Feature Extraction from Rural Satellite Imagery Using Color Based CBIR Techniques

Mamatha Y.N[1] and Dr.A.G..Ananth[2]

[1] Lecturer, R V College of Engineering, Bangalore-560059.
[2] Professor, R V College of Engineering, Bangalore-560059.
`mamatharaj_76@rediffmail.com` , `antisro@yahoo.com`

Abstract:

The ability to organize and retrieve visual information such as images is becoming a crucial problem for specialists, because processing visual information requires perceptual abilities not yet known to exist in computational form. Therefore the ability to retrieve visual information is a complex problem. The main retrieval pattern include the features of an image such as colour, and shape. The algorithms used to calculate the similarity between extracted feature with an image database, that utilizes colour features of the images as the basis of comparison and retrieval. An object-oriented definition also involves prescribing a set of pertinent features or pixels as well as a method such as a classification algorithm with the appropriate training data. A low resolution satellite image of a rural scenario has been selected for the analysis. The entire image scene is divided into four equal quadrants. The Content-based retrieval of image features have been carried out for the selected two quadrants I and II. The color classification of the image in two quadrants clearly indicates that there are four important features that are common to the images[1] The features identified include, Natural Vegetation, Water Bodies, Land, and Housing associated with the image. Further using the L*a*b color spectral distribution and histogram techniques an estimation of the different features identified in the image has been made for all the four quadrants. The results are presented in the paper.

Keywords: Bayesian Semantic, segmentation, Satellite image, Features

Introduction:

An image feature is a distinguishing primitive characteristics or attribute of an image. Some features such as luminance, shape descriptor, and gray scale texture are natural since they correspond to visual appearance of an image. Other features such as amplitude histogram, color histogram, and spatial frequency spectra lack a natural correspondence. Different features are often grouped into feature vectors. Images in an archive can be segmented into regions characterized by homogeneous feature vectors [1,2] Similarity search in the n-dimensional feature space thus consists of comparing the target feature vector with the feature vectors stored in the database. An object-oriented definition of a feature object involves prescribing a set of pertinent features as well as a method (such as a clustering algorithm with the appropriate parameters) which characterizes the homogeneity of the object. Feature objects can be predefined and pre-extracted, user-defined and constructed at query time using pre-extracted features, or even user-defined and extracted at query time. Various spatial indexing schemes such as R-Trees can be used to facilitate feature space indexing [3,4.]

Digital images databases however, open the way to colour based searching. In this paper technical aspects of current content-based image retrieval systems are discussed. This will provide an overview of the functionality of temporary image retrieval systems in terms of technical aspects: querying, relevance feedback, features, matching measures, indexing data

structures, and result presentation. It compares specific systems, rather than general architectures. It also is a thorough foundation for claims that most systems use low level features, and few use high level semantic meaningful features.

Through the database one by one, another way is to specify the image in terms of keywords, or in terms of image features that are extracted from the image, such as a color histogram. Yet another way is to provide an image or sketch from which features of the same type must be extracted as for the database images, in order to match these features. Relevance feedback is about providing positive or negative feedback about the retrieval result, so that the system can refine the search. Consideration is of several classes of features that are used to specify queries: color, texture, shape, spatial layout, and faces. Color features are often easily obtained directly from the pixel intensities, e.g. color histogram over the whole image, or a fixed sub image or a segmented region is often used.

Bayesian Inference:

The objective is to determine the "target" image $T \in I$ that the user requires. T could be a specific known image in the database or more generally that image in I which best satisfies the user's subjective search criteria. The determination of T is accomplished by displaying a set of ND images from I, from which the user picks one that best satisfies what is being looked for. The system uses this information to select another image set, from which the user picks one, and so on. We define $Di \subseteq I$ to be the set of displayed images at the ith iteration of this process, and $Ai \in Di$ to be the image picked. We define $H_t = \{D_1, A_1, D_2, A_2, \ldots, D_t, A_t\}$ to be the history of displayed images and user actions up to the tth iteration. The learning algorithm is based around the model for the probability of which image a user picks from D_k given the target image T

$$P(A_k \mid D_k, T, \sigma, F) = \frac{\exp\left(-d_F(A_k, T)/\sigma\right)}{\sum_{T_j \in D_k} \exp\left(-d_F(T_j, T)/\sigma\right)}, \ A_k \in D_k,$$

Where σ is a precision parameter and d_F is a normalized distance measure in the set of image features F. The parameter σ is a measure of the performance of the feature space in describing the user's query; a small value of σ implies that there is a small region of feature space that contains images that satisfy the user's query, whereas a large value implies that images satisfying the query are not well clustered in feature space. In this case we have 3 sets of features: global color, texture and segmentation features, so $F \in \{GC, TX, SG\}$. The idea is that users may search according to different feature types, which should be learnt through inferring F. The unknowns are T, σ and F. Given H_t, our knowledge about these unknowns is given by the posterior distribution:

$$P(T, \sigma, F \mid H_t) \propto \left(\prod_{k=1}^{t} P(A_k \mid D_k, T, \sigma, F)\right) P(T) P(\sigma) P(F)$$

Where P(T), P(σ) and P(F) are prior distributions that we assume are uniform: $P(T = Ti) = N^{-1}$, $i = 1, \ldots, N$, $P(\sigma) = 1$, $0 \leq \sigma \leq 1$ and $P(F) = 1/3$, $F \in \{GC, TX, SG\}$.

We discrete the range of σ into 20 values 0.025, 0.075, . . . , 0.975, from which we can compute this posterior exactly at all combinations of (T,F, σ) without the need for any approximations. The interest in this paper is the marginal posterior distribution of T:

$$P(T = T_i \mid H_t) = \int_0^1 \sum_{F \in \{\text{GC,TX,SG}\}} P(T_i, \sigma, F \mid H_t), \ i = 1, \ldots, N.$$

Which can also be computed exactly using the discretisation of σ. This is used to decide which set of images to display at the next iteration. In this system the ND images with the highest marginal posterior probability are selected for D_{t+1}.

Data Analysis:

For a satellite image for rural scene is selected for the detection and estimation of natural resources. The image is subjected to various color based classification and image processing techniques such as enhancement and contrast stretching to derive the various features present in the image. Based on the color content of the various features the L*a*b images has been constructed to identify and estimate the extent of the resources available in the rural scene. By subdividing the original image into four quadrants the histograms have been constructed for two of the selected sample images for estimating the natural resources available in the region. The image in Figure 1(a) shows 995 x 571 size of a satellite Image of a rural scene. The image is subjected to image processing enhancement techniques using contrast stretching is shown in fig 1 (b). We have also tried to identify number of natural features in the images by color identification.

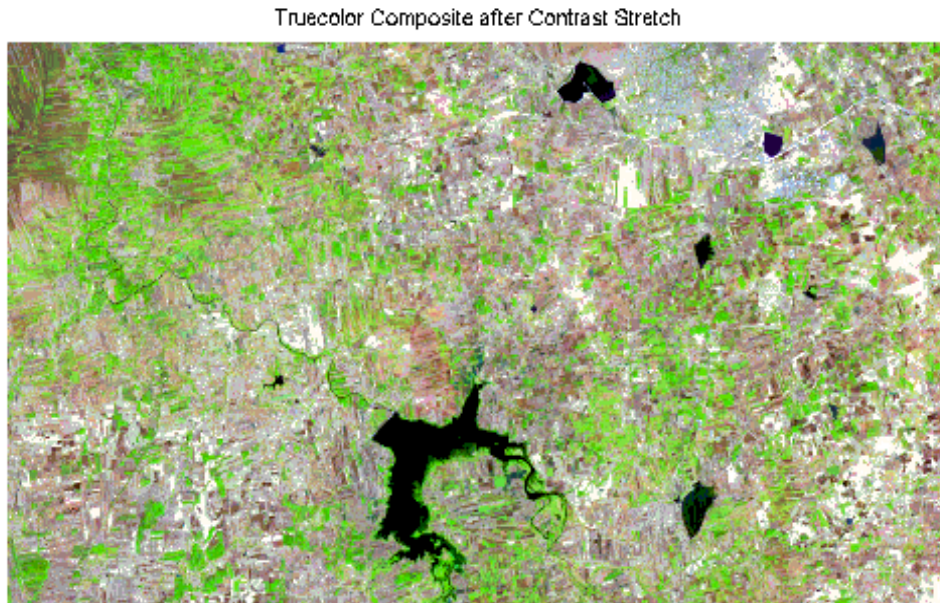Figure 1(a) : Original satellite image

Figure 1(b) : Enhanced satellite image

This is carried out by converting RGB image into L*a*b color space in order to find the sample region of each color which helps in finding the pixel distribution of individual colors computed using MATLAB tool is shown in figure1 (c).
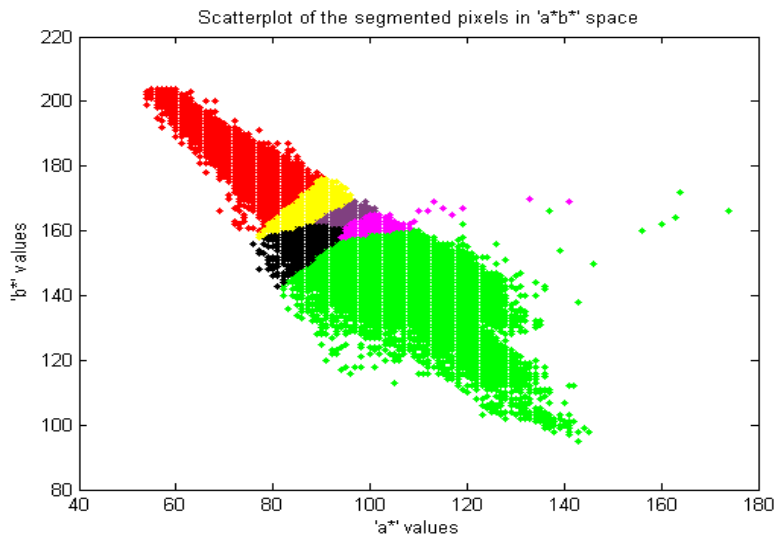
Figure 1(c):L*a*b color diagram of original image

Then order to make a better estimation of the natural resources the original image has been divided into four quadrants as shown in Table 1. The sub images in Quadrants I and II have been selected for further image processing. Fig 2(a) and 2 (b) shows the original two sub-images selected for detailed analysis.

Quadrant I
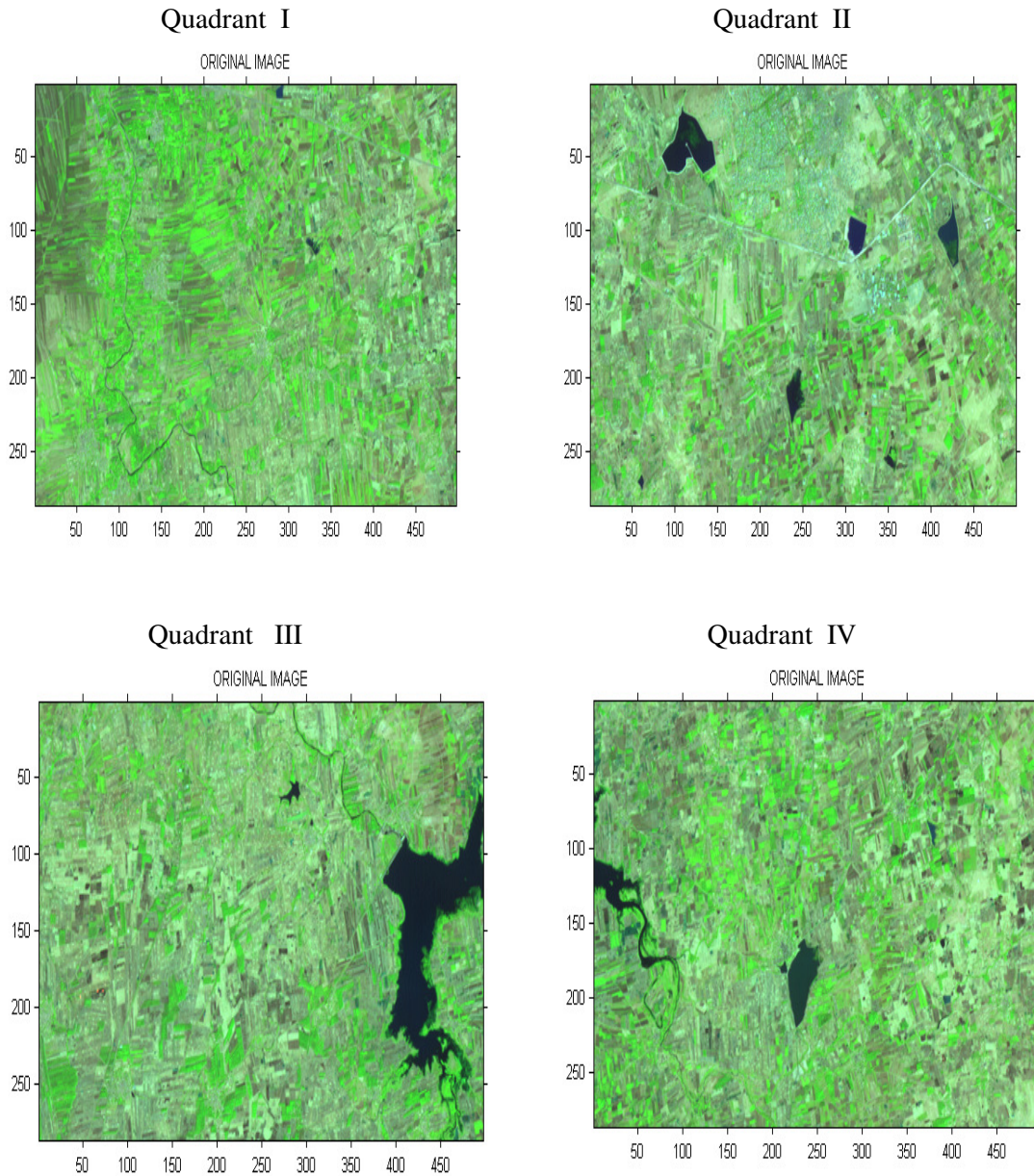
Quadrant II

Quadrant   III

Quadrant IV

Figure 2. The original image divided into four Quadrants(Quadrant I,II,III,IV).

The two sub-images are subjected to image enhancement techniques by contrast stretching. The enhanced image is shown in Fig(3a) and   3(b) The natural features are more clearly evident from the figures and each image reflect different color contents for the natural resources identified in the original image.

Further it can be noted that the   image in figure 3 (a) indicates the presence of only land and Vegetation features and   absence of water bodies. But Fig 3 (b) strongly indicates in addition to Vegetation and land, the presence of water bodies (Dark Blue) and houses (Light Blue) which are new features in the figure.

Truecolor Composite after Decorrelation Stretch

Figure 2 (a) :1st quadrant of original image
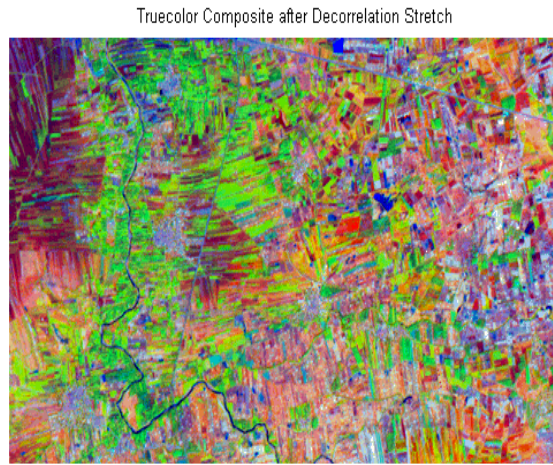
Figure 2 (a):Contrast stretch enhanced image of 2(a)

Truecolor Composite after Decorrelation Stretch
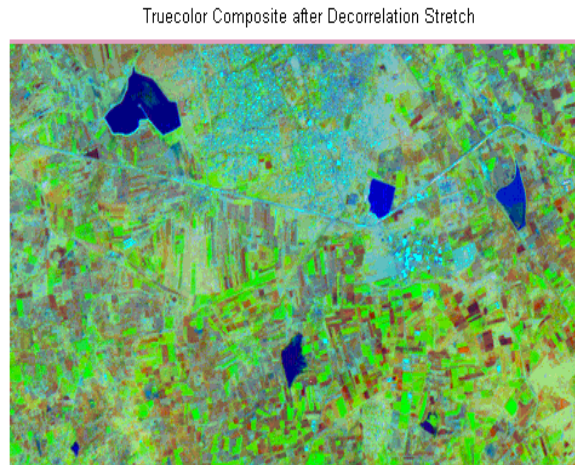
Figure 2 (b) :2nd quadrant of original image

Figure 2 (b):Contrast stretch enhanced image of 2(b)

In order to estimate the extent of features present in the images the L*a*b color diagram has been derived and plotted for the two sub images. Figure 4 (a) shows for Quadrant I the color spectral distribution. It is evident in the figure that there are two important features that are predominant in this region are the Land (Red) and Water Bodies (Black). The other features representing Vegetation (Green), Housing (Cyan) are present to a lesser extent in the region

Figure 4 (b) shows L*a*b color diagram for Quadrant II.. It is evident from the figure that there are two dominant features present in the image are Vegetation (Green) and Housing

covered by Vegetation (Cyan). The features such as Land (Red) and Water bodies (Black) are present to a lesser extent in the region.
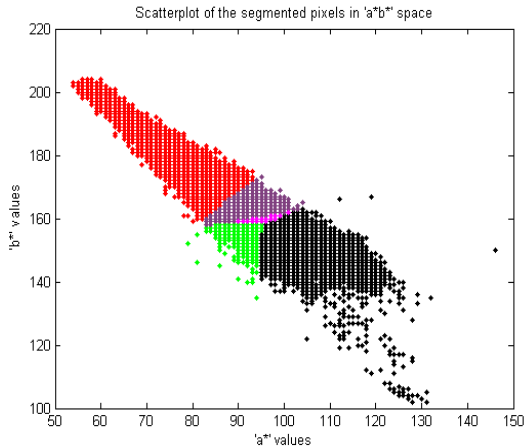

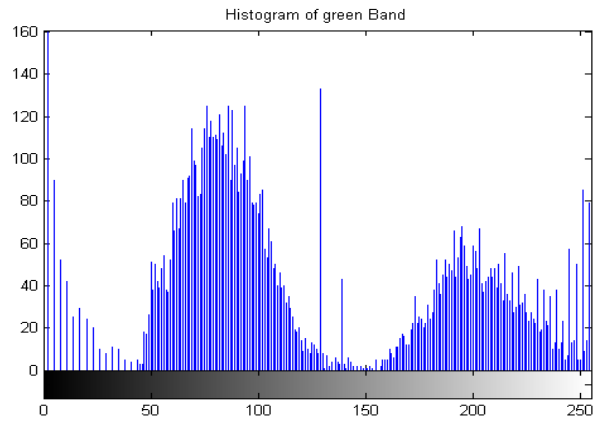Figure 3 (a):L*a*b color diagram for Quadrant I

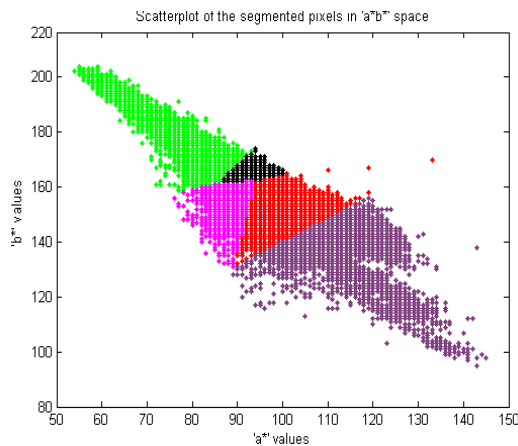Figure 4 (a)Histogram plot for image of 2(a)
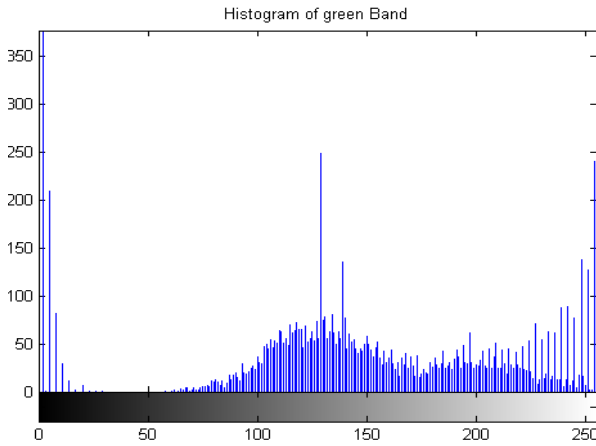

Figure 3 (b):L*a*b color diagram for Quadrant II

Figure 4 (a)Histogram plot for image of 2(b)

In order to emphasize the observations, the distribution of RGB colors in the form of a histogram is plotted in figure 5 (a) and 5(b). It is very clearly seen from the figure 4(a) that the there is a wide peak corresponding dark regions centered around 0 pixel and strong peaks around pixel 75 corresponding to vegetation and peak at pixel 200 corresponding to soil distribution. Where as in Figure 4 (b) shows strong peaks corresponding to water bodies centered on pixel 0 and a wide flat distribution pixel 200 - 250 corresponding to housing distribution. The results of the analysis color diagrams for the satellite Rural image clearly shows that in a Rural image it is possible to identify different features which are clearly distinguishable from one another using L*a*b color diagrams. The proper identification and estimation of different features present in different regions of the Rural image will be important for estimating the extent of natural resources available in different parts of the Rural image. The

relative estimation of natural resources available in a Rural region and very useful for Resource development and   Management in the Rural regions.

IMAGE (RURAL)	RED (LAND)	GRN (VEG)	BLK (WAT)	CYN (HOU+VEG)
Qradrant I	Max	Min	Max	Min
Qradrant II	Min	Max	Min	Max
Qradrant III	Max	Min	Max	Min
Qradrant IV	Min	Max	Max	Max

Table 2: Shows the extent of natural resources extracted in the Rural image LAN: Land, VEG: Vegetation ,WAT: Water Bodies and HOU+VEG: Housing covered by Vegetation

Conclusions:

From the results and discussions presented it can be concluded that

The natural colors in the satellite imageries of the rural scene can be subjected to color based image processing for identifying natural resources. The Image enhancement techniques are effective in identification of the major features such as Vegetation, Water bodies, Land and Housing present in the satellite imageries.   The color spectral analysis   carried out by using L*a*b system bring out distinct features present in the image and   useful for estimating the extent of the   features present in a rural image. Histogram techniques complement the identification of color features   and assist in   identifying and estimating the features   present in a rural image.

Acknowledgement:

References:

[1] Beng Chin Ooi, Kian-Lee Tan, Tat Seng Chua, Wynne Hsu   *"Fast image retrieval using color-spatial information"* Department of Information systems & Computer Science, National University of Singapore, Lower Kent Ridege Singapore -119260-1998
[2] Cyrus Shahabi and Yi-Shin Chen " *Soft Query in Image Retrieval Systems"* Integrated Media systems and Computer Science Department University of Southern California. Los Angeles, CA90089-2564- 2000

[3] Guodong Guo, Stan Z.Li, and Kap Luk Chan **"Learning Similarity for Texture Image Retrieval"** *School of IEEE, Nanyang Technological University , Nanyang Avenue, Singapur- 2001*
[4] Aditya Vailaya, Mario A.T Figueiredo, Anil.K Jain, and Hong-Jiang **"Image Classification for Content Based Indexing"** *Publised in IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL-10,NO.1, 2001*
[5] Napoleon H.Reyes, and Elmer P.Dadios **"Dynamic Color Object Recognition Using Fuzzy Logic"** *Published in the Journal of Manufacturing Engineering and Management in 2003*
[6] Eduardo Hart, Sung-Hyuk Cha, Charled Tappert   (2004) **"Interactive Flag Identification Using a Fuzzy-Neural Technique"** *Published in the proceeding of Student/Faculty Research Day, CSIS University 2004.*
[7] Shan Li, M.C.Lee   "Rotation and Scale Invariant Color Image Retrieval Using Fuzzy Clustering" Published in Computer Science Journal , Chinese University of Hong Kong 2004

 [8] Arti Khaparde, B.L Deekshatulu, M.Madhavilatha, Zakira Farheen, Sandhya kumari.V **"Content Based Image Retrieval Using Independent Component Analysis"** *Published in IJCSNS International Journal of Computer Science and Network Security. Vol.8 No.4, 2008*
[9]Yo-Ping Huang , Tsun-Wei Chang and and Li-Jen Kao (2008) **"Using Fuzzy SOM Stategy for Satellite Image Retrieval and Image Retrieval and Information Mining "** *Department of Computer Science and Engineerig, Tatung University, Taipei, Taiwan-2008.*

About the Author1: Mamatha Y.N *was born in Mysore on 25/05/1975 , Karnataka. She received B.E (Electraonics & Communication) degree in 2000 from Mysore University and M.Tech(Digital Communication) degree in 2004 from VTU, Belgaum, and Perusing PhD in VTU. Presently working as a lecturer in Dept of Telecommunication, RVCE. Bangalore, Karnataka India. Research interests include Image process, Wireless Networking*

About the Author2: Dr.A.G..Ananth *was born on Bangalore on 3 November 1947 at Bangalore Karnataka,. India .He received M.Sc degree in 1969 in Nuclear Physics from Bangalore University. In 1975 Physical Research Laboratory Ahmedabad awarded him Ph.D degree in Space physics . He served as Deputy Director in ISRO, currently working as   Professor in Telecommunication Department of R.V College of Engineering, Bangalore. His research interests include Space physics, Biomedical signal processing, Image processing and MIMO systems.*