SEMANTIC IMAGE RETRIEVAL USING MULTIPLE FEATURES

Nishant Singh[1], Shiv Ram Dubey[2], Pushkar Dixit[3], Jay Prakash Gupta[4]

Dept. of Computer Engg. and Applications, GLA University, Mathura, India
[1]singhnishant88@gmail.com, [2]shivram1987@gmail.com,
[3]pushkardixit1987@gmail.com, [4]jayprakash619@gmail.com

ABSTRACT

In Content Based Image Retrieval (CBIR) some problem such as recognizing the similar images, the need for databases, the semantic gap, and retrieving the desired images from huge collections are the keys to improve. CBIR system analyzes the image content for indexing, management, extraction and retrieval via low-level features such as color, texture and shape. To achieve higher semantic performance, recent system seeks to combine the low-level features of images with high-level features that conation perceptual information for human beings. Performance improvements of indexing and retrieval play an important role for providing advanced CBIR services. To overcome these above problems, a new query-by-image technique using combination of multiple features is proposed. The proposed technique efficiently sifts through the dataset of images to retrieve semantically similar images.

KEYWORDS

Content Based Image Retrieval, Feature Extraction, Similarity Matching & Image Retrieval.

1. INTRODUCTION

"A Picture is Worth One Thousand Words". It is one of the famous proverbs. The meaning of this proverb is that pictures can replace words, which means images play a much more important role than the text, the information which is contained in the image even the words cannot described them. 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]. There are two main steps in CBIR – (i) Feature Extraction (ii) Matching the features of database images with the query image [2]. Carson et al. [3] have retrieved the images from large database using the basic CBIR technique. They have transformed the image from raw pixel data to small set of coherent region based on color and texture space. The one main drawback in their approach was the shape feature. The images which are similar in terms of shape are not retrieved. Rui et al. [4] has given a CBIR system that uses many visual features like Shape, Texture and Color. They have, however, ignored two main characteristics: (1) the semantic gap between the high level and low level features, and (2) description of human perception of visual content.

In the development of a real time CBIR system, feature evaluation time and query response time should be effective and optimized. A better performance can be obtained if feature-dimensionality and space complexity of the algorithms are optimized. In this paper, a three phase methodology is

Natarajan Meghanathan, et al. (Eds): SIPM, FCST, ITCA, WSE, ACSIT, CS & IT 06, pp. 277–284, 2012.
 DOI : 10.5121/csit.2012.2327

proposed for the extraction of semantic information from visual data. In the first phase database of images with its calculated feature are created. In the second phase, images related to the query image given by the user are retrieved by applying individual feature process. In the third phase, 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 accuracy for an image retrieval method is very difficult to define because it is very subjective and user dependent. Even if there is exactly same input given to the user, different human beings may probably have different views about the similarity of the images. So, it is very important that a retrieval system adapts to different user requirements. Many systems have been developed to improve the accuracy and efficiency of image retrieval systems. All the approaches aim to refine the features and to improve the similarity matching [5].

The rest of the paper is organized as follows. In section II, we give the solution for the problem of semantic–based image retrieval and the details of our proposed system are described. Section III, is devoted to the experiments and results discussion and finally, section IV concludes the paper.

2. PROPOSED METHOD

The proposed method for image retrieval is composed of three phases. In first phase, database of images with its calculated feature are created. In second phase, images related to the query image are retrieved by applying individual feature process. In third phase, we retrieve the images by combining all the features of the images which are retrieved in second phase and get the final results in a form of images which are semantically similar to the query image.

2.1. Preparing the Database (phase 1)

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 color, texture and shape feature and store these entire calculated feature with its corresponding images in the database.

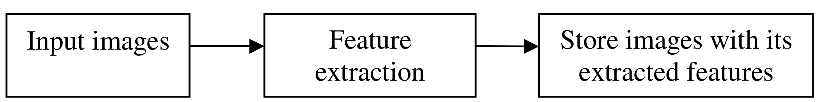


Figure 1. Preparing the database

2.1.1. Color Features

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.

2.1.2. Texture Features

In this paper we have used Tamura texture feature (i.e. coarseness, contrast and directionality) [6],

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. Coarseness can be computed by the following algorithm

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

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

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

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

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

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

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

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

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

Contrast

Contrast represents the quality of picture in an image. It can be influenced by the four factors - (1) Sharpness in various edges, (2) Repetition of regular patterns, (3) Dynamic range of gray-levels, and (4) Polarization of the distribution. In this paper, the contrast of an image is calculated by the following equation

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

Where, μ_4 is the 4^{th} 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.

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

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

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

-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

2.1.3. Shape Features

In this paper, we have used Zernike [7] moment to demonstrate shape feature. We first apply Scale Invariant feature Transform (SIFT) [8] method to all images and then apply shape feature to avoid any scaling effect and to produce better results. Zernike is useful information of shape that can be used effectively in the image retrieval because they have used moment which is invariant to the translation and rotation.

2.2. Image Retrieval by using each Feature individually (phase 2)

First we calculate the feature of query image and stored it in some variable so that we use variable for further access rather than calculating the feature of query image many times. The procedure for the image retrieval by considering each feature separately is depicted in Figure 3. 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. 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.

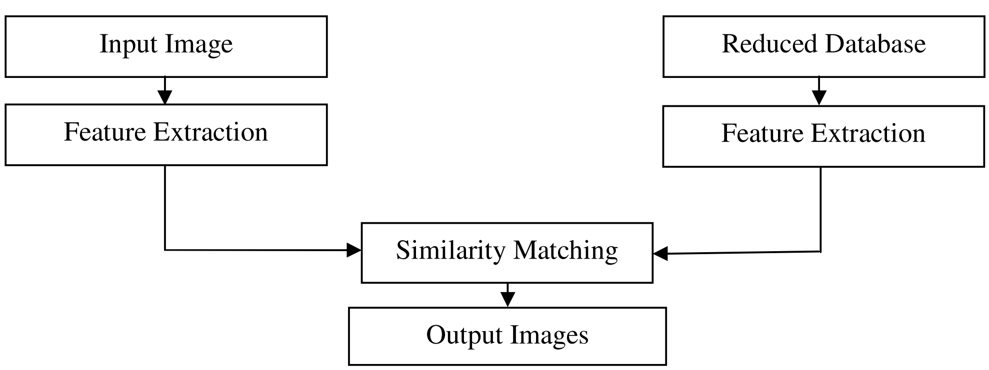


Figure 3. Image retrieval by using each feature individually

2.3. Image Retrieval by using all the Features simultaneously (phase 3)

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 phase 2. 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 of semantic retrieval. The approach to combine the features is shown in Figure 4. If we use phase 3 without using phase 2 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.

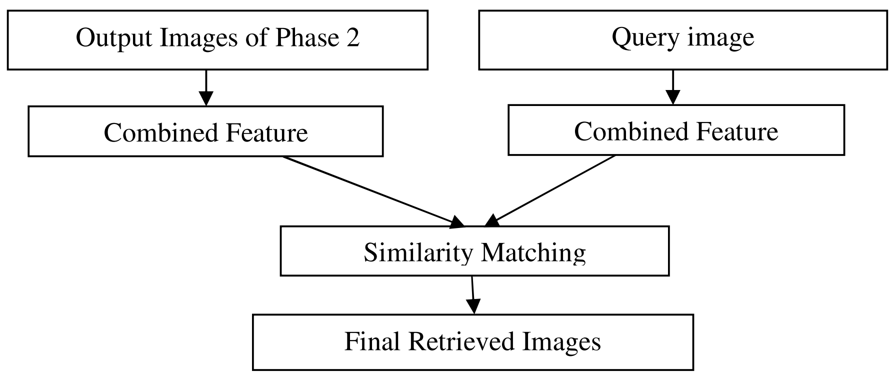


Figure 4. Image retrieval by combining all the features

3. 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.

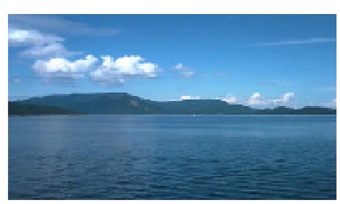
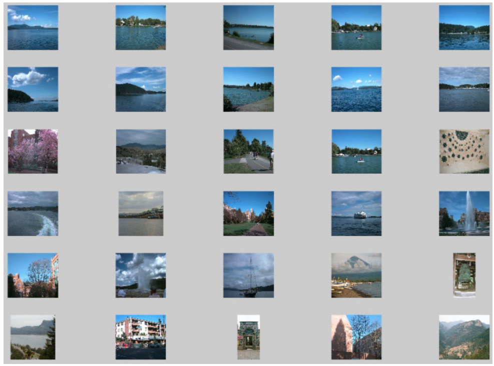
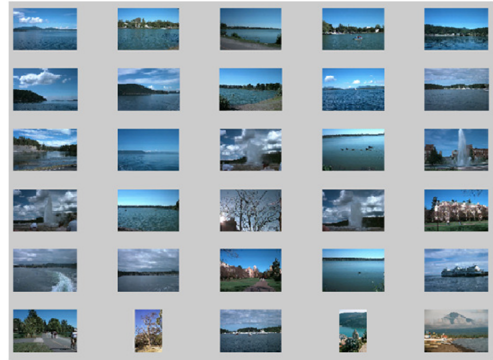


Figure 5. Example of a query image from inside the database

Figure 5 shows the query image from inside the database for which we have to retrieve the most similar images semantically. The information in the image is the presence of sky above the mountain with sea which is semantic information. Figure 6(a) shows the most similar images for the queried image by considering standard CBIR system. In this figure, there are ten images which are semantically similar to the input image. Figure 6(b) 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.



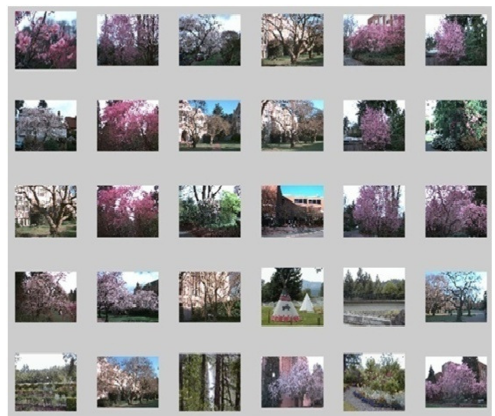
(a)



(b)

Figure 6. Image retrieval on the basis of (a) standard CBIR and (b) proposed method

(a)



(b)

Figure 7. (a) Example of a query image from outside the database, and (b) Image retrieval by our proposed method

Figure 7(a) shows the query image from outside the database for which we have to retrieve the most similar images semantically. The information in the image is the presence of flower which is look like tree which is semantic information. Figure 7(b) shows the most similar images for the queried image by our proposed method. In this figure, there are eighteen images which are semantically similar to the input image. Figure 8(a-b) 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 8(a-b). Figure 8(c) 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. Table 1 shows a comparison of precision among CBIR and our proposed approach. Our introduced method shows fairly more accurate result and outperforms the existing image retrieval methods.

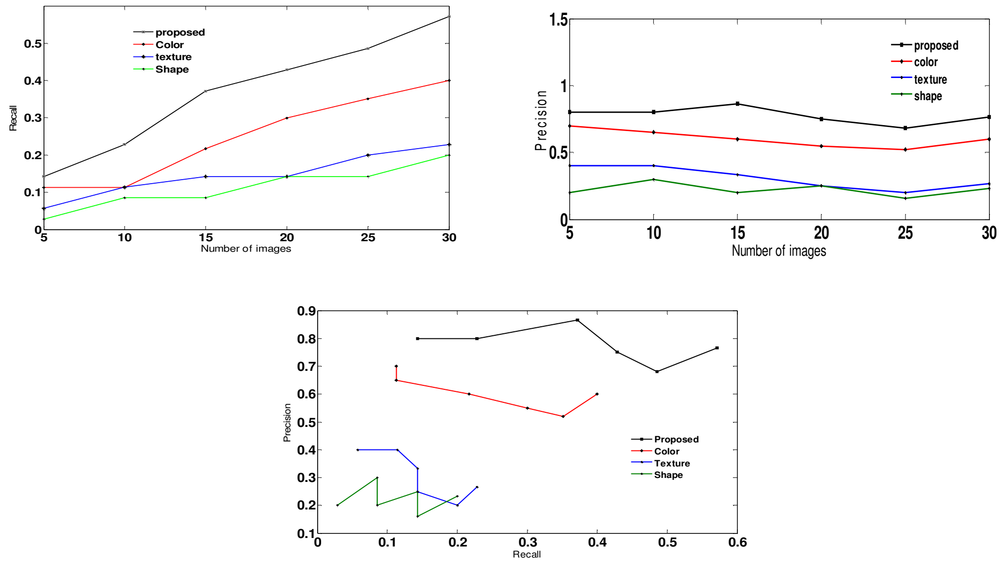


Figure 8. (a) Recall versus number of images, (b) Precision versus number of images, and (c) Precision versus Recall

Table 1. Precision computed at different number of retrieved images for various systems.

Method	Number of retrieved images					
	5	**10**	**15**	**20**	**25**	**30**
Proposed	.79	.79	.87	.78	.70	.79
CBIR	.35	.35	.33	.32	.31	.30

4. CONCLUSIONS AND FUTURE WORK

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 three phases, in the first phase, a feature database is created, in the second phase we retrieve images by using each feature individually and in third phase we retrieve images by using all the features simultaneously. 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. The work done in proposed CBIR system can be used fully in the new version of the system based on the key-frames extracted from the video sequences. These algorithms can benefit very much from the combination of features extracted from different content. Considerable improvements can thus still be expected from such multi modal features.

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

[4] 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.

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

[6] 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.

[7] 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.

[8] H. Zhou, Y. Yuan and C. Shi, “Object tracking using SIFT features and mean shift,” Computer Vision and Image Understanding, vol. 113, no. 3, pp. 345-352, 2009.