

## Content-based Image Retrieval by Information Theoretic Measure

M. Hanmandlu\* and Anirban Das#

\*Indian Institute of Technology Delhi, New Delhi-110016, India

#R Systems International Ltd., Gurgaon -122002, India

E-mail: mhmandlu@gmail.com

### ABSTRACT

Content-based image retrieval focuses on intuitive and efficient methods for retrieving images from databases based on the content of the images. A new entropy function that serves as a measure of information content in an image termed as ‘an information theoretic measure’ is devised in this paper. Among the various query paradigms, query by example (QBE) is adopted to set a query image for retrieval from a large image database. In this paper, colour and texture features are extracted using the new entropy function and the dominant colour is considered as a visual feature for a particular set of images. Thus colour and texture features constitute the two-dimensional feature vector for indexing the images. The low dimensionality of the feature vector speeds up the atomic query. Indices in a large database system help retrieve the images relevant to the query image without looking at every image in the database. The entropy values of colour and texture and the dominant colour are considered for measuring the similarity. The utility of the proposed image retrieval system based on the information theoretic measures is demonstrated on a benchmark dataset.

**Keywords:** Image retrieval, fuzzy features, descriptors, entropy, indexing

### 1. INTRODUCTION

The use of entropy in an image retrieval system is not as popular as compared to other methods which utilize wavelets or colour, texture, and shape descriptors. But the potentiality of entropy as a descriptor cannot be ignored and in the recent past, many researchers have started exploring the possibility of using entropy in different domains. Entropy has been found effectively useful in image indexing<sup>4,16,11,18</sup> and in similarity measures<sup>12,16,20</sup>.

The term entropy as a scientific concept was first used in thermodynamics by Clausius<sup>17</sup>. Its probabilistic interpretation in the context of statistical mechanics is attributed to Boltzmann<sup>19</sup>. Shannon<sup>1</sup> has used this concept to describe the properties of long sequences of symbols, and applied the results to a number of basic problems in coding theory and data transmission. Later the definition of entropy was extended to the field of information theory. The entropy of a system as defined by Shannon<sup>1</sup> gives a measure of uncertainty about the actual structure of the image. Shannon’s definition based on the information gain from an event is inversely proportional to its probability of occurrence.

The entropy of an image is used for different applications in image processing. The interpretation of entropy in an image depends on how an event is defined and also the definition of its posterior probability. In general, gray level is considered as an event and colour histogram as its probability density function. In addition to this, it is assumed that gray levels are statistically independent. Pun<sup>5</sup> and Kapur<sup>6</sup>, *et al.* have used the Shannon’s concept to define the entropy of an image assuming that an image is entirely represented by its gray level histogram

only. Leung<sup>7</sup>, *et al.* have attempted to isolate an object from the background by using the Gray-scale Image Entropy. The entropy (Shannon’s) of the histogram may be taken as a measure of information content in an image; such entropy is also called global information measure of the image. A low value of entropy indicates the skewness of the distribution of gray values, while a high value may be taken as an indicator of nearly uniform distribution of gray values. Definitely the histogram and the global entropy are not dependent on the spatial distribution of gray values in the image. The co-occurrence matrix<sup>2</sup> captures the spatial details of an image to some extent. The entropy of the co-occurrence matrix gives another measure of image information known as local entropy or second order entropy. Likewise the conditional entropy of a partitioned image can also be defined.

In texture analysis, an important approach to region description is to quantify its texture content. The randomness is the basic property of texture. By exploiting this fact, entropy as a statistical descriptor of texture could measure the variability in the image.

Pal and Pal<sup>8</sup> define a new entropy function based on the exponential behavior of information gain and applied co-occurrence based entropy methods for image thresholding. Zachary<sup>16</sup> attempts to use entropy as a visual feature of an image and showed how effectively entropy can be employed for indexing and also as a similarity measure of images in an image retrieval system. In IKONA<sup>11</sup>, a region-based image retrieval system, human faces are identified and preprocessed using the entropy map which assigns a saliency for each pixel in the face. This saliency is expressed as the entropy of a local

gray level distribution in a region around each pixel.

A new entropy function is presented in this paper. This entropy function is so designed that it can be efficiently used for image analysis. The properties and proofs of the proposed entropy function are presented. This new function is compared with the well known entropy functions. Two entropy based features with one feature representing the colour information and the other representing the texture information are derived from an image. Apart from this, another visual descriptor of the image called the dominant colour is described. These entropy based features are further used in the multi dimensional indexing technique. An interim result set is created using the indices of images for improving the performance of the retrieval system. The results of image retrieval will be compared with those of the classical model proposed by Swain and Ballard<sup>15</sup>.

**2. NEW ENTROPY FUNCTION**

Most of the entropy functions are not suitable for representing the information in a fuzzy set. These include Shannon<sup>1</sup>, Renyi<sup>9</sup> and Pal<sup>10</sup> *et al.* entropy functions. These entropy functions are generalised by introducing a polynomial in the exponential gain function. The proposed entropy function is shown to satisfy the basic properties of entropy and then Pal and Pal's entropy function is proved to be the special case of this function.

The function involved in the entropy need not be a membership function; it could be any feature. It may be noted that when we use a membership function, the unknown parameters of this membership function will parameterize the entropy function indirectly. However, the choice of a suitable membership function is not easy. Hence, the main motivation behind development of the entropy function is our concern to represent the information/uncertainty contained in a fuzzy set.

Here, authors are mainly concerned with a single fuzzy set. However in a fuzzy rule usually many fuzzy sets are encountered, but this case will be addressed in the future work. The definition of Pal and Pal<sup>10</sup> entropy function is now extended considering the exponential behavior of the gain function. This will pave the way to devise a new entropy function for representing the information in a fuzzy set.

**2.1 Definition and Properties**

The information gain corresponding to the occurrence of the *i*<sup>th</sup> event is defined as

$$I(p_i) = e^{-(ap_i^3 + bp_i^2 + cp_i + d)} \tag{1}$$

for all probabilities  $p_i \in [0,1]$ ,  $H_{Sn}^1(P_n) = -\sum_{i=1}^n p_i \log p_i$  and  $a, b, c$  and  $d$  are the real-valued parameters.

The entropy is expressed by

$$H = E(I(p_i)) = \sum_{i=1}^n p_i I(p_i) \tag{2}$$

Note that the above entropy function is an expectation of the information gain function. Some of the important properties associated with this function are stated now.

Property 1:  $I(p_i) = e^{-(ap_i^3 + bp_i^2 + cp_i + d)}$  is a continuous function

for all  $p_i \in [0,1]$

Property 2:  $I(p_i)$  is bounded

Property 3: With the increase in  $p_i$ ,  $I(p_i)$  decreases.

Property 4:  $H(P) = \sum p_i * I(p_i)$  is a continuous function for all  $p_i \in [0,1]$  real valued  $a, b, c$  and  $d$  parameters.

Property 5: If  $p_1 = p_2 = \dots = p_n = 1/n$  then  $H(P)$  is an increasing function of  $n$ .

Property 6: 
$$H(P) = \sum_{i=1}^n p_i * e^{-(ap_i^3 + bp_i^2 + cp_i + d)}$$

where,  $0 \leq p_i \leq 1$  and  $\sum p_i = 1$  is a concave function.

Property 7: Consider a partition  $A = [A_1, A_2, \dots, A_n]$  with the probabilities,  $p_i = P_r[A_i]$  and assume that  $p_1 < p_2$ . If  $p_1$  is increased and  $p_2$  is decreased by an equal amount, say  $\delta$  ( $\delta > 0$  and  $\delta \leq (p_2 - p_1) / 2$ ), then the entropy increases.

Property 8: Entropy is maximum when all  $p_i$ 's are equal. In other words,  $H(P) \leq H(1/n, 1/n, \dots, 1/n)$

Property 9: Entropy is minimum if and only if all  $p_i$ 's except one are zeros and that single  $p_i$  is equal to 1.

Property 10: Consider the partition of the event space as  $A = [A_1, A_2, \dots, A_n]$  and the probability  $p_i = P_r(A_i)$ . If a new partition  $B$  is formed by subdividing one of the sets of  $A$ , then  $H(B) \geq H(A)$ .

The proofs of some of the above properties are consigned to *Appendix 'A'*.

**2.1.1. Normalised Entropy**

The normalised entropy  $H_N$  can be defined as

$$H_N = (H - e^{-(a+b+c+d)}) / \lambda \tag{3}$$

where, The constant  $\lambda = e^{-\left(\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d\right)} - e^{-(a+b+c+d)}$ ;  $a, b, c$  and  $d$  are real valued parameters; and  $n$  is the number of events in the probabilistic experiment (or the number of states in the system).

The normalised entropy satisfies all the properties of an entropy function.

**2.1.2. Conditional Entropy**

Consider two partitions  $A = [A_1, A_2, \dots, A_n]$  and  $B = [B_1, B_2, \dots, B_m]$  and let us define that the product of two partitions,  $A = [A_i]$  and  $B = [B_j]$  is a partition whose elements are all intersections  $A_i B_j$  and the product of partitions is denoted by  $A.B = [A_i B_j]$ . Let  $p_{ij}$  be the probability of the event  $A_i B_j$ , i.e.,  $p_{ij} = Pr [A_i B_j]$  and the marginal probabilities  $p_i = P_r[A_i]$  is defined as,

$$p_i = \sum_{j=1}^m p_{ij}$$

Similarly the marginal probability,  $q_j = P_r [B_j]$  can be enumerated as,

$$q_j = \sum_{i=1}^n p_{ij}$$

The conditional entropy of  $A_i$  given that  $B_j$  has occurred is denoted by

$$\Pr[A_i | B_j] = \Pr[A_i B_j] / \Pr[B_j] = p_{ij} / q_j = p_{ij} \quad (4)$$

Similarly,

$$\Pr[B_j | A_i] = \Pr[A_i B_j] / \Pr[A_i] = p_{ij} / p_i = q_{ji} \quad (5)$$

Therefore the entropy of a partition  $A$ , given that  $B_j$  has occurred is given by  $H[A|B_j]$  as:

$$H[A | B_j] = \sum_{i=1}^n p_{ij} e^{-[ap_{ij}^3 + bp_{ij}^2 + cp_{ij} + d]}$$

Thus the conditional entropy of  $A$  assuming  $B$  is therefore:

$$H[A | B] = \sum_{j=1}^m q_j H[A | B_j] = \sum_{j=1}^m \sum_{i=1}^n q_j p_{ij} e^{-[ap_{ij}^3 + bp_{ij}^2 + cp_{ij} + d]} \quad (6)$$

$$H[B | A] = \sum_{i=1}^n p_i \sum_{j=1}^m q_{ji} e^{-[aq_{ji}^3 + bq_{ji}^2 + cq_{ji} + d]} \quad (7)$$

Now the entropy of the product of the partitions,  $A.B$  is easily found to be

$$H[A.B] = \sum_{i=1}^n \sum_{j=1}^m p_{ij} e^{-[ap_{ij}^3 + bp_{ij}^2 + cp_{ij} + d]} \quad (8)$$

### 2.2 Comparison with other Entropy Functions

In the last few decades treatment of uncertainty is one of the concerns in the research circles. Shannon's entropy<sup>1</sup> is the pioneering work on the information measure. This entropy function is defined in the domain of probability with  $n$ -states and in that the information gain is inversely related to its probability of occurrence.

$$H_{Sn}^1(P_n) = - \sum_{i=1}^n p_i \log p_i \quad (9)$$

Renyi<sup>9</sup> has extended the definition of Shannon's entropy to an incomplete probability distribution. The Renyi's entropy of order  $\alpha$  is of the form

$$H_{Rn}^1(P_n, \alpha) = \frac{1}{1-\alpha} \log \left[ \frac{\sum_{i=1}^n p_i^\alpha}{\sum_{i=1}^n p_i} \right], \alpha > 0, \alpha \neq 1 \quad (10)$$

It is noted that as  $\alpha$  tends to 1, Renyi's entropy matches with Shannon's entropy. In contrast to Shannon's entropy, Pal and Pal's entropy function considers the information gain as an exponential function:  $e^{(1-p_i)}$

$$H_{PPn}^1(P_n) = \sum_{i=1}^n p_i e^{(1-p_i)} \quad (11)$$

The normalised Pal and Pal's Entropy is given by

$$H_{Nor}^1(P_n) = \frac{H_{PPn}^1(P_n) - 1}{(e^{\frac{1}{n}} - 1)} \quad (12)$$

Following the above entropy function, an exponential entropy function is devised with an eye on incorporating information into it. A significant feature of this entropy function is that it captures the inherent uncertainty in the fuzzy sets.

The uncertainty in a fuzzy set is non-statistical in nature and it plays an important role in fuzzy image processing. Moreover, the use of a polynomial in the exponential function yields the gain function bestowed with four tunable parameters. The parameters of the new entropy function provide controls on the information gain and proper tuning of these parameters by way of optimisation leads to the correct measure of uncertainty.

Recall the new entropy function defined above as

$$H = \sum_{i=1}^n p_i e^{-(ap_i^3 + bp_i^2 + cp_i + d)} \quad (13)$$

and its corresponding normalised entropy is as follows

$$H_N = (H - e^{-(a+b+c+d)}) / (e^{-\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d} - e^{-(a+b+c+d)}) \quad (14)$$

This entropy function behaves like Pal and Pal's entropy function for  $a = 0, b = 0, c = 1$  and  $d = -1$ . The curves in **Figs 1** and **2** show Shannon's entropy, Normalised Pal and Pal's entropy and Normalised new entropy for a two-state system for the two sets of parameter values. It may be observed that though the new entropy function behaves like the Shannon's and Pal and Pal's entropy functions yet it expands and contracts for varying values of the parameters.

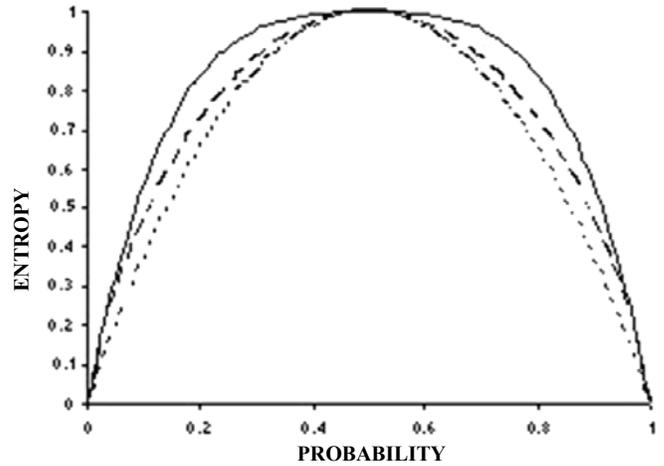


Figure 1. Curves for Shannon's, normalised Pal and Pal's & normalised fuzzy entropy ( $a = 0, b = 1, c = 3.2$ ).

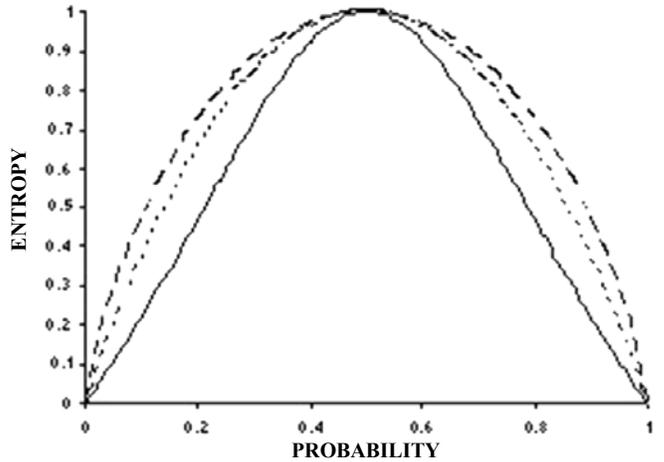


Figure 2. Curves for Shannon's, normalised Pal and Pal's & normalised fuzzy entropy ( $a = 5, b = 0, c = 0$ ).

### 2.3 Possible Applications

- **Biometrics:** In finger print identification, we can model the minutiae points by finding their distances with respect to a reference point which may be selected as the one with the highest curvature. In the palm print, we can model the points on the primary lines (there are mainly three such lines) and use this information for authentication.
- **Medical Image processing:** Breast cancer and brain tumor detection can be easily attempted by the entropy based modeling. In both these problems, issues in the affected organs undergo textural changes on the infringement of malignancy. The tumour regions can be extracted by applying the entropy function in a shifting window. The tissues in the affected portions naturally show entropy values different from those of the benign tissues for a chosen set of coefficients of the entropy function.
- **Cryptography:** The power of the proposed entropy can be best utilized in the encryption of secrete keys for secure transmission. In the above we have cited a few problems but one can harvest many fields should the ingenuity and imagination permit. However, in this paper its application to image retrieval will be discussed by way of indexing.

### 3. FEATURE EXTRACTION AND INDEXING

Many visual features have been explored in the literature on content-based image retrieval (CBIR) for the purpose of depicting colour, texture, shape and other properties of an image. Combining the various features usually achieves better performance in retrieval. Although a large number of features could represent the image very accurately, the inherent problem with this approach is the shortage of storage capacity for large image database. Hence an efficient multidimensional indexing technique is required for dimension reduction.

The objective of this section is to reduce the dimensionality of the feature space and to improve the performance of the retrieval process. For this, the entropy-based features are extracted from the images. The entropy of an image, by definition, is the measure of information content in the image<sup>8</sup>. As will be seen, the entropy function maps an n-dimensional vector to a single real number (i.e. one dimensional space) and so it can be regarded as a dimension reduction operator.

The extracted features from an image are stored to serve as an index of that image. Since the expensive management of storage and comparison time is less significant than the retrieval accuracy, only three concise and precise features are used to describe the contents of the image. These are colour entropy, texture entropy and dominant colour of the image.

#### 3.1 Entropy-based Image Features

The entropy of a system as defined by Shannon<sup>1</sup> gives a measure of uncertainty about its actual structure. Shannon's definition based on the information gain from an event is inversely proportional to its probability of occurrence. Pun<sup>5</sup> and Kapur<sup>6</sup>, *et al.* have used Shannon's concept to define entropy of an image assuming that the image is entirely represented by its

gray level histogram only. Unlike the logarithmic behavior of Shannon's entropy, the gain function in our entropy definition is of exponential nature as discussed in Section II. Two entropy based image features – colour entropy and texture entropy are now presented.

##### 3.1.1 Colour Entropy

The RGB colour space is chosen to represent the image. Let  $F = \{f(x, y)\} M \times N$  be an image of size  $M \times N$  where  $f(x, y)$  is the colour vector  $(r, g, b)$  in the RGB space at  $(x, y)$  point and  $N(r, g, b)$  is the frequency of the colour vector  $(r, g, b)$ .

$$\sum_r \sum_g \sum_b N_{(r,g,b)} = M \times N$$

Then,

In a natural image, it has been observed that out of 2563 different colour levels, a small fraction of different colour levels are actually used. So instead of considering all colour levels, the colour levels of an image are quantized adaptively. To achieve this, the available colour levels are clustered into colour bins.

The number of bins is not preset; it depends on the distribution of colour in that particular image. The above algorithm adaptively clusters all the colour levels into bins. Pun<sup>5</sup> and Kapur<sup>6</sup>, *et al.* use the gray-level histogram to represent the image; here the histogram is extended to the colour bins. It is generally assumed that the distribution of colours across an image follows the uniform distribution, i.e. each colour has a  $1/(M \times N)$  probability where the image size is  $M \times N$ . The algorithm for forming bins is given Algorithm 1.

Let,  $Z = \{z_1, z_2, \dots, z_n\}$  be the set of events that corresponds to the colour bins. Let us consider the probability of the colour bin  $z_i$  as  $p(z_i) = (\text{Number of pixels in the colour bin } Z_i) / M \times N$ . In this context, the colour entropy is taken in the normalised form,  $H_{N, \text{Colour}}$  given by

$$H_{N, \text{Colour}} = (H_{\text{Colour}} - e^{-(a+b+c+d)}) / \lambda \tag{15}$$

$$H_{\text{Colour}} = \sum_{i=1}^n p(z_i) e^{-\{a \cdot p^3(z_i) + b \cdot p^2(z_i) + c \cdot p(z_i) + d\}}$$

with

where,  $\lambda = [e^{-\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d} - e^{-(a+b+c+d)}]$  and  $n$  is the number of colour bins.

##### 3.1.2 Entropy Optimisation

The entropy optimisation is resorted to estimate the four tunable parameters:  $a, b, c$  and  $d$ . The derivatives of  $H_N$  with respect to  $a, b, c$  and  $d$  are obtained from:

$$\frac{\partial H_N}{\partial a} = \Lambda_1 + \frac{[-\sum_{i=1}^n p^4(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}}] + H_1}{H_2 - H_1}$$

$$\frac{\partial H_N}{\partial b} = \Lambda_2 + \frac{[-\sum_{i=1}^n p^3(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}}] + H_1}{H_2 - H_1}$$

$$\frac{\partial H_N}{\partial c} = \Lambda_3 + \frac{[-\sum_{i=1}^n p^2(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}}] + H_1}{H_2 - H_1}$$

**Algorithm 1. Algorithm for bin formation**

- 
1. for (All available colour levels)
    - 2. {
      - 3. for (All available Bins)
        - 4. {
          - 5. if ( any two colours of Colour\_level (r, g, b) matches with the respective two colours of Bin\_Center (r, g, b) and the third colour of Colour\_level lies in the predefined interval of the third colour of Bin\_Center )
            - 6. {
              - 7. Include the Colour\_level into the respective Bin
            - 8. }
              - 9. if (Colour\_level is not included in any Bin)
                - 10. {
                  - 11. Create a new Bin and include the Colour\_level into that Bin;
                  - 12. Set Colour\_level as Bin\_Center;

---

$$\frac{\partial H_N}{\partial d} = \Lambda_4 + \frac{[-\sum_{i=1}^n p(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}} + H_1]}{H_2 - H_1} \quad (16)$$

where,

$$\Lambda_1 = \frac{[\sum_{i=1}^n p(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}} - H_1]}{(H_2 - H_1)^2} \left( \frac{H_2}{n^3} - H_1 \right)$$

$$\Lambda_2 = \frac{[\sum_{i=1}^n p(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}} - H_1]}{(H_2 - H_1)^2} \left( \frac{H_2}{n^2} - H_1 \right)$$

$$\Lambda_3 = \frac{[\sum_{i=1}^n p(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}} - H_1]}{(H_2 - H_1)^2} \left( \frac{H_2}{n} - H_1 \right)$$

$$\Lambda_4 = \frac{[\sum_{i=1}^n p(z_i) * e^{-\{ap^3(z_i)+bp^2(z_i)+cp(z_i)+d\}} - H_1]}{(H_2 - H_1)^2} \left( \frac{H_2}{1} - H_1 \right)$$

$$\text{and } H_1 = e^{-(a+b+c+d)}, \quad H_2 = e^{-\left(\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d\right)}$$

Then the parameters are updated using gradient descent learning as

$$\begin{aligned} a^{New} &= a^{Old} - \varepsilon_1(i) \frac{\partial H_N}{\partial a} \\ b^{New} &= b^{Old} - \varepsilon_2(i) \frac{\partial H_N}{\partial b} \\ c^{New} &= c^{Old} - \varepsilon_3(i) \frac{\partial H_N}{\partial c} \\ d^{New} &= d^{Old} - \varepsilon_4(i) \frac{\partial H_N}{\partial d} \end{aligned} \quad (17)$$

where the learning rates  $\varepsilon(i)$ . As this learning has convergence problems we propose reinforcement learning.

**3.1.3 Reinforcement Learning**

The reinforcement learning requires some re-use policy of how to use the past information.

*Exploitation:* In this we exploit the Reuse policy which requires integrating knowledge of the past policy into the current learning process. Here we need to bias the exploratory process of the new policy with the past one. We have used here the sigmoid function for  $\varepsilon$  in which cumulative of the past errors is biased by the term  $k_2$  and the slope or gain of the function is changed by the term  $k_1$ .

*Exploration:* In the earlier work  $k_1$  and  $k_2$  were incremented by constant, but they are updated here by random numbers thus boosting exploration. The proposed<sup>11</sup> reinforcement learning gets stuck up in the local minima. This has been modified to incorporate evolutionary feature<sup>27</sup>. In this each parameter is updated by a number of updating laws called a population each having a pair of random numbers for  $k_1$  and  $k_2$ . The law yielding the minimum is considered as the global solution. Because of this it overcomes the drawback of local minima by taking into account both exploitation and exploration strategies.

If  $n$  is the population size, then we make use of the Reinforced learning law given by Hanmandlu and Murthy<sup>27</sup>

$$\varepsilon_i(i) = e^{-(k_{1i}(i) \sum err(j,i) + k_{2i}(i))} \quad (18)$$

where,  $err(j,i) = H_N(j,i) - H_N^{new}(j-1,i) = \Delta H_N(j,i)$ ;  $l=1,1,\dots,4$ ,  $i=1$  to  $n$  and  $j$  is the current iteration.

The initial values of  $k_{1i}(i)$  and  $k_{2i}(i)$  are taken both as random numbers. We use the following policy to adjust the values of  $k_{1i}(i)$  and  $k_{2i}(i)$ . This consists of the following steps:

If  $\sum err(j,i)$  is increasing, then  $\varepsilon(i)$  must decrease. So  $k_{1i}(i)$  should increase.

(a) If  $\sum err(j,i)$  is decreasing, then  $\varepsilon(i)$  need not change. So  $k_{2i}(i)$  should increase.

(b) If  $\sum err(j,i)$  is constant, then  $\varepsilon(i)$  should not change. So  $k_{1i}(i)$  and  $k_{2i}(i)$  are not changed.

If  $H_N(j,i)$  is not differentiable then parameter of interest, say  $a$ , is updated as follows:

$$\begin{aligned}
 a(i+1) &= a(i) - \varepsilon(i) \frac{H_N(a_{old}(i) + p) - H_N(a_{old}(i))}{p} \\
 &= x(i) + \varepsilon(i) \frac{\Delta f}{p}
 \end{aligned}
 \tag{19}$$

where,  $p = 0.0001$ .

The evolution is bestowed by the following procedure: We create a population of solution processes (particles) for each parameter. Let this population be denoted as  $n$  (taken to be 10). The function values of all particles  $H_N(j, i)$  are arranged in the ascending order. After that half of the populations with high values of function are eliminated as they are unfit to propagate to new generation. Thereafter reproduction is carried out by the rest half of the population. This procedure continued until we reach the global minimum of the function.

### 3.1.4 Colour Entropy Results

In this sub-section, colour entropy values are computed for different natural images. The parameters  $a, b, c, d$  are initialized to one and then the entropy optimisation technique is applied to find the normalised entropy value termed as colour entropy. Table 1 shows the colour entropy values and the corresponding optimised parameter values.

## 3.2 Texture Entropy

The ambiguities in texture arising due to the fuzzy nature of image function allow us to devise fuzzy texture features. So rather than using the colour histogram to represent an image, the fuzzy features that capture the fuzziness in the texture property are extracted about the neighborhood of each pixel. Since texture is region based, arrangement of image functions (i.e., intensities) of pixels in a local region, say, a window is made use of in order to characterize the texture using the Gaussian type membership function. The cumulative response about the central pixel from this window replaces the pixel intensity giving rise to the texture image when taken over the entire image.

### 3.2.1 Extraction of Fuzzy Features

To convert the spatial domain image into the fuzzy domain, the spatial arrangement of gray levels of pixels over a window is utilised. The fuzzy property can be expressed in terms of a membership function. A membership function to this effect is represented by the Gaussian type function.

$$\mu_{(k,l)}(i, j) = \exp\left[-\left\{\frac{x(i, j) - x(k, l)}{\tau}\right\}^2\right]
 \tag{20}$$

where,  $x(i, j)$  is the gray level of the pixel at the  $(i, j)$ <sup>th</sup> position and  $\tau$  is the fuzzifier which is specified to be the window size ( $\tau$  is taken as 5)

We note that

$$\mu_{(i,j)}(k, l) = 1 \quad \text{if } x(i, j) = x(k, l)
 \tag{21}$$

To consider the response from the neighboring pixels, the cumulative response of  $(i, j)$ <sup>th</sup> pixel is obtained as

$$y(i, j) = \sum_{k,l} \mu_{(k,l)}(i, j) * x(k, l) / \sum_{k,l} \mu_{(k,l)}(i, j)
 \tag{22}$$

This is the defuzzified response of the  $(i, j)$ <sup>th</sup> pixel over the window of size 5. This process is repeated for all pixels in the image resulting in a texture feature image consisting of all the defuzzified values. For convenience of notation, the matrix formed by  $y(i, j)$  is designated as the response matrix,  $Y$  which in turn would represent the texture feature image.

Let  $Z = \{z_1, z_2, \dots, z_n\}$  be the set of distinct responses of  $Y$ . Let us consider the probability of the texture response  $z_i$  as  $p(z_i) = (\text{Number of pixels having the texture response } z_i) / M \times N$ . On the heels of the normalised colour entropy, here comes the normalised texture entropy  $H_{N, \text{Texture}}$  as

$$\begin{aligned}
 H_{N, \text{Texture}} &= (H_{\text{Texture}} - e^{-(a+b+c+d)}) / \lambda \\
 \text{With } H_{\text{Texture}} &= \sum_{i=1}^n p(z_i) * e^{-\{a * p^3(z_i) + b * p^2(z_i) + c * p(z_i) + d\}}
 \end{aligned}
 \tag{23}$$

where,  $\lambda = [e^{-\left(\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d\right)} - e^{-(a+b+c+d)}]$  and  $n$  is the number of distinct responses.

Table 1. Colour entropy

| Image   | Colour entropy | Optimised parameter values |          |          |          |
|---|----------------|----------------------------|----------|----------|----------|
|   |                | <i>a</i>                   | <i>b</i> | <i>c</i> | <i>d</i> |
| <br>828  | 0.997806       | 0.999999                   | 1        | 1.00002  | 1        |
| <br>1739 | 0.99871        | 0.999999                   | 0.999999 | 1.00001  | 1        |
| <br>8658 | 0.999563       | 1                          | 1        | 1        | 1        |

In the entropy function (22, 23),  $a, b, c$  and  $d$  are four tunable parameters with  $0 \leq H_{N,Texture} \leq 1$ . Here the texture entropy is optimised to estimate these parameters using the optimisation technique described earlier.

In order to compute the texture entropy values for different natural images, the values of the parameters -  $a, b, c, d$  are initially set at 1 and then the normalised entropy value is calculated using the entropy optimisation technique. The texture entropy values and the corresponding optimised parameter values are shown in Table 2.

3.2.2 Visual Image Feature

It may be noted that the images might have the similar statistical features yet they are visually different. The choice of an appropriate visual feature plays an important role in classifying the visually similar images. There are several visual features – colour, texture, shape, contrast, coarseness, normalised area of an object etc. proposed by earlier researchers<sup>13,14</sup>. In the present work, we make use of another visual feature – dominant colour of an image.

Dominant colour of an image is the particular colour level that has the highest frequency in the image. It is a perceptual property of an image. As the CIE Lab colour space is perceptually uniform, the dominant colour is transformed to the lab colour space consisting of luminance component  $L$  and chrominance components  $C = \{a, b\}$ . Table 3 shows the RGB values and the CIE Lab values of the dominant colour and the percentage area covered by this colour.

Table 2. Texture entropy

| Image   | Texture entropy | optimised parameter values |          |         |     |
|---|-----------------|----------------------------|----------|---------|-----|
|   |                 | $a$                        | $b$      | $c$     | $d$ |
| <br>1574 | 0.997884        | 0.999999                   | 0.999999 | 1.00002 | 1   |
| <br>4585 | 0.997416        | 0.999999                   | 0.999999 | 1.00002 | 1   |
| <br>5237 | 0.9352          | 0.999977                   | 1.00004  | 1.00048 | 1   |

Table 3. Dominant colour

| Image           | <br>739 | <br>3078 |
|-----------------|--|---|
| RGB             | 229, 107, 0  | 85, 100, 129  |
| CIE lab         | 76.244, 20.2398, 125.801   | 68.4903, 3.45487, -5.34797  |
| Dominant colour |         |          |

The images of sea and sunset category, shown in Table 3, have distinct dominant colour, which is necessarily an important visual feature. The images of the same category will have nearly the same dominant colour, therefore in the retrieval system, CIE Lab values of the dominant colour could be used to classify similar images.

It is possible that the images of different categories have the same dominant colour, so the dominant colour is appropriate to be one of the visual features for classification.

3.2.3 Indexing through Feature Vector

As the image collection is getting larger and larger, the retrieval speed is becoming a bottleneck. Hence effective high dimensional indexing techniques need to be explored. The high dimensionality of feature vectors is normally of the order of  $10^2$ . Applying the dimension reduction on the feature vectors, an embedded dimension much lower than the initial dimension though still higher for linear ordering is obtained. Therefore a multi-dimensional indexing scheme is required to index the reduced embedded dimensional feature vectors. From this discussion it follows that the multi-dimensional indexing owes its allegiance to the high dimensionality of the feature space.

As there are only three features, viz., colour entropy, texture entropy and dominant colour used in this work, the dimension of feature space is too small to apply any dimension reduction algorithm. An entropy-based two dimensional indexing is employed here using both colour and texture entropy values. In our method, the dimension of the feature vectors is not dynamic; i.e., it remains fixed as there is no need for any sophisticated tree-data structure for multidimensional indexing.

A feature vector of an image indexed by a pair of entropy values of colour and texture becomes a point in the two-dimensional feature plane whose axes are colour entropy and texture entropy.

The concept of interim result is invoked here. The interim result set is a subset of image search space containing images semantically similar to the query image. The above indexing scheme would improve the retrieval performance by creating an interim result set. For a particular query image, the colour entropy and texture entropy, say ( $qColourEnt, qTextureEnt$ ) can be easily calculated. Now here is the algorithm for creating an interim result set (Algorithm 2).

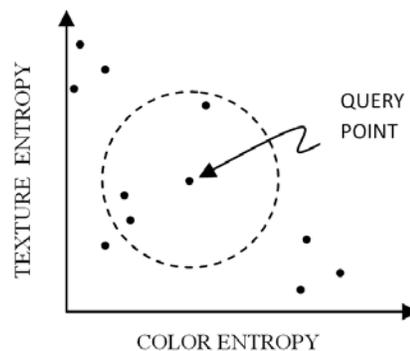


Figure 3. Image feature plane.

**Algorithm 2**


---

 Algorithm for interim result set

1. for( Each feature point in Feature Database)
  2. { Let cEnt = Colour Entropy of a feature point and
  3.     tEnt = Texture Entropy of a feature point
  4.     if( Euclidean\_distance( (qColourEnt, qTextureEnt), (cEnt, tEnt)) <  $\epsilon$ )
  5.     {
  6.         Include Image (cEnt, tEnt) into the interim result set
  7.     } // End of if ( )
  8. } // End of for ( )
- 

In the above procedure, the choice of  $\epsilon$  is very crucial, because the size of the interim result set is directly proportional to the value of  $\epsilon$ . It is important to note that the value of  $\epsilon$  should depend on the size of the feature space and also on the entropy values of colour and texture of the query image.

Let  $\xi_{Color}^2$  and  $\xi_{Texture}^2$  represent the second order moments of entropy values of colour entropy and texture of the query image respectively. The expressions for the statistical descriptors are as follows:

$$\xi_{Color} = \sqrt{\frac{1}{n} \sum_{i=1}^n (cEnt_i - qColorEnt)^2} \quad (24)$$

$$\xi_{Texture} = \sqrt{\frac{1}{n} \sum_{i=1}^n (tEnt_i - qTextureEnt)^2} \quad (25)$$

where,  $n$  is the size of feature space,  $cEnt_i$  and  $tEnt_i$  denote the colour entropy and texture entropy of  $i^{\text{th}}$  image respectively.

An expression for  $\epsilon$  is coined by

$$\epsilon = k * \text{Max}(\xi_{Color}, \xi_{Texture}) \quad (26)$$

where,  $k$  is any positive real number.

The interim result set contains all the feature points having a distance of less than or equal to  $\epsilon$  from the query image. One of the purposes of creating the interim result set is to reduce the search space without sacrificing the correctness of results of the retrieval process.

### 3.2.4 Experimental Results

A set of 9907 images from low resolution web-crawled miscellaneous database<sup>26</sup> used in WBIS<sup>25</sup> is our benchmark dataset for generating the query results. Next the interim result set for different query images is created. The recall for the interim result set denotes the percentage of correct retrieved images out of all correct images in the image database. The ratio of the size of the interim result set to the original size of the image database gives the reduction factor. Therefore a large recall value with high reduction factor indicates the efficiency of the algorithm.

It may be observed from Table 4 that the reduction factor increases by decreasing the value of  $\epsilon$ . For lower value of  $\epsilon$  naturally recall also decreases. The higher recall value indicates the correctness of the system; hence the reduction factor is increased so as to keep the recall value higher. Here are the results for other query images on the same image database.

## 4. IMAGE RETRIEVAL PROCESS

Any search engine, text-based or otherwise, is plagued by the problem of un-related matches. Often in the case of text based search engines, this hitch arises from the use of ambiguous keywords, such as bank, interest. Content-based image retrieval system allows a user to set query by an example image<sup>21</sup>, which ideally removes the ambiguity in setting up of queries. But the success of the image retrieval depends on the extraction of image features and measurement of similarity of query image with the images in a large data collection. In continuation of the feature extraction and the effective indexing methodology in Section 3, the distance measures between any two images in the interim result set are the necessary ingredients for comparison. Furthermore, the performance of the retrieval process will be compared with the traditional colour histogram matching<sup>15</sup>.

### 4.1 Image Distance Measure

To represent an image the required are the entropy based feature descriptors: colour entropy and texture entropy and a visual colour feature – dominant colour. These feature descriptors allow us to define a distance metric that closely matches the human perception. The idea is that the similarity between the two images should be measured in terms of not only the closeness of colour and texture distributions but also the closeness of the dominant colour of the image.

Consider two images, a query image  $I_q$  and a target image  $I_t$  to measure the distance. The dominant colours of both the images are denoted by  $CQ(L_q, a_q, b_q)$  and  $CT(L_t, a_t, b_t)$  and the other two entropy based features – colour entropy and texture entropy, for query image and target image by  $ColourEnt_q$ ,  $ColourEnt_t$  and  $TexEnt_q$ ,  $TexEnt_t$  respectively. Then the distance between the images is computed as a linear combination of two  $L_2$  distances. The first  $L_2$  distance is measured between the two dominant colours; the lower value of this distance indicates the visual closeness in terms of colour ignoring the information in the image, which is given by the entropy. So, the distance between the dominant colours is:

$$D_1 = \sqrt{(L_q - L_t)^2 + (a_q - a_t)^2 + (b_q - b_t)^2} \quad (27)$$

The second  $L_2$  distance is measured between the entropy based feature vectors (colour entropy, texture entropy). The entropy is a real number which indicates the amount of information contained in an image in terms of colour distribution and textural pattern. The lower value of second distance indicates the similarity of images in terms of content.

**Table 4. Interim result set for car query on the database of size 9907**

| Query image   | Interim result set size | k    | $\epsilon$ | Recall (%) | Reduction (%) |
|---|-------------------------|------|------------|------------|---------------|
| <br>1649 | 8703                    | 1.5  | 0.170843   | 100        | 12.15         |
|   | 8405                    | 1    | 0.113895   | 100        | 15.16         |
|   | 8183                    | 0.75 | 0.085421   | 100        | 17.40         |
|   | 7833                    | 0.5  | 0.056947   | 98.98      | 20.93         |

**Table 5. Interim result set for rose query on the database of size 9907**

| Query image  | Interim result set size | k    | $\epsilon$ | Recall (%) | Reduction (%) |
|--|-------------------------|------|------------|------------|---------------|
| <br>817 | 8700                    | 1.5  | 0.172226   | 99.03      | 12.18         |
|  | 8398                    | 1    | 0.114817   | 99.03      | 15.23         |
|  | 8168                    | 0.75 | 0.086113   | 99.03      | 17.55         |
|  | 7820                    | 0.5  | 0.057408   | 97.09      | 21.07         |

**Table 6. Interim result set for seaside query on the database of size 9907**

| Query image  | Interim result set size | k    | $\epsilon$ | Recall (%) | Reduction (%) |
|--|-------------------------|------|------------|------------|---------------|
| <br>1792 | 8696                    | 1.5  | 0.173055   | 100        | 12.22         |
|  | 8387                    | 1    | 0.11537    | 99.64      | 15.34         |
|  | 8154                    | 0.75 | 0.086527   | 99.64      | 17.69         |
|  | 7801                    | 0.5  | 0.057685   | 98.94      | 21.26         |

**Table 7. Interim result set for pattern on the database of size 9907**

| Query image   | Interim result set size | k    | $\epsilon$ | Recall (%) | Reduction (%) |
|---|-------------------------|------|------------|------------|---------------|
| <br>2766 | 8688                    | 1.5  | 0.173493   | 97.53      | 12.30         |
|   | 8381                    | 1    | 0.115662   | 95.62      | 15.40         |
|   | 8147                    | 0.75 | 0.086746   | 94.16      | 17.76         |
|   | 7780                    | 0.5  | 0.057830   | 91.58      | 21.46         |

$$D_2 = \sqrt{(ColorEnt_T - ColorEnt_Q)^2 + (TexEnt_T - TexEnt_Q)^2} \quad (28)$$

The distance between two images  $I_Q$  and  $I_T$  is defined as

$$Dist(I_Q, I_T) = \alpha * D_1 + (1 - \alpha) * 100 * D_2 \quad (29)$$

where,  $0 \leq \alpha \leq 1$

The value of  $D_2$  lies between 0 to 1 and it is also very low in comparison to  $D_1$ . To scale both the distances at the same level, a weight of 100 is assigned to  $D_2$ . The value of  $\alpha$  in the distance measure formula can be interpreted as the relative importance between the dominant colour and the information content. The range of  $\alpha$  varies from 0 to 1. The higher value of  $\alpha$  assigns

more importance to the dominant colour than the information content in the retrieval of similar images. In sunset image, dominant colour has more importance, so the user could set the value of  $\alpha$  more than 0.5 (which stands for equal importance) in forming the query for the image retrieval. It is important to note that the background colour or dominant colour might be irrelevant in many images but this would mislead the retrieval process. The value of  $\alpha$  should be set to 0 for the queries where the dominant colour does not have any importance.

Table 8 shows that the distances between the similar images are comparatively less than the distances between the different categorical images. In this table, the distances between the query image and sample images are measured for  $\alpha = 0.5$ .

## 4.2 Image Retrieval by Querying

A search engine whether text-based or otherwise is prone to unwanted matches. There are several reasons to get incorrect matches; one of them is due to the use of ambiguous keywords like interest, bank, etc., and another due to the use of inappropriate words to describe the desired images. In the proposed content-based image retrieval (CBIR) system, the user can set the query by an example from the query image set. In our system, the natural and real image databases are categorized into different classes of images, namely, animal, aircraft, architecture, landscape, sea shores, vehicles, flowers, human activities, etc. A small set of query images is presented in Table 9.

## 4.3 Colour Histogram Matching

The axes used for the histograms are the three opponent colour axes, assigned as follows:

$$r_g = r - g \quad (30)$$

$$b_y = 2 \times b - r - g \quad (31)$$

$$w_b = r + g + b \quad (32)$$

Here  $r$ ,  $g$ , and  $b$  represent red, green, and blue signals, respectively. The  $r_g$ ,  $b_y$ , and  $w_b$  axes are analogous to the opponent colour axes used by the human visual system<sup>22</sup>. They are used here simply to allow the intensity ( $w_b$ ) axis to be more coarsely sampled than the other two, because the intensity axis is more sensitive to the variation in lighting from shadows and distance from the light source. The  $w_b$  axis is divided into 8 sections while the  $r_g$  and  $b_y$  axes are each divided into 16 sections, for a total of  $8 \times 16 \times 16 = 2048$  bins. As the total intensity limits the colour differences possible, only a fraction of them can actually receive counts. To clarify this point, suppose that the camera outputs a maximum number of intensity levels,  $M$  on each channel. Letting  $w_b = 0$  ( $r = g = b = 0$ ) or  $w_b = 3M$  ( $r = g = b = M$ ), both axes  $b_y$  and  $r_g$  turn out to be 0. Thus the maximum intensity level restricts the available axes.

Several measures have been proposed for the dissimilarity between two histograms,  $H = \{h_i\}$  and  $K = \{k_i\}$ . The bin-by-bin dissimilarity is determined by comparing

**Table 8. Comparison of distances between the different categorical images**

|             |   | Sample Images   |   |   |   |
|-------------|---|---|---|---|---|
|             |   |  |  |  |  |
|             |   | 1603  | 1623  | 828   | 842   |
| Query Image |  | 4.95312   | 4.89965   | <b>0.225879</b>   | <b>0.181743</b>   |
|             | 840   |   |   |   |   |
| Query Image |  | <b>0.340588</b>   | <b>0.399056</b>   | 5.14536   | 5.09391   |
|             | 1599  |   |   |   |   |

**Table 9. Sample query set for query by example**

|   |   |   |
|---|---|---|
|  |  |  |
| Car – 1599  | Rose - 805  | Mountain – 125  |
|  |  |  |
| Seaside – 1713  | Sunset - 128  | Pattern – 444   |

the contents of the corresponding histogram bins, i.e.  $h_i$  and  $k_i$  for all  $i$  but not  $h_j$  and  $k_j$  for  $i \neq j$ . The dissimilarity between the two histograms is determined as a sum of all pair-wise comparisons and it implies a binary ground distance with a threshold depending on the bin size.

Minkowski-Form Distance: This distance is commonly defined as:

$$d_{L_r}(H, K) = \left( \sum_i |h_i - k_i|^r \right)^{1/r} \quad (33)$$

The  $L_1$  distance is often used for computing dissimilarity between colour images<sup>15</sup>.

Kullback-Leibler Divergence and Jeffrey Divergence: The Kullback-Leibler divergence appears as:

$$d_{KL}(H, K) = \sum_i h_i \log \frac{h_i}{k_i} \quad (34)$$

From the information theory point of view, the K-L divergence measures how inefficient it could be to code one histogram using the other as the code-book<sup>23</sup>. However, the K-L divergence is non-symmetric and is sensitive to histogram binning. The empirically derived Jeffrey divergence is a modification of the K-L divergence that is numerically stable and robust with respect to noise and the size of histogram bins<sup>24</sup>. The Jeffrey divergence is determined from:

$$d_J(H, K) = \sum_i \left( h_i \log \frac{h_i}{m_i} + k_i \log \frac{k_i}{m_i} \right) \quad (35)$$

$\chi^2$  statistics:

$$d_{\chi^2}(H, K) = \sum_i \frac{(h_i - k_i)^2}{m_i} \quad (36)$$

$$\text{where, } m_i = \frac{h_i + k_i}{2}$$

Given a query image with histogram  $H$ , each database image with histogram  $K$  receives a measure of dissimilarity for the query image. It is then easy to rank the database images based on their dissimilarity measures and return the best matches.

## 5. RESULTS AND DISCUSSION

The experimental configuration for a process as subjective as computing similarity between the images must be carefully set up to gauge the results with other methods and we need to remove any perceptual biases of an experimenter. To minimize the human subjectivity, the random samples of different sizes are used and also a large sample space is taken as the image database. The large image database rules out the possibility of having dominance of any particular category of images.

Our retrieval system is tested using 9907 images from the low resolution database<sup>26</sup> as mentioned above. This image collection contains different category of images like cars, roses, mountains, patterns, animal, landscape, seaside, flowers, human activities, etc. To provide numerical results, 8 sample images are taken randomly selected from four categories prior to the manual determination of the correctness of retrieved images. Each category contains at least 100 relevant images. A retrieved image is considered a match if it belongs to the same category of the query image.

The query results obtained by our entropy-based matching are compared with those obtained from colour histogram matching using different dissimilarity measures –  $\chi^2$  statistics, Jeffrey divergence and  $L_1$  norm. Our content based image retrieval system is developed in Visual C++ 6.0 as an offline system on Intel Celeron machine with 1.40GHz processor and 256 MB RAM. The average retrieval time is 1 second per 1000 images. The process of image retrieval involves the following steps:

- Create the interim result set using the proposed indexing scheme.
- Extract feature values (colour entropy, texture entropy and dominant colour) for the query image.
- Calculate the image distance of the query image with all images in the interim result set.
- Rank the images according to the distance measured from the query image.
- Set the highest rank of the retrieved image which has the minimum distance with respect to the query image.

The results of retrieval by our entropy-based system are compared with those from colour histogram matching<sup>15</sup>. Testing is carried out on a set of query images from the benchmark image<sup>26</sup> database of size 9907.

It may be observed from Fig. 4 that the performance of the proposed system on the car query (car Id: 1639) outperforms

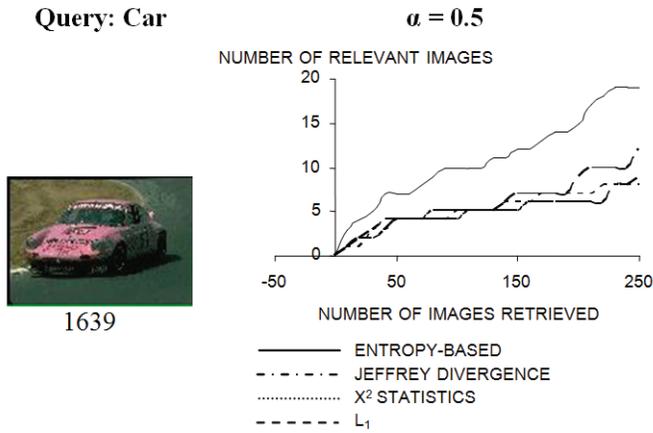


Figure 4. Performance of the car query on the data set of size 9907.

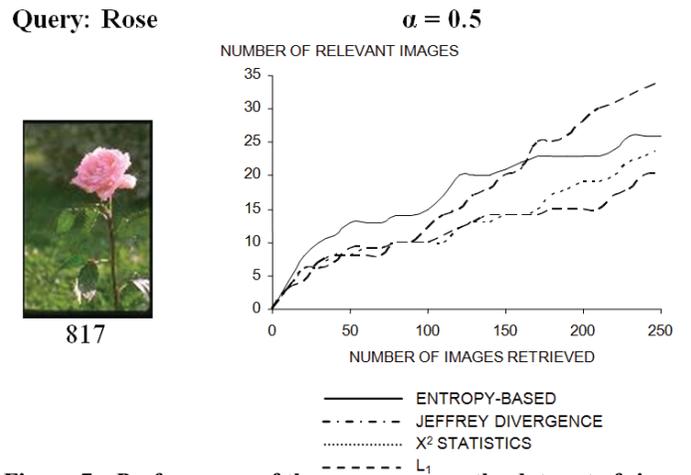


Figure 7. Performance of the rose query on the data set of size 9907.

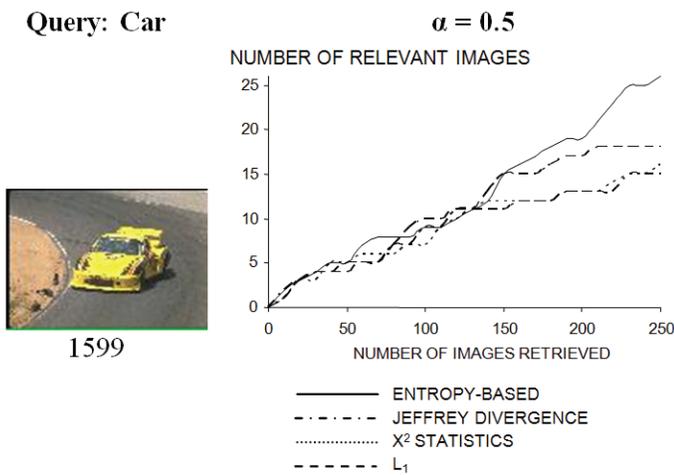


Figure 5. Performance of the car query on the data set of size 9907.

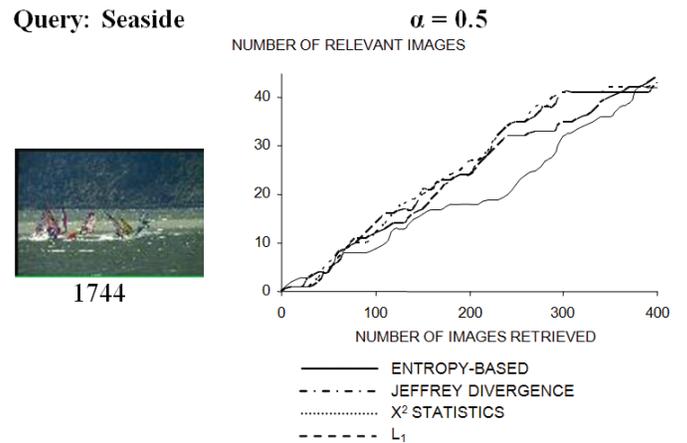


Figure 8. Performance of the seaside query on the data set of size 9907.

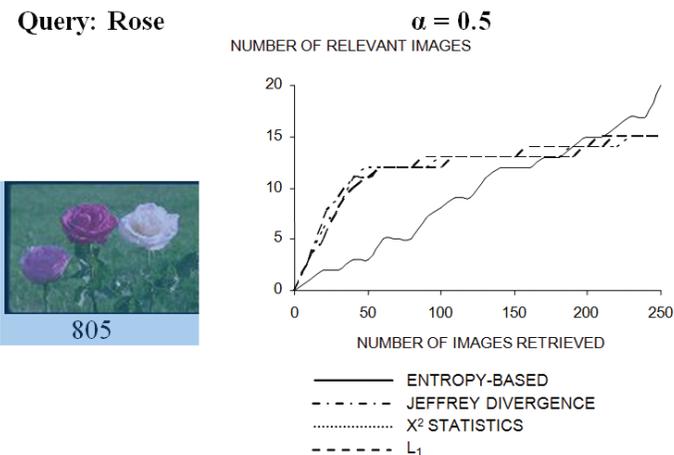


Figure 6. Performance of the rose query on the data set of size 9907.

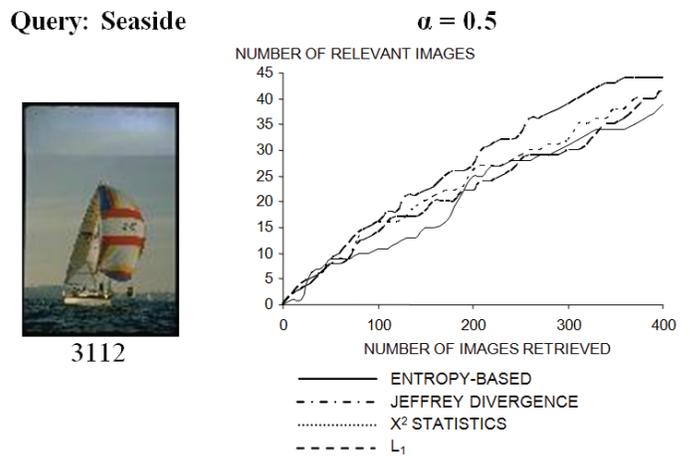


Figure 9. Performance of the seaside query on the data set of size 9907.

## Query: Pattern

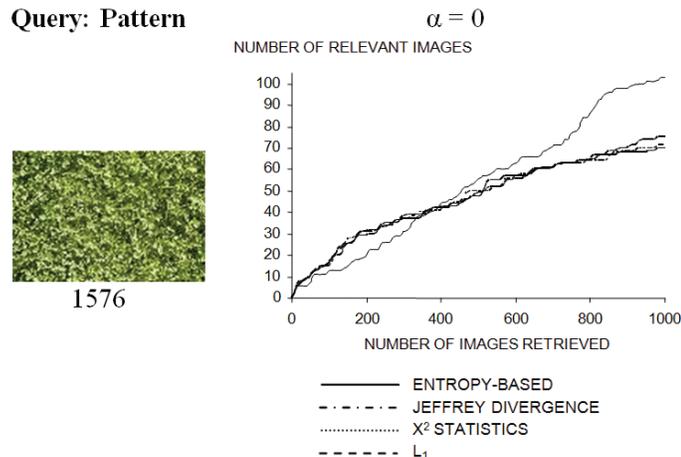


Figure 10. Performance of the pattern query on the data set of size 9907.

Jeffrey divergence,  $\chi^2$  statistics and  $L_1$  norm, but the results from Fig. 5 show that entropy based matching is comparable with others. The entropy-based method retrieves the correct images using only 3 features – colour entropy, texture entropy and dominant colour thus enhancing the processing speed of the system with a reduced number of feature values.

The performance of other query images of the categories: Rose, seaside and pattern are now discussed.

Figures 10 and 11 both show the performance of the image retrieval for query images from the pattern category. The value of  $\alpha$  is set to 0 to rule out the dominant colour information. The entropy based method retrieves the images purely based on entropy values and it is found to perform reasonably better than other histogram matching methods for the pattern category thus demonstrating its ability to retrieve semantically similar images.

It may be noted from the results of experiments that the proposed system performs better than the histogram matching for the query images, which possess distinct features in terms of texture or colour. In the rose, seaside category images, dominant colour and colour distribution are the prominent features and in these cases the proposed system performs reasonably well, but our proposed CBIR system has stooped to the subdued performance in the face of histogram matching for some cases. If the pattern images have the distinct texture features, then our system comes out with significantly better performance as shown in Figs 10 and 11.

## 6. CONCLUSIONS

A new entropy function that is aimed at representing the information in a fuzzy set is presented along with some important properties. The proposed entropy function has four tunable parameters that can be estimated by optimising the entropy function itself. For obtaining the global solution reinforcement learning along with population based approaches is used.

Two types of entropy based image features – colour entropy and texture entropy are utilized in this work. The colour entropy is described in terms of randomness in the distribution

## Query: Pattern

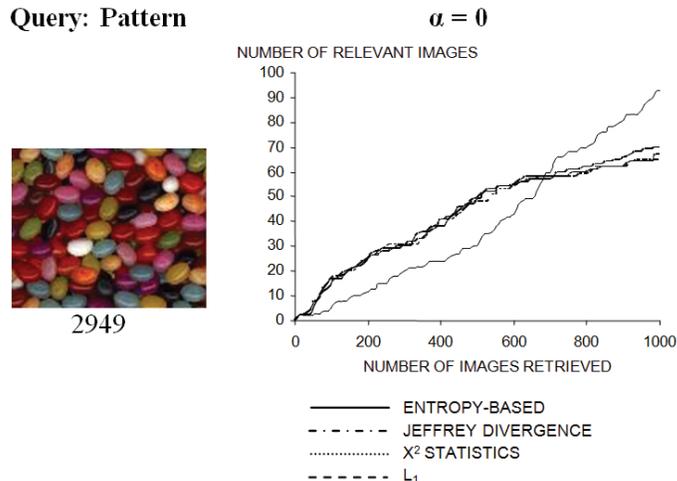


Figure 11. Performance of the pattern query on the data set of size 9907.

of colours in an image. For this, the newly devised entropy function comes handy and resorting to the optimisation of this entropy function yields the optimised colour entropy value. The resulting entropy values uniquely classify the non textured image. But for the textured image first the fuzzy texture features of the image are extracted and then the new entropy function is applied to the features for optimising the entropy values of texture.

Both the entropy based image features are derived in such a manner that they rely only on the colour and texture distributions of an image. In the course of research, the need for visual colour information is realized, consequently the dominant colour value is considered as the third image feature. Despite the efficiency of the RGB colour space, the pixels are transformed from the RGB colour space to the CIELAB colour space because the latter one has the property of perceptual uniformity. Falling in line, the CIELAB value of the dominant colour is also used as the visual feature.

Indices permit the computer in finding the images relevant to a query without looking at every image in the database. The indexing of the colour feature vectors is investigated to speed up the atomic queries. Dimensionality is one of the major concerns in the indexing scheme. The index speed degrades as the dimensionality of the data indexed increases. In the experiments carried out, images are indexed by the entropy values of colour and texture. The uniqueness of the entropy values of an image and their lower dimensionality rule out the possibility of using any sophisticated and computationally taxing multidimensional indexing data structures. The normalised entropy values of colour and texture are found to be more suitable as a pair to index an image. This indexing scheme is found to be effective in creating an interim result set for a query image. The algorithm for the interim result set and the results obtained ensure that the size of the database is reduced considerably without affecting the correctness of the results. The objective of an image retrieval system is to identify images from the database, which are similar to the query image. Similarity or distance measurement between the two

images is derived from the empirical estimates of the image features. The three image features – colour entropy, texture entropy and dominant colour in Lab colour space are selected to formulate the similarity measure. Two distances:  $L_2$  distance between the dominant colours of two images and  $L_2$  distance of between the two colour entropy values and also between the texture entropy values are explored. These two  $L_2$  distances are combined in linear form to give the measure of entropy. Table 8 demonstrates that the distance between semantically similar images are significantly low in comparison to the distance between the two dissimilar images. This measure is used to rank the retrieved images by their distances from the query image.

In the proposed retrieval system, the query paradigm is employed for setting up the query as ‘Query by Example’ (QBE). In that WBIIS test image dataset of 9907 colour miscellaneous images is organized into several categories with at least 100 images per category. The features of the images are extracted off-line and the similarity measures along with their ranking are computed at the time of executing the query. The query results so obtained are compared with those from the colour histogram matching using  $\chi^2$  Statistics, Jeffrey Divergence and  $L_1$  norm. The results of our system are very promising in the retrieval of the similar images from the database. Exploring the entropy function for fuzzy modeling of the image retrieval system is the next phase of this work.

## REFERENCES

- Shannon, C.E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, 379-423.
- Jain, K. Fundamental of digital image processing. Prentice Hall of India, 2000.
- Manjunath, S. & Ma, W.Y. Texture features for browsing and retrieval of image data. *IEEE Trans. Pattern Anal. Mach. Intelligence*, 1996, **18**(8), 837-42.
- Koskela, M.; Laaksonen, J. & Oja, E. Entropy-based measures for clustering and SOM topology to content-based image indexing and retrieval. In the Proceedings of the 17<sup>th</sup> International Conference on Pattern Recognition (ICPR'04), 2004, edited by J. Kittler, M. Petrou, and M.S. Nixon, IEEE CS Press, US. 2004, **2**, pp. 1005-1008.
- Pun, T. A new method for gray-level picture thresholding using the entropy of the histogram. *Signal Processing*, 1980, **2**, 223-37.
- Kapur, J.N.; Sahoo, P.K. & Wong, A.K.C. A new method for gray-level picture thresholding using the entropy of the histogram. *Compu. Graph. Vision Image Proc.*, 1985, **29**, 273-85.
- Leung, K.; Lam, F.K. & To, J.T.P. Entropy-based multiscale image segmentation with edge refinement. In the Proceedings of 1<sup>st</sup> International Conference on Information Technology and Application (ICITA 2002), 2002, Bathurst, NSW Australia, 24-28 November 2002.
- Pal, Nikhil R. & Pal, Sankar K. Entropy: A new definition and its applications. *Syst. Man Cyber.*, 1991, **21**(5), 1260-270.
- Renyi, A. On measure of entropy and information. In the 4<sup>th</sup> Berkeley Symposium of Mathematics Statistics and Probability, Berkeley, California, 1961, **1**, pp. 547-61.
- Pal, N.R. & Pal, S.K. Some Properties of the Exponential Entropy. *Information Sciences*, 1992, **66**, 113-17.
- Boujemaa, N.; Fauqueur, J.; Ferecatu, M.; Fleuret, F.; Gouet, V.; LeSaux, B. & Sahbi, H. IKONA: interactive specific and generic image retrieval. In International Workshop on Multimedia Content-based Indexing and Retrieval (MMCBIR'2001), France, 2001.
- Moshfeghi, M.; Saiz, C. & Yu, H. Content-based Retrieval of medical images with relative entropy. In Proceedings of SPIE, 2004, **5371**, pp. 259-267.
- Byoung, K.; Jing, P. & Hyeran, B. Region-based image retrieval using probabilistic feature relevance learning. *Pattern Anal. Appl.*, 2001, **4**(2-3), 174-84.
- Battiato, S.; Gallo, G. & Nicotra, S. Perceptive visual texture classification and retrieval. In the 12<sup>th</sup> International Conference on Image Analysis and Processing, 2003, Mantova, Italy, 17-19 September 2003, pp. 524-29.
- Swain, M. & Ballard, D. Colour indexing. *Inter. J. Comp. Vision*, 1991, **7**(1), 11-32.
- Zachary, J. M. Jr. Louisiana State University, 2000. PhD Dissertation.
- Yagi, E. Analytical approach to Clausius's first memoir on mechanical theory of heat (1850). *Historia Sci.* 1981, **20**, 77-94.
- Wang, J. Z.; Wiederhold, G.; Firschein, O. & Wei, S.X. Content- based image indexing and searching using daubechies' wavelets. *Int. J. Digital Libr.*, 1997, **1**(4), 311-28.
- Boltzmann, Ludwig (1896, 1898). Lectures on gas theory. Translated by Stephen G. Brush, Berkeley: University of California Press, New York, 1995.
- Jain, R.; Murthy, S.N.J.; Chen, P.L.J. and Chatterjee, S. Similarity measures for image databases. In Proceedings of SPIE, 1995, **2420**, pp. 58-65.
- Faloutsos, C.; Barber, R.; Flickner, M.; Hafner, J.; Niblack, W.; Petkovic, D. & Equitz, W. Efficient and effective querying by image content. *J. Intell. Inform. Syst.*, 1994, **3**(3-4), 231-62.
- Lennie, E, & D'Zmura, M. Mechanisms of colour vision. *CRC Crit. Rev. Neurobiol.* 1988, **3**, 333-400
- Cover, T.M. & Thomas, J.A. Elements of information theory. Wiley Series in Telecommunications. John Wiley & Sons, New York, NY, USA. 1991.
- Puzicha, J.; Hofmann, T. & Buhmann, J. Non-parametric similarity measures for unsupervised texture segmentation and image retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 1997. San Juan, Puerto Rico, pp. 267-72.
- Wang, WBIIS J.Z.; Wiederhold, G.; Firschein, O. & Sha, X.W. Content-Based Image Indexing and Searching Using Daubechies' Wavelets. *Int. J. Digital Libraries*, 1998, **1**(4), 311-28.
- Download databases for research comparison (low resolution web-crawled misc database used in WBIIS). <http://wang.ist.psu.edu/docs/related.shtml> (Accessed on :

September 2010)

27. Hanmandlu, M. & Murthy, O.V.R. Reinforcement learning in the entropy based recognition of handwritten hindi numerals. *In* the 10<sup>th</sup> World Multi-conference on Systemics, Cybernetics and Informatics, 16-19 July 2006, Orlando, Florida, USA.

modeling for dynamic systems and applications of fuzzy logic to image processing, document processing, medical imaging, multimodal biometrics, surveillance and intelligent control. He has authored a book on Computer Graphics and published more than 220 publications in both conferences and journals. He has guided 15 PhDs and 100 MTech students.

**Contributors**



**Dr Madasu Hanmandlu** received his MTech (Power systems) from REC Warangal, Jawaharlal Nehru Technological University (JNTU), India, in 1976, and PhD (Control Systems) from Indian Institute of Technology (IIT), Delhi, India, in 1981. Presently working as Professor in Department of Electrical Engineering, IIT, Delhi. His current research interests mainly include: Fuzzy



**Dr Anirban Das** received MCA from Jawaharlal Nehru University, New Delhi in 1999, MTech (Computer Sci. Engg) from IIT Delhi in 2002 and PhD (Computer Sci.) from Jamia Millia Islamia University, New Delhi in 2008. He is currently working with R Systems International Ltd. as Technical Architect for biometric system development. His research interests include:

Image processing, computer vision, pattern recognition, biometric authentication, and artificial intelligence.

**APPENDIX ‘A’**

**Proofs of some important properties of the new entropy function**

We will now provide the proofs of some important properties presented in Section II-A.

*Proof of Property 2*

$$I(p_i) \rightarrow e^{-d} \text{ as } p_i \rightarrow 0 \text{ and } I(p_i) \rightarrow e^{-(a+b+c+d)} \text{ as } p_i \rightarrow 1.$$

We consider  $k_1 = e^{-d}$  and  $k_2 = e^{-(a+b+c+d)}$ . Since a, b, c and d are real so  $k_1$  and  $k_2$  are finite. Hence  $I(p_i)$  is bounded.

*Proof of Property 3*

We have  $k_1 = e^{-d}$  and  $k_2 = e^{-(a+b+c+d)}$  as defined in the Property 2. The ratio,  $k_1 / k_2 = e^{-d} / e^{-(a+b+c+d)} = e^{(a+b+c)} > 1$  for  $(a+b+c) > 0$  (37)

To prove that  $I(p)$  is a decreasing function, we need to show that the derivative of  $I(p)$  with respect to  $p$  is always negative or zero.

$$\text{The derivative, } \frac{d}{dp} I(p) = \frac{d}{dp} e^{-(ap^3+bp^2+cp+d)} = -e^{-(ap^3+bp^2+cp+d)} (3ap^2 + 2bp + c) \tag{38}$$

Definitely for any  $p, 0 \leq p \leq 1, e^{-(ap^3+bp^2+cp+d)} > 0$ . Hence,

$$\frac{d}{dp} I(p) = \frac{d}{dp} e^{-(ap^3+bp^2+cp+d)} = -e^{-(ap^3+bp^2+cp+d)} (3ap^2 + 2bp + c) \tag{39}$$

So we can conclude that for  $a \geq 0, b \geq 0$  and  $c \geq 0, I(p)$  always decreases for  $0 \leq p \leq 1$ .

*Proof of Property 5*

Consider the case where  $p_1 = p_2 = \dots = p_n = 1/n$  and  $n \geq 1$ . Then

$$H(P) = \sum_{i=1}^n p_i e^{-(ap_i^3+bp_i^2+cp_i+d)} = \sum_{i=1}^n (1/n) e^{-\left(\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d\right)} = \sum (1/n) h(n) \tag{40}$$

where,  $h(n) = e^{-\left(\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d\right)}$

To prove that  $H(P)$  is an increasing function, it is sufficient to prove that  $h(n)$  is an increasing function.

$$\frac{d}{dn}h(n) = \frac{d}{dn}e^{-\left(\frac{a}{n^3}+\frac{b}{n^2}+\frac{c}{n}+d\right)}\left[\frac{3a}{n^4}+\frac{2b}{n^3}+\frac{c}{n^2}\right]e^{-\left(\frac{a}{n^3}+\frac{b}{n^2}+\frac{c}{n}+d\right)} = (1/n^4)(3a+2bn+cn^2)e^{-\left(\frac{a}{n^3}+\frac{b}{n^2}+\frac{c}{n}+d\right)} \quad (41)$$

For  $n \geq 1$  and  $a \geq 0, b \geq 0, c \geq 0$   $\frac{d}{dn}h(n) \geq 0$ .

Therefore  $H(P)$  is an increasing function for  $a, b, c \geq 0$ .

#### Proof of Property 7

For the sake of simplicity, consider  $p_1 < p_2$ . Let us now construct a new partition  $B = [B_1, B_2, A_3, \dots, A_n]$  with  $P_r[B_1] = p_1 + \delta, P_r[B_2] = p_2 - \delta$  and  $\delta > 0$  and  $\delta \leq (p_2 - p_1)/2$ .

Now,  $H(B) - H(A) = H(p_1 + \delta, p_2 - \delta, p_3, \dots, p_n) - H(p_1, p_2, p_3, \dots, p_n)$

$$\begin{aligned} &= (p_1 + \delta)e^{-[a(p_1 + \delta)^3 + b(p_1 + \delta)^2 + c(p_1 + \delta) + d]} - p_1e^{-[ap_1^3 + bp_1^2 + cp_1 + d]} \\ &+ (p_2 - \delta)e^{-[a(p_2 - \delta)^3 + b(p_2 - \delta)^2 + c(p_2 - \delta) + d]} - p_2e^{-[ap_2^3 + bp_2^2 + cp_2 + d]} \end{aligned} \quad (42)$$

Since  $\phi(p) = pe^{-[ap^3 + bp^2 + cp + d]}$  is convex, the condition  $\phi(p_1) + \phi(p_2) < \phi(p_1 + \delta) + \phi(p_2 - \delta)$  is easily satisfied if  $p_1 < p_1 + \delta < p_2 - \delta < p_2$ . Hence,  $H(B) - H(A) > 0$ , i.e.  $H(B) > H(A)$ , thus completing the proof.

#### Proof of Property 8

The proposed entropy function for the probability distribution  $P = [p_1, p_2, \dots, p_n]$  is defined as follows:

$$H(P) = \sum_{i=1}^n p_i e^{-(ap_i^3 + bp_i^2 + cp_i + d)}$$

Now,

$$\begin{aligned} \frac{\partial H}{\partial p_i} &= e^{-(ap_i^3 + bp_i^2 + cp_i + d)} - p_i e^{-(ap_i^3 + bp_i^2 + cp_i + d)}(3ap_i^2 + 2bp_i + c) \\ &= e^{-(ap_i^3 + bp_i^2 + cp_i + d)}[1 - 3ap_i^3 - 2bp_i^2 - cp_i] \end{aligned} \quad (43)$$

$$\begin{aligned} \frac{\partial^2 H}{\partial p_i^2} &= e^{-(ap_i^3 + bp_i^2 + cp_i + d)}(-9ap_i^2 - 4bp_i - c) - e^{-(ap_i^3 + bp_i^2 + cp_i + d)}(1 - 3ap_i^3 - 2bp_i^2 - cp_i)(3ap_i^2 + 2bp_i + c) \\ &= e^{-(ap_i^3 + bp_i^2 + cp_i + d)}\{-9a^2p_i^5 - 12abp_i^4 - (4b^2 + 6ac)p_i^3 + (12a - 4bc)p_i^2 + (6b - c^2)p_i + 2c\} \end{aligned} \quad (44)$$

$$\begin{aligned} \left. \frac{\partial^2 H}{\partial p_i^2} \right|_{p_i=1/n} &= -(1/n^5)e^{-\left(\frac{a}{n^3}+\frac{b}{n^2}+\frac{c}{n}+d\right)}\{2cn^5 + (6b - c^2)n^4 + (12a - 4bc)n^3 - (4b^2 + 6ac)n^2 - 12abn - 9a^2\} \\ &= -\beta \end{aligned} \quad (45)$$

where,

$$\beta = (1/n^5)e^{-\left(\frac{a}{n^3}+\frac{b}{n^2}+\frac{c}{n}+d\right)}\{2cn^5 + (6b - c^2)n^4 + (12a - 4bc)n^3 - (4b^2 + 6ac)n^2 - 12abn - 9a^2\}$$

and

$$\frac{\partial^2 H}{\partial p_i \partial p_j} = 0 \quad \text{for } i \neq j \text{ and } i, j = 1, 2, \dots, n$$

The Hessian matrix is of the form

$$H = \begin{pmatrix} -\beta & 0 & 0 & \dots & 0 \\ 0 & -\beta & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\beta \end{pmatrix} \quad (46)$$

For the point  $P_0 = (1/n, 1/n, \dots, 1/n)$  to attain the maximum,  $H|_{P_0}$  should be negative definite. Moreover,  $H$  is negative definite if the determinant value of  $k^{\text{th}}$  principal minor of  $H$  has the sign of  $(-1)^k, k = 1, 2, \dots, n$ .

Therefore, determinant of  $k^{\text{th}}$  principal minor of  $H$  is

$$\begin{pmatrix} -\beta & 0 & 0 & \dots & 0 \\ 0 & -\beta & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & -\beta \end{pmatrix} = \begin{cases} (-1)k\beta k \\ \beta k \text{ if } k \text{ is even} \\ -\beta k \text{ if } k \text{ is even} \end{cases} \tag{47}$$

Here, a need arises to show that  $\beta > 0$ , which implies that

$$(1/n^5)e^{-\left(\frac{a}{n^3} + \frac{b}{n^2} + \frac{c}{n} + d\right)} \{2cn^2 + (6b - c^2)n^4 + (12a - 4bc)n^3 - (4b^2 + 6ac)n^2 - 12abn - 9a^2\} > 0$$

This yields

$$n^3 > \frac{(4b^2 + 6ac)n^2 + 12abn + 9a^2}{2cn^2 + (6b - c^2)n + (12a - 4bc)}$$

The above is always true if  $a \geq 0, b \geq 0, c \geq 0$  and  $a, b, c$  must satisfy (14).

So,  $H(P) \leq H(1/n, 1/n, \dots, 1/n)$ .

*Proof of Property 9*

Suppose that  $p_i = 0$  for all  $i$  except  $p_k$  where  $p_k = 1$ . Next, we prove that the entropy  $H$  is the minimum by contradiction. Consider that there are at least two non-zero probabilities say  $p_i$  and  $p_j$  for the minimum value of  $H$ . Now using the Property 7, we can write that

$$H(p_1, p_2, \dots, p_i + \delta, \dots, p_j - \delta, \dots, p_n) > H(p_1, p_2, p_3, \dots, p_n)$$

where,  $\delta > 0$  and  $\delta \leq |p_i - p_j|/2$

In our case,  $H(0, 0, \dots, \delta, \dots, 1 - \delta, \dots, 0) > H(0, 0, \dots, 1, \dots, 0)$ . This contradicts the fact that  $H$  is the minimum. Hence  $H$  is minimum only when all  $p_i$ 's except one are zeros.

*Proof of Property 10*

The partition  $A = [A_1, A_2, \dots, A_n]$  is changed to  $B = [B_a, B_b, A_2, \dots, A_n]$  where  $A_1$  is subdivided into  $B_a$  and  $B_b$  and  $p_a = P_r(B_a), p_b = P_r(B_b)$  and  $p_1 = p_a + p_b$

Now  $H(A) = \sum_{i=1}^n p_i e^{-(ap_i^3 + bp_i^2 + cp_i + d)}$  and let us consider  $\phi(p) = p e^{-(ap^3 + bp^2 + cp + d)}$ .

We can write

$$H(A) - \phi(p_a + p_b) = H(B) - \phi(p_a) - \phi(p_b)$$

$$\Rightarrow H(A) - H(B) = \phi(p_a + p_b) - \phi(p_a) - \phi(p_b)$$

$$H(A) - H(B) = (p_a + p_b) e^{-[a(p_a + p_b)^3 + b(p_a + p_b)^2 + c(p_a + p_b) + d]} - p_a e^{-[ap_a^3 + bp_a^2 + cp_a + d]} - p_b e^{-[ap_b^3 + bp_b^2 + cp_b + d]}$$

$$\Rightarrow H(A) - H(B) = p_a e^{-[ap_a^3 + bp_a^2 + cp_a + d]} [e^{-[3ap_a^2 p_b + 3ap_a p_b^2 + ap_b^3 + 2bp_a p_b + bp_b^2 + cp_b]} - 1] + p_b e^{-[ap_b^3 + bp_b^2 + cp_b + d]}$$

$$[e^{-[3ap_a^2 p_b + 3ap_a p_b^2 + ap_b^3 + 2bp_a p_b + bp_b^2 + cp_b]} - 1]$$

$$\Rightarrow H(A) - H(B) \leq 0$$

This completes the proof.