a better segmentation. It has further been revealed that the exponential entropic measure is more effective than the logarithmic (Shannon's) entropy in extracting thresholds. The χ^2 statistic can be regarded as a very much effective tool for segmentation when Poisson distribution is used.

It is to be mentioned here that the algorithms described here are all based on global information (i.e., histogram) of an image. Usually, the histogram based algorithms are less effective than local (spatial) information based algorithms. However, the segmentation produced here by the Poisson distribution based methods is no way worse than that of the spatial information based methods [30,31], described in earlier chapters.

Chapter 6

HIGHER ORDER FUZZY ENTROPY AND HYBRID ENTROPY

6.1 INTRODUCTION

In Chapter 4 (section 4.3.5) we defined the entropy of a fuzzy set (equation 4.105) considering an exponential information function. This definition and the definition of Deluca and Termini (equation 1.28), reflect an amount of difficulty in deciding whether an element would be considered a member of a fuzzy set or not. This type of information is not always sufficient for dealing with practical problems. In many occasions, we may be interested in evaluating the behavior of average uncertainty (ambiguity) associated with a collection of supports.

A new concept called "higher order fuzzy entropy", $H^{(r)}$, $(r \geq 1)$ has been introduced here in this context. When $r = 1$, $H^{(r)}$ reduces to equation 4.105 or equation 1.28 depending on the type of information function used. Various entropy definitions have also been discussed in this regard.

Furthermore, there have been several attempts to combine the probabilistic and fuzzy (possibilistic) entropy to obtain a measure of total uncertainty involved in a system. These measures have been reviewed and a new definition, called "hybrid entropy", H_{hy} has been suggested in the second part of this chapter. This measure can be regarded as a generalised entropy of a set such that when the fuzziness is removed, the measure turns out to be the classical entropy of the set.

Various properties of these measures have been stated and proved. The higher order entropy, H^r, is found to possess some interesting properties which can be used in feature evaluation and image segmentation problems. As expected, H^r conveys more information regarding the actual structure of the set than H^1 (conventional fuzzy entropy) does. H_{hy}, on the other hand, can be considered an objective function for proper defuzzification (enhancement) of a set. These characteristics have been further demonstrated with examples. An index of similarity of supports has also been proposed.

Finally, the higher order fuzzy entropy has been applied to develop algorithms for edge detection and segmentation of an image.

6.2 CLASSICAL ENTROPY

It has already been discussed in chapter 4 that Shannon [109,110] defined the term entropy to represent a quantitative measure of information of a probabilistic system. If p_i is the probability of occurrence of the i-th state of an n-state system, $0 \leq p_i \leq 1$, $\Sigma\, p_i = 1$, then the entropy of the system is defined as

$$H = -\sum_{i=1}^{n} p_i \log p_i \qquad (6.1)$$

In chapter 4 a new definition of entropy [30] has also been introduced as

$$H = \sum_{i=1}^{n} p_i \, e^{1-p_i} \qquad (6.2)$$

It is to be mentioned here that both of the above measures give the average amount of information that would be gained from the occurrence of any arbitrary state of a probabilistic system. In this chapter, we shall concentrate on the entropy measures of a fuzzy set. Before describing the new measures, let us first critically anlysis, some of the existing fuzzy entropy measures

6.3 ENTROPY MEASURES OF FUZZY SETS

Zadeh [117] defined the entropy of a fuzzy subset A for the finite set $\{x_1, x_2, \ldots, x_n\}$ with respect to the probability distribution $\{p_1, p_2, \ldots, p_n\}$ as

$$H_P = - \sum_{i=1}^{n} \mu_A(x_i) \; p_i \; log \; (p_i) \qquad (6.4)$$

where μ_A is the membership function of A, and p_i is the probability of occurrence of x_i. Zadeh in [117] did not clarify the physical significance of H_P, even he did not mention whether H_P can be used as a measure of fuzziness or not .

This H_P can be viewed as a weighted Shannon entropy where the membership values are taken as the weight. Let us now see how the use of H_P, as a measure of fuzziness, leads to an undesirable situation. Consider a fuzzy set with two elements x_1, x_2 with probabilities p_1, p_2 as shown in Table-6.1.

TABLE -6.1 :Value of Zadeh's entropoy

CASE	μ_i	p_i	H^P
	0.5	0.01	
1			0.028007
	0.5	0.99	
	0.8	0.01	
2			0.44800
	0.9	0.99	

Since probabilities are same for both the cases, the fuzziness is expected to be higher for case-1; but this is not reflected in H_p. This example, therefore, reveals the difficulty in using H_p as a measure of fuzziness.

Kaufmann [119] defined the entropy of a fuzzy set with n supports as

$$H^k = \{-1/log\ (n)\ \} \sum_{i=1}^{n} \phi_i\ log\ (\phi_i) \qquad (6.5)$$

where $\phi_i = \mu_i \ / \sum_{i=1}^{n} \mu_i \quad ; \ i=1,2,\ldots,n.$

The drawback of this measure is that it does not depend on the absolute values of μ_i, but on their relative values. Thus a set with μ_i = 0 or 0.1 or 0.8 or 1 for all of its elements, would have same entropy (equal to 1). This is intuitively unappealing.

Deluca and Termini [120] have used a different expression (based on the Shannon's function) to define the entropy of a fuzzy set as follows:

$$H = -k \sum_{i=1}^{n} \{\mu_i\ log(\mu_i)\ + (1 - \mu_i)\ log\ (1-\mu_i)\}. \qquad (6.6)$$

Here k is a normalising constant. Equation (6.6) is claimed to express an average amount of fuzziness(ambiguity) present in a set A and it has the following desirable properties. This properties have already been stated in chapter-4, for the convenience of the reader it is presented again.

P1: H is minimum iff $\mu_i = 0$ or 1 for all i.

P2: H is maximum iff $\mu_i = 0.5$ for all i.

P3: $H \geq H^*$ where H^* is the entropy of A^*, a sharpened version of A.

(A^* is a sharpened version of A if $\mu^* \leq \mu$ for μ in [0,0.5] and $\mu^* \geq \mu$ for μ in [0.5,1].)

P4: H = H' where H' is the entropy of the complement set.

(for the sake of notational simplicity $\mu(x_i)$ has been written as μ_i.

Any measure of fuzziness should possess these four properties. Note that equation 6.4 does not satisfy all these properties. Bart Kosko [122] defined the entropy of a fuzzy set as the ratio of distances between the fuzzy set and its nearest and furthest nonfuzzy neighbours. This measure also satisfies properties P1 thru P4.

Pal and Pal [30] have also given a measure of fuzziness in a set (which has been discussed in chapter 4) as

$$H = k \sum_{i=1}^{n} \{\mu_i \, e^{1-\mu_i} + (1-\mu_i) \, e^{\mu_i}\} \qquad (6.7)$$

It has been proved in chapter 4 that equation 6.7 satisfies properties P1 thru P4.

It is to be noted here that the meaning of the entropy of a fuzzy set is quite different from that of the probabilistic entropy (equation 6.1 and 6.2). The former gives, as mentioned before, the average amount of ambiguity (difficulty) in deciding

whether an element belongs to a set or not, while the later gives the average gain in information from the occurrence of an event. Since an ordinary set is a special case of a fuzzy set , a relationship between these two types of information is expected. There have been several attempts in this regard to combine the probabilistic (Shannon's) and possibilistic (fuzzy) entropy. These are explained below.

Deluca and Termini [120] attempted to do so in the following manner. Consider an experiment in which the elements $x_1, x_2, \ldots, x_n$ may occur, once and only once in each trial with probabilities

$$p_1, p_2, \ldots, p_n \quad (p_i \geq 0, \sum_{i=1}^{n} p_i = 1).$$

Shannon's entropy of the probability distribution is given by

$$H(p_1, p_2, \ldots, p_n) = - \sum_{i=1}^{n} p_i \log(p_i) \tag{6.8}$$

This gives the average amount of information gained from the knowledge of the occurrence of an element. Suppose there exists a difficulty in the interpretation of x_i, the out come of a trial, as 0 or 1. The amount of ambiguity involved in the interpretation of x_i is given by [120]

$$S(\mu_i) = -\mu_i \log(\mu_i) - (1-\mu_i) \log(1-\mu_i) \tag{6.9}$$

The statistical average (m) of $S(\mu_i)$ is given by

$$m = \sum_{i=1}^{n} p_i \, S(\mu_i) \tag{6.10}$$

This m is claimed to give the average amount of difficulty in taking a decision (0 or 1) on the elements x_i, $i=1,2,\ldots,n$. Combining equations 6.8 and 6.10 the total entropy is defined as follows.

$$H_{tot} = H(p_1,p_2,..,p_n) + m\ (\mu,p_1,p_2,\ldots,p_n) \qquad (6.11)$$

According to Deluca and Termini H_{tot} gives the total average uncertainty that one may have in making a prevision about the elements of $\{x_1,x_2,\ldots,x_n\}$ which occur as a result of the experiment, and in taking a decision 0 or 1 on their values.

Xie and Bedrosian [121] defined the total entropy of a fuzzy set in a little different way. Consider a set A′ containing only 0 and 1 with probabilities p_0 and p_1 ($p_0 + p_1 = 1$, $0 \le p_0,p_1 \le 1$). Suppose, due to some reason the sharpness in the set has changed and resulted in a fuzzy set A. The membership value of an element is changed to an arbitrary value in the range [0,0.5] from 0 and in [0.5,1] from 1. In this way the ordinary set A′ has been changed to a fuzzy set A. Thus the fuzzy set A has two types of uncertainties, one due to the random uncertainty in the ordinary set and the other is the fuzzy uncertainty arising due to the fuzziness in the set. They defined the total entropy of a fuzzy set as follows.

$$H_{tot} = H(p_0,p_1) + (1/n) \sum_{i=1}^{n} S(\mu_i) \qquad (6.12)$$

This entropy reduces to Shannon's entropy of a two state system

when the fuzziness is removed (i.e., when the second part vanishes).

6.4 JUSTIFICATION FOR NEW DEFINITIONS [62,126,127]

In this section we shall be justifying the need for two new definitions namely, **hybrid** and **higher order** entropy of fuzzy sets. Before explaining their necessity let us, first critically analyse the aforementioned measures.

Regarding equation 6.11 the following points are in order. Deluca and Termini [120] presented equation 6.11 as if m in equation 6.11 is different from equation 6.6. But this is not the case. A critical analysis of the aforesaid example considered to derive equation 6.11 shows that m is, in fact, equal to H in equation 6.6. Suppose, the said experiment is repeated N times, the entropy of the resulting fuzzy set, as given by equation 6.6 would then be

$$H = \{-1/N\} \sum_{i=1}^{N} \{\mu_i \log(\mu_i) + (1-\mu_i) \log(1-\mu_i)\} \qquad (6.13)$$

Let n_i be the number of times x_i has occurred, then $\sum_{i=1}^{n} n_i = N$.

Thus, $H = (-1/N) \sum_{i=1}^{n} n_i \{\mu_i \log(\mu_i) + (1-\mu_i) \log (1-\mu_i)\}$

or $H = - \sum_{i=1}^{n} (n_i/N) \{\mu_i \log(\mu_i) + (1-\mu_i) \log(1-\mu_i)\}$

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

or $H = \sum_{i=1}^{n} p_i S(\mu_i)$

or $H = m(\mu, p_1, p_2, \ldots, p_n)$

Therefore, m in equation 6.11 is not different from H in equation 6.6.

Referring back to the same experiment of Deluca and Termini, there will not be any difficulty in interpreting an outcome x_i as either 0 or 1 when the associated fuzziness is removed. In other words, the entropy of the system should boil down to that of a two state system, $H(p_0, p_1)$, $p_0 + p_1 = 1$. But this is not supported by equation 6.11 which, in such a case, reduces to $H(p_1, p_2, \ldots p_n)$. This has also been pointed out by Xie and Bedrosian [121] whose measure (equation 6.12) is free from this drawback.

However, the equation 6.12 has the following unappealing implications. Xie and Bedrosian established an equivalence between fuzzy information and Shannon information because both of the information measures have the same mathematical form. If $p_i = \mu_i$, they inferred that the average amount of fuzzy information yielded by a fuzzy set with n elements is "equivalent" to the average amount of Shannon information yielded by n independent binary Shannon information sources. Based on this, it has been concluded that fuzzy information can be transferred to Shannon information and inversely [121]. This type of equivalence is physically meaningless except that both of the measures yield same numerical value.

Moreover, fuzzy information is conceptually different from the probabilistic information. Their arithmatic sum may not yield any meaningful quantity. In that sense, it is difficult to interpret H_{tot}. Furthermore, if fuzziness is removed, equation 6.12 always reduces to $H(p_0,p_1)$ irrespective of the defuzzification process. For example, consider the following two cases. In the first case, the symbols in the range [0,0.5] are converted to zero and the remaining to unity. In the second case, some of the symbols in [0,0.5] are wrongly mapped to 1. In both cases equation 6.12 will yield the same entropy value. This is not at all a desirable property.

It is to be mentioned here that since a fuzzy set is a generalised version of an ordinary set, the entropy of a fuzzy set deserves to be a generalised version of classical entropy by taking into account not only the fuzziness of the set but also the underlying probability structure . In other words, it should be such that the classical entropy becomes its special case when fuzziness is removed. It is also not necessary that, with the removal of fuzziness the value of the generalised entropy decreases. Furthermore, in the example of 0 and 1 considered by Xie and Bedrosian, there would be one and only one type of uncertainty (difficulty) which is associated with the interpretation of an incoming symbol as 0 or 1. Of course, this difficulty depends on two factors namely, the probability distribution and the possibility distribution; but this can not be the sum of two uncertainties (e.g. probabilistic and fuzzy) as

done in equation 6.12. Such a measure may be called **Hybrid Entropy.**

The entropy of a fuzzy set as given by either equation 6.6 or 6.7 gives the average ambiguity in taking the decision whether an element belongs to the set or not. Consider a fuzzy set " good foot ball players" with 20 members. For any player x , $\mu(x)$ gives the degree of goodness of the player x. The entropy as given by equation 6.6 or 6.7 for such a set gives an average amount of difficulty in taking a decision whether an individual player is good or not. This type of information is not always sufficient. Often we are interested to know, if a team of, say eleven, players is formed by selecting any eleven players from the twenty, to what extent on an average, the team can be called "good" . This raises two important issues. Firstly, how to measure the degree to which a collection of objects, as a whole, possesses the property of "goodness" . Secondly, how to get a measure of average amount of uncertainty related to such collections.

Let us now consider the first issue. Suppose $x_1, x_2, .., x_n$ are n supporting elements with membership values $\mu_1, \mu_2, \ldots, \mu_n$ with respect to some property P. Then to what extent $\{ x_1, x_2, \ldots, x_n \}$ as a whole (collectively) possesses the property P. This obviously depends on the problem at hand. For example, in a quiz team, if μ_i is the ability of the i-th member then the ability of the team as a whole would be $max\ (\mu_i)$, because if one member succeeds the team succeeds. On the other hand, suppose a group of acrobats are

standing in such a fashion that all of them will fall if any one of them falls. Under this situation, if μ_i is the stability of the i-th member, then the stability of the team as a whole would be $min\ (\mu_i)$. Allen [135] has addressed this issue of properties of sets.

In order to get an answer to the second problem, a new definition of entropy of a fuzzy set is required which will give a measure of average uncertainty associated with any arbitrary subset with r supports. Such an entropy may be called the r-th **Order Entropy** of a fuzzy set(r th order fuzzy entropy or higher order fuzzy entropy). For $r = 1$, the new definition should correspond to a measure satisfying the properties **P1** thru **P4**.

6.5 NEW DEFINITIONS [62,126,127]

In this section we shall define the higher order entropy and the hybrid entropy of a fuzzy set.

6.5.1 Higher Order fuzzy Entropy

Let P be a fuzzy property set with a finite number of supports n, i.e., $P = \{\mu_i/x_i,\ i=1,2,..,n\ \}$ where μ_i denotes the degree to which x_i possesses the property P. Out of n elements consider a combination of r elements. Let S_i^r denote the i-th such combination and $\mu(S_i^r)$ denote the degree to which the combination S_i^r , as a whole, possesses the property P. There are $\binom{n}{r}$ such combinations. The entropy of order r of the fuzzy set A is defined as

$$H^r = 1 / \binom{n}{r} \sum_{i=1}^{\binom{n}{r}} [\mu(S_i^r)\, e^{1-\mu(S_i^r)} + \{1-\mu(S_i^r)\}\, e^{1-\mu(S_i^r)}] \quad (6.14)$$

Similar definition using the logarithmic function can also be given as follows.

$$H^r = -1 / \binom{n}{r} \sum_{i=1}^{\binom{n}{r}} [\mu(S_i^r)\, log\{\mu(S_i^r)\} + \{1-\mu(S_i^r)\}\, log\{1-\mu(S_i^r)\}] \quad (6.15)$$

In our subsequent discussion unless stated explicitly, H^r will refer to the equation 6.14. Therefore, H^r will give a measure of the average amount of difficulty in taking a decision on any subset of size r with respect to the property P. In the example of Acrobats, H^r will denote an average amount of ambiguity (difficulty) in deciding a random team of r acrobats as "stable". If $r = 1$, H^r in equations 6.14 and 6.15 reduces to equation 6.7 and 6.6, respectively. H has the following properties.

Pr1. H attains a maximum if $\mu_i = 0.5$ for all $i = 1,2,\ldots,n$.

Pr2. H attains a minimum if $\mu_i = 0$ or 1 for all $i = 1,2,\ldots,n$.

Pr3. $H^r \geq H^{*r}$, where H^{*r} is the r-th order entropy of a sharpened version of the fuzzy set.

Pr4. H is, in general, not equal to $H^{r'}$, where $H^{r'}$ is the r-th order entropy of the complement set.

Pr5. $H^r \geq H^{r+1}$ when all $\mu_i \in [0,0.5]$ and

$H^r \leq H^{r+1}$ when all $\mu_i \in [0.5,1]$.

Note that the property **P4** of equation 6.6 or 6.7 is not, in general, valid here. The additional property **Pr5** implies that H is a monotonically nonincreasing function of r for $\mu_i \in [0,0.5]$ and a monotonically nondecreasing function of r for $\mu_i \in [0.5,1]$ (when *min* operator has been used to get the group membership value). Proofs of **Pr1** thru **Pr3** for H^1 (i.e., for $r=1$) are given in chapter - 4; when the exponential behavior of gain function is considered in the entropy measure. Their proofs for H^r ($r \geq 1$) along with the proof of **Pr5** are given below.

Proof of Pr1.

If $\mu_i = 0.5$ for all i, then $\mu(S_j^r) = 0.5$ for $j = 1,2, \ldots,\binom{n}{r}$ irrespective of whether the *max* or *min* operator is used to get the group membership function. This fact and the proof of **P1** in section 4.3.6 prove the property .

Proof of Pr2.

If $\mu_i = 0$ or 1 for all i , then $\mu(S_j^r) = 0$ or 1 for $j = 1,2,..,\binom{n}{r}$ in both cases when the *max* or *min* operator is used to arrive at the group membership function. This fact in conjunction with the proof of **P2** in section 4.3.6 , proves the property.

Proof of Pr3.

If $\mu(S_j^r) \geq 0.5$ then $\mu(S_j^{r*}) \geq 0.5$ and $\mu(S_j^{r*}) \geq \mu(S_j^r)$. Again when $\mu(S_j^r) \leq 0.5$ then $\mu(S_j^{r*}) \leq 0.5$ and $\mu(S_j^{r*}) \leq \mu(S_j^r)$. Note that the above two conditions are true both for the *max* and *min* operators. This fact and the proof of **P3** (in section 4.3.6) prove this property.

Proof of Pr4.

Proof of this property is obvious.

Proof of Pr5.

Let the property set P be defined from the set of supports S_i^r as

$$P_i^r = \{(\mu_j^i / x_j^i),\ j = 1,2,....,r;\ i=1,2,...,\binom{n}{r}\}$$

where $x_j^i \in S_i^r$ and μ_j^i denotes the degree of possessing the property P by the j-th element of S_i^r. Let $\mu(S_i^r) = \min_j \{\mu_j^i,\ j=1,2,...r\}$. Now there are $(n-r)$ subsets of size $(r+1)$ such that

$$S_i^r \subset S_i^{r+1}$$

Let us denote these subsets by $S_{i,j}^{r+1}$, $j=1,2,....,n-r$;

i.e., $S_{i,j}^{r+1}$ is the j th subset of P of size $(r+1)$ such that

$$S_i^r \subset S_{i,j}^{r+1}.$$

Now $S_i^r \subset S_{i,j}^{r+1}$ (6.16.1)

$$\Longrightarrow \min\{\mu_1^i,\mu_2^i,....,\mu_r^i\} \geq \min\{\mu_1^{i,j},\ \mu_2^{i,j},....,\mu_{r+1}^{i,j}\} \quad (6.16.2)$$

$$\Longrightarrow \mu(S_i^r) \geq \mu(S_{i,j}^{r+1});\ j = 1,2,.....,n-r \quad (6.16.3)$$

Thus, for every S_i^r, $i = 1,2,..,\binom{n}{r}$, there are $(n-r)$ subsets of size $(r+1)$ such that condition 6.16.3 is true. Thus for every $\mu(S_i^r)$, there are $(n-r)$ times $\mu(S_i^{r+1})$ such that

$$\mu(S_i^r) \geq \mu(S_{i,j}^{r+1}),\ j = 1,2,.....,(n-r). \quad (6.16.4)$$

Since from a set of size $(r+1)$ we can generate exctly $(r+1)$ distinct set of size r, we can infer that in the sequence

$$\mu(S_{i,j}^{r+1}) \ , \ i = 1,2,\ldots,\binom{n}{r} \ ; \ j = 1,2,\ldots,(n-r)$$

each of

$$\mu(S_i^{r+1}) \ , \quad i = 1,2,\ldots,\binom{n}{r+1}$$

has occurred exactly $(r+1)$ times.

Thus,

$$(n-r) \sum_{i=1}^{\binom{n}{r}} \mu(S_i^r) \geq (r+1) \sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1}) \qquad (6.16.5)$$

$$\Longrightarrow \frac{(n-r) \sum_{i=1}^{\binom{n}{r}} \mu(S_i^r)}{\binom{n}{r}} \geq \frac{(r+1) \sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1})}{\binom{n}{r}} \qquad (6.16.6)$$

$$\Longrightarrow (n-r)\,\bar{\mu}_r \geq (r+1) \frac{\sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1})}{\binom{n}{r}} \qquad (6.16.7)$$

$$\Longrightarrow \bar{\mu}_r \geq \frac{\sum_{i=1}^{\binom{n}{r+1}} \mu(S_i^{r+1})}{\binom{n}{r+1}} \qquad (6.16.8)$$

$$\Longrightarrow \bar{\mu}_r \geq \bar{\mu}_{r+1} \qquad (6.17)$$

Thus, the average value of all $\mu(S_i^r)$ is not less than that of $\mu(S_i^{r+1})$.

In order to prove **Pr5**, let us investigte the behavior of an entropy function with respect to its membership values. Consider the entropic function(equation 6.7).

In chapter - 4 (section 4.3.6) we have seen that $f(\mu) = \mu\, e^{1-\mu} + (1-\mu)\, e^{\mu}$ monotonically increases with μ over $\mu \in [0,0.5)$, attains a maximum at $\mu = 0.5$ and then monotonically decreases over $\mu \in (0.5,1]$.

This fact and inequation 6.17 prove the property **Pr5**.

6.5.2 Illustration.

Let us consider the problem of selecting a team of five acrobats from a group of ten. Define a fuzzy set A called "stable acrobats" with ten supports , i.e., $A = \{\mu_i / x_i ; i = 1,2,..,10\}$, where μ_i denotes the stability of the i-th individual x_i. The fuzzy entropy of order 5 gives the ambiguity (difficulty) in deciding any arbitrary team of 5 players from the group of 10 as stable or unstable. The values of H^1 thru H^6, presented in Table 6.2 highlight the key features of higher order entropy as compared to H^1.

Based on Table-6.2 the following observations can be made.

When all the μ_i values are same, H^1 to H^6 , as expected become same. This is because of the fact that the difficulty in

TABLE-6.2: Higher order entropy

CASE	μ_A	H^1	H^2	H^3	H^4	H^5	H^6	S
1	{1,1,1,1,1,1,1,1,1,1}	0	0	0	0	0	0	1.0
2	{.5,.5,.5,.5,.5,.5,.5,.5,.5,.5}	1	1	1	1	1	1	1.0
3	{1,1,1,1,1,.5,.5,.5,.5,.5}	0.5	.7777778	.9166667	.9761905	.9960318	1	.6428570
4	{.5,.5,.5,.5,.5,.6,.6,.6,.6,.6}	.9809018	.9915119	.9968170	.9990906	.9998485	1	.9892983
5	{.6,.6,.65,.9,.9,.9,.9,.9.9,.915}	.5384138	.6785914	.7810661	.8553280	.9059016	.9373114	.7934287
6	{.8,.8,.8,.8,.8,.8,.9,.9,.9,.9}	.5388815	.6135760	.6415336	.6495325	.6508657	.6508657	.8782636
7	{.5,.5,.5,.5,.5,.5,.9,.9,.9,.9}	.7483621	.9161205	.9790324	.9970038	1	1	.8168817
8	{.7,.7,.7,.7,.7,.8,.8,.8,.8,.8}	.7485938	.8028872	.8300339	.8416682	.8455463	.8463219	.9323772

taking a decision regarding possession of a property on an individual is same as that of a group selected therefrom. The value of H would, of course, be dependent on the μ_i values. Since all the μ_i in case-2 are 0.5 (most ambiguous) the corresponding H value is higher than that of case-1.

Consider the cases 3 and 4. In case 3 the difference between H^1 and H^2 is very high indicating a higher dis-similarity within the singletons of the fuzzy set with respect to the possession of property P. On the other hand, the case 4 for which H^1 is very close to H^2 reflects that the fuzzy singletons of the set are more alike (similar) with respect to the property P. It is also seen that the higher the similarity among singletons the quicker is the convergence to the limiting value of H.

Based on the above observation, let us define an **Index of similarity** of supports of a fuzzy set as $S = H^1 / H^2$ (when $H^2 = 0$, H^1 is also zero and S is taken as 1). Obviously, when $\mu_i \in [0.5,1]$ and the *min* operator is used to assign the degree of possession of the property by a collection of supports, S will lie in [0,1], as $H^r \leq H^{r+1}$ .Similarly, when $\mu_i \in [0,0.5]$, S may be defined as H^2 / H^1 so that S lies in [0,1]. Higher the value of S the more alike (similar) are the supports of the fuzzy set with respect to the property P. This index of similarity can, therefore, be regarded as a measure of the degree to which the members of a fuzzy set are alike.

Let us now consider the cases 5 and 6. In both cases the H^1

values are almost same but the index of similarity is lower for case 5 than that of case 6. This indicates that the case 6 has more similar supports than that of the case 5.

Note that the cases 7 and 8 also have almost identical values for H^1 , but these values are much higher than those of the cases 5 and 6. Based on this observation we can infer that fuzzy sets corresponding to cases 7 and 8 have more number of supports with lower degree of possession in [0.5,1] of the property P than those of the cases 5 and 6 . Again the index of similarity for case 8 is higher than that of case 7 indicating that the members of the fuzzy set corresponding to case 8 are more alike than those of case 7.

Therefore, the value of conventional fuzzy entropy (H^1) can only indicate whether the fuzziness in a set is low or high. In addition to this, the value of H^r also enables one to infer whether the fuzzy set contains similar supports (or elements) or not. The similarity index thus defined can be successfully used for measuring interclass and intraclass ambiguity (i.e., class homogenity and contrast) in pattern recognition and image processing problems.

Let us consider the work of Pal and Chakraborty [136] who have used the measure H^1 for evaluating feature importance in Pattern recognition problem . They used Π - type membership function to represent a pattern class. They made the membership values lie in the range [0.5,1] with 0.5 corresponding to the

boundary elements of the pattern class and 1 at the central point (the point corresponding to the average value of the feature). This type of representation is a very natural one. They decided a feature to be of high importance , if it has a low H^1 value. We have already shown (e.g., cases 5 and 6, or 7 and 8 of Table-6.2) that it is possible to have two fuzzy sets with significantly different compactness, but almost identical values for entropy . It is therefore, not appropriate to evaluate the importance of a feature just on the basis of H^1 . A better evaluation may possibly be done by considering H^1 and S together. Obviously, a low value of H^1 together with a high value of S will indicate that the feature is more important.

Table -6.2 also reveals the validity of the property Pr5, i.e., $H^r \leq H^{r+1}$ when $\mu_i \in [0.5,1]$.

6.5.3 HYBRID ENTROPY

In section 6.3 we have discussed the various attempts that have been made to combine the probabilistic and fuzzy uncertainties and their associated problems. Since an ordinary set is a special case of a fuzzy set, it is logical to think of a generalised definition of entropy of a fuzzy set which reduces to the probabilistic entropy in absence of fuzziness. Let us refer back to the example (mentioned in section 6.3) of digital communication over a noisy channel. For this type of example Xie and Bedrosian [121] assumed two types of uncertainties. But one

can visualise that there is one and only one type of difficulty and that is in the interpretation of an incoming symbol as 0 or 1. Of course, this is dependent on two different factors: the probability of generation of 0 and 1 by the source, and the transfer function (channel noise), which makes them fuzzy.

Let p_0 and p_1 be the probabilities of occurrence of the 0 and 1 symbols respectively, and μ_i denotes the membership for the fuzzy set "symbol close to 1".

Let us consider an expression

$$E_1 = (1/n) \sum_{i=1}^{n} \mu_i \, e^{1-\mu_i} \qquad (6.18)$$

Differentiating E_1 with respect to μ_i we get

$$E_1'(\mu_i) = (1/n)\,(1-\mu_i)\, e^{1-\mu_i} \; ; \; i = 1,2,..,n$$

$$\Longrightarrow \quad E_1'(\mu_i) \geq 0 \text{ for } 0 \leq \mu_i \leq 1 \; .$$

(i.e., the rate of change of E_1 is seen to be non negative)

Thus E_1 is a monotonically increasing function of μ_i for $\mu_i \in [0,1]$. In other words, as μ_i increases from 0 to 1, E_1 also increases. Now for an incoming zero(0) symbol, if μ_i increases the difficulty in the correct interpretation of the symbol increases and hence the difficulty in the interpretation of a '0' as '1' decreases; i.e, a wrong interpretation of a '0' symbol becomes favourable. On the other hand, for an incoming

'1' symbol if μ_i increases the difficulty in its correct interpretation decreases. The higher the value of E_1, the more favourable will be the interpretation of an arbitrary incoming symbol as 1. Therefore, E_1 can be taken as the average likeliness (possibility) of interpreting a received symbol as '1'.

Similarly,

$$E_0 = (1/n) \sum_{i=1}^{n} (1-\mu_i)\, e^{\mu_i} \qquad (6.19)$$

can be viewed as the average likeliness of interpreting a received symbol as 0.

Since p_0 and p_1 are respectively the probabilities of occurrence of 0 and 1 the **Hybrid Entropy** of the fuzzy set A (symbols close to 1) may be defined as

$$H_{hy} = - p_0 \log (E_0) - p_1 \log (E_1) \qquad (6.20)$$

(using logarithmic behavior)

or

$$H_{hy} = p_0 \; e^{1-E_0} + p_1 \; e^{1-E_1} \qquad (6.21)$$

(using exponential behavior)

H_{hy} has the following properties.

Prop 1.

In the absence of fuzziness when $n\, p_0$ received symbols take the value zero(0) and $n\, p_1$ symbols take the value of one(1) then E_0 and E_1 reduce to p_0 and p_1

respectively and equations 6.20 or 6.21 boil down to classical entropy. Thus, we see that H_{hy} boils down to the classical entropy when a proper defuzzification process is applied to detect (restore) correctly the incoming symbol.

The proof of this property is trivial.

This property has significant application in automatic image enhancement, segmentation and noise reduction process where the difference between the classical entropy and H_{hy} can act as an objective criterion. The lower the difference, the lesser is the fuzziness associated with the individual symbol and higher will be the accuracy in classifying them as 0 or 1.

Prop 2.

If $\mu_i = 0.5$ for all $i = 1,2,....,n$ then $E_0 = E_1$, and equations 6.20 and 6.21 become independent of p_0 and p_1. In this case H_{hy} assumes a constant value.

Proof

If $\mu_i = 0.5$ for $i = 1,2,..,n$; then equation 6.20 takes the form,

$$H_{hy} = - p_0 \log \{(1/n) \sum_{i=1}^{n} \frac{1}{2} e^{1/2} \} - p_1 \log \{(1/n) \sum_{i=1}^{n} \frac{1}{2} e^{1/2}\} \quad (6.22)$$

or

$$H_{hy} = - \log \{\frac{1}{2} e^{1/2} \} \quad (6.23)$$

i.e., H_{hy} takes a constant value and becomes independent of p_0 and p_1. Similarly it can also be shown for equation 6.21.

This is, indeed a desirable property, as when all $\mu_i = 0.5$ the receiver is unable to distinguish among the received symbols. Note that this property is not satisfied by equations 6.11 and 6.12.

Prop 3.

For a given p_0 and p_1 ($p_0 + p1 = 1$, $0 \leq p_0, p_1 \leq 1$), of all defuzzified versions, the hybrid entropy H_{hy} is minimum for the properly defuzzified version.

proof

Suppose some defuzzification process has resulted in Q_0 times zero(0) symbols and Q_1 times one (1) symbols ($Q_0 + Q_1 = n$).

Then,

$$E_0 = (1/n) \{ \sum_{i=1}^{Q_0} (1-\mu_i)\, e^{\mu_i} + \sum_{i=Q_0+1}^{n} (1-\mu_i)\, e^{\mu_i} \} \qquad (6.24)$$

Without loss of generality we can assume that first Q_0 symbols are mapped to zero(0) and the remaining symbols to one(1).

Thus,

$$E_0 = (1/n) \quad Q_0 \qquad (6.25)$$

Similarly,

$$E_1 = (n-Q_0)/n \qquad (6.26)$$

and

$$E_0 + E_1 = Q_0 / n + (n - Q_0) / n = 1 \qquad (6.27)$$

Thus, we see that for any defuzzified version $E_0 + E_1 = 1$.

Now suppose under any arbitrary defuzzification process E_0 and E_1 take the values q_0 and q_1 $(q_0 + q_1 = 1)$, respectively. Under this situation the hybrid entropy takes the form,

$$H_{hy} = -p_0 \log q_0 - (1-p_0) \log (1-q_0). \qquad (6.28)$$

We have already seen that a proper defuzzification will make $q_0 = p_0$ and $q_1 = p_1$. Therefore, in order to prove the property, we need to show that

$$-p_0 \log p_0 - (1-p_0) \log (1- p_0) \leq$$
$$-p_0 \log q_0 - (1-p_0) \log (1- q_0) \qquad (6.29)$$

with equality if and only if $p_0 = q_0$.

We know that $\log x$ is a concave function, therefore, $\log x$ always lies below its tangent. Now consider the equation of tangent at $x = 1$; which is $y = x-1$. Thus, we can write $\log x \leq x-1$, with equality if and only if $x=1$.

Therefore,

$$\log (q_0 / p_0) \leq q_0 / p_0 - 1$$

or
$$p_0 \log (q_0 / p_0) \leq q_0 - p_0 \qquad (6.30)$$

Similarly,

$$(1-p_0) \log \{ (1-q_0) / (1-p_0) \} \leq (1-q_0) - (1-p_0) \qquad (6.31)$$

Now from equations 6.30 and 6.31 we get,

$$p_0 \log (q_0 / p_0) + (1 - p_0) \log \{ (1-q_0) / (1-p_0) \} \leq$$
$$q_0 - p_0 + (1-q_0) - (1-p_0) \qquad (6.32)$$

or $p_0 \log q_0 - p_0 \log p_0 + (1-p_0) \log (1-q_0) -$

$$(1-p_0) \log (1-p_0) \leq 0 \qquad (6.33)$$

or $-p_0 \log q_0 - (1-q_0) \log (1-q_0) \geq$

$$-p_0 \log p_0 - (1-p_0) \log (1-p_0) \qquad (6.35)$$

Hence the proof.

Prop 4.

E_0 and E_1 can not take any arbitrary values. They satisfy the following conditions:

$$\frac{1}{2} e^{1/2} p_1 \leq E_1 \leq \frac{1}{2} e^{1/2} + (1- \frac{1}{2} e^{1/2}) p_1 \qquad (6.36)$$

and

$$(1-p_1) \frac{1}{2} e^{1/2} \leq E_0 \leq 1-p_1 (1 - \frac{1}{2} e^{1/2}) \qquad (6.37)$$

Proof

Let us consider equation 6.19 which is

$$E_1 = (1/n) \sum_{i=1}^{n} \mu_i e^{1-\mu_i}$$

$$\Longrightarrow \quad E_{1max} = (1/n) [n p_0 * \frac{1}{2} * e^{1/2} + n p_1]$$

(Since E_1 is found to be monotonically increasing function of μ_i and the expected number of symbol 0 (1) is $n p_0$ ($n p_1$) and each of them can take a maximum value of 1/2 (1)).

$$= p_1 + (1-p_1) * \frac{1}{2} e^{1/2}$$

$$= \frac{1}{2} * e^{1/2} + (1- \frac{1}{2} * e^{1/2}) p_1 \qquad (6.38)$$

and

$$\Longrightarrow \quad E_{1min} = (1/n)\,[n\,p_0 * 0 + n\,p_1 * \tfrac{1}{2} * e^{1/2}]$$

$$= p_1 * \tfrac{1}{2} * e^{1/2} \qquad (6.39)$$

Combining equations 6.38 and 6.39 we get

$$\tfrac{1}{2} * e^{1/2} * p_1 \leq E_1 \leq \tfrac{1}{2}\, e^{1/2} + (1 - \tfrac{1}{2}\, e^{1/2})\, p_1 \qquad (6.40)$$

$$\Longrightarrow \quad 0.82436 * p_1 \leq E_1 \leq 0.82436 + 0.17564 * p_1$$

Similarly, from equation 6.20 we have

$$E_{0max} = (1/n)\ \{ n\,p_0 * 1 + n\,p_1 * \tfrac{1}{2}\, e^{1/2} \}$$

$$= 1 - p_1\ \{ 1 - \tfrac{1}{2}\, e^{1/2} \} \qquad (6.41)$$

and

$$E_{0min} = (1/n)\ \{ n\,p_0 * \tfrac{1}{2} * e^{1/2} + n\,p_1 * 0 \}$$

$$= (1-p_1) * \tfrac{1}{2} * e^{1/2} \qquad (6.42)$$

Combining equations 6.41 and 6.42 one can write the following condition.

$$(1-p_1) * \tfrac{1}{2} * e^{1/2} \leq E_0 \leq 1 - p_1\,(1 - \tfrac{1}{2} * e^{1/2}) \qquad (6.43)$$

$$\Longrightarrow (1-p_1) * 0.82436 \leq E_0 \leq 1 - 0.17564 * p_1 \qquad (6.44)$$

This completes the proof.

Prop 5.

$E_0 + E_1$ attains the maximum value of $e^{1/2}$ when all μ_i s assume a value of 1/2.

Thus,

$$E_0 + E_1 \leq e^{1/2} \qquad (6.45)$$

Proof

In chapter 4 it has been shown that equation 6.7 attains the maximum value of $e^{1/2}$ when all $\mu_i = 0.5$. Hence the above condition is true.

As mentioned in Prop 1, when a proper defuzzification process is applied to result in $n\ p_0$ times 0 symbol and $n\ p_1$ times 1 symbol, then E_0 and E_1 reduce to p_0 and p_1 respectively to make $E_0 + E_1 = 1$.

6.5.4 Interpretation of H_{hy} in Image Processing

Let us consider an example of a digital image in which, say, 70 % pixels look white, while the remaining 30 % look dark. Thus the probability of a white pixel p_w is 0.7 and that of a dark pixel p_b is 0.3. Suppose, the whiteness of the pixels is not constant, i.e., there is a variation (grayness) and similar is the case with the black pixels. The hybrid entropy of such an image can be written as follows.

$$H_{hy} = - p_w \log \{(1/n) \sum \mu_i\ e^{1-\mu_i} \} - p_b \log \{(1/n) \sum (1-\mu_i)\ e^{\mu_i} \}$$

where μ_i gives the degree of whiteness of the i-th pixel. Now H_{hy} will be close to the classical entropy if there is not much variation in the grayness within the white and dark portions. Table 6.3 demonstrates the use of H_{hy} in acting as an

TABLE-6.3 : Hybrid entropy

(with $p_0 = 0.3$ and $p_1 = 0.7$)

CASE	μ_A	E_0	E_1	H_{hy}	$\lvert H - H_{hy} \rvert$	H_{tot}	$\lvert H - H_{tot} \rvert$
1	{.9,.8,.8,.8,.8,.7,.6,.45,.4,.3}	0.6061934	0.8856076	0.2352039	.3756607	1.169123	.55825830
2	{.99,.9,.9,.9,.8,.79,.7,.2,.1,.05}	0.5247224	0.7700894	0.3763399	.2345245	0.9789118	.36804750
3	{.999,.99,.99,.99,.98,.9,.9,.1,.05,0}	0.3622024	0.73642	0.5188341	.0920302	.7556359	.1447716
4	{1,1,1,1,1,1,1,0.0004,0,0 }	0.30	0.7001086	0.6107557	.00010866	0.6112173	.00035295
5	{1,1,1,1,1,1,1,0,0,0}	0.70	0.30	0.6108643	.0	0.6108643	.0

objective criterion for a proper defuzzification process.

Each fuzzy set of the Table - 6.3 is a sharpened (defuzzified) version of the previous one such that μ value increases (decreases) in the sharpened version for those $\mu > 0.5$ (< 0.5). One can observe that with proper defuzzification E_1 approaches 0.7 and E_0 approaches 0.3; in fact they monotonically decrease to 0.7 and 0.3 respectively from their initial values. As a result, = $|H-H_{hy}|$ decreases to zero.

Let us now consider the effect of improper defuzzification on the pattern shown in case 1 of the Table 6.4 . Two types of defuzzifications are considered here. In cases 2-4 all the symbols with $\mu = 0.5$ are transformed to zero when some of them were actually generated from symbol '1'. In cases 5-6 of Table 6.4 some of the μ values greater than 0.5 which were generated from symbol 1 (or belong to the white portion of the image) are wrongly defuzzified and brought down towards zero (instead of 1). In both situations, it is to be noted that $|H - H_{hy}|$ does not reduce to zero.

Let us now consider the column no. 7 in Tables 6.3 and 6.4 which indicates that the measure H_{tot} of Xie and Bedrosian does not have the aforesaid properties.

Now, in image processing the process of defuzzification can be viewed as a contrast enhancement operation. Therefore, the measure $|H - H_{hy}|$ can be regarded as an objective criterion for an appropriate enhancement of an image.

TABLE-6.4 : Effect of wrong defuzzification

(with $p_0 = 0.3$ and $p_1 = .7$)

CASE	μ_A	E_1	E_0	H_{hy}	$\lvert H - H_{hy} \rvert$	H_{tot}	$\lvert H - H_{tot} \rvert$
1	{.9,.9,.8,.8,.7,.6,.5,.5,.4,.3}	.6203827	.8765243	.2354792	.3753851	1.171366	.5605017
2	{.999,.999,.9,.8,.7,.7,.3,.3,.2,.1}	.5766330	.7760898	.3426057	.2682586	1.021888	.4110242
3	{1,1,1,.99,.9,.9,.1,.1,0,0}	.4508141	.6481178	.5425881	.0682762	0.7464977	.1356333
4	{1,1,1,1,1,1,0,0,0,0}	.40	.60	.6324652	.0216008	0.6108643	.00
5	{.99,.99,.1,.1,.9,.8,.7,.2,.1,.1}	.6304756	.6345526	.4567688	.1540955	0.945773	.3349087
6	{1,1,0,0,1,1,1,0,0,0}	.5	.5	.6931472	.08228284	0.6108643	.00

6.6 APPLICATIONS

This section attempts to illustrate some applications of H^r to edge detection and image segmentation of digital images. For detecting edges we have considered H^1 whereas, for segmentation H^1 and H^2 both have been used as objective criteria.

6.6.1 Application of H^1 to Contour Extraction

Edginess Measure

Let $N^d_{x,y}$ be a neighbourhood of order d of the pixel (x,y), $d \geq 2$. For example,

$$N^2_{x,y} = \{ (x,y),(x,y-1),(x,y+1),(x+1,y),(x-1,y)\} \qquad (6.46)$$

and

$$N^3_{x,y} = \{ (x,y),(x-1,y),(x+1,y),(x,y-1),(x,y+1),(x-1,y-1), (x-1,y+1),(x+1,y-1),(x+1,y+1)\} \qquad (6.47)$$

The edge-entropy, $H^E_{x,y}$, of the pixel (x,y), giving a measure of edginess at (x,y) may be computed as follows. For every pixel (x,y) having intensity $f(x,y)$, compute the average, maximum and minimum values of gray levels over $N^d_{x,y}$. Let us denote the average, maximum and minimum values by *Avg*, *Max* and *Min*, respectively. Now define the following quantities.

$$D = max \{ Max - Avg , Avg - Min \} \qquad (6.48)$$

$$B = Avg \qquad (6.49)$$

$$A = B - D \qquad (6.50)$$

$$C = B + D \qquad (6.51)$$

A Π-type membership function is then used to compute $\mu(x,y)$ for all $(x,y) \in N^d_{x,y}$, such that $\mu(A) = \mu(C) = 0.5$ and $\mu(B) = 1$. It is to be noted that $\mu(x,y) \geq 0.5$. Such a $\mu(x,y)$, therefore, gives the degree to which a gray level is close to the average value computed over $N^d_{x,y}$. In other words, it represents a fuzzy set "pixel intensity $f(x,y)$ close to its average value", averaged over $N^d_{x,y}$. When all $f(x,y)$ over $N^d_{x,y}$ are either equal or close to each other(i.e., they are within the same region), such a transformation will make all $\mu(x,y)=1$ or close to 1. In other words, if there is no edge, pixel values will be close to each other and the μ values will be close to one(1); thus resulting in a low value of H^1. On the other hand, if there is an edge (dissimilarity in gray values over $N^d_{x,y}$),then the μ values will be more away from unity; thus resulting in a high value of H^1. Therefore, the entropy H^1 over $N^d_{x,y}$ can be viewed as a measure of edginess($H^E_{x,y}$) at the point (x,y). The higher the value of $H^E_{x,y}$, the stronger is the edge intensity and the easier is its detection. There are several ways [102] in which one can define a Π- type function as shown in Fig. 6.1. Zadeh's standard Π function [105] may also be used here.

The above mentioned concept has been used to formulate an algorithm for edge detection of an image. The algorithm using $N^3_{x,y}$ neighbourhood system is explained below.

Algorithm 1

S0. For each $(x,y) \in F$, repeat steps S1 thru S4.

S1. Compute *Max*, *Min* and *Avg* for $f(x,y)$ over $N^3_{x,y}$.

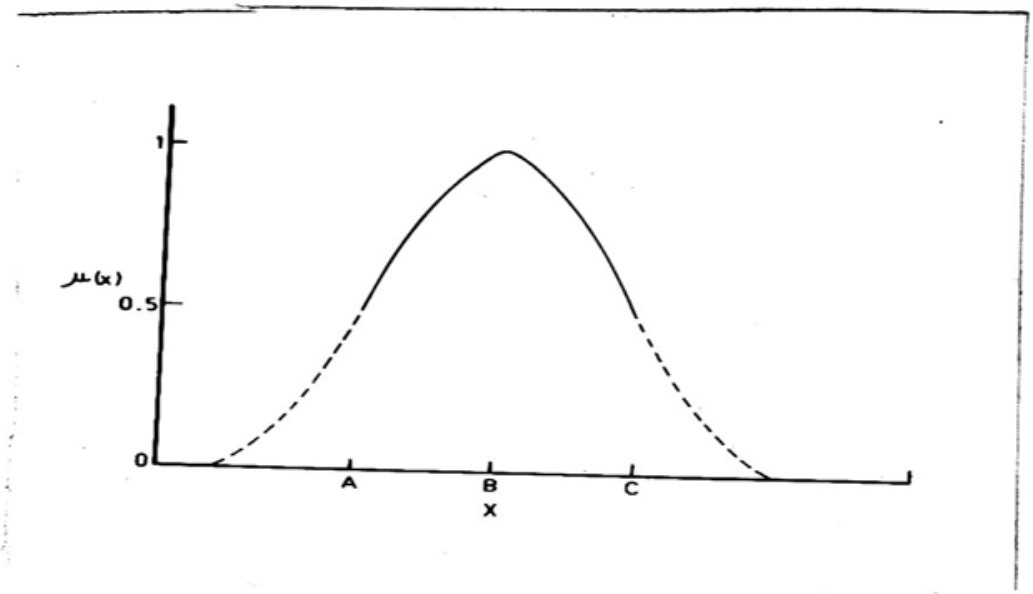


Fig. 6.1 Plot of Π function

S2. Compute D = Maximum { *Max* - *Avg* , *Avg* - *Min* }

$B = Avg$

$A = B - D$

$C = B + D$

S3. Apply Π - function [fig.6.1] to compute $\mu(x,y)$ for all $(x,y) \in N^3_{x,y}$ such that $\mu(B)=1$, $\mu(A)=\mu(C) = 0.5$

S4. Compute the edge entropy $H^E_{x,y}$ at (x,y) by H^1 (equation 6.6 or 6.7) over $N^3_{x,y}$.

S5. Normalise the edge entropy values in the range (1-L) to get the edge image.

6.6.2 Application of H^r to Object Extraction

Before describing the methodology, let us define an image from the fuzzy set theoretic point of view.

Image definition [101]:

An image F of size P*Q and L levels can be considered an array of fuzzy singletons, each having a value of membership denoting its degree of belongingness to a region in F. In the notation of fuzzy sets, we may, therefore, write

$$F = \bigcup_m \bigcup_n \{ \mu(m\ n) / f(m\ n) \} \qquad (6.52)$$

$m = 1,2,....,P$; $n = 1,2,...,Q$

where $f(m,n)$ denotes the gray value at the (m,n)th pixel and $\mu(m,n)$ $(0 \leq \mu(m,n) \leq 1)$ denotes the degree of belonging to some region(or grade of possessing some property) by the (m,n)th pixel

intensity, $f(m,n)$. The way of computing $\mu(m,n)$ from $f(m,n)$ depends on the problem at hand.

Selection of Membership Function for Segmentation

Let B be an assumed threshold which partitions the image F into two parts namely, object and background. Suppose the gray level ranges [1-B] and [B+1 - L] denote, respectively, the object and background of the image F. An inverse Π-type function as shown by the solid line in the Fig.6.2, is used here to obtain $\mu(m,n)$ values of F. The inverse Π-type function is seen (from Fig.6.2) to be generated by taking union of S(x ; (B-(L-B)),B,L) and 1 - S(x;1,B,(B + B - 1)), where S denotes the standard S function defined by Zadeh[105].

The resulting function as shwon by the solid line, makes μ lie in [0.5,1]. Since the ambiguity(difficulty) in deciding a level as a member of the object or the background is maximum for the boundary level B, it has been assigned a membership value of 0.5(i.e., Cross-over point). Ambiguity decreases (i.e., degree of belongingness to either object or background increases) as the gray value moves away from B on either side. The $\mu(m,n)$, thus obtained denotes the degree of belongingness of a pixel $f(m,n)$ to either object or background.

Since B is not necessarily the mid point of the entire gray scale, the membership function(solid line in Fig.6.2) may not be a symmetric one.

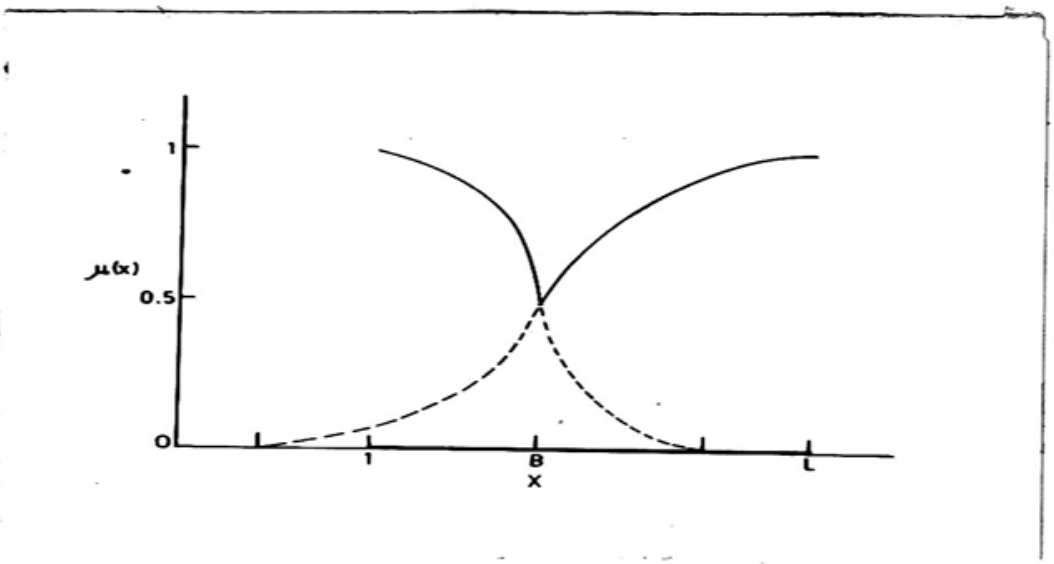


Fig. 6.2 Plot of inverse Π function

Minimisation of fuzzy Entropy for Thresholding

In order to explain the criterion for thresholding, let us consider H^1 and the Fig. 6.2. The entropy H^1 of the image F is defined as

$$H_s^1(F) = 1 / (P*Q) \sum_{m=1}^{P} \sum_{n=1}^{Q} \{ \mu(m,n)\ e^{1-\mu(m,n)} + (1-\mu(m,n)\ e^{\mu(m,n)} \}$$

As explained in section 6.3, H_s^1 (F) is maximum when $\mu(m,n)=0.5$ for all (m,n). Therefore, higher the number of pixels having $\mu(m,n) \cong 0.5$, the greater will be the value of $H_s^1(F)$.

Now modification of the cross over point B will result in different segmented images with varying H_s^1 (F). When B corresponds to the appropriate boundary between two regions, there will be a few pixels in F having $\mu(m,n) \cong 0.5$ and the rest will have values away from 0.5; thus contributing least towards $H_s^1(F)$. Therefore, $H_s^1(F)$ will show a minima at the appropriate boundary.

Let us now consider the second order entropy H^2 as a measure . Here H^2 reflects the similarity(homogenity) among the gray levels in a region. Therefore, considering the previous framework of H^1, minimisation of H^2 will also result in a proper segmentation of the object and background.

To explain the merit of H^2 over H^1, let us reconsider the cases 5 and 6 of Table-6.2. Here both cases have almost same H^1 value but case 6 has significantly lower value of H^2 than that of

case 5. It is also seen that the supports in case 6 are more similar than those of case 5. Therefore, H^2 is seen to reflect the homogenity among the supports in a better way than H^1 does.

Thus, considering the problem of object-background classification, H^2 seems to be more sensitive to the selection of appropriate threshold; i.e., the improper selection of the threshold is more strongly reflected by H^2 than H^1. In other words, the thresholds obtained by H^2 measure have more validity than those by H^1 (which only takes into account the histogram information). Similar arguments hold good for even higher order (r > 2) entropy.

Algorithm 2

Assume a threshold s , $1 \leq s \leq L$, and execute the following steps.

S1. Apply an inverse $\prod$ -type function [Fig.6.2] to get the fuzzy image $\mu(x,y)$, with $\mu(x,y) \in [0.5,1]$. (The membership function is in general asymmetric)

S2. Compute the r th order fuzzy entropy of the object (H_O^r) and the background (H_B^r) considering only the spatially adjacent sequences of pixels present within the object and background respectively. Use the *min* operator to get the membership value of a sequence of pixels.

S3. Compute the total rth order fuzzy entropy of the

partitioned image as $H_s^r = H_O^r + H_B^r$.

S4. Minimise H_s^r with respect to s to get the threshold for object background clasification.

6.7. RESULTS

The edge detection algorithm described in the previous section has been implemented on two images. In order to compare the relative performance of the proposed method , the well known difference operatros, namely, Robert's gradient, Prewitt's gradient and the Sobel's gradient have also been implemented. Fig. 6.3.1 represents the input image of a biplane, while Figs. 6.3.2 and 6.3.3 represent the edge detected output produced by the proposed method with exponential and logarithmic entropy, respectively. From these two figures, it is clearly visible that the exponential entropy has produced a less noisy and clear boundary of the plane than the logarithmic one. Figs. 6.3.4-6.3.6 depict the edge detected outputs generated by Robert's , Sobel's and Prewitt's gradients, respectively. From the results , one can see that both of the proposed entropic methods have resulted in much better results than these difference operators, which are found to generate many spurious edges. Fig. 6.4.1 displays the input image of Lincoln, while Figs. 6.4.2-6.4.6 represent the output images generated by the exponential entropic, logarithmic entropic methods , Robert's , Sobel's and Prewitt's difference operators respectively. Here again, we see that the proposed methods (exponential and logarithmic) have resulted in a much better edge detected output than the difference operators.

See Fig. 2.7.1
Page -85

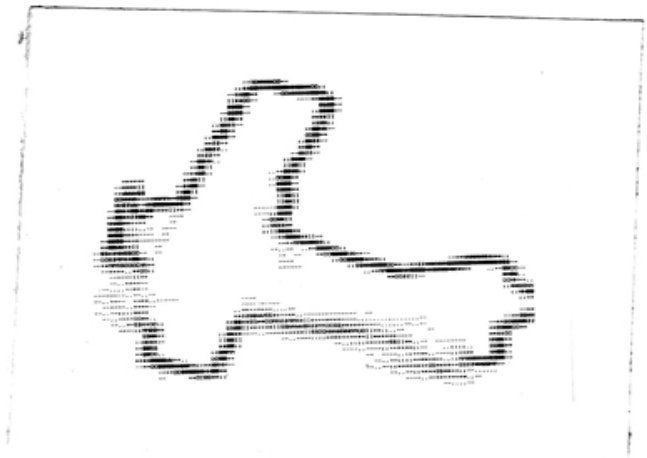


Fig. 6.3.1 Input

Fig. 6.3.2 Edge deteected output by exp. ent.

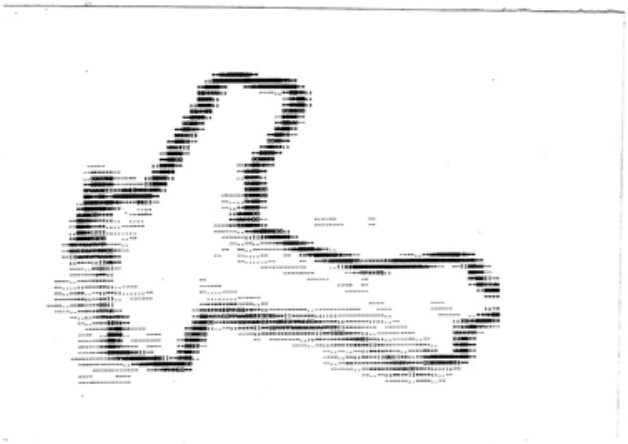


Fig. 6.3.3 Edge detected output by log. ent.

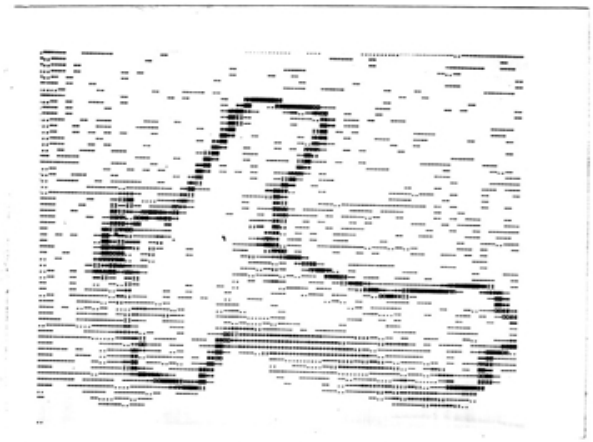


Fig. 6.3.4 Edge detected output by Robert's operator

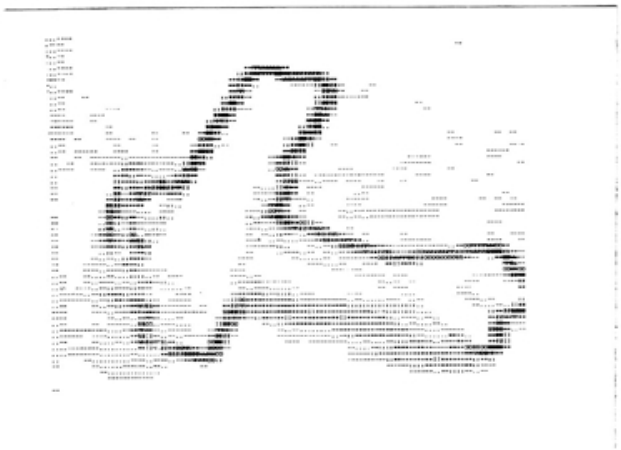


Fig. 6.3.5 Edge detected output by Sobel's operator

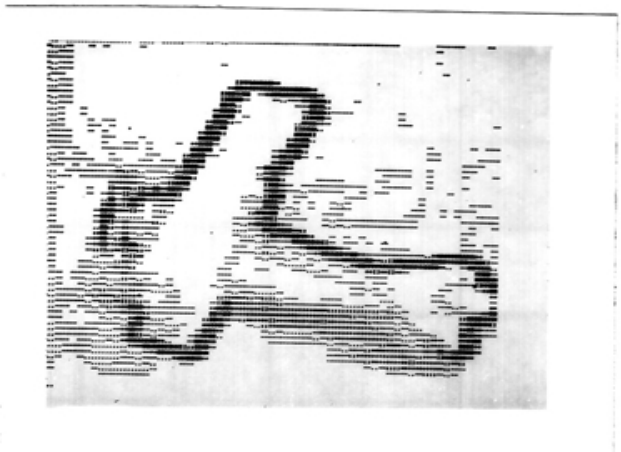


Fig. 6.3.6 Edge detected output by Prewitt's operator

Fig. 6.3 Input and edge detected output of biplane

See Fig. 2.5.1
Page - 81

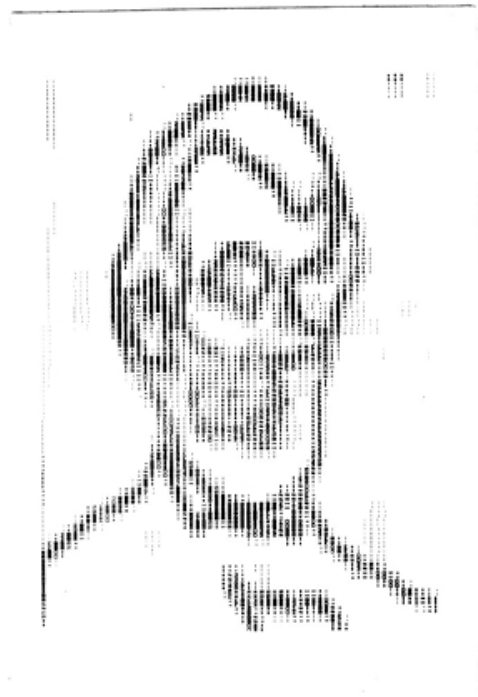


Fig. 6.4.1 Input

Fig. 6.4.2 Edge deteected output by exp. ent.

Fig. 6.4.3 Edge detected output by log. ent.

Fig. 6.4.4 Edge detected output by Robert's operator

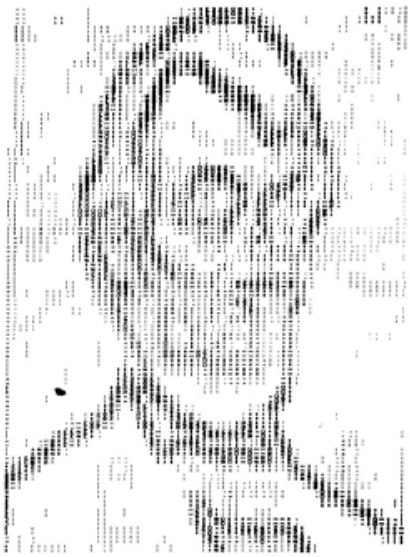


Fig. 6.4.5 Edge detected output by Sobel's operator

Fig. 6.4.6 Edge detected output by Prewitt's operator

Fig. 6.4 Input and edge detected output of Abraham Lincoln

The proposed entropic methods are found to be less sensitive to noise, because of the use of a dynamic membership function based on a local neighbourhood. The methods are not sensitive to the direction of edges. Moreover, any fuzziness measure, like index of fuzziness [102] can also be used in the proposed algorithm.

The proposed segmentation method (Algorithm 2) has been implemented for r = 1 and 2, and applied on the images of biplane , Lincoln and boy. The thresholds produced are shown in Table 6.5 and the segmented images are displayed in Figs 6.5 - 6.7. It is observed that both H^1 and H^2 have resulted in same thresholds (except for the image of boy).

6.8 CONCLUSION

New entropy measures such as higher order fuzzy entropy and hybrid entropy of a set have been introduced. The higher order entropy (H^r) is found to possess some interesting properties which can be used in feature evaluation and image segmentation problem. It leads to define a measure, called "Index of similarity" of supports of a set. As expected, H conveys more information about the actual structure of a set than H^1 does. H_{hy}, on the other hand, can be used as an objective measure for proper defuzzification (enhancement) of a set.

As an attempt to demonstrate their applicability ,H^1 has been used for edge detection while both H^1 and H^2 have been used for

TABLE - 6.5

Thresholds Produced by H^1 and H^2

IMAGES	H^1	H^2
Biplane Fig. 6.5	12	12
Lincoln Fig. 6.6	08	08
Boy Fig. 6.7	13	12

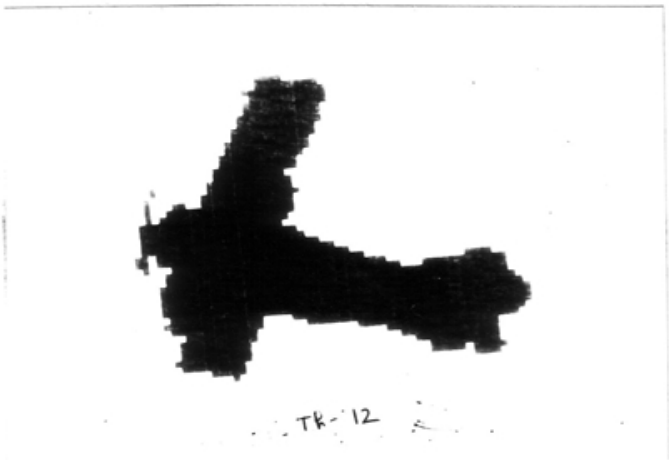


Fig. 6.5.1 & 6.5.2 Segmented output by H^1 and H^2

Fig. 6.5 Segmented image of biplane

Fig. 6.6.1 & 6.6.2 Segmented output by H^1 and H^2

Fig. 6.6 Segmented image of Abraham Lincoln

Fig. 6.7.1 Segmented output by H^1

Fig. 6.7.2 Segmented output by H^2

Fig. 6.7 Segmented images of boy

image segmentation. Their computation is based on an adaptive membership function which is automatically selected by the system without human intervention. The edge detected output produced by the proposed method is found to be better than those of the Robert, Sobel and Prewitt difference operators.

Chapter 7

CONCLUSION AND FURTHER SCOPE OF WORK

7.1 Conclusion

The work presented in this thesis consists mainly of two parts namely, providing various measures of information of a set, in general, and of an image in particular, and demonstrating their success in image segmentation problems. Theory of probability and fuzzy set theory have been used in this regard.

Based on the co-occurrences of pixels in an image, a measure of homogeneity within a region and a measure of contrast between regions have been proposed. The contrast measure makes use of the behavior of human reaction to brightness variation (the logarithmic response of human visual system to brightness). The higher order entropy and the conditional entropy of a partitioned image have been defined considering the interdependency of pixels. Both of these measures are dependent on the probability distribution of sequences of pixels. The higher order entropy provides an objective measure of image information taking into account the spatial distribution of gray levels. On the other hand, the conditional entropy gives a measure of contrast of a partitioned image. Attempt has also been made to define a measure, 'positional entropy', which gives information about the location of an object within a scene.

Besides giving all these information measures on an image, a general definition of classical entropy considering an exponential

behavior of information gain has been provided. In this context, the mathematical limitations of the well known Shannon's entropy, have been discussed. Various properties of the new exponential entropy have been established. This has also been extended for defining the aforementioned information measures of an image.

Need for defining the higher order fuzzy entropy and hybrid entropy of a set has been established . The higher order fuzzy entropy has been defined in order to have an estimate of the ambiguity associated with an arbitrary collection of supports. Based on this measure an index of similarity of supports has then been defined. In order to establish a bridge between the probabilistic and possibilistic uncertainties, a new measure called "hybrid entropy" of a set has been introduced. This may be viewed as a generalised entropy of a set such that it boils down to the classical entropy of a two state system when the associated fuzziness in the set is removed. Its application in contrast enhancement of an image is indicated.

To demonstrate the applicability of these measures, several algorithms for image segmentation have been formulated. These algorithms have been implemented on a set of images with widely different types of histogarm and the results are compared with those of several existing methods.

Two algorithms have been developed based on the measures of contrast and homogeneity for multi thresholding. Provisions have

been kept for removing the undesirable segments. This is achieved either in a separate merging phase or by taking into account the relative region size while formulating the objective criterion. The higher order entropy and the conditional entropy in their both logarithmic and exponential versions have been used to formulate various algorithms for object-background segmentation. The methods based on the conditional entropy are found to be much effective in extracting an object. As illustrations of the applicability of the higher order fuzzy entropy, an attempt has been made to use H^1 and H^2 as objective crirerion for edge detection and object extraction. These algorithms use adaptive (dynamic) membership functions which are determined from the image data without requiring any choice of parameter by the user.

Justification for using Poisson distributions to describe the gray level variation has been given based on the physics of image formation. A set of parametric object extraction algorithms have been formulated based on the maximum entropy principle and minimum χ^2 principle. The algorithms formulated need only global information. The superiority of algorithms using Poisson distribution over those using normal distribution has been established. The exponential entropic thresholding is seen to perform better than that of logarithmic entropy. Furthermore, the χ^2 - statistic is found to be a very useful criterion for object extraction.

In conclusion, the thesis provides various information measures of a set, in general, and image information in particular, and demonstrates their success in image segmentation problems. These measures include exponential classical entropy, higher order fuzzy entropy, hybrid entropy, local entropy, conditional entropy, positional entropy and contrast-homogeneity measures. Exponential entropy represents image information in a better way than the Shannon's entropy. Gray level distribution is more appropriately represented by Poisson distributions than by normal distributions. The conditional entropic method is found to be the best local information based method. Whereas, methods using Poisson distribution are seen to be best among the global information based methods.

7.2 Further Sccpe of Work.

Some more properties, of the exponential entropy , higher order fuzzy entropy, and the hybrid entropy of a set, need to be investigated. Application of these measures to object extraction problem has only been considered here. Further investigation of these measures to other areas e.g., feature evaluation , enhancement, restoration , noise reduction may be done. These measures are also useful in defining various quantitative indices for processing and analysis of image in robot vision and remote sensing applications where frequent human intervention (and subjective judgement on image quality) is not possible.

The superiority of the use of Poisson distribution to approximate the probability distribution of gray levels in an image has been established over the commonly used noraml distribution in image segmentation problem. This concept of Poisson distribution can be extended to many other occasions which use normal distribution.

The noise sensitivity of all the proposed algorithms needs to be studied. Further, the the application has been restricted here over only 2-dimensional gray tone image. Their extension to 3-dimensional object extraction and to colour image space may also be done. Development of quantitative measures for objective evaluation of the thresholded image and establishment of a link between the entropies presented here and some of the Markovian models may constitute a part of future investigation.

REFERENCES

1. A.Rosenfeld and A.C.Kak, Digital Picture Processing, *Academic Prees*, New York, 1982.

2. E. L. Hall, Computer Image Processing and Recognition , *Academic Press*, New York, 1979.

3. R.C. Gonzalez and P. Wintz, Digital Picture Processing, *Addison-Wesley*, Mass, London, 1977.

4. L. S. Davis and A. Rosenfeld, Iterative histogram modification, *IEEE* Trans. Syst. man and Cybern., 8,300-302,1978.

5. L. S. Davis and A. Rosenfeld, Noise cleaning by iterated local averaging, *IEEE Trans. Syst. Man Cybern.*,705-710, 1978.

6 T. S. Huang, Image enhancement : a review, *Opto- Elect.* 1, 49-59, 1969.

7. R. Hummel, Histogram modification techniques, *Comp. Graph. Image. Proc.*, 4, 209-224, 1975.

8. R. Hummel, Image enhancement by histogram transform, *Compt. Graph.. Image Proc.*, 6,184-195,1977.

9. A. Lev, S. W. Zuker and A Rosenfeld, Iterative enhancement of noisy images, *IEEE Trans. Syst. Man Cybern.*, 7, 435-442,1977.

10 J. M. S. Prewitt, Object enhancement and extraction in picture processing and psychopictorics(B. S. Lipkin and A rosenfeld Eds),*Academic press*, NY, 75-149, 1970.

11 S. K. Pal, Image enhancement and a quantitative index using fuzzy sets, Int. Jour. Systems Sc., 18,No. 9,1783-1797,1987.

12. E.S.Angel and A.K.Jain, Restoration of images degraded by spatially varying point spread functions by a conjugate gradient method, *Appl. Opt.* 17, 1978, 2186-2190.

13. T.M.Cannon, H.J.Trussell and B.R.Hunt, Comparison of image restoration methods, *Appl. Opt.* 17, 1978, 3384-3390.

14. N.Y.Chu and C.D.McGillem, Image restoration filters based on a 1-0 weighting over the domain of support of the PSF, *IEEE Trans. Acoust. Speech Signal Processing* ASSP-27, 1979, 457-464.

15. K.A.Dines and A.C.Kak, Constrained least squares filtering, *IEEE Trans. Acoust. Speech Signal Processing,* ASSP-25, 1977,346-350.

16. B.R.Friden and J.J.Bruke, Restoring with maximum entropy II : Superresolution of photographs of diffraction-blurred images, *J. Opt. Soc. Amer.* 62, 1972, 1207- 1210.

17. B.R.Friden, Image enhancement and restoration ,in Picture Processing and Digital Filtering (T.S.Huang ed.), Springer-Verlag, Berlin and New York, 1975.

18. B.R.Hunt, The application of constrained least squares estimation to image restoration by digital computer, *IEEE Trans. Comput.* c-22, 1973, 805-812.

19. B.R.Hunt, Bayesian methods in nonlinear digital image restoration, *IEEE Trans. Comput.* C-26, 1977, 219-229.

20. A.K.Jain and E.Angel, Image restoration, modeling, and reduction of dimensionality,*IEEE Trans. Comput.* C-23, 1974, 470-476.

21. D. L. Barnea and H. F. Silverman, A class of algorithms for fast digital image registration, *IEEE Trans. Compt.*, C-21,1972, 179-186.

22. L. S. Davis and A. Rosenfeld, An application of relaxation labelling to spring-loaded template matching, *Proc. Int. Joint Conf. Patt. Recogn., 3rd*, 1976, 591-597.

23. R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, *Wiley, NY*, 1973.

24. R. N. Nagel and A. Rosenfeld, Ordered search techniques in template matching, *Proc. IEEE* 60, 1972,242-244.

25. K. Price and D. R. Reddy, Matching segments of images, *IEEE Tran. Pattern Anal. Machine Intelligence* 1, 1979,110-116.

26. Z. W. Ziang, W. Lu and D. Hu, A new digital image registration algorithm on the double spatial intensity gradients using pyramids, *patt. Recogn. Lett.* 9,335-340,1988.

27. G. Stockman and S. Kopstein, Image registration from edge content, *Proc. Symp. Automatic Imagery Patt. Recogn.* 8th, 139 - 157, 1978.

28. S. L. Horowitz and T. Pavlidis, Picture segmentation by directed split and merge procedure, *Proc. 2nd Int. Joint Conf. Patt. Recogn.*, 424-433, 1974.

29. S.K. Pal & N.R. Pal, 'Segmentation Based on Measures of Contrast, Homogeneity, and Region Size', *IEEE Trans. Syst. Man Cybern.*, Vol. SMC-17, No.5, pp.857-868, 1987.

30. N.R.Pal and S.K.Pal, Object-Background segmentation Using new definitions of entropy, *IEE Proc.*, Pt,E, July,284-295, 1989.

31. N.R.Pal and S.K.Pal, Entropic Thresholding, *Signal Processing*, 16, February, 97-108,1989.

32 N. R. Pal and S. K. Pal, Image model, Poisson distribution and object extraction, *Int. Jour. Patt. Recogn. and Artificial Intell.*(communicated).

33. J,N Kapur , P.K.Shaoo and A.K.C Wong, A new method for gray level picture thresholding using the entropy of histogram, *Compt. Graph., Vision and Image Processing*, vol-29, pp 273-285, 1985.

34. L. S. Davis, A. Rosenfeld, and J. S. Weszka, Region extraction by averaging and thresholding, *IEEE Trans. Systems Man Cybern*, 5, 1975,383-388.

35. R. Nevatia, Locating object boundaries in textured environments,*IEEE Trans. Systems Man Cybern.* 25,1976,1170-1175.

36. N. Otsu, A threshold selection method from gray level histograms, *IEEE Trans. Systems Man Cybern.* 9, 1979, 62-66.

37. N. Ahuja and A. Rosenfeld, A note on the use of second order gray level statistics for threshold selection, *IEEE Trans. Systems,Man and Cybern.* 8,1978,895-898.

38. L. S. Davis, A survey of edge detection techniques, *Compt. Graph. Image Processing* 4, 1975, 248-270.

39. J. M. Beanliew and M. Goldberg, Hierarchy in Picture Segmentation : A Stepwise Optimization approach , *IEEE Trans.Patt. Analysis Machine Intell.*, Vol.11, No.2, Feb. 89, 150-163.

40. P.A.Dondes and A Rosenfeld, Pixel classification on gray level and local "Busyness" , *IEEE Trans. on Patt. Analysis Machine Intell.*, Vol.4, No.1, Jan. 82, 79-84.

41. Angela Y.WU, T. Hong , and A. Rosenfeld, Threshold Selection using Quadtrees, *IEEE Patt. Analysis Machine Intell.*, Vol.4, No.1, Jan.82, 90-94.

42. A.K. Jain, S.P. Smith, L.E.Backer, Segmentation Of Muscle Cell Pictures : A Preliminary Study, *IEEE Patt. Analysis Machine Intell.*, Vol.2, No.3, May. 80,232-242.

43. J. Mantas, *Methodologies* in *Patt. Recogn. and Image Analysis* : A brief Survey, *Patt. Recogn.*, Vol.20, No.1, 1987, 1-6.

44. M.Pietikainen, A.Rosenfeld and I.Walter, Split and Link algorithm for Image Segmentation, *Patt. Recogn.*, Vol.15, No.4, 1982, 287-298.

45. M.Amadasun and R.A.King, Low level Segmentation of Multispectral Images via Agglomerative Clustering of Uniform Neighbours, *Patt. Recogn.*, Vol.21, No.3, 1988, 261-268.

46. B. Bhanu and B.A.Rarvin, Segmentation of Natural Scene, *Patt. Recogn.*, Vol.20, No.5, 487-496.

47 S. Basu and K. S. Fu, Image segmentation by syntactic method, *Patt. Recogn.*, Vol.20, No. 1, 1987, 33-44.

48. K.V. Mardia and T.J.Hainsworth, A Spatial Thresholding Method for Image Segmentation, *IEEE Patt. Analysis Machine Intell.*, Vol.10, No.6, nov.88, 919-927.

49. P.C.Besl and R.C.Jain, Segmentation using Variable Surface Fitting, *IEEE Patt. Analysis Machine Intell.*,10, No.2, 167-192,198

50. A. Rosenfeld and R.C.Smith, Thresholding using Relaxation, *IEEE Patt. Anal. Machine Intell.*, 3, No.5, Sept.81, 598-606.

51. R. Hoffman and A.K. Jain, Segmentation and Classification of Range Images, *IEEE Patt. Analysis Machine Intell.*, Vol.9, No.5, Sept.87, 608-620.

52. A.Perez, R.C.Gonzalez, An Iterative Thresholding Algorithm for Image Segmentation, *IEEE Patt. Analysis Machine Intell.*, Vol.9, No.6, Nov.87, 742-751.

53. A.B.Brink, Gray level Thresholding of Images using a Correlation criterion, *Patt. Recogn. Lett.*, Vol.9, No.5, June 89, 335-341.

54. T. Hong, K. A .Narayanan, S.pelag, A.Rosenfeld and T. Silberberg, Image Smoothing and Segmentation by Multiresolution Pixel linking : Further Experiments and Extensions, *IEEE Syst. Man Cybern.*, Vol.12,No.5, sept/oct.82, 611-622.

55. J.Kittler and J.Illingworth, Minimum Error Thresholding, *Patt. Recogn.*, Vol.19, No.1, 1986, 41-47.

56. M.Suk and S.M.Chung, A New Image Segmentation Technique based on Partition mode Test, *Patt. Recogn.*, 16, No.5, 1983,469-480.

57. H.S.Don and K.S.Fu, A Syntactic Method for Image Segmentation and Object recognition, *Patt. Recogn.*, Vol.18, No.1, 1985,73-87

58 T. Pun, Entropic Thresholding, a new approach, *Compt. Graph. Image Processing,* Vol. 16, 210-239, 1981.

59 T. Pun, A new method for gray level picture thresholding using the entropy of the histogram, *Signal processing,* Vol. 2, 223-237, 1980.

60 A.Rosenfeld, Survey,Image analysis and Computer Vision, *Comp. Vision Graph. Image Processing*, Vol.46, No.2, May.89, 196-250.

61 S. K. Pal and N. R. Pal, Segmentation using contrast-Homogeneity measure, *Patt. Recogn. lett.*, Aprl. 1987, 239-304.

62. N. R. Pal and S. K. Pal, Some information measures on fuzzy sets and their application to image processing, NACONECS-1989,94-96, 1989.

63 S. K. Pal and N. R. Pal, Object extraction using higher order entropy, *Proc. Int. conf. Patt. Recogn.*, Rome, Italy, Oct. 88, 348-350.

64 A. K. Jain, Advances in mathematical models for image processing, *Proc. IEEE,* Vol. 69, 502-528, May 1981.

65 F. R. Hansen and H. Elliott, Image segmentation using simple Markov random field models, *Compt. Graph. Image Processing,* Vol. 20, 101-132, 1982.

66 H. Elliott, H. Derin, R. Cristi and D. Geman, Application of the Gibbs distribution to image segmentation, *IEEE Int. Conf. Acoust. Speech, Signal Processing,* San Diego, CA, Mar. 1984.

67 H. Derin, H. Elliott, R. Cristi and D.Geman, Bayes Smoothing algorithms for segmentation of binary images modelled by Markov Random fields, *IEEE Trans. Patt. Anal. Machine Intell.*, PAMI 6, No. 6, 707-720, Nov. 1984.

68 S. K. Pal and N. R. Pal, Object background classification using a new definition of entropy, *IEEE 1988 Int. Conf. Syst., man Cybern.*, China, Vol. 2, 773-776, 1988.

69 N. R. Pal and S. K. pal, New entropic thresholding, *Proc. Seminar on Parallel Processing systems and their applications,* calcutta, 120-123, 1988.

70. J.S. Weszka, 'Survey of threshold selection techniques', *Compt. Graph. and Image Processing,* Vol.7, 259-265, 1978.

71. A. Rosenfeld and L.S. Davis, 'Iterative histogram modification, *IEEE Trans. Syst. Man Cybern.*, Vol.SMC-8,No.4, 300-302, 1978.

72. S. Peleg, 'Iterative histogram modification', *IEEE Trans. on Syst. Man Cybern.*, Vol.SMC-8, No.7, 555-556, 1978.

73. J.S. Weszka and A. Rosenfeld, 'Threshold evaluation techniques', *IEEE Trans. Syst. Man Cybern.*, Vol.SMC-8, 622-629, 1978.

74. F.Deravi and S.K.Pal, 'Gray level thresholding using second-order statistics', *Patt. Recogn. Lett.*, Vol.1, No.5, 417-422, 1983.

75. B. Chanda, B.B. Chaudhuri and D. Dutta Majumder, 'On image enhancement and threshold selection using the gray level co-occurrence matrix', *Patt. Recogn. Lett.*, Vol.3, No.4, 243-251, 1985.

76. Y.Nakagawa and A.Rosenfeld, Some experiments on variable Thresholding, *Patt.Recogn.*, 11, 191-204, 1979

77. S.Peleg, A New Probabilistic Relaxation Scheme, *IEEE Trans Patt. Analysis Machine Intell.*, 2, 1980, 362-369.

78. A.Rosenfeld and L.S.Davis, Image segmentation and image models, *Proc.IEEE* 67, 1979, 764-772.

79. A.Rosenfeld, R.A.Hummel and S.W.Zucker, Scene labeling by relaxation operatios, *IEEE Trans Syst. Man Cybern.*, 6, 1976, 420-433.

80. J.S.Weszka and A.Rosenfeld, Histogram modification for Threshold selection, *IEEE Trans Syst. Man Cybern.*, 9, 1979, 38-52.

81. S.D.Yanowitz and A.M.Bruckstein, A New Method for Image Segmentation, *Comp. Vision Graph. Image Processing*, 46, No.1, April.89, 82-95.

82. A.S.Abutaleb, Automatic Thresholding of gray level pictures using Two - dimensional entropy, *Compt. Graph. Image Processing*, 47, 22-32, 1989.

83. D.J.Bartliff, An Automatic Procedure for Image Segmentation, *Patt. Recogn. Lett.*, Vol.1, No.5-6, July.83, 435-442.

84 S. K. Pal and N. R. pal, Two stage segmentation algorithm incorporating psychovisual phenomena in contrast homogeneity measure, Plat. Jub. Conf. Syst. and Signal Processing, 366-369, Dec. 1986, Bangalore, India.

85 N.R. Pal and S. K. Pal Entropy: a new definition and its applications, *IEEE Trans. Syst., Man and Cybern.*(accepted for publication).

86. Y.T.Zhou, V.Venkateswar and R.Chellappa, Edge detection and linear Feature extraction using a 2-D Random field Model, *IEEE Trans. Patt. Analysis Machine Intell.*, Vol.11, No.1, Jan.89, 84-95.

87. J.Kittler, J.Eggleton, J.Illingworth and K.Paler, An Average Edge Detector, *Patt. Recogn .Lett.*, Vol.6, No.1,Jun.87, 27-32.

88. C. J. Jacobus and R. T. chein, Two New Edge Detectors, *IEEE Trans. Patt. Analysis Machine Intell.*, Vol.3, No.5, Sept.81, 581-592.

89. W.H.H.J.Lunscher and M.P.Beddoes, Optimal Edge Detector Design I : Parameter selection and noise effects, *IEEE Patt. Anal. Machine Intell.*, 8, No.2, Mar.86, 164-177.

90. W.H.H.J.Lunscher, M.P.Beddoes, Optimal Edge Dectector Design II : Coefficient Quantization, *IEEE Patt. Analysis Machine Intell.*, Vol.8, No.2, Mar.86, 178-187.

91. R.M.Haralick, Digital step edges from zero crossing of second directional derivatives, *IEEE Patt. Analysis Machine Intell.*, Vol.6, No.1, Jan.84, 58-68.

92. T.Peli and D.Malah, A Study of Edge Detection Algorithms, *Compt. Graph. and Image proce.*, Vol.20, No.1, Sept.82, 1-21.

93. A.Ikonomopoulos, An Approach to Edge Detection Based on the Direction of edge elements, *Comp. Graph. Image Proc.*, Vol.19, No.2, June.82, 179-195.

94. D.Marr and E.Hildreth, Theory of edge detection, *Proc. roy. soc. London,* Vol.B 207, 187-217, 1980.

95. M.K. Kundu and S.K. Pal, 'Thresholding for edge detection using human psychovisual phenomena', *Patt. Recog. Lett.*, Vol.4, No.6, pp.433-441, 1986.

96. R.M.Haralick, Edge and Region Analysis for digital image data, *Compt. Grapg. Image Proc.*, 12, 60-73, 1980.

97. Y.Lu and R.C.Jain, Behavior of edges in Scale Space, *IEEE Trans, Patt. Analysis Machine Intell.*, Vol.11, No.4, April.89, 337-356.

98. K.Paler and J.Kittler, Gray level Edge Thinning : A new method, *Patt. Recogn. Lett.* , Vol.1, No.5-6, July.83, 409-416.

99. J.F.Haddon, Generalised Threshold Selection for Edge Detection, *Patt. Recogn.*, Vol.21, No.3, 1988, 195-203.

100 S. K. Pal and R. A. King, Image enhancement using fuzzy sets, *Electronic Lett.*, Vol. 16,376-378,1980.

101. S.K. Pal, R.A. King and A.A. Hashim, 'Automatic grey level Thresholding through index of fuzziness and entropy', *Patt. Recog. Lett.*, Vol.1, pp.141-146, 1983.

102 S. K. pal and D. Dutta Majumder, Fuzzy methematical approach to pattern recognition, *Wiley Eastern Limited,* 1985.

103 R. Jain and H. H. Nagel, Analysing a real world scene sequence using fuzziness, *Proc. IEEE Conf. Decision and Control,* New Orleans, 1367-1372, 1977.

104 B. L. Deekshatulu, G. Kashipati Rao and R. Krishnan, A quantitative index of information and a method of enhancement for digital imaging using fuzzy sets, *Proc. IEEE Int. Conf. Syst. Man Cybern.*, New Delhi, Vol. 2, 1105-1109, 1983.

105 L. A. Zadeh, K. S. Fu, K. Tanaka and M Shimara (Eds), Fuzzy sets and their applications to cognitive process and decision process, *Academic press,* London, 1975.

106 S. K. Pal and A. Rosenfeld, Image enhancement and thresholding by optimization of fuzzy compactness. *Patt. Recogn. Lett.* 7, 77-86, 1988.

107 T.S.Huntsberger, C.L.Jacobs, R.L.Canon, Iterative fuzzy Image Segmentation, *Patt. Recogn.*, Vol.18, No.2, 1985, 131-138.

108 J. C. Bezdek, Fuzzy Mathematics in Pattern Recognition, Ph. D. Dissertation, *Cornell University,* NY, 1973.

109. C.E Shannon, A Mathematical Theory Of Communication, *Bell Syst. Tech. Journal*, Vol 27 pp 379-423, July, 1948.

110. C.E Shannon and W. Weaver, *The Mathematical Theory of Cpmmunication,* Urbana, The University of Illinois, 1949.

111 A. Renyi, On measures of entropy and information, *Proc. Fourth Berkeley Symposium on Mathematical Statistics and Probability,* Vol. 1, 541-561, University of California Press, Berkeley, 1961.

112 R. S. Verma, Generalization of Renyi's entropy of order α, *Jour. Math. Sci.,* Delhi, Vil. 1, 34-38, 1986.

113 J. N. Kapur, Generalised entropy of order α and type β, *Math. Seminar,* Delhi, Vol. 4, 78-94, 1967.

114 P. N. Rathie, On a generalized entropy and a coding theorem, *Jour. Appl. Probability,* Vol. 17, 124-133,1970.

115 S. Guiasu, Information Theory with applications, *Mc-Graw Hill International Book Company,* 1977.

116 A. Kandel, Fuzzy mathematical techniques with applications, *Addision-Wesley Publishing Company,* Massachusetts.

117 L . A. Zadeh, Probability measures of fuzzy sets, *Jour. Math. Analysis and Application,* 23,421-427,1968.

118 E. Backer, Cluster analysis by optimal decomposition of induced fuzzy sets, *Delft University Press,* 1978.

119 A . Kaufmann, Introduction to the theory of fuzzy subsets, Fundamental theoretical elements, Vol. 1, *Academic,* New York, 1975.

120 A Deluca and S. termini, A definition of nonprobabilistic entropy in the setting of fuzzy sets theory, *Information and Control*, 20,301-312,1972.

121 W. X. Xie and S. D. Bedrosian, An information measure for fuzzy sets, *IEEE Trans. Syst. Man and Cybern.*, 14, No. 1 Jan/Feb 1984,151-156

122 B. Kosko, Fuzzy entropy and conditioning, *Information Sciences*, 40, 165-174, 1986.

123 J. Knopfmacher, On measures of fuzziness, *Jour. Math. analysis and application*, 49, 529-534, 1975.

124 C. Dujet, Separation and measures of fuzziness, *Fuzzy Sets and Systems*, 28, 245-262, 1988.

125 W. Sander, On measures of fuzziness, *Fuzzy Sets and Systems*, 29, 49-55, 1989.

126 N.R. Pal and S. K. Pal, Higher order fuzzy entropy and hybrid entropy of a set, *Information Science* (to appear).

127 S. K. Pal and N. R. Pal, Higher order entropy, hybrid entropy and their applications, Proc. INDO-US workshop on spectrum analysis in one and two dimensions. Nov. 27-29, 1989, New Delhi, NBH OXFORD Publishing Company, New Delhi.

128. G. Buchsbaum, 'An analytical derivation of visual nonlinearity' *IEEE Trans. Biomedical Engineeging*, Vol. BME-27, pp.237-242, 1980.

129. P. Zuidema et.al. 'A mechanistic approach to threshold behaviour of visual systems', *IEEE Trans. Syst. Man Cybern.*, Vol.SCM-13, pp.923-934, 1983.

130. T.A. Hantea and V.R.Algazi, Perceptual Models and the filtering of High Contrast Achromatic Images, *IEEE Syst. Man Cybern.*, Vol.14, No.2, Mar/April.84.

131 R. Haralick, K. Shanmugam and I. Dinstein, Textural features for image classification, *IEEE Trans. Syst. Man Cybern.*, Vol. 3, No. 6, Nov. 73, 610-621.

132 D. K. Pickard, A Curious binary lattice process, *Jour. Appl. Prob.*, Vol. 14, 717-731, 1977.

133. J.C. Dainty and R. Shaw, *Image Science*, Academic Press, New York, 1974.

134. C.R Rao, *Linear Statistical Inference and its Applications*, Wiley Eastern Private Limited, New Delhi-1974.

135 A. D. Allen, Measuring the Empirical Properties of Sets, *IEEE Trans. Syst., man Cybern.*, SMC 4, No. 1, 66-73,1974.

136 S. K. Pal and B. Chakraborty, Fuzzy set theoretic measure for automatic feature evaluation, *IEEE Trans. Syst. Man Cybern.*, Vol -16, No. 5, Sept/Oct 1986, 174-760

137 H.G.Barrow, A.P.Ambler and R.M.Burstall, Some techniques for recognizing structure in pictures, *Frontiersof Pattern Recognition* (S.Watanabe, ed.), 1-29, Academic Press, New York, 1972.

138 H.G.Barrow and R.J.Popplestone, Relational descriptions in picture processing, in *Machine Intelligence* (B.Meltzer, D.Michie, eds.) Vol.6, 377-396, Edinburgh Univ. Press, Edinburgh, Scotland, 1971.

139 V.Claus, H.Ehrig and G.Rozenberg(eds), Graph -Grammars and their Application to Computer Science and Biolo Springer-Verlag, Berlin and New York, 1979.

140 K.S.Fu, Syntactic Methods in Pattern Recognition, Academic Press, New York, 1974.

141 K.S.Fu (ed.), Syntactic Pattern Recognition, Applications, Springer-Verlag, Berlin and New York, 1977.

142 R.C.Gonzalez, M.G.Thomason, Syntactic Pattern Recognition : An Introduction , Addision -Wesley,Reading, Massachusetts, 1978.

143 R.M.Haralick, Statistical and structural approaches to texture, *Proc. IEEE* 67, 1979, 786-804.

144 L.Kitchen and A.Rosenfeld, Discrete relaxation for matching relational structures, *IEEE Trans Syst. Man Cybern.* 9, 1979, 869-874.

145 A.Rosenfeld, Array and Web languages : an overview, in Automata, Languages, Development (A.Lindenmayer and G.Rozenberg,eds.) pp. 517-529, North-Holland Publ., Amsterdam, 1976.

146 A.Rosenfeld(ed.), Image Modeling, Academic Press, New York, 1981.

147 A.Rosenfeld and L.S.Davis, Image segmentation and image models, *proc. IEEE* 67, 1979, 764-772.

148 L.G.Shapiro, Data structures for picture processing : a survey, *Comput. Graph. Image Processing* 11, 1979, 162-184.

149 P.H.Winston, Learning structural descriptions from examples, in The Psycology of Computer Vision (P.H.Winston, ed.), pp 157-209, McGraw-Hill, New York, 1975.

LIST OF AUTHOR'S PUBLICATION ON WHICH THE THESIS IS BASED

1. S.K. Pal & N.R. Pal, 'Segmentation Based on Measures of Contrast, Homogeneity, and Region Size', *IEEE Trans. Syst. Man Cybern.*, Vol. SMC-17, No.5, pp.857-868, 1987.

2. N.R.Pal and S.K.Pal, Object-Background segmentation Using new definitions of entropy, *IEE Proc.*, Pt,E, July,284-295, 1989.

3. N.R.Pal and S.K.Pal, Entropic Thresholding, *Signal Processing*, 16, February, 97-108,1989.

4. N. R. Pal and S. K. Pal, Image model, Poisson distribution and object extraction, *Int. Jour. Patt. Recogn. and Artificial Intell.* (communicated).

5. S. K. Pal and N. R. Pal, Segmentation using contrast-Homogeneity measure, *Patt. Recogn. lett.*, Aprl. 1987, 239-304.

6. N. R. Pal and S. K. Pal, Some information measures on fuzzy sets and their application to image processing, NACONECS-1989,94-96, 1989.

7. S. K. Pal and N. R. Pal, Object extraction using higher order entropy, *Proc. Int. conf. Patt. Recogn.*, Rome, Italy, Oct. 88, 348-350.

8. S. K. Pal and N. R. Pal, Object background classification using a new definition of entropy, *IEEE 1988 Int. Conf. Syst., man Cybern.*, China, Vol. 2, 773-776, 1988.

9. N. R. Pal and S. K. pal, New entropic thresholding, *Proc. Seminar on Parallel Processing systems and their applications*, calcutta, 120-123, 1988.

10 S. K. Pal and N. R. pal, Two stage segmentation algorithm incorporating psychovisual phenomena in contrast homogeneity measure, Plat. Jub. Conf. Syst. and Signal Processing, 366-369, Dec. 1986, Bangalore, India.

11 N.R. Pal and S. K. Pal, Entropy: a new definition and its applications, *IEEE Trans. Syst., Man and Cybern.* (accepted for publication).

12 N.R. Pal and S. K. Pal, Higher order fuzzy entropy and hybrid entropy of a set, *Information Science* (to appear).

13 S. K. Pal and N. R. Pal, Higher order entropy, hybrid entropy and their applications, Proc. INDO-US workshop on spectrum analysis in one and two dimensions. Nov. 27-29, 1989, New Delhi, NBH OXFORD Publishing Company, New Delhi.