Smart Computing Review

Content-based Image Retrieval Using Haar Wavelet Transform and Color Moment

Md. Iqbal Hasan Sarker [1] and Md. Shahed Iqbal [2]

[1] Department of Computer Science & Engineering, Chittagong University of Engineering & Technology / Chittagong-4349, Bangladesh / iqbal@cuet.ac.bd

[2] Department of Computer Science & Engineering, Chittagong University of Engineering & Technology / Chittagong-4349, Bangladesh / cooling_dew@yahoo.com

* Corresponding Author: Md. Iqbal Hasan Sarker

Received March 8, 2013; Revised May 17, 2013; Accepted May 24, 2013; Published June 30, 2013

Abstract: Content-based image retrieval (CBIR) deals with the retrieval of most similar images corresponding to a query image from an image database by using visual contents of the image itself. It requires feature extraction and computation of similarity. In this paper, we propose a content-based image retrieval method that uses a combination of color and texture features. The Haar wavelet transform is used for texture feature extraction, and for color feature extraction we use color moments. The distance between the query image features and the database images' features is computed by using Canberra distance. We assign weights to texture feature distance and color feature distance and calculate the similarity with combined feature distance. Experiment results reflect the importance of the Haar wavelet transform and color moments in the performance of our proposed CBIR method.

Keywords: CBIR, Haar wavelet transform, Color moment, Canberra distance, F-norm theory

Introduction

Content-based image retrieval (CBIR) [1] is an image search technique where images are selected from an image database by using a reference image rather than metadata, such as keywords, tags and descriptions associated with that image. Here, input for the search is an image, and the output is similar images from the database. The similarity

DOI: 10.6029/smartcr.2013.03.002

between two images is measured by calculating the distance between the two images. That distance is calculated from feature vectors, and the feature vectors are constructed from the content of the image. Here, content refers to color, texture and shape of the image.

Kherfi et al. mentioned the two drawbacks in keyword annotation image retrieval. First, images are not always annotated, and manual annotation is expensive and time consuming. Second, the same image may be annotated differently by different observers [2]. Sometimes it is very difficult and time consuming to give a proper name to an image, and sometimes a lot of information is needed to properly describe an image. That means people need to tag every image properly to ensure good results from metadata-based image retrieval, which is obviously a time consuming and inefficient procedure. Unlike the traditional approach of using keyword annotation as a method to search images, a CBIR system performs retrieval based on the similarity of feature vectors such as color, texture, shape and other image content.

Most CBIR systems work in the same way. A feature vector is extracted from each image in the database and the set of all feature vectors is organized as a database index. When similar images are searched with a query image, a feature vector is extracted from the query image and is matched against the feature vectors in the index. Differences between the various systems lie in the features they extract and the algorithms used to extract those features. The block diagram of a basic CBIR system is shown in Figure 1.

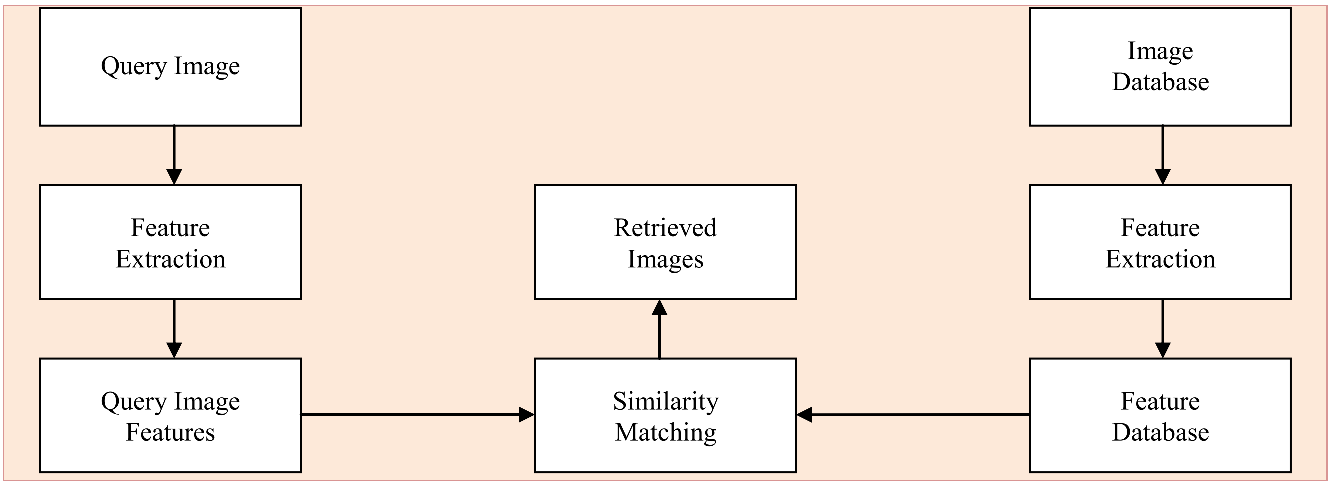

Figure 1. A CBIR system block diagram

There is a lot of research being done in the field of CBIR in order to generate better methodologies for feature extraction. In this paper, we use the Haar wavelet transform and color moments to extract features from images. First, we apply the Haar wavelet transform to the images, then extract features by using F-norm theory; using color moments, we extract color features. We combine these two features by adjusting their weights, and similarity is measured using the Canberra distance. Efficiency of our proposed method is measured in terms of recall.

The rest of the paper is organized as follows. In Section 2, we discuss previous work in CBIR. In Section 3, we explain our proposed method. Section 4 describes the implementation and experimental results, and finally, conclusions are given in Section 5.

Related Work

Content-based image retrieval has become a prominent research topic, and researchers have proposed different methods to improve the system.

Color features are the most widely used visual features in CBIR systems. The color indexing work of Swain and Ballard [5], which is based on color histograms, has demonstrated the potential of using color for indexing. Stricker and Orengo [6] have shown that moment-based color distribution features can be matched more robustly than color histograms because histograms do not capture spatial relationships of color regions. Hence, in our proposed method, color moments are used for color feature extraction.

Texture is an important feature for CBIR systems. Various techniques have been developed for measuring texture similarity. Tamura et al. took the approach of devising texture features that correspond to human visual perception [7]. They defined six textural features (coarseness, contrast, directionality, line-likeness, regularity and roughness) and compared them with psychological measurements for human subjects. The first three components of Tamura features have been used in some early well-known image retrieval systems, such as QBIC [8]. Wavelet transform provides a multi-resolution approach to texture analysis and classification [9]. Khan et al. [10] used the Haar wavelet transform for texture feature extraction.

Combination of features is also used in content-based image retrieval. Choras et al. [11] proposed an integrated color, texture and shape feature extraction method in which Gabor filtration is used to determine the number of regions of interest (ROIs). They calculated texture and color features from the ROIs based on threshold Gabor features and histograms, color moments in luminance-bandwidth-chrominance space, and shape features based on Zernike moments.

Proposed Method

■ Wavelet Transformation

Wavelet transforms provide a multi-resolution approach to texture analysis and classification. The computation of the wavelet transforms of a two-dimensional signal involves recursive filtering and sub-sampling. At each level, the signal is decomposed into four frequency sub-bands, LL, LH, HL, and HH, where L denotes low frequency and H denotes high frequency. Figure 2 shows level 1 of the 2D wavelet transform.

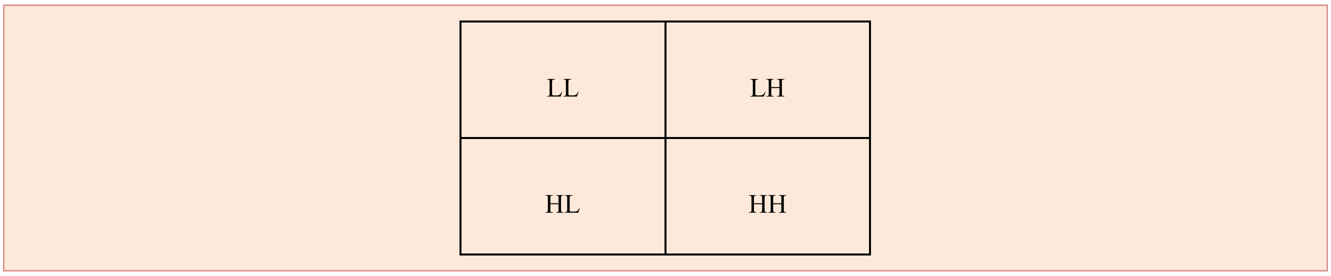

Figure 2. Level 1 of the 2D wavelet transform.

We used Haar wavelets for image decomposition.

■ Haar Wavelet Transform

If a data set $X_0, X_1 \ldots X_{N-1}$ contains N elements, there will be N/2 averages and N/2 wavelet coefficient values. The averages are stored in the first half of the N element array, and the coefficients are stored in the second half of the N element array. The averages become the input for the next step in the wavelet calculation.

The Haar equations to calculate an average a_i and a wavelet coefficient c_i from an odd and even element in the data set are:

$$a_i = \frac{X_i + X_{i+1}}{2} \tag{1}$$

$$c_i = \frac{X_i - X_{i+1}}{2} \tag{2}$$

Steps for a 1D Haar transform of an array of N elements are as follows:

1. Find the average of each pair of elements using Equation 1. (N/2 averages)
2. Find the difference between each pair of elements and divide it by 2. (N/2 coefficients)
3. Fill the first half of the array with averages.
4. Fill the second half of the array with coefficients.
5. Repeat the process on an average part of the array until a single average and a single coefficient are calculated.

Steps for a 2D Haar transform are:

1. Compute 1D Haar wavelet decomposition of each row of the original pixel values.
2. Compute 1D Haar wavelet decomposition of each column of the row-transformed pixels.

Red, green and blue values are extracted from the images. Then we apply the 2D Haar transform to each color matrix.

Figure 3 shows level 1of the 2D Haar decomposition of the image.

Figure 3. Level 1 of a 2D Haar decomposition

Wavelet Feature Extraction

We apply Haar wavelet decomposition of an image in the RGB color space. We continue decomposition up to level 4, and with F-norm theory [12] we decrease the dimensions of image features and perform highly efficient image matching.

Suppose A is a square matrix and A_i is its i^{th} order sub-matrix where

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \dots & \dots & \dots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}, \qquad A_i = \begin{bmatrix} a_{11} & \dots & a_{1i} \\ \dots & \dots & \dots \\ a_{i1} & \dots & a_{ii} \end{bmatrix} \quad (i = 1 \sim n)$$

The F-norm of A_i is given as:

$$\left\|A_i\right\|_F = \left(\sum_{k=1}^{i} \sum_{l=1}^{i} \left|a_{kl}\right|^2 \right)^{1/2} \tag{3}$$

If $\Delta A_i = \left\|A_i\right\|_F - \left\|A_{i-1}\right\|_F$ and $\left\|A_0\right\|_F = 0$, we can define the feature vector of A as:

$$V_{AF} = \{\Delta A_1, \Delta A_2 \dots \Delta A_n\} \tag{4}$$

Color Feature

Color is one of the most commonly used visual features in content-based image retrieval. Color features have been found to be effective for measuring similarity between images. One of the main aspects of color feature extraction is the choice of a color space. A color space is a multidimensional space in which the different dimensions represent the different components of color. An example of a color space is RGB. In the RGB color space, a color is represented by a triplet (R,G,B), where R gives the intensity of the red component, G gives the intensity of the green component and B gives the intensity of the blue component. The space spanned by the R, G, and B values completely describes visible colors, which are represented as vectors in the 3D RGB color space. As a result, the RGB color space provides a useful starting point for representing color features of images. However, the RGB color space is not perceptually uniform.

We can use the RGB color space to generate other color spaces. The idea for color space transformation is to develop a model of color space that is perceptually similar to human vision. Color spaces such as HSV, CIE 1976 (L*, a*, b*), and CIE 1976 (L*, u*, v*) are generated by nonlinear transformation of the RGB space. The CIE color spaces represent the three characteristics that best characterize color perceptually: hue, lightness, and saturation. However, the CIE color spaces are inconvenient because of the calculation complexities of the transformation to and from the RGB color space. The HSV color space is also a nonlinear transformation of RGB, but it is easily invertible. The HSV color space is approximately perceptually uniform. The components of the HSV color space are Hue, Saturation and Value. In this paper, we use the HSV color space for color feature extraction.

The RGB values can be converted into HSV values according to the following formula:

$$H = \begin{cases} \theta & \text{if B} \le \text{G} \\ 360 - \theta & \text{if B} > \text{G} \end{cases} \tag{5}$$

where:

$$\theta = \cos^{-1}\left\{ \frac{\frac{1}{2}[(R-G)+(R-B)]}{[(R-G)^2 + (R-B)(G-B)]^{1/2}} \right\}$$

$$S = 1 - \frac{3}{(R+G+B)}[\min(R,G,B)] \tag{6}$$

$$I = \frac{1}{3}(R+G+B) \tag{7}$$

Color Moment

We use color moments for color feature extraction. The basis of color moments lies in the assumption that the distribution of color in an image can be interpreted as a probability distribution. Probability distributions are characterized by a number of unique moments (e.g., normal distributions are differentiated by their mean and variance). Computing color moments is done in the same way as computing moments of a probability distribution.

The first color moment can be interpreted as the average color in the image and can be calculated by using the following formula:

$$E_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij} \tag{8}$$

where:

N = number of pixels in the image
p_{ij} = value of the j-th pixel of the image at the i-th color channel

The second color moment is the standard deviation, which is obtained by taking the square root of the variance of the color distribution.

$$\sigma_i = \sqrt{\left(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - E_i)^2\right)} \tag{9}$$

where:
E_i = mean value, or first color moment, for the i-th color channel of the image

The third color moment is the skewness.

$$s_i = \sqrt[3]{\left(\frac{1}{N} \sum_{j=1}^{N} (p_{ij} - E_i)^3\right)} \tag{10}$$

We get nine numbers—three moments for each color channel as color features for each of the image.

Similarity Measure

The similarity between two images is computed by calculating the distance between feature representation of the query image and feature representation of the image in the dataset. We use Canberra distance (Equation 11) for distance calculation of the feature vectors.

$$dis(q,d) = \sum_{i=1}^{n} \frac{|q_i - d_i|}{|q_i| + |d_i|} \tag{11}$$

where :

$q = (q_1, q_2 ... q_n)$ is the feature vector of the query image,

$d = (d_1, d_2 ... d_n)$ is the feature vector of the image in the database, and

n = number of elements of the feature vector.

If the distance between feature representation of the query image and feature representation of the database image is small, then it is considered similar.

Combining the Features

During our experiment, we worked with two feature extraction methods, Haar wavelets for texture feature extraction, and color moments for color feature extraction. Using only a single feature for image retrieval may be inefficient. To produce efficient results, we combined the two features by adjusting appropriate weights.

Similarity between two images is computed by calculating the distance between the feature vectors of the two images. In our proposed method, we calculate two distances from two feature vectors for each image. Then the final distance is calculated by combining weights with those two distances. The final distance between the query image and the image in the database is calculated as follows:

$$d = d_1 * w_1 + d_2 * w_2 \tag{12}$$

where:

d_1 = calculated distance using Haar wavelet features

w_1 = weight for Haar wavelet features

d_2 = calculated distance using color features

w_2 = weight for color features

We experiment with different combinations of weights by increasing and decreasing the weight values. But we get better image retrieval when we set weights $w_1 = 0.65$ and $w_2 = 0.35$. The distance d is calculated for each query image with all images in the database. The image that has a lower distance value is considered the similar image.

Experiments and Results

In our experiment, we selected four types of image, 100 images in each category, and 400 images in total from Wang's database [13]. The images were resized to 256 x 256 pixels.

The performance of a CBIR system can be measured in terms of its precision and recall. Precision measures the retrieval accuracy; it is the ratio between the number of relevant images retrieved and the total number of images retrieved. Recall measures the ability to retrieve all relevant images in the database. It is the ratio between the number of relevant images retrieved and all of the relevant images in the database. They are defined as follows:

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \tag{13}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} \tag{14}$$

We extracted the Haar wavelet features and color features separately, and then we experimented with the CBIR system by using the Haar wavelet features, the color features, and a combination of these features. Table 1 shows the average recall results for Haar wavelet, color moment and the combination of Haar wavelet and color moment.

Table 1. Average recall results

Category	Haar	Color Moment	Haar + Color Moment
Buses	0.71	0.79	0.90
Dinosaurs	0.68	0.91	0.94
Roses	0.79	0.57	0.82
Horses	0.67	0.81	0.86
Average Recall (%)	71.25	77.0	88.0

Figure 4 shows the comparison of the average recall for Haar wavelet, color moment and combination of Haar wavelet and color moment.

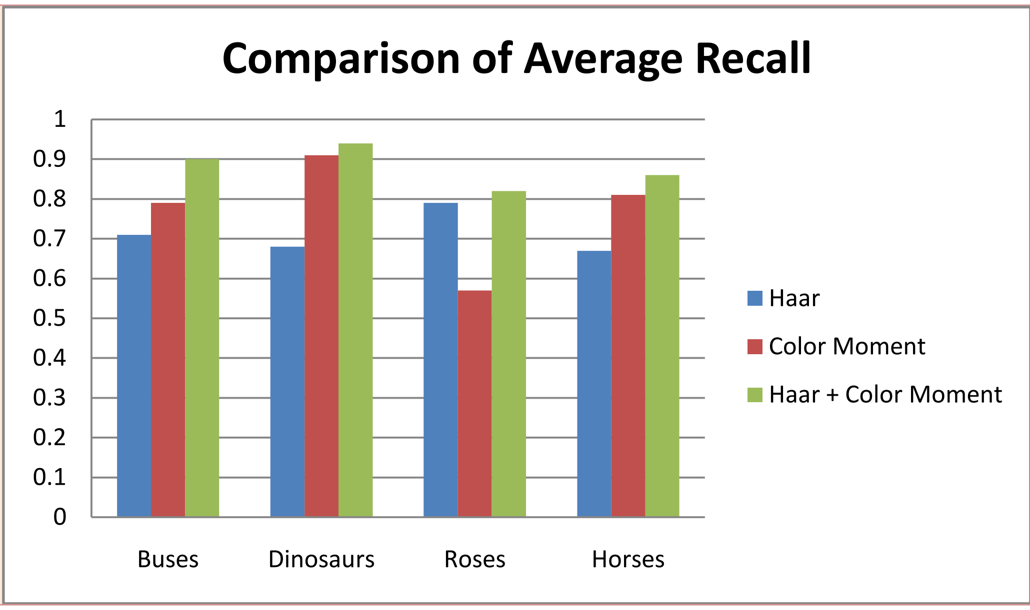

Figure 4. Comparison of average recall

From Table 1 and Figure 4, we see that the average recall (%) based on Haar wavelet is 71.25, the average recall (%) based on color moment is 77.0 and the average recall (%) based on the combination of Haar wavelet and color moment (the proposed method) is 88.0. That means, our experiment shows that retrieval efficiency increases when we combine the two features.

We implemented our proposed method in Java. Some screenshots from our application for sample query images are shown below. Figure 5 shows retrieval results for the query image "bus" based on the Haar wavelet feature. Figure 6 shows retrieval results for the same query image based on color moment, and Figure 7 shows retrieval results for the same query image based on our proposed method. Figure 8, Figure 9, and Figure 10 show the retrieval results for the query images "dinosaur," "rose," and "horse," respectively, based on our proposed method.

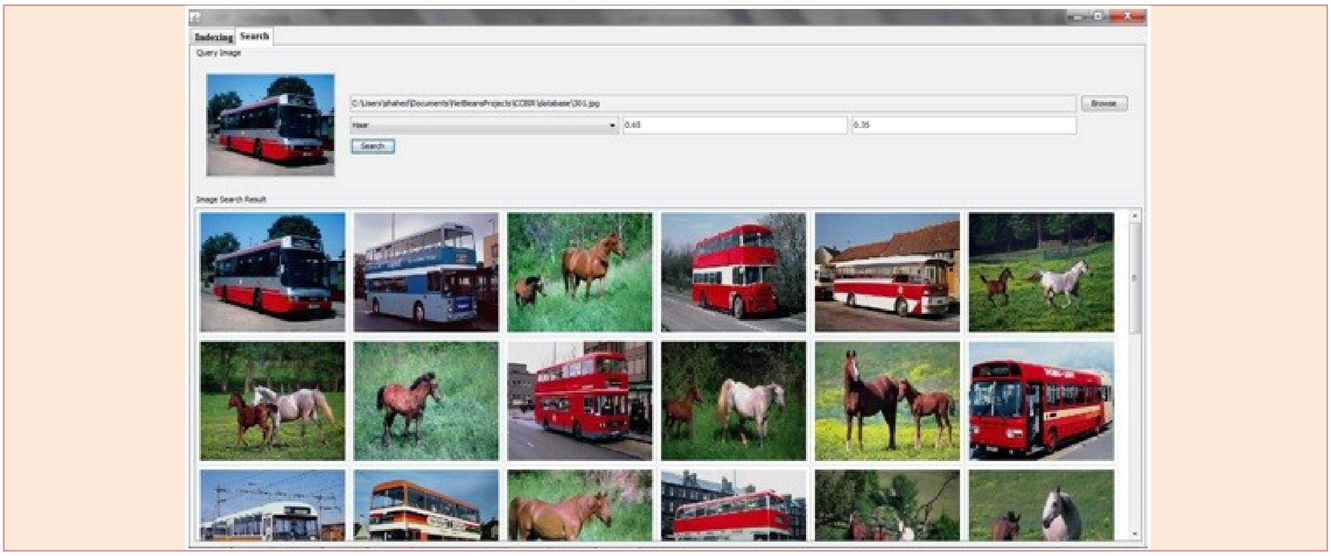

Figure 5. Retrieved images based on Haar wavelet for the bus as the query image

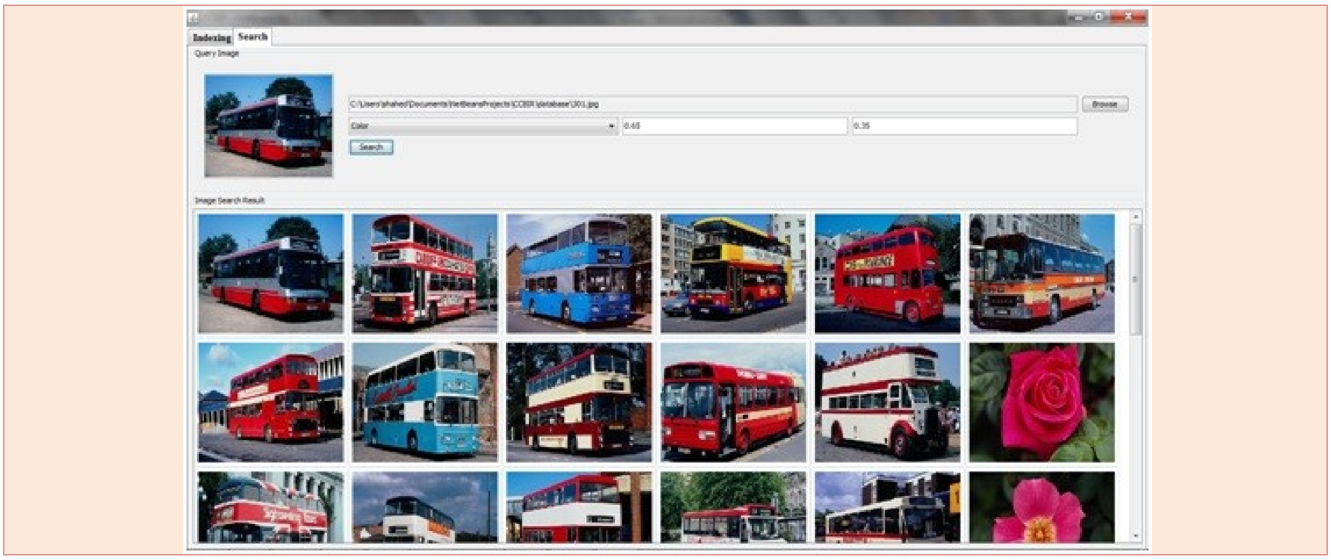

Figure 6. Retrieved images based on color moment for the bus as the query image

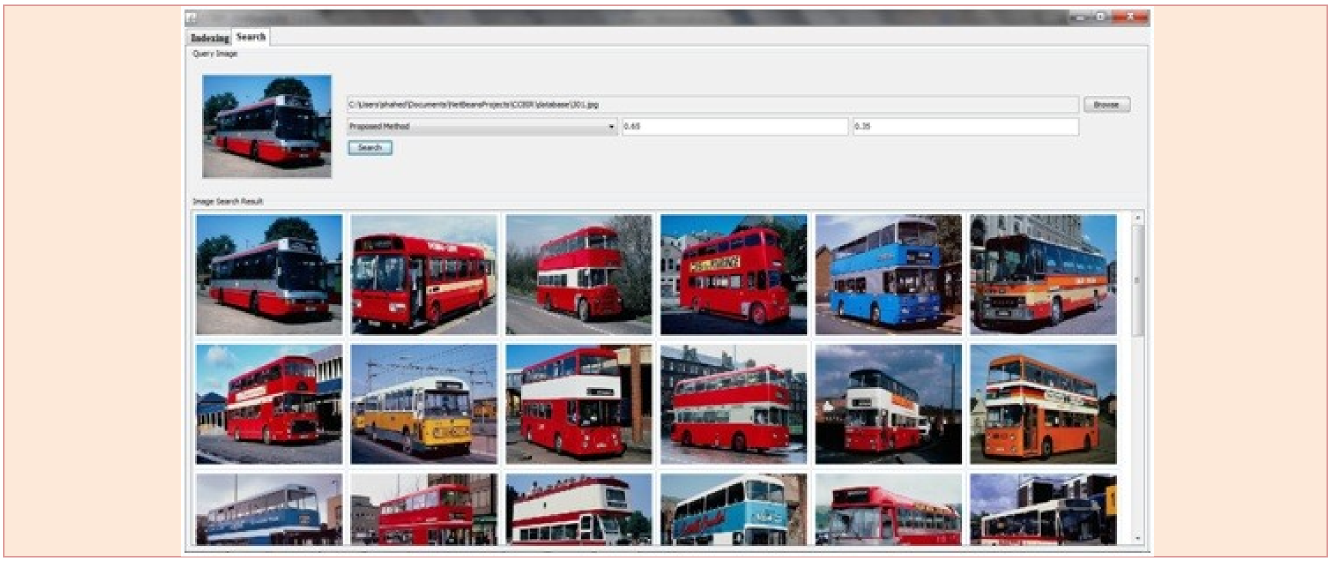

Figure 7. Retrieved images based on our proposed method for the bus as the query image

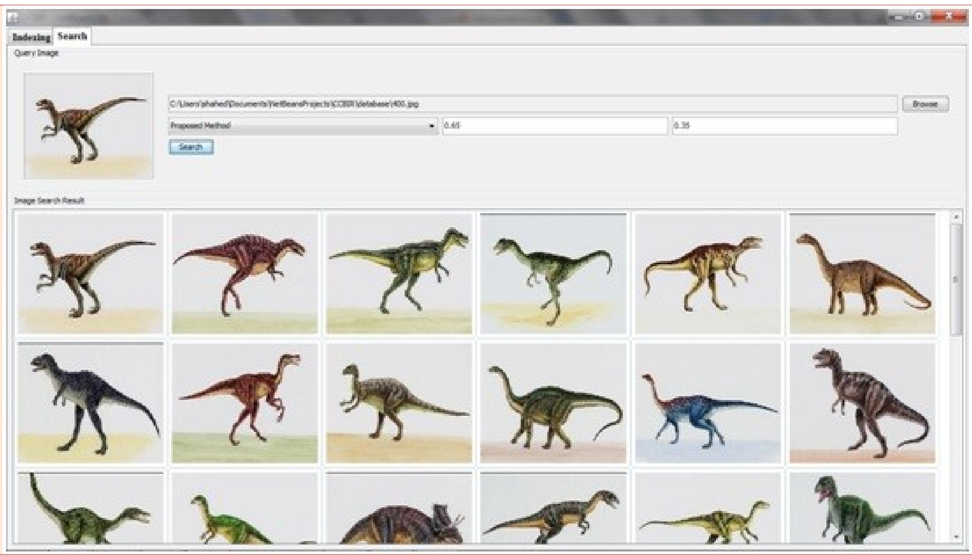

Figure 8. Retrieved images based on our proposed method for the dinosaur as the query image

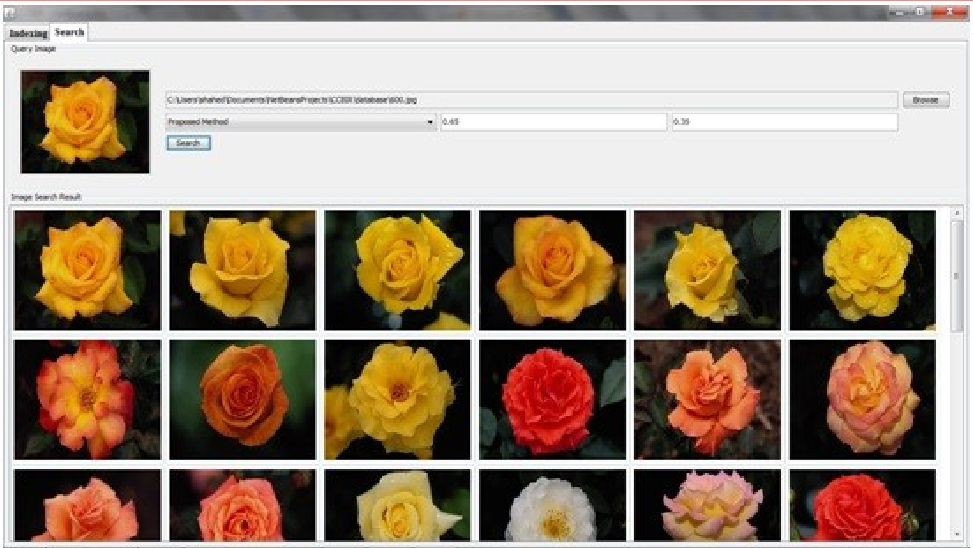

Figure 9. Retrieved images based on our proposed method for the rose as the query image

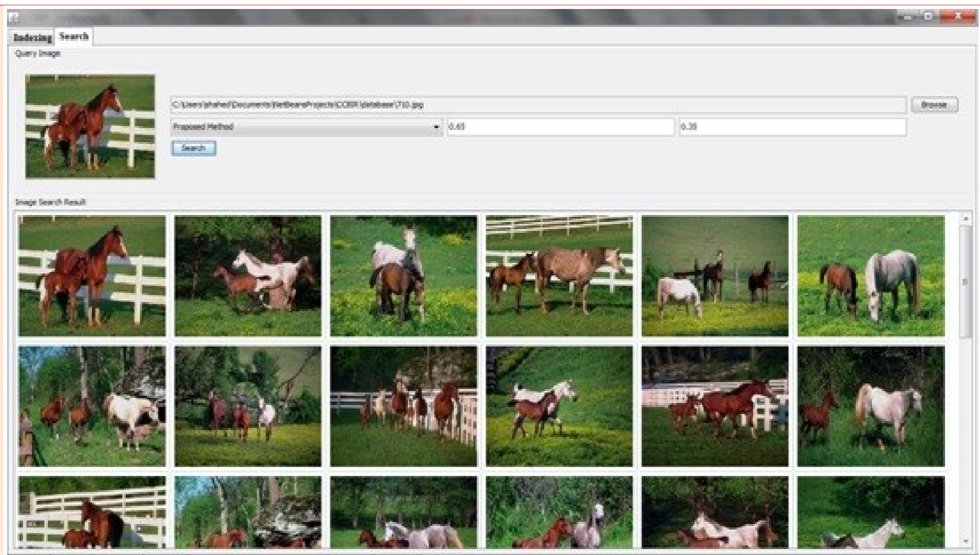

Figure 10. Retrieved images based on our proposed method for the horse as the query image

Conclusion

In this paper, we have proposed an efficient CBIR method based on the Haar wavelet transform and color moments. First we apply the Haar wavelet transform to images in the database, then we apply F-norm theory to reduce the dimensions of the feature vectors, and thus a feature vector for texture is extracted using the Haar wavelet transform. For color feature extraction we use color moments and get nine floating point numbers as features. All the features for the database images are pre-calculated and stored in the database. When similar images are searched using a query image, we calculate the features for that image, and the distance between the query image features and the database image features is computed using Canberra distance. After adjusting weights, the similar images according to their distance values are displayed. Our experiment results demonstrate that the proposed method has higher retrieval accuracy than the other methods based on single feature extraction.

References

[1] R.Datta, D. Joshi, J.Li, J.Z. Wang, "Image retrieval: ideas, influences, and trends of the new age," *ACM Computing Surveys,* vol. 40, no. 2, article 5, pp. 1-60, Apr. 2008.
[2] M. L. Kherfi , D. Brahmi and D. Ziou "Combining visual features with semantics for a more effective image retrieval," in *Proc. of 17th Int. Conf. on Pattern Recognition*, pp.961 -964 , 2004.
[3] B.Ramamurthy and K.R.Chandran, "Content based Image Retrieval for Medical Images using Canny Edge Detection Algorithm," *International Journal of Computer Applications*, vol.17, no.6, pp. 32-37, Mar. 2011.
[4] A. Ross, A. Jain and J. Reisman, "A hybrid fingerprint matcher," in *Proc. of Int. Conf. on Pattern Recognition, Quebec City*, vol. 3, pp. 795-798, 2002.
[5] M.Z. Swain, and D.H. Ballard, "Color Indexing," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 11-32, Nov. 1991. Article (CrossRef Link)
[6] M. Stricker, and M. Orengo, "Similarity of color images," in *Proc. of SPIE Conference on Storage and Retrieval for Image and Video Databases* , vol. 2420, pp. 381-392, Mar. 1995. Article (CrossRef Link)
[7] H. Tamura, S. Mori, and T. Yamawaki, "Texture features corresponding to visual perception," *IEEE Trans. on Systems, Man, and Cybernetics*, vol. 8, No. 6, June 1978. Article (CrossRef Link)
[8] W. Niblack, R. Barber,W. Equitz, M. Flickner, E. Glasman, D. Pektovic, P. Yanker, C. Faloutsos, and G. Taubin,, "The QBIC Project: Querying Images by Content using Color Texture and Shape," in *Proc. SPIE Int. Soc. Opt. Eng., in Storage and Retrieval for Image and Video Databases*, vol. 1908, pp. 173-187, 1993.
[9] T. Chang and C.C.J. Kuo, "Texture analysis and classification with tree-structured wavelet transform," *IEEE Trans. on Image Processing*, vol. 2, no. 4, pp. 429-441, Oct. 1993. Article (CrossRef Link)
[10] W. Khan, S. Kumar, N. Gupta and N. Khan, "Signature Based Approach For Image Retrieval Using Color Histogram And Wavelet Transform," *International Journal of Soft Computing and Engineering*, vol.1, no. 1, Mar. 2011.
[11] R.S. Choras, T. Andrysiak and M. Choras, "Integrated color, texture and shape information for content-based image retrieval," *Pattern Analysis & Applications,* vol. 10, no. 4, pp. 333-343, Oct. 2007. Article (CrossRef Link)
[12] H. Huang, W. Huang, Z. Liu, W. Chen and Q. Qian, "Content-based color image retrieval via lifting scheme," in Proc. of *ISADS Conference*, pp. 378 – 383, 4-8 Apr. 2005.
[13] http://wang.ist.psu.edu/docs/related.shtml

Md. Iqbal Hasan Sarker received the B.Sc degree in Computer Science & Engineering from Chittagong University of Engineering & Technology (CUET), Bangladesh in 2009. Currently, he is pursuing an M.Sc in Computer Science & Engineering at the same University. Since 2010, he has been serving as a faculty member in the same department and university. His research interest includes Digital Watermarking, Computer Vision, Digital Image Processing, Cryptography and data mining.

Md. Shahed Iqbal is pursuing a B.Sc in Computer Science & Engineering from Chittagong University of Engineering & Technology (CUET). His research interests include computer vision, image processing and Human Computer Interface.