An Efficient Content-based Image Retrieval System Integrating Wavelet-based Image Sub-blocks with Dominant Colors and Texture Analysis

Sherin M. Youssef
College of Engineering
Arab Academy for Science and Technology
Alexandria, Egypt

Saleh Mesbah
College of Computer Science and IT
Arab Academy for Science and Technology
Alexandria, Egypt

Yasmine M. Mahmoud
College of Engineering
Arab Academy for Science and Technology
Alexandria, Egypt

***Abstract*— There is a great need of developing efficient content-based image retrieval systems (CBIR) because of the availability of large image databases. Three new image retrieval systems to retrieve the images using color and texture features are proposed. The image is divided into equal sized non-overlapping tiles. The discrete wavelet transform, HSV color feature, cumulative color histogram, dominant color descriptor (DCD) and Gray level co-occurrence matrix (GLCM) are applied to image partitions. An integrated matching scheme based on Most Similar Highest Priority (MSHP) principle is used to compare the query and database images. The adjacency matrix of a bipartite graph is formed using the sub-blocks of query and images in the database. The proposed techniques indeed outperform other retrieval schemes in terms of average precision and average recall. The developed techniques are able to perform scale, translation, and rotation invariant matching between images. In the future, we need to reduce the semantic gap between the local features and the high-level user semantics to achieve higher accuracy.**

Keywords-component; image retrieval; Dominant color; Gray level co-occurrence matrix.

I. INTRODUCTION

Content-based image retrieval is the application of computer vision techniques to the image retrieval problem of searching for digital images in large databases. Nowadays, fast and effective searching for desired images from large scale image database became an important and challenging research topic [1, 2]. Because of the increase of the bandwidth availability and rapid developments in processor, memory and storage technologies and the propagation of video and image data in digital form, this made the CBIR technology an important alternative to traditional text-based image searching, the CBIR systems can greatly enhance the accuracy of the information being returned [3]. The query to the database can be of various types, some of them are: Query-by-sketch, where the user provides a sketch of the image he is looking for, and Query-by-example, in which the user gives an example image similar to the one he is looking for. In this paper, we focus on the query-by-example approach.

CBIR technology makes full use of image content features such as, color, texture and shape, etc which are analyzed and extracted automatically to achieve the effective retrieval [4, 5]. Using a single feature for image retrieval cannot be a good solution for the accuracy and efficiency. High-dimensional feature vector will reduce the query efficiency; low-dimensional feature vector will reduce query accuracy, so it may be a better way using multi features for image retrieval.

The discrete wavelet transform provides a sufficient way for high-accuracy retrieval system. Color is one of the most widely used low-level visual features and is invariant to image size and orientation. The HSV color space is closer to human conceptual understanding of colors, so it gave good results in many CBIR systems [6, 7]. The DCD describe the salient color distributions in an image or a region of interest, and provides an effective, compact, and intuitive representation of colors presented in an image [3, 8]. The Color Histogram is effective in characterizing both the global and local distribution of colors in an image. It gives accurate results as proved in [3]. Texture is from the most important visual features, it refers to the visual patterns of the homogeneity of regions. Many objects in an image can be distinguished solely by their textures without any color information. Local features based methods proved good results [9, 14]. Our proposed CBIR system is based on the wavelet transform, HSV color feature, Dominant Color Descriptors, Cumulative Color Histogram as color features and lastly, Gray-Level Co-occurrence Matrix (GLCM) as a texture feature. The objective of this work is to develop a technique which captures local color and texture descriptors in a coarse segmentation framework of image sub-blocks. The image is partitioned into equal sized non-overlapping tiles. The features computed on these tiles serve as local descriptors of color and texture. An integrated matching procedure based on adjacency matrix of a bipartite graph between the image tiles is provided. The combination of these features forms a robust feature set in retrieving applications. The rest of the paper is organized as follows. The section II outlines the proposed models. Section III explores the features used in this work. Section IV presents the similarity matching method. The section V deals with the experimental setup, results and analysis. The section VI presents conclusions and future work.

II. THE PROPOSED SYSTEM

The main phases of the proposed system include: Image partitioning into equal-sized sub-blocks, applying the discrete wavelet transform to each sub-block, converting the sub-blocks to HSV color space, Non-uniform quantization, color histogram generation for image sub-blocks, extraction of dominant color descriptors and texture analysis of each image sub-block, and finally, the similarity matching using MSHP principle. The block diagram of the proposed model is illustrated in Figure 1. In the next sections, each phase will be described in details.

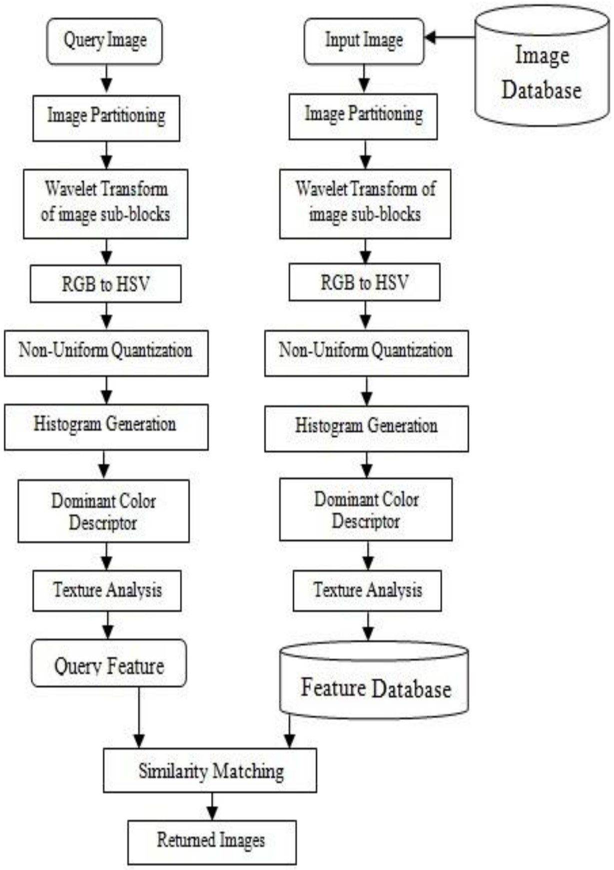

Figure 1. The Proposed System.

III. PROPOSED MODEL PHASES DESCRIPTION

A. *Image Pre-processing*

Firstly the image is partitioned into 6 (2X3) equal sized sub-blocks as shown in Figure 2. The size of the sub-block in an image of size 256X384 is 128X128. The images with other than 256X384 size are resized to 256X384.

Figure 2. Partitioned images.

B. *Discrete Wavelet Transform*

Wavelet analysis has been widely used in the image processing because of the unique characteristics and advantages of signal analysis, especially in the field of image retrieval. The powerful time-frequency analysis ability of the wavelet makes the image characteristics can be well described and provides a feasible way for high-accuracy retrieval system [11].

A k-level two-dimensional wavelet transform makes image decomposed into four sub-bands, are known as LL, LH, HL and HH according to the frequency characteristics, each band can be used for each decomposition level. We apply the wavelet transform to decompose the image. Afterward, we continue our work on the low frequency sub-band which contains most of the energy of the image as shown in Figure 3. Figure 4 shows the low frequency sub-band of images sub-blocks.

Figure 3. 1-level two dimensional wavelet transform on sample images in the database.

Figure 4. 1-level two dimensional wavelet transform on images sub-blocks in the WANG database.

C. *Conversion to HSV Color Space*

HSV color space is widely used in computer graphics, visualization in scientific computing and other fields [6, 13]. In this space, hue is used to distinguish colors, saturation is the percentage of white light added to a pure color and value refers to the perceived light intensity. The advantage of HSV color space is that it has the ability to separate chromatic and achromatic components. HSV color space is superior to other color spaces for image processing. The values of H in an HSV color space are the angles of the cylinder, and the values for S are the radii [12].

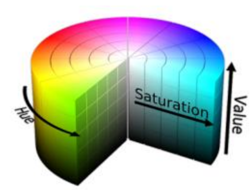

Figure 5. The HSV Color Space [17].

Because of a large range of each component, it is essential to quantify HSV space component to reduce computation and improve efficiency. At the same time, because the human eye to distinguish colors is limited, do not need to calculate all segments. Unequal interval quantization according the human color perception has been applied on H, S, and V components. Based on the color model of substantial analysis, we divide color into eight parts. Saturation and intensity is divided into three parts separately in accordance with the human eyes to distinguish [7]. In accordance with the different colors and subjective color perception quantification, quantified hue (H), saturation (S) and intensity (V) are showed as equation (1).

$$H=\begin{cases}0 \text{ if } h \in [316, 20]\\ 1 \text{ if } h \in [21, 40]\\ 2 \text{ if } h \in [41, 75]\\ 3 \text{ if } h \in [76, 155]\\ 4 \text{ if } h \in [156, 190]\\ 5 \text{ if } h \in [191, 270]\\ 6 \text{ if } h \in [271, 315]\end{cases} \quad S=\begin{cases}0 \text{ if } s \in [0, 0.2]\\ 1 \text{ if } s \in [0.2, 0.7]\\ 2 \text{ if } s \in [0.7, 1]\end{cases} \quad V=\begin{cases}0 \text{ if } v \in [0, 0.2]\\ 1 \text{ if } v \in [0.2, 0.7]\\ 2 \text{ if } v \in [0.7, 1]\end{cases} \tag{1}$$

In accordance with the quantization level above, the H, S, V three-dimensional feature vector for different values of with different weight form one-dimensional feature vector named P:

$$P = \alpha H + \beta S + V \tag{2}$$

Where α is quantified series of H and β is quantified series of S. By experimental studies, α=9 and β=3 produced good quantization levels, then

$$P = 9H + 3S + V \tag{3}$$

In this way, three-component vector of HSV form one-dimensional vector, which quantize the whole color space for the 72 kinds of main colors. This quantification can be effective in reducing the images by effects of light intensity, but also reducing the computational time and complexity.

D. Color Histogram Generation

The color Histogram is derived by first quantizing colors in the image to 72 bins in HSV color space, and counting the number of image pixels in each bin. The color histogram is easy to compute and effective in characterizing both the global and local distribution of colors in an image. In addition, it is robust to translation, rotation and scale. However, one of the weaknesses of color histogram is that when the characteristics of images should not take over all the values, the statistical histogram will appear in a number of zero values. The emergence of these zero values would make similarity measure that does not accurately reflect the color difference between images and statistical histogram method to quantify more sensitive parameters. Therefore, we used a one dimensional vector P by constructing a cumulative histogram of the characteristics of image after using non-interval HSV quantization for P.

As the components of feature vector may have different physical meaning entirely, their rate of change may be very different. It will be much of the deviation of the calculation of the similarity if we do not normalize, so we must normalize the components to the same range.

E. Dominant Color Descriptor

The dominant color descriptor (DCD) is widely applied in the image retrieval taken as one of MPEG-7 color descriptors. DCD describes the representative color distributions and features in an image or a region of interest through an effective, compact and intuitive format. An image retrieval method based on the fixed number's MPEG-7 dominant color descriptor is employed. It describes color features with a smaller number of features. This method is expected to effectively shorten image retrieval time and enhance retrieval performance. The structure of a DCD, F, is defined as:

$$F = \{pi, ci, vi, s\}, i=1, 2\ldots NDCD \tag{4}$$

Where NDCD is the number of dominant colors, s is the spatial coherency value that represents the overall spatial homogeneity of the dominant colors, pi is the percentage of pixels in the image corresponding to the ith dominant color, vi is a vector representing the ith dominant color, and the ci is optional which is the variation of the dominant color values pixels around vi. In the most of application, s is set to zero [8].

F. Gray-level Co-occurrence Matrix (GLCM)

Gray-level Co-occurrence Matrix creates a matrix with the directions and distances between pixels, and then extracts meaningful statistics from the matrix as texture features. GLCM is composed of the probability value, it is defined by P (i, j | d, θ) which expresses the probability of the couple pixels at θ direction and d interval. When θ and d is determined, P (i,j| d,θ) is showed by Pi, j. Distinctly GLCM is a symmetry matrix[15]. Its level is determined by the image gray-level. Elements in the matrix are computed by the equation shown below:

$$P\ (i, j|d,\ \theta) = \frac{P(i,j|d,\theta)}{\Sigma i\ \Sigma j\ \ P(i,j|d,\theta)} \tag{5}$$

GLCM expresses the texture feature according the correlation of the couple pixels gray-level at different positions. The GLCM is computed in four directions for θ=0o, θ=45o, θ=90o, θ=135o. The GLCM four statistical parameters energy, contrast, entropy and inverse difference are computed for the four directions matrices [9, 19].

IV. Intergrated Image Matching

In this paper, we used the same matching procedure proposed in [10]. In our system, a sub-block from query image is allowed to be matched to any sub-block in the target image. However a sub-block may participate in the matching process only once. A bipartite graph of sub-blocks for the query image and the target image is built as shown in Figure 6. The labeled edges of the bipartite graph indicate the distances between sub-blocks. A minimum cost matching is done for this graph.

Since, this process involves too many comparisons; the method has to be implemented efficiently. To this effect, we have designed an algorithm for finding the minimum cost matching based on most similar highest priority (MSHP) principle using the adjacency matrix of the bipartite graph.

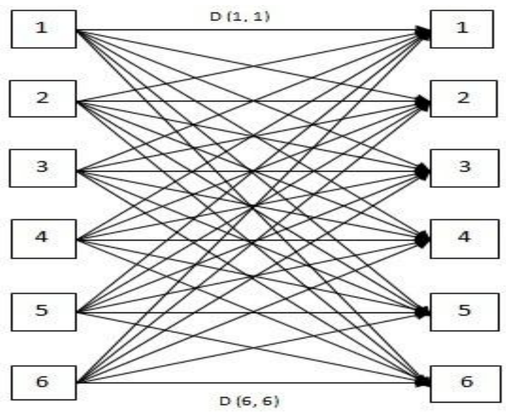

Figure 6. Bipartite graph showing 6 sub-blocks of target and query images.

This will prevent sub-block i of query image and sub-block j of target image from further participating in the matching process. The distances, between i and other sub-blocks of target image and, the distances between j and other sub-blocks of query image, are ignored (because every sub-block is allowed to participate in the matching process only once). This process is repeated till every sub-block finds a matching. The complexity of the matching procedure is reduced from O (n2) to O (n), where n is the number of sub-blocks involved. The integrated minimum cost match distance between images is now defined as:

$$Dqt=\Sigma i \;\; \Sigma j \; dij \qquad (6)$$

Where i=1, 2 …n j=1, 2 ...n. And dij is the best-match distance between sub-block i of query image q and sub-block j of target image t and Dqt is the distance between images q and t.

V. Experimental Setup

A. Experimental Setup

Data set: Wang's [16] dataset comprising of 1000 Corel images with ground truth. The image set comprises100 images in each of 10 categories. The images are of the size 256 x 384.

Figure 7: One example image from each of the 10 classes of the WANG database.

The WANG database is a subset of the Corel database of 1000 images which have been manually selected to be a database of 10 classes of 100 images each. The images are subdivided into 10 classes such that it is almost sure that a user wants to find the other images from a class if the query is from one of these 10 classes. This is a major advantage of this database because due to the given classification it is possible to evaluate retrieval results. One example of each class can be seen in Figure 7. This database was used extensively to test the different features because the size of the database and the availability of class information allows for performance evaluation.

Feature set: The feature set comprises color and texture descriptors computed for each sub-block of an image as we discussed in section III.

B. Results and Analysis

Three proposed scenarios are presented. In the First Scenario, The original image is converted into HSV color space and the partitioned sub-blocks proceed to the dominant color extraction and texture analysis. In the second scenario, a k-level wavelet transform is applied to the whole image and the low frequency sub-band is fed into the RGB to HSV converter, then the produced HSV sub-band proceeds to color and texture analysis. In the third Scenario, the discrete wavelet transform is applied to each image sub-block producing low frequency sub-band for every tile, each tile proceeds to color and texture feature extraction.

The results are benchmarked with some of the existing systems using the same database. This experiment used each image in each class as a query image. It was carried out with the number of retrieved images set as 20 to compute the precision P of each query image and finally obtain the average precision P/100 (100 images of a class).

The experiments were carried out on a Core i3, 2.2 GHz processor with 4GB RAM using MATLAB.

The precision and recall measurements of Mehtre et al. [20] are often used to describe the performance of an image retrieval system. The precision and recall are defined as follows:

$$\text{Precision}=\frac{\text{Number of relevant images retrieved}}{\text{Total number of images retrieved}} \qquad (7)$$

$$\text{Recall}=\frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}} \qquad (8)$$

The proposed scenarios are compared with other retrieval systems. The First system is image retrieval based on HSV color and GLCM texture features of the whole image, the second one is image retrieval based on HSV color and GLCM texture features of an image sub-blocks with one to one matching principle, the third system is image retrieval based on HSV color and GLCM texture features of an image sub-blocks with MSHP principle, those comparisons appear in Tables 1, 2, 3 and 4. Another comparison of average precision is carried out with the system of Jhanwar et.al. [21] and the retrieval system Haung and Dai's [22] which are examined on the same WANG database. The comparison appears in Table V.

It is obvious that the proposed models have achieved better average precision of various images than the other systems illustrated.

TABLE I. THE AVERAGE PRECISION OF DIFFERENT METHODS.

Class	Average Precision					
	HSV + GLCM	HSV + GLCM of sub-blocks	HSV + GLCM of sub-blocks + MSHP	Proposed Scenario 1	Proposed Scenario 2	Proposed Scenario 3
Africa	0.34	0.41	0.44	0.52	0.52	0.66
Beaches	0.21	0.32	0.5	0.57	0.59	0.71
Building	0.24	0.37	0.45	0.5	0.53	0.61
Bus	0.51	0.66	0.75	0.81	0.77	0.83
Dinosaur	0.39	0.43	0.61	0.67	0.71	0.94
Elephant	0.26	0.39	0.39	0.45	0.48	0.59
Flower	0.81	0.87	0.87	0.92	0.93	0.98
Horses	0.28	0.35	0.35	0.38	0.41	0.57
Mountain	0.2	0.34	0.34	0.41	0.43	0.6
Food	0.25	0.31	0.31	0.37	0.35	0.54
Average	0.349	0.445	0.501	0.56	0.572	0.703

TABLE II. THE AVERAGE RECALL OF DIFFERENT METHODS

Class	Average Recall					
	HSV + GLCM	HSV + GLCM of sub-blocks	HSV + GLCM of sub-blocks + MSHP	Proposed Scenario 1	Proposed Scenario 2	Proposed Scenario 3
Africa	0.22	0.28	0.28	0.31	0.5	0.76
Beaches	0.15	0.19	0.27	0.35	0.42	0.72
Building	0.2	0.22	0.24	0.32	0.39	0.61
Bus	0.31	0.32	0.32	0.36	0.43	0.57
Dinosaur	0.23	0.36	0.38	0.38	0.51	0.52
Elephant	0.21	0.29	0.27	0.3	0.45	0.51
Flower	0.45	0.48	0.49	0.56	0.71	0.88
Horses	0.19	0.25	0.25	0.33	0.36	0.56
Mountain	0.23	0.24	0.25	0.28	0.52	0.55
Food	0.18	0.23	0.21	0.25	0.3	0.54
Average	0.237	0.286	0.296	0.344	0.459	0.622

TABLE III. PRECISION VS. NUMBER OF RETURNED IMAGES

Number of of returned images	Precision					
	HSV + GLCM	HSV + GLCM of sub-blocks	HSV + GLCM of sub-blocks + MSHP	Proposed Scenario 1	Proposed Scenario 2	Proposed Scenario 3
20	0.38	0.42	0.5	0.55	0.57	0.71
40	0.32	0.37	0.46	0.51	0.56	0.69
60	0.27	0.35	0.43	0.47	0.52	0.63
80	0.2	0.31	0.38	0.43	0.45	0.59

TABLE IV. RECALL VS. NUMBER OF RETURNED IMAGES

Number of of returned images	Precision					
	HSV + GLCM	HSV + GLCM of sub-blocks	HSV + GLCM of sub-blocks + MSHP	Proposed Scenario 1	Proposed Scenario 2	Proposed Scenario 3
20	0.22	0.25	0.28	0.31	0.5	0.76
40	0.15	0.19	0.27	0.35	0.42	0.72
60	0.2	0.22	0.24	0.32	0.39	0.61
80	0.31	0.32	0.33	0.36	0.43	0.57

The Figure 8 is a graph showing the comparison of average precision obtained by proposed systems with other retrieval systems.

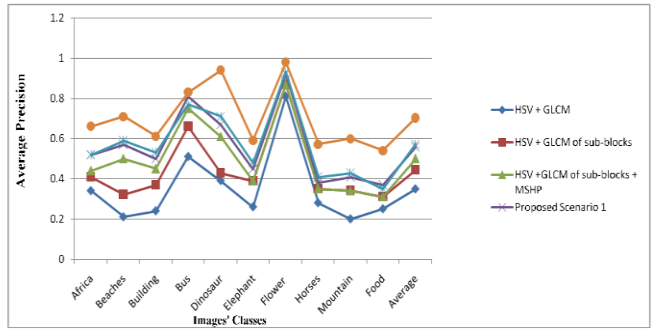

Figure 8. The average precision of different methods.

The Figure 9 is a graph showing the comparison of average recall obtained by proposed systems with other retrieval systems.

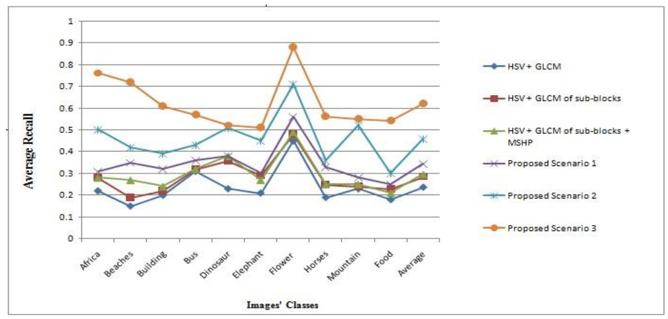

Figure 9. The average recall of different methods.

Figure 10 is a graph showing the comparison of precision versus the number of images returned obtained by proposed systems with other retrieval systems.

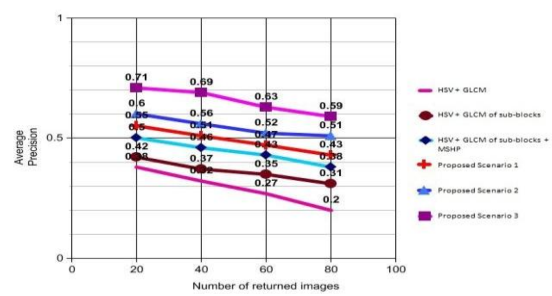

Figure 10. Precision vs. Number of images returned.

The Figure 11 is a graph showing the comparison of recall versus the number of images returned.

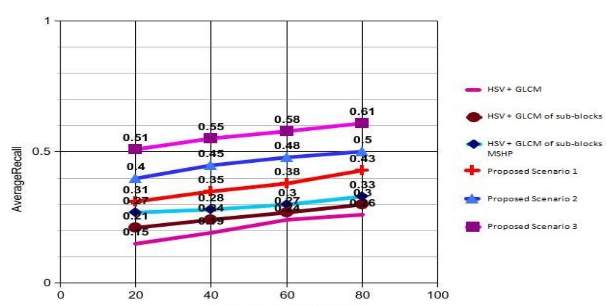

Figure11. Recall vs. No. of images returned.

TABLE V. THE AVERAGE PRECISION OF DIFFERENT SYSTEMS

Class	Average Precision				
	Jhanwar et.al.[21]	Huang and Dai's[22]	Proposed Scenario 1	Proposed Scenario 2	Proposed Scenario 3
Africa	0.45	0.42	0.52	0.52	0.66
Beaches	0.39	0.45	0.57	0.59	0.71
Building	0.37	0.41	0.5	0.53	0.61
Bus	0.74	0.85	0.81	0.77	0.83
Dinosaur	0.91	0.59	0.67	0.71	0.94
Elephant	0.3	0.42	0.45	0.48	0.59
Flower	0.85	0.89	0.92	0.93	0.98
Horses	0.57	0.59	0.38	0.41	0.57
Mountain	0.29	0.27	0.41	0.43	0.6
Food	0.37	0.43	0.37	0.35	0.54
Average	0.524	0.532	0.56	0.572	0.703

The Figure 12 is a graph showing the comparison of average precision obtained by proposed systems with other retrieval systems illustrated in [20,21].

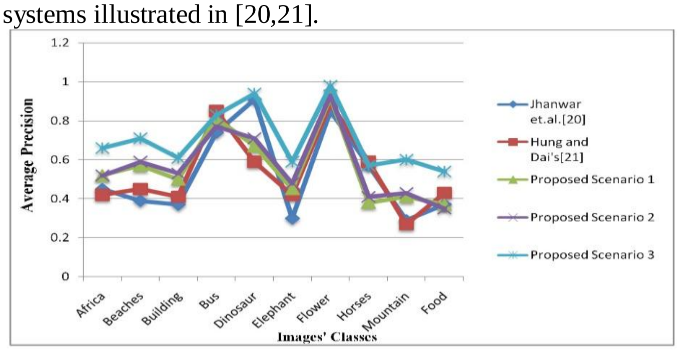

Figure 12. The average precision of different systems.

VI. CONCLUSION

The proposed systems use color and texture features in addition to discrete wavelet transform to characterize a color image for image retrieval. HSV, color histogram and DCD can describe color features of the pixels in an image, while, GLCM can describe texture distribution. These features are invariant to scale, translation and rotation. Since the features used can describe different properties of an image, the proposed systems integrate these features to retrieve accurate images. The experimental results show that the proposed systems outperform the other retrieval systems stated. As a future work, we plan to reduce the semantic gap between the local features and the high-level user semantics to achieve higher accuracy.

REFERENCES

[1] Y. Rui, T.S. Huang, S. Mehrotra, “Retrieval with relevance feedback in MARS,” in Proc. of the IEEE Int'1 Conference on image Processing, NewYork, pp. 815–818, 1997.

[2] H.T. Shen, B.C. Ooi, K.L. Tan, “Giving meanings to www images,” in Proc. of ACM Multimedia, pp. 39-48,2000.

[3] M. B. Rao, B.P. Rao, A. Govardhan, “Content-based Image Retrieval using Dominant Color and Texture features,” pp. 118-123, Feb 2011.

[4] Y. Rui, C. Alfred, T.S. Huang, “Modified descriptor for shape representation, a practical approach,” in Proceedings of First Int's workshop on Image Database and Multimedia Search, 1996.

[5] C. Lihua, L. Wei, L. Guohui, “Research and Implementation of an Image Retrieval Algorithm based on Multiple Dominant Colors,” Journal of Computer Research & development, Vol. 36, pp. 96-100, 1999.

[6] JI-QUAN MA, “Image Retrieval with HSV Color Space and Texture Features,” proceedings of the international conference on Web Information Systems and Mining, 2009.

[7] J. ma, “Content-Based Image Retrieval with HSV Color Space and Texture Features,” International Conference on Web Information Systems and Mining, pp. 61-63, 2009.

[8] H. Shao, Y. Wu, W. Cui, J. Zhang, “Image Retrieval Based on MPEG-7 Dominant Color Descriptor,” The 9th International Conference for Young Computer Scientists, pp. 753-757, 2008.

[9] R. Mehta, N. Mishra and S. Sharma, “Color - Texture based Image Retrieval System,” International Journal of Computer Applications, Vol. 24, pp.24-29, June 2011.

[10] J. Li, J.Z. Wang, and G. Wiederhold, “IRM: Integrated Region Matching for Image Retrieval,” in Proc. of the 8th ACM Int. Conf. on Multimedia, pp. 147-156, Oct. 2000.

[11] P.S. Hiremath , Jagadeesh Pujari, “Content Based Image Retrieval using Color, Texture and Shape features,” 15th International Conference on Advanced Computing and Communications, IEEE Computer Society, pp. 780-784, 2007.

[12] R.C. Gonzalez, R.E. Woods, “Digital Image Processing,” second ed., Prentice-Hall, Upper Saddle River, NJ, 2002.

[13] Y. Rui, T.S. Huang, S.F. Chang, “Image retrieval: current techniques, promising directions, and open issues,” Journal of Visual Communication and Image Representation 10, pp 39-62., 1999.

[14] C.R. Shyu, "Local versus Global Features for Content-Based Image Retrieval," IEEE Workshop on Content-Based Access of Image and Video Libraries, 1998.

[15] B. S. Manjunath, W.Y. Ma, “Texture feature for browsing and retrieval of image data,” IEEE Transaction on PAMI, vol.18, no.8, pp. 837-842, 1996.

[16] Wang's dataset http://wang.ist.psu.edu/

[17] http://en.wikipedia.org/wiki/HSV_color_space

[18] M.R. Mustaffa, F. Ahmad, R.W. Rahmat, R. Mahmod, “Content-Based Image Retrieval Based On Color-Spatial Features,” Malaysian Journal of Computer Science, Vol. 21, pp. 1-12, 2008.

[19] P. Howarth and S. Ruger, “Robust texture features for still-image retrieval,” IEE. Proceedings of Visual Image Signal Processing, Vol. 152, No. 6, December 2005.

[20] G.P. Babu, B.M. Mehtre, M.S. Kankanhalli, “Color indexing for efficient image retrieval,” Multimedia Tools and Applications, pp. 327–348, 1995.

[21] N. Jhanwar, S. Chaudhuri, G. Seetharaman, B. Zavidovique, “Content based image retrieval using motif co-occurrence matrix,” Image and Vision Computing 22, pp.1211–1220, 2004.

[22] P.W. Huang, S.K. Dai, “Image retrieval by texture similarity,” Pattern Recognition 36, pp. 665–679, 2003.