2012 International Conference on Computing Sciences

Semantic Image Retrieval by Combining Color, Texture and Shape Features

1st Nishant Singh, 2nd Shiv Ram Dubey, 3rd Pushkar Dixit, 4th Jay Prakash Gupta
Department of Computer Engineering and Applications
GLA University, Mathura, Uttar Pradesh, India
[1]singhnishant88@gmail.com, [2]shivram1987@gmail.com, [3]pushkardixit1987@gmail.com, [4]jayprakash619@gmail.com

***Abstract*— The volume of digital images generated and uploaded on the internet every day by the scientific, medical, educational, industrial and other communities are very large. The problem of retrieving the desired images from huge collections is a major problem. The user queries are becoming very specific and traditional text-based methods cannot efficiently handle them. The subjectivity of human perception and the rich contents of the images further aggravate the problem. To overcome this problem, a new query-by-example technique using multiple color, texture and shape features is proposed and evaluated in this paper. The experimental results suggest that our proposed technique is efficient and retrieves semantically more similar images.**

Keywords-CBIR; Feature Extraction; Similarity Matching; Image Retrieval;

I. INTRODUCTION

In the last few decades there has been enormous growth in the field of internet as well as digital multimedia. Many methods have been developed in the literature to store and manage these kinds of data. Content based image retrieval (CBIR) system analyses image contents via the low-level features for indexing and retrieval, such as color, shape and texture. In order to achieve higher semantic performance, these systems seek to fuse low-level features with high-level features that contain perceptual information of human beings [1]. However, such fusion increases the feature extraction time and the memory requirement as well as the retrieval complexity. CBIR, also known as query by image content, does not consider human-input metadata such as the keywords and texts, whereas query by text considers it. CBIR system is composed two main steps. First step extract various features of the images and store in a feature database. Second step match the feature of query image with the features stored in the feature and find the images which are visually similar to the input image [2].

Color, texture and shape are the primary characteristic of an image which carries useful information that can be used effectively in the image retrieval. Color histogram is the simplest method to code the color information. A major advantage of using histograms, apart from their ease to generate, is that they are invariant to image rotation and translation. However, histograms are not able to convey the spatial information, so it is very difficult to differentiate the images with the same colors but different color distribution [2]. Pass and Zabih [3] divide the histogram in regions by applying the mean of color clustering. Hsu et al. [4] have selected some set of colors and analyzed spatial information on them using entropy quantization including event covering method. Geometric moments, legendre moments, zernike moments and pseudo zernike moments are some region based methods to represent the shape information using moment description [5].

Carson et al. [6] have retrieved the images from large database using the basic CBIR technique. They transformed the image from raw pixel data to small set of coherent region based on color and texture but they have not included shape of the objects in the image that leads to the absence of the images in the retrieved images which are similar in shape. Rui et al. [7] have given a CBIR system that uses many visual features like shape, texture and color. They have, however, ignored the semantic gap between the high level and low level features and they have also not considered the description of human perception of visual content.

The accuracy for an image retrieval system is difficult to define as it is user dependent and very subjective [8]. For the same output different human beings may probably have different views about the similarity. So, it is important that a retrieval system adapt the views of different users.

In this paper, a two phase methodology is proposed for the extraction of semantic information from visual data. In the first phase feature database of images is created. In the second phase, images related to the query image given by the user are retrieved. For image retrieval, first the database is sieved very coarsely. This is done using the hue histogram technique. Feature matching is done on this reduced dataset. At the end of this step, for each feature, a set of image is obtained. Finally, we retrieve the images by combining all the features which results in a set of images which are semantically more similar to the query image.

The rest of the paper is organized as follows. In section II, we present a solution for the semantic based image retrieval by considering the color, texture and shape features jointly. In section III, we show the experimental result and discuss the behavior of the proposed system by considering various issues. Finally, section IV of this paper made the conclusions and highlights the scope of the future work.

II. PROPOSED METHOD

The proposed method for image retrieval is composed of two phases. In the first phase, we construct a feature database of the images. We use algorithm for the image retrieval in the second phase of the proposed work.

978-0-7695-4817-3/12 $26.00 © 2012 IEEE
DOI 10.1109/ICCS.2012.64

Phase 1: Preparing the feature database

In this phase, we store the images in the database with the features extracted from it. Figure 1 illustrates the preparation of the feature database. We extract some color, texture and shape features and store in the feature database.

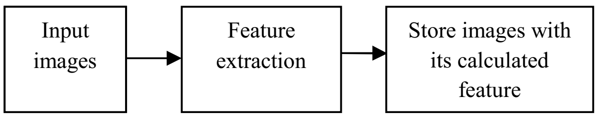

Figure 1: Preparing the database

1) Color Feature Extraction

Color Histogram is an often used feature to extract color information of an image and is frequently used in CBIR system, which contains frequency of each color. A color histogram is basically a distribution of colors in any digital image. We extract color histogram in RGB and HSV color spaces. For multi-spectral images, in which each pixel represents an arbitrary number of measurements, the color histogram in general is N-dimensional, with N being the number of measurements taken. We reduce the number of bins by quantization for being the computationally efficient.

2) Texture Features

In this paper we have used four texture features, Tamura (i.e. coarseness, contrast and directionality) [9], Energy.

a. Coarseness

Coarseness is the information of the size of texture elements. The value of the coarseness will be lower in the smoother areas and vice versa. In general, coarseness is the measurement of roughness in the image.

Algorithm to find coarseness--

1. For every pixel (n_0, n_1) in an image we calculate the average over neighborhood pixels. The size of the neighborhood is the powers of two, e.g.:1*1, 2*2, 4*4... 32*32.

$$C_k(n_0,n_1) = \frac{1}{2^k}\sum_{i=1}^{2^{2k}}\sum_{j=1}^{2^{2k}} I(n_0 - 2^{k-1} + i, n_1 - 2^{k-1} + j) \qquad (1)$$

2. For every pixel (n_0, n_1) we calculate the differences between the non overlapping neighborhoods on opposite sides of the pixel in vertical direction as well as in horizontal direction.

$$D_k^v(n_0,n_1) = | C_k(n_0, n_1 + 2^{k-1}) - C_k(n_0, n_1 - 2^{k-1}) | \qquad (2)$$

$$D_k^h(n_0,n_1) = | C_k(n_0 + 2^{k-1}, n_1) - C_k(n_0 - 2^{k-1}, n_1) | \qquad (3)$$

3. At each pixel (n_0, n_1) we select the size which is leading to the highest difference value.

$$A(n_0,n_1) = \arg(\max_{k=1\ldots5} \max_{d=h,v} D_k^d(n_0,n_1)) \qquad (4)$$

4. In last we take the average to find the coarseness value of the image.

$$F_{crs} = \frac{1}{n_0 n_1}\sum_{n_0=1}^{n_0}\sum_{n_1}^{n_1} 2^{A(n_0,n_1)} \qquad (5)$$

b. Contrast

Contrast represents the quality of picture in an image. It can be influenced by the following four factors:

(1) Sharpness in various edges

(2) Repetition of regular patterns

(3) Dynamic range of gray-levels

(4) Polarization of the distribution

In this paper, the contrast of an image is calculated by the equation

$$F_{con} = \frac{\sigma}{\alpha_4^x} \text{ with } \alpha_4 = \frac{\mu_4}{\sigma_4} \qquad (6)$$

Where, μ_4 is the 4th moment of the mean μ, σ^2 is the variance of the gray values in an image, and x is a constant calculated as 0.25 from the empirical observations.

$$\mu_4 = \frac{1}{n_0 n_1}\sum_{n_0=1}^{n_0}\sum_{n_1=1}^{n_1}(X(n_0,n_1) - \mu)^4 \qquad (7)$$

c. Directionality

Directionality is the presence of orientation in the texture of an image. Two textures differing only in the orientation are considered as same directionality. In the semantic image retrieval, the use of directionality is of great significance. The directionality of the horizontal derivatives ΔH and vertical derivatives ΔV are determined by the convolution of the image X(n_0,n_1) with the 3 × 3 operators shown in Figure 2a and 2b respectively.

-1	**0**	**1**
-1	**0**	**1**
-1	**0**	**1**

(a)

-1	**-1**	**-1**
0	**0**	**0**
1	**1**	**1**

(b)

Figure 2: Operators used for the convolution in (a) Horizontal Derivative and (b) Vertical derivative

Directionality for every pixel (n_0, n_1) is calculated by the following equation

$$\theta = \frac{\pi}{2} + \tan^{-1}\frac{\Delta v(n_0,n_1)}{\Delta h(n_0,n_1)} \qquad (8)$$

d. Energy

Energy for an image can be calculated by the following formulae

$$E = \sum_{i=1}^{M} \sum_{j=1}^{N} P^2(i, j) \quad (9)$$

3) Shape features

In this paper, we have used edge and zernike moment to demonstrate shape feature. Edge is useful information of shape that can be used effectively in the image retrieval. We have used zernike moment because it is invariant to the translation and rotation.

a. Edge detection

We have used 'sobel' operator to find the edge feature of any image. This method finds edges using the 'sobel' approximation to the derivative and returns the edges at the points where the gradient is maximal.

b. Zernike moments

Zernike Moments [10] are useful tools in pattern recognition and image analysis due to their orthogonality and rotation invariance property. It is calculated as the correlation value of the shape with zernike basis function.

Phase 2: Image retrieval

Step1: Feature extraction of the input image

In this phase the system extract all the features of the input image namely color histogram for the color feature, tamura, entropy and energy for the texture feature and zernike moments and edge for the shape feature by using the same technique as given in phase 1.

Step2: Reducing the feature database

We store the highest three values of the hue histogram of all the images in the feature database in phase 1 with Hd1 is the highest one, Hd2 is the second highest and Hd3 is the third highest value. The entries in the feature database is stored in sorted order and sorting is applied on Hd1 first then on Hd2 and then on Hd3.

Table 1: Example of sorting on the highest three values of the hue histogram

Image ID	**Hd1**	**Hd2**	**Hd3**
5	0.110	0.100	0.100
2	0.110	0.105	0.090
7	0.120	0.120	0.110
4	0.130	0.125	0.110
1	0.150	0.140	0.120
6	0.150	0.140	0.130
3	0.150	0.145	0.110

An example of sorting is shown in Table 1, all the images is sorted according to the highest three values of the hue histogram with Hd1 having the highest priority and Hd3 having the least priority. We reduce the database by comparing highest three values of hue histogram of input image with the highest three values of the hue histogram stored in the feature database for each images. For example, let for any input image the value of Hd1 0.140 and the value of threshold is 0.010 then we consider the those images which have the Hd1 value between the range of 0.140±0.010. The same process is applied to Hd2 and Hd3 also. This approach greatly reduces the number of image that must be considered in the image retrieval process and the result of the system is not affected because hue is invariant to the illumination and color differences. We use the different values of the threshold for Hd1, Hd2 and Hd3 and calculated it empirically. Figure 3 shows a framework for the reduction of feature database on the basis of hue histogram.

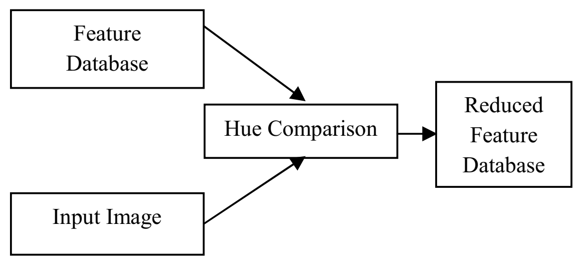

Figure 3: Reduction of feature database

Step3: Image retrieval by using each feature individually

The procedure for the image retrieval by considering each feature separately is depicted in Figure 4. In this step, we extract most similar images to the input image from the reduced feature database considering color, texture and shape feature individually using similarity matching.

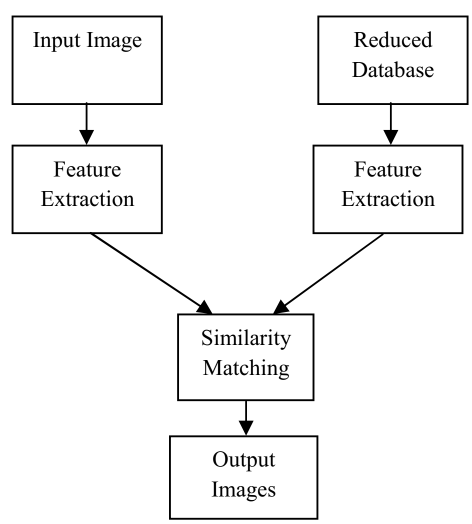

Figure 4: Image retrieval by using each feature individually

We consider each feature individually such that we have all the images which are similar to input images on the basis of either color, texture or shape that are required to be able to perform semantic retrieval.

Step4: Image retrieval by using all the features simultaneously

In this step, we retrieve the images using similarity matching on those images which are similar to the input image either by color, texture or shape considering all the features simultaneously. We combine all the features to obtain a single feature vector for the input image and for each image which are returned in the step 3. By doing this we are able to retrieve all those images which are similar in all respect to the input image and provide a scope to semantic retrieval. The approach to combine the features is shown in Figure 5.

If we use step 4 without using step 3 then there is a possibility that some images will not come in the result which are semantically similar due to the dimensionality of the combined feature. The problem of dimensionality occurs due to lack of proper normalization of all the features.

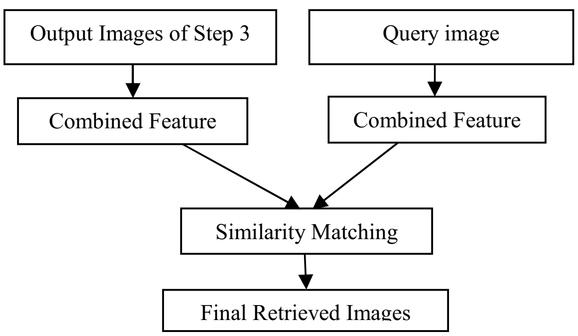

Figure 5: Image retrieval by combining all the features

III. EXPERIMENTAL RESULT

In order to validate the proposed approach for the semantic image retrieval by considering multiple features, we have used a database of 2135 images having various types of images. The query image can be selected either from the dataset or from the outside.

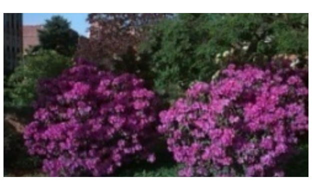

Figure 6: Example of a query image

Figure 6 shows the query image for which we have to retrieve the most similar images semantically. The information in the image is the presence of flowers which is semantic information. Figure 7 shows the most similar images for the queried image by considering shape feature. In this figure, there are three images which are semantically similar to the input image.

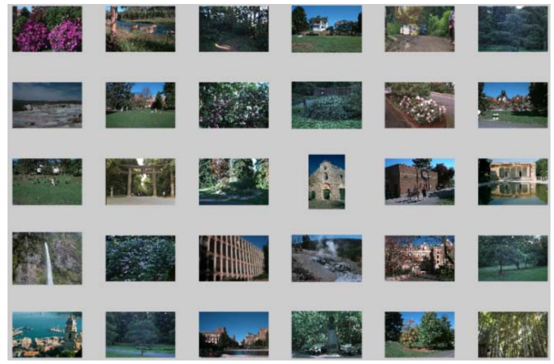

Figure 7: Image retrieval on the basis of shape

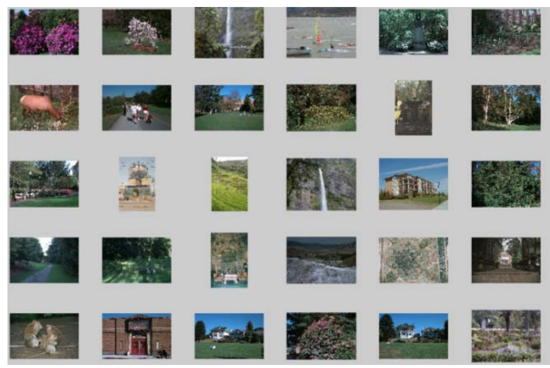

Figure 8: Image retrieval on the basis of texture

Figure 8 shows the most similar images using texture feature having four semantic similar images. We have 13 semantic similar images when color feature is used as shown in the Figure 9. This information lead to conclusion that color feature have more information than texture and shape feature, but alone color is not able to find out all the semantic images. So combining the multiple features is necessary to retrieve more semantic similar images. Figure 10 shows the result of retrieval process when all the features are combined. This result includes 19 images that are similar to the input image semantically.

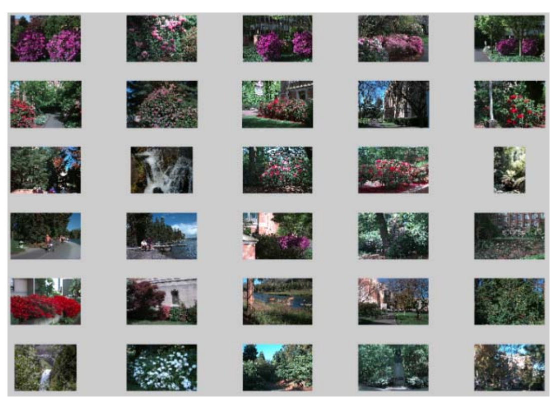

Figure 9: Image retrieval on the basis of color

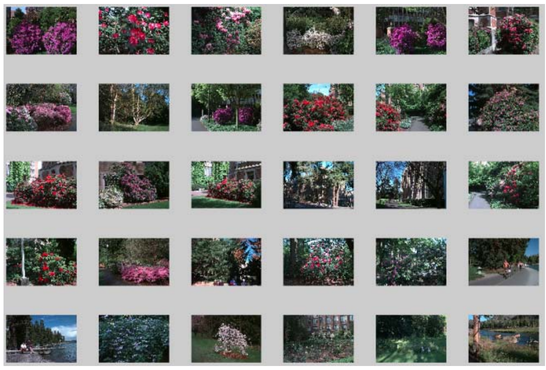

Figure 10: Image retrieval by combining all the features

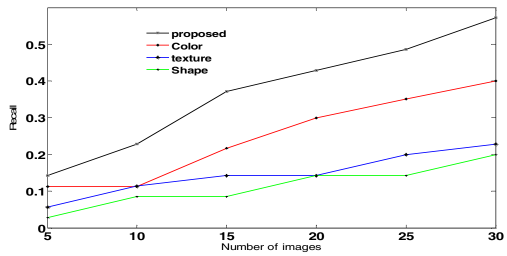

Figure 11: Recall versus number of images

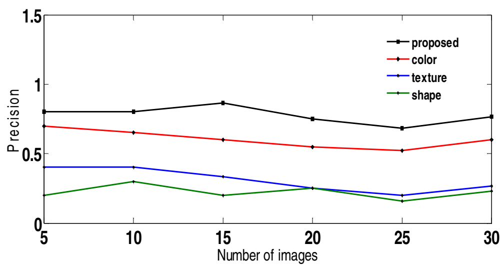

Figure 12: Precision versus number of images

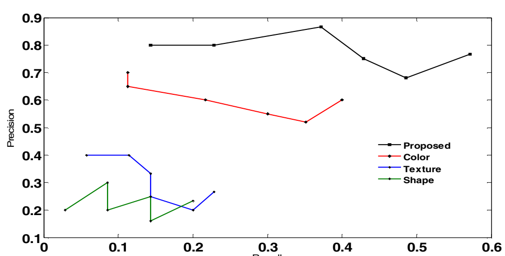

Figure 13: Precision versus Recall

$$\text{Precision} = \frac{\text{number of relevant image retrieved}}{\text{total number of image retrieved}}$$

$$\text{Recall} = \frac{\text{number of relevant image retrieved}}{\text{total number of relevant image}}$$

Figure 11 and 12 shows the recall and precision while considering different number of images retrieved. Precision and recall is the quantitative measurement of performance. For the proposed method the precision and recall is higher than any of individual feature as shown in Figure 11 and 12. Figure 13 compares the recall with the precision while number of images is different. Our proposed approach greatly improves the precision-recall performance of the image retrieval system.

IV. CONCLUSIONS

A semantic image retrieval approach using multiple features is proposed and experimentally evaluated in this paper. We have used color, texture and shape features to improve the performance of content based image retrieval semantically. The presented work operates in two phases, in the first phase, a feature database is created and stored according to the highest three values of the hue histogram to reduce the processing time, in the second phase, first we have applied the similarity matching using each feature individually and then jointly. Our experimental results suggest that the proposed approach matches those images which are more semantically similar with the query image and it is able to improve the precision and recall of the image retrieval system.

REFERENCES

[1] N. Singhai and S. Shandilya, "A Survey On: Content Based Image Retrieval Systems," *IJCA*, vol. 4, no. 2, pp. 22-26, 2010.

[2] W. Xiaoling and M. Hongyan, "Enhancing Color histogram for Image Retrieval," In *International Workshop on Information Security and Application*, 2009, pp. 622-625.

[3] G. Pass and R. Zabih, "Histogram refinement for content based image retrieval," In *WACV*, Sarasoto, 1996, pp. 96-102.

[4] H. wynne, S. T. Chua and H. H. Pung, "Integrated color-spatial approach to content-based image retrieval," In *ACM International Multimedia Conference & Exhibition*, 1995, pp. 305-313.

[5] X. Simon and M. Pawlak, "On Image Analysis by Moments," *IEEE Trans. On Pattern Analysis and Machine Intelligence*, vol. 18, no. 3, pp. 254-266, 1996.

[6] C. Carson, S. Belongie, H. Greenspan, "Region-Based Image Querying," *Workshop on Content-Based Access of Image and Video Libraries*, 1997, pp. 42-49.

[7] Y. Rui, T. Huang and S. Mehrotra, "Relevance feedback techniques in interactive content-based image retrieval," In *Proc. Storage and Retrieval of Image and Video Databases*, 1998.

[8] C. Zhang and T. Chen, *An Active Learning Framework for Content Based Information Retrieval*, Carnegie Mellon University Pittsburgh, Pennsylvania, USA

[9] H. Tamura, S. Mori and T. Yamawaki, "Textural features corresponding to visual perception," *IEEE Trans on Systems, Man and Cybernetics*, vol. 8, no. 6, pp. 460–472, 1978.

[10] G. R. Amayeh, A. Erol, G. Bebis and M. Nicolescu, "Accurate and efficient computation of high order zernike moments," In *ISVC*, 2005, Lake Tahoe, NV, USA, pp. 462-469.