Image Mining using Content Based Image Retrieval System

Rajshree S. Dubey
M.tech (CSE)
BIST
Bhopal, India

Niket Bhargava
Astt.Prof. CSE
BIST
Bhopal, India

Rajnish Choubey
Astt.Prof. CSE
TCT
Bhopal, India

Abstract: The image depends on the Human perception and is also based on the Machine Vision System. The Image Retrieval is based on the color Histogram, texture. The perception of the Human System of Image is based on the Human Neurons which hold the 10^{12} of Information; the Human brain continuously learns with the sensory organs like eye which transmits the Image to the brain which interprets the Image. The research challenge is that how the brain processes the information in the semantic manner is hot area of research till date. Hence we can say that a single set of Algorithm cannot define the complete set of operations on the field of Image Mining.

This Paper examines the State-of – art technology-Image mining techniques which is based on the Color Histogram, texture of him Image. The query Image is taken then the Color Histogram and Texture is taken and based on this the resultant Image is output.

Keywords: Image retrieval, pattern recognition, Image Mining, Color Histogram.

I. INTRODUCTION

Since the storage technology is being cheaper and cheaper the storage capacity is also increased, hence and so forth the storage of Image Databases is also increased. The analysis of the Images on these databases is very useful and gives result which can be applied to different applications. The growth of Image Mining is on the base of Image Data relationship, Patterns and implicit knowledge of Image which is not defined on the Image dataset.

This domain is called the Image Mining. It is the interdisciplinary field which needs the effort from all the sectors like Image Processing, Image Retrieval, Machine Learning, and Artificial Intelligence and last but not the least the Data Mining. This Paper examines the State-of – art technology, Image mining techniques and discusses the related features of Image Retrieval Systems.

The largest storage is the World Wide Web which can be termed as the Largest Image repository which encompasses different types of data and also holds the different types of Multimedia data sets. Now a day the Information is communicated is in the form of Images, which can also be termed as the age of Visual Communications. Image databases are typically very large in size. We have witnessed an exponential growth in the generation and storage of digital images in different forms, because of the advent of electronic sensors (like CMOS or CCD) and image capture devices such as digital cameras, camcorders, scanners, etc.

There is an advancement in the text based search Engines but the advancement on the Image based search engines is negligible. To make the data mining technology successful, it is very important to develop search engines in other multimedia data types, especially for image data types. Mining of data in the imagery domain is a challenge. Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images. It is more than just an extension of data mining to the image domain. Image mining is an interdisciplinary endeavor that draws upon expertise in computer vision, Pattern recognition [5], image processing, image retrieval, data mining, machine learning, database, artificial intelligence, and possibly compression.

Most of the activities in mining image data have been in the search and retrieval of images based on the analysis of similarity of a query image or its feature(s) with the entries in the image database. The image retrieval systems can be broadly categorized into two categories based on the type of searches, using either description of an image or its visual content.

II.CONTENT BASED IMAGE RETRIEVAL

The Content Based Image Retrieval System (CBIR) [6] is based on the visual content of the Image. Desirable features can extracted based on the visual content of the Image. The development of the Image Mining technique is based on the Content Based Image Retrieval system. Colour, texture, pattern, image topology, shape of objects and their layouts and locations within the image, etc are the basis of the Visual Content of the Image and they indexed.

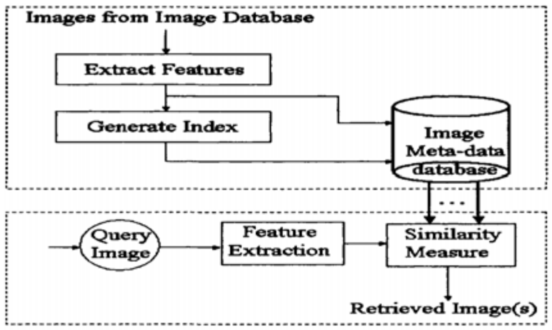

Fig- (1). Content Based Image Retrieval System Architecture.

The Several features that are used in the Image Retrieval system are the following. The popular amongst them are colour of the Image, What is the Texture, The Contours i.e. (shape), image topology, on which layout or the Background is Image depicted, region of interest etc.

2.1Colour Features

One of the features that are used is the colour of the Image, since it has very strong co-relation with the underlying objects of the Image. The colour of an Image is very robust to background, orientation, scaling and texture of the Image. In the terms of Digital Imagery the Colour Pixels is the Combination of (RED, BLUE and GREEN). It is well known fact that any colour combination of these three colours can produce any colour. The HSV colour space is depicted below, because that is widely used in the CBIR community.

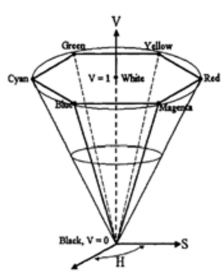

Fig- 2. Perceptual representation of HSV colour space.

The Red, Blue and Green pixels can be considered as r, g, b and holds in the range of closed intervals [0,1] and these normalized values r, g, and *b* are used to transform into the HSV space. Transformation of RGB to HSV colour space can be accomplished by the following set of equations [1]:

$$V = \max(r, b, g)$$

2.1.1Colour histogram

The overall distribution of colours can be done with the help of Colour Histogram which is the basis of characterization of the Image. Color histogram is a representation of the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges, that span the image's color space, the set of all possible colors. The color histogram can be built for any kind of color space, although the term is more often used for three-dimensional spaces like RGB or HSV

2.1.2Colour coherence vector

The colour coherence vector [7]-based approach was designed to accommodate the information of spatial colour into the colour histogram. Here we can classify each pixel in an image, based on whether it belongs to a large uniform region or not. For example, we can consider a region to be uniformly coloured if it consists of the same colour and the area of the region is above a certain threshold (say, 2%) of the whole image area. We refer to the pixels in these regions as *coherent* pixels.

2.2Texture Features

Texture is a very interesting image feature that has been used for characterization of images, with application in content-based image retrieval. a major characteristic of texture is the *repetition of a pattern or patterns over a region* in an image. The elements of patterns are sometimes called *textons.* The size, shape, colour, and orientation of the textures can vary over the region. The difference between two textures can be in the degree of variation of the textures. It can also be due to spatial statistical distribution of the textons in the image. Texture is an innate property of virtually all surfaces, such as bricks, fabrics, woods, papers, carpets, clouds, trees, lands, skin, etc. It contains important information regarding underlying structural arrangement of the surfaces in an image. When a small area in an image has wide variation of discrete tonal features, the dominant property of that area is *texture.*

In this method the co-occurrence matrix is constructed based on the orientation and distance between image pixels. Meaningful statistics are extracted from this co-occurrence matrix, as the representation of texture. Since basic texture patterns are governed by periodic occurrence of certain Gray levels, co-occurrence of Gray levels at predefined relative positions can be a reasonable measure of the presence of texture and periodicity of the patterns. Using this co-occurrence matrix, the texture features metrics are computed as follows [9].

$$\text{Entropy} = -\sum_i \sum_j C(i,j)\log C(i,j),$$

$$\text{Energy} = \sum_i \sum_j C^2(i,j),$$

$$\text{Contrast} = \sum_i \sum_j (i-j)^2 C(i,j),$$

$$\text{Homogeneity} = \sum_i \sum_j \frac{C(i,j)}{1+|i-j|}$$

Practically, the co-occurrence matrix C (i, j) is computed for several values of displacement dx, dy, and the one which maximizes a statistical measure is used.

III. ALGORITHM

The algorithm for content based image retrieval system using multiple Feature given as below[

Step 1: Preprocessing

Input various object Images O_i Create B_m Block Matrix. Calculate Mean µ of Block Matrices. Concatenate all Block Matrices obtain.

Step 2: Feature Extraction

Convert $B^m_{4\times4}$ Block matrices "r & s" RGB from space to HSV from space where r and s represent the average values of vector. Extract feature vector Vj from HSV space. The major statistical data that are extracted are histogram mean, standard deviation, and median for each color channel i.e. Red, Green, and Blue. So totally 3 × 3 = 9 features per segment are obtained.

Extract texture feature such as entropy, energy, contrast, and homogeneity from the co-occurrence matrix of the gray levels of the image. Texture are taken into account which is the combined values of entropy , (1), energy(2), contrast(2), co-relation of pixel values(2), and homogeneity(2) which combines to the total value of (9).

Step 3: Similarity Measurement

Calculate Euclidean Distance then get Euclidean (D)

$$Deucld(r,s) = \sqrt{\sum_{i=1}^{N}(r_i - s_i)^2}$$

(Where r and s represent the average values of feature vectors respectively)

While (Ti==Oi) // where Ti is the test (query) Image & Oi is the object images.
Repeat above procedure for n object images.
Now we have 'N' object image and its Euclidean distance matrices.

IV. EXPERIMENTAL SETUP AND RESULTS

The result analysis of the histogram and texture based matching technique is given below. The graph is also shown for the texture and histogram

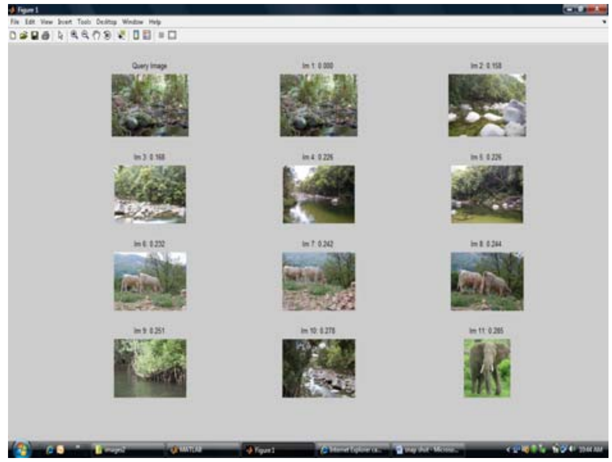

Fig-1.1 Histogram Based Matching

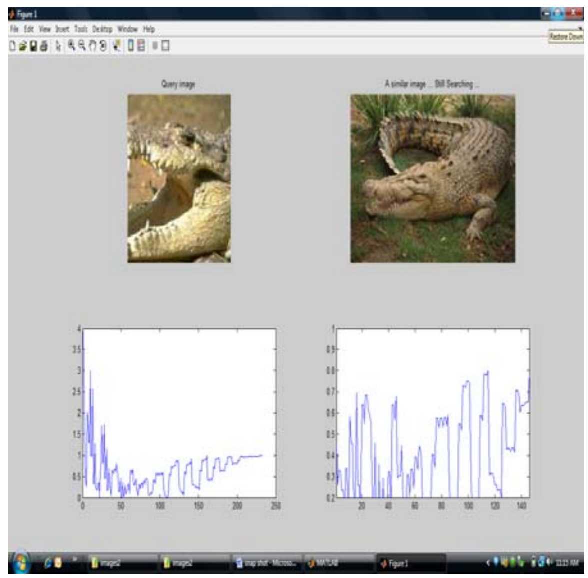

Fig-1.2 Histogram Based Matching

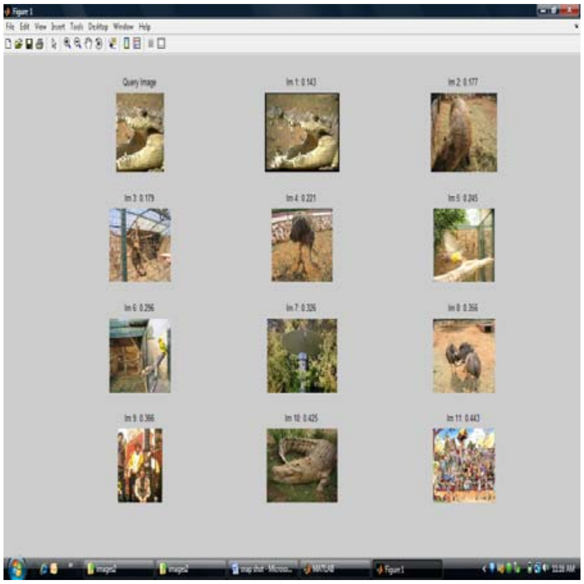

Fig-1.3 Histogram Based Matching

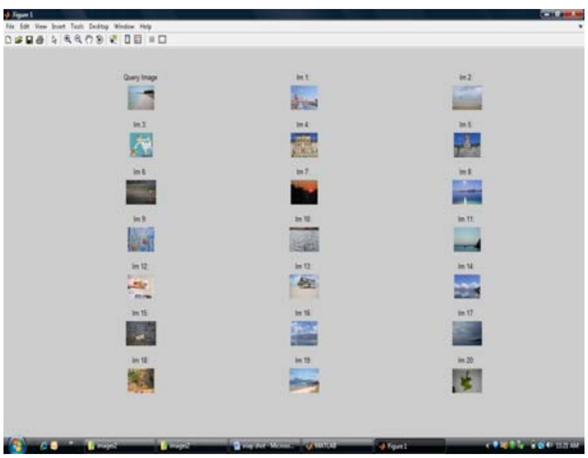

Fig-1.4 Texture Based Matching

Test Image	Match Image	Unmatched Image	Percentage of Matching
img1	8	3	73%
img2	7	4	64%
img3	5	6	46%
img4	7	4	64%
img5	6	5	54%
img6	6	5	54%
img7	6	5	54%
img8	6	5	54%
img9	6	5	54%
img10	7	4	64%

Table: For Query image

V. CONCLUSION AND FUTURE WORK

In this Paper, a histogram-based search methods and color texture methods were investigated in two different color spaces, RGB and HSV. Histogram search characterizes an image by its color distribution, representation is that information about object location, shape, and texture is included. Thus this dissertation showed that images retrieved by using the global color histogram may not be semantically related even though they share similar color distribution in some results. An image retrieval demo system was built to make it easy to test the retrieval performance and to expedite further algorithm investigation. In general, histogram-based retrievals in HSV color space showed better performance than in RGB color space since it showed higher precision values for the same recall values. For graphic images, all three histogram-based image retrieval methods in RGB color space showed better performance than those in HSV color space. And it is found that the histogram Euclidean distance and histogram intersection distance in HSV color space are most useful among histogram distance measures in the average sense. In a viewpoint of computation time, using HSV color space is faster than using RGB color space and using Intersection method is faster than using Euclidean or Quadratic method. Thus, in conclusion, the Histogram Intersection-based image retrieval in HSV color space is most desirable among the retrieval methods mentioned in considering both computation time and retrieval effectiveness

REFERENCE

[1] S. K. Saha, A. K. Das, and B. Chanda, "Content-based image retrieval: An experiment with shape and texture features," *Technical Report No. BEC09-04-2002*, Department of Computer Science and Technology, Bengal Engineering College (Deemed University), India, May 2003.

[2] Bishnu, P. Bhunre, B. B. Bhattacharya, M. K. Kundu, C. A. Murthy, and T. Acharya, "Content based image retrieval: Related issues with Euler Vector," in *Proceedings of the IEEE Int. Conf. on Image Processing* (Rochester, NY), pp. 11:585-11:588, September, 2002.

[3] A. Bishnu, B. B. Bhattacharya, M. K. Kundu, C. A. Murthy, and T. Acharya, "Euler vector: A combinatorial signature of gray-tone images," in *Proceedings of the IEEE Int. Conf. on Information Technology: Coding & Computing* (Las Vegas), pp. 121-126, April, 2002

[4] Q. Iqbal and J. K. Aggarwal, "Combining structure, color and texture for image retrieval: A performance evaluation," in *Proceedings of International Conference on Pattern Recognition* (Quebec City, Canada), 2002

[5] D. Luo, *Pattern Recognition and Image Processing.* New York:Horwood Publishing, 1998.

[6] H. R. Kang, *Color technology for electronic imaging devices.* Bellingham, Washington, USA: SPIE Optical Engineering Press, 1997

[7] G. Pass and R. Zabith, "Histogram refinement for content-based image retrieval," in *Proceedings of of IEEE Workshop and Applications of ComputerVision*, pp. 96-102, 1996

[8] M. Strieker and M. Orengo, "Similarity of color images," in *Proceedings of SPIE Storage and Retrieval for Image and Video Databases III*, vol. 2185 (San Jose, CA), pp. 381-392, February 1995

[9] R. Jain, R. Kastuni, and B. G. Schunck, *Machine Vision.* New York: McGraw-Hill, 1995.

[10] R. C. Gonzalez and R. E. Woods, *Digital Image Processing.* Reading, MA: Addison-Wesley, 1993.