Exploring a New Direction in Colour and Texture based Satellite Image Search and Retrieval System

Amruta Buch, Swati Jain, and S.N. Pradhan

***Abstract*-- Content based Image Retrieval systems (CBIR) have become a reliable tool for many image database applications. Today, the need for reliable, automated satellite image classification and browsing systems is more than ever before. Everyday there is a massive amount of remotely sensed data being collected and sent by terrestrial satellites for analysis. The use of automated tools for this analysis has become imperative due to the large volumes required to be processed on a daily basis. Our goal is to develop a prototype of Content based Image Retrieval Systems for Satellite Images, wherein the user enters query in form of a sample image and the relevant similar images based on image content are displayed as an output. The Work presents a novel approach of satellite image search and retrieval system based on colour and texture. Texture information is obtained using co-occurrence matrix of the gray scaled images. The properties like correlation and homogeneity are good parameters to measure the textural property; these are evaluated at proper offset and angle. Colour is inevitable feature that dominates the human perception most. The Hue Saturation and Value (HSV) colour space is quite similar to the way in which humans perceive colour. Making this as the foundation, the work presented here uses the combination of the HSV colour and texture feature. Results have shown that the retrieval based on combination of texture and colour based features match better with the human perception.**

***Index Terms*-- Content based Image Retrieval, Gray level co-occurrence matrix, Hue Saturation and Value colour model, K-means algorithm.**

I. Introduction

Today, the need for consistent, automated satellite image classification and browsing systems is more than ever before. Everyday there is a massive amount of remotely sensed data being collected, processed and sent by terrestrial satellites for analysis. The metadata based search and retrieval systems for accessing remote sensing data and images, in particular, allow only queries by geographical coordinates, date and time of acquisition, sensor type, image description, image size and categories of image. Satellite image metadata is the fundamental for the usage and understanding of the image it describes. The shortcoming with metadata based search and retrieval are [3], [5]: Considerable amount of time and effort are needed to manually annotate each individual image. Imprecision is associated with the subjective human perception of the contents being annotated. At times, textual information is not enough to describe the complete image object. It requires the image content analysis to understand the intricate details of the object. Moreover, the access to image archives becomes more difficult when the data quantity acquired by a new generation of high-resolution satellite sensors is enormous. As a consequence, new technologies are needed to easily and selectively access the information content of image archives and finally to increase the actual exploitation of satellite observations. This growing need for Content based Image Retrieval System for Satellite Images motivated us to introduce an efficient and effective method for searching and retrieving satellite images based on their features. The system facilitates the interaction between a user and an image database by automatic analysis of the image content such as texture and HSV colour [3]. It typically relies on feature extraction and similarity measure algorithms to match.

II. Texture Feature Extraction

Texture refers to the arrangement of the basic constituents of a material [9]. In a digital image, texture is depicted by spatial interrelationships between, spatial arrangements of the image pixels. Adjectives often used to describe texture are smooth (uniform, homogeneous), intermediate, and rough (coarse, heterogeneous). Colour and texture both can be used to determine percentage of area occupied by object like water body, vegetation. Different approaches like statistical, geometrical, model based and signal processing can be applied to evaluate texture feature [9]. One of the defining qualities of texture is the spatial distribution of gray values. The use of statistical features is therefore one of the early methods proposed in the machine vision literature. Statistical technique outshines the remaining techniques in terms of time and space complexity.

Co-occurrence matrix is the statistical technique used in this work to evaluate the texture feature [2], [8]. Spatial gray level co-occurrence estimates image properties related to second-order statistics. Second order statistics are defined as the likelihood of observing a pair of gray values occurring at the end-points of a dipole of random length placed in the image at a random location and orientation. These are properties of pairs of pixel values.

The co-occurrence matrix features suffer from a number of difficulties. There is no well established method of selecting the displacement vector and computing co-occurrence matrices for different values. For a given image, a large number of features can be computed from the co-occurrence matrix. This means that some sort of feature selection method must be used to select the most relevant features. Correlation, Homogeneity, Contrast, Energy and Entropy are the features which can be computed from the co-occurrence matrix. M.A.Shaban and Onkar Dikshit [6] have proposed an optimum window size for all the features. Results indicated that for Contrast and Entropy, the optimum window size is 7 and for Correlation it is 9. Any further increase in window size more than this optimum value not only adds to computational burden without any significant improvement in accuracy but also provides inaccurate textural properties due to the window encompassing more than one class. In the following, we will use below notation to denote N x N image with gray levels,

$$\{I(x,y), 0 \le x \le N-1, 0 \le y \le N-1\} \quad (1)$$

The G x G gray level co-occurrence matrix P_d for a displacement vector d= (dx, dy) is defined as follows. The entry (i, j) of P_d is the number of occurrences of the pair of gray levels i and j which are 'd' distance apart [3]. Formally, it is given as,

$$P_d(i,j) = \left|\{((r,s),(t,v)) : I(r,s) = i, I(t,v) = j\}\right| \quad (2)$$

TABLE I
TEXTURE FEATURES EVALUATED FROM GRAY LEVEL CO-OCCURRENCE MATRIX

Texture Feature	Formula
Energy	$\sum_i \sum_j P_d^2(i,j)$
Entropy	$-\sum_i \sum_j P_d(i,j) \log P_d(i,j)$
Contrast	$\sum_i \sum_j (i-j)^2 P_d(i,j)$
Homogeneity	$\sum_i \sum_j \frac{P_d(i,j)}{1+\|i-j\|}$
Correlation	$\frac{\sum_i \sum_j (i-\mu_x)(j-\mu_y) P_d(i,j)}{\sigma_x \sigma_y}$

In this work, we have finalized on using Gray level co-occurrence matrix which is the traditional method with low computational cost to be used for texture determination. The limitation of most of the textural models is that they are for gray scale images and not for colour images [3]. In this work, texture features are evaluated at different offset and angles from the gray level co-occurrence matrix. Thereafter, similar images were retrieved and visually assessed on basis of these five individual features and it was observed that Correlation and Homogeneity are adequate enough to distinguish the texture of satellite images. Calculation of texture feature from gray level co-occurrence matrix will result in feature vector comprising 39 attributes each per image.

$$txt\_feat = (image\_name, corr\_1, corr\_2, \ldots, corr\_16, avg\_corr, med\_corr, stddev\_corr, homo\_1, homo\_2, \ldots, homo\_16, avg\_homo, med\_homo, stddev\_homo)$$

Here *txt_feat* feature vector comprises of name of image (image_name), correlation and homogeneity values calculated at various offset and angles (corr_i and homo_j , where $1 \le i,j \le 16$), mean (avg_corr), median (med_corr), standard deviation value of the above correlation values (stddev_corr) and mean (avg_homo), median (med_homo) and standard deviation value of the above calculated homogeneity values (stddev_homo). Average, median and standard deviation values of correlation and homogeneity texture features are used to nullify the noise component of an image. The Euclidean distance measure is used for evaluating similarity measure in this texture based satellite image search and retrieval procedure where images with similar texture have a smaller distance value when Euclidian distance between the query and candidate target image is calculated.

III. COLOR FEATURE EXTRACTION

HSV (Hue, Saturation and Value) defines a type of colour space. The HSV colour space has three components: hue, saturation and value. In HSV, hue represents pure colour. In this model, hue is an angle from 0 degrees to 360 degrees. Saturation indicates the amount of white added to pure colour. It ranges from 0 to 1. When the value is 0, the colour is white and when the value is 1, the colour is a primary colour. Value is the brightness of the colour and varies with colour saturation. It ranges from 0 to 100 %. When the value is 0 the colour space will be totally black. With the increase in the value, the colour space brightens up and shows various colours [5], [10]. The HSV colour space is quite similar to the way in which humans perceive colour which is not always the case with RGB or CMYK.

Below mentioned steps explain the procedure to extract and store the colour features. It also explains the distance measure used for image retrieval procedure.

Steps for Feature Generation:

- Select the folder consisting of satellite images.
- Convert RGB coloured image to HSV coloured image.
- Perform Non Uniform Colour Quantization. (Divide Hue into 36 intervals using K-means Algorithm).
- Calculate appearance probability (q_k) of each hue value in each I_k interval.
- Calculate the mean value of Hue, Saturation and Lightness in I_k interval using below mentioned formula.

$$\bar{h}_k = \sum_{h_i \in I_k} \frac{h_i}{n_i} \quad (3)$$

$$\bar{s}_k = \sum_{h_i \in I_k} \frac{s_i}{n_i} \quad (4)$$

$$\bar{l}_k = \sum_{h_i \in I_k} \frac{l_i}{n_i} \quad (5)$$

where n_i is the number of unique hue values in I_k interval.

- Store the feature vector (q_k), (h_k), (s_k) and (l_k) in separate tables of the database.

The **distance measure** used for this colour based image retrieval technique is as follows:

Let A and B be target and query image respectively each consisting 36 attributes in feature vector.

$$|A - B| = \sum_{k=1}^{36} |b_k - a_k| \quad (6)$$

$$|b_i - a_j| = (0.001 * |q_i^b - q_j^a| + 1) * \sqrt{(h_i^b - h_j^a)^2 + (s_i^b - s_j^a)^2 + (l_i^b - l_j^a)^2} \quad (7)$$

Appearance probability and average hue component is given more weightage, as hue is an important factor to identify the dominant colours. Hue describes the pure basic colours. The basic colours present in satellite images are green, blue, brown and yellow.

Texture and colour feature independently give satisfactory results. But in certain cases, when texture of two visually different query and target images are similar, then image retrieval using colour feature becomes inevitable. However, it is possible that the similar colour composition of two visually different images may comprise of different texture features. This can result in display of irrelevant images in the top similar outcomes. The best method to overcome the above limitation is to combine colour and texture features and then proceed with the image retrieval. In order to boost the image retrieval efficiency we have performed image retrieval based on combination of texture and HSV colour features.

IV. Combined Texture And Color Based Image Retrieval

Most of the satellite images retrieval system use texture features but color features are very important and can add to very vital information to the textural information hence combination will have rich information about the content of the images. Moreover, it is observed from previous topics, that texture and colour feature independently have some limitations. The efficiency of retrieval technique with combined texture and colour feature is evaluated after visually assessing the retrieved results for a set of 22 test query images.

Steps for Feature extraction (Offline procedure):

- Select folder consisting of satellite images.
- Convert RGB coloured image into a gray scaled image.
- Extract texture features of each gray scaled image and store it in the database.
- Convert RGB coloured image into a HSV coloured image.
- Extract colour features and store it in the database.

Image Retrieval procedure:

- Select Query image.
- Convert RGB coloured query image into a gray scaled image.
- Extract texture features.
- Evaluate and store the similarity measure of query image with respect to all other target images based on texture.
- Obtain the HSV colour feature from the query image.
- Evaluate and store the distance measure (as specified in colour section) of the query image's colour feature from each and every target image's colour feature.
- Normalize colour and texture feature in order to avoid dominance of one feature over another.
- Find out the combined similarity measure by summation of normalized texture and colour distance measure.
- Rank all the images in ascending order of their similarity measure.
- Find out the top 'n' resultant images in order of their similarity and display them in sequence.

V. Results And Analysis

In the work presented here, Matlab 7 is used for algorithm development and implementation. Postgresql 9.0.2-1; windows open source database is used for managing various databases. Dataset consists of Google Map images, Quickbird and Ikonos satellite images. Total number of images in the archive is 216. 22 Test images are considered for testing purpose and are stored in a separate folder. This work also includes implementation of image retrieval using RGB color [1] combined with texture feature based retrieval procedure which is later used as benchmark for comparing the proposed technique.

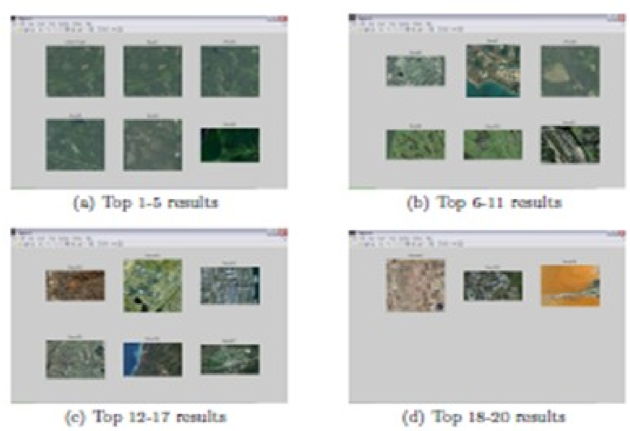

Fig 1. Top 20 images retrieved with respect to query image comprising vegetation using combination of texture and HSV colour feature.

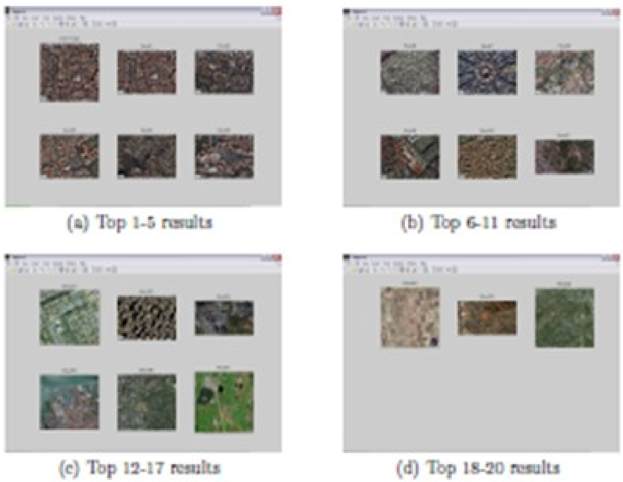

Fig 2. Top 20 images retrieved with respect to query image comprising city using combination of texture and HSV colour feature.

In subfigure (a) of above results, 1st image in topmost left corner indicates the query image. The images followed by the query image are the top 20 results retrieved from the image database. Here query images comprises of variety of images such as vegetation, land, city and waterbody. For Query image 1, a vegetation image, out of 20 resultant images, 13 images are highly similar to query image, 3 images i.e. image number 11,13 and 19 are moderately similar (as they comprise part of vegetation and part of city image) and rest of images seem to be irrelevant when visually verified. They are listed in top 20 results because of their overlapping colour and texture feature. For Query Image 2, a city image, except for resultant image 17 and 18 rest all images are highly similar to the query image.

TABLE II
RETRIEVED IMAGE RELEVANCY OF THE IMAGE RETRIEVAL TECHNIQUE USING COMBINATION OF TEXTURE AND COLOUR FEATURE

Technique	Retrieved Image Relevancy (%)
Combined texture and RGB colour [1]	70.29
Combined texture and HSV colour	72

TABLE III
TIME TAKEN BY THE IMAGE RETRIEVAL TECHNIQUE USING COMBINATION OF TEXTURE AND COLOUR FEATURE

Technique	Time taken (sec)
Combined texture and RGB colour [1]	192.8327
Combined texture and HSV colour	22.1841

Here, texture combined with RGB or HSV colour feature is giving almost similar results which are much better than considering texture or colour feature independently. Image retrieval technique based on combined texture and HSV colour feature is more preferable than the other combined technique because of the following reasons:

- We can observe from the above table that combined texture and HSV colour based technique takes less time to retrieve similar images given a query image compared to the other technique because of the less complex computations involved in the retrieval process.

The colours used in HSV can be clearly defined by human perception, which is not always the case with RGB or CMYK.

VI. CONCLUSION

Today, the need for consistent, automated satellite image classification and browsing systems is more than ever before. Everyday there is a massive amount of remotely sensed data being collected, processed and sent by terrestrial he experiments were conducted with the database of visible spectrum satellite images. Image Retrieval based on texture or colour individually gives satisfactory results whereas combined texture and colour based Image Retrieval shows comparatively better results.

Out of top 20 results, the percentage of relevant images obtained only using texture feature was 69.37% and only using HSV colour feature was 67.49%. Texture feature falls short to distinguish images having similar texture composition like satellite images comprising land and images comprising waterbody. Both the images have smooth texture but different colour distribution. This resulted in retrieval of land images given a waterbody query image. In case of Colour feature, two entirely different textured images may have similar HSV colour distribution resulting in retrieval of higher number of incorrect images in the top 'n' results. Thus a combination of texture and colour feature can help in overcoming the limitation of the individual features and achieve improved relevant image efficiency. Percentage of relevant images retrieved and visually verified in top 20 target images using combined HSV colour and Texture is 72 %. The limitation of this work is the candidate search space for searching and retrieving similar images is the entire dataset. This may affect the search time. Hence the proposed technique can be extended in a way where concepts of clustering can be inculcated. This will improve the search time and percentage of relevant images by reducing the search space.

VII. REFERENCES

[1] Chin Chen Chang, Wen Chuan We, and Yu Chen Hu. "Content based color image retrieval system using colour difference features", Second International Conference on Future Generation Communication and Networking Symposia, 2008.
[2] Mohamed Gebril, Eric Kihn, and Mikhail Zhizhin. "Structural indexing of satellite images using texture feature extraction for retrieval", IEEE, 2010.
[3] Swati Jain and S.N.Pradhan, "Content based image retrieval system using colour, shape and texture", Dissertation thesis, Nirma Univesity, 2007.
[4] Giovanni B. Marchisio, Wen Hao Li, Michael Sannella, and Jill Goldschmeider. Geobrowse: An integrated environment for satellite image retrieval and mining. IEEE, 1998.
[5] Oge Marques and Borko Furht, "Content based Image and Video Retrieval", Kluwer,Academic Publishers, 2002.
[6] M.A.Shaban and Onkar Dikshit, "Textural classification of high resolution digital satellite imagery", IEEE, 1998.

[7] DU Peijun, CHEN Yunhao, TANG Hong, and FANG Tao, “Study on content based remote sensing image retrieval”, IEEE, 2005.
[8] Erika Podest and Sasan Saatchi, “Application of texture to jers-1 sar imagery for tropical forestland cover classification”, IEEE, 1999.
[9] Mihran Tuceryan and Anil K. Jain, “Handbook of Pattern Recognition and Computer Vision”, World Scientific Publishing Co., January 1998.
[10] Chen Xianqiao, Yan Xinping, and Chu Xiumin, “Research on image content retrieval with colour features”, Second International Conference on Computational Intelligence and Natural Computing 2011.