DOI: 10.24996/ijs.2020.61.8.26

ISSN: 0067-2904

Image Retrieval Using Data Mining Technique

Sarmad T. Abdul-Samad*, Sawsan Kamal
Department of Computer Science, College of Science, AL-Nahrain University, Baghdad, Iraq

Received: 15/7/2019 Accepted: 19/11/2019

Abstract
Even though image retrieval is considered as one of the most important research areas in the last two decades, there is still room for improvement since it is still not satisfying for many users. Two of the major problems which need to be improved are the accuracy and the speed of the image retrieval system, in order to achieve user satisfaction and also to make the image retrieval system suitable for all platforms. In this work, the proposed retrieval system uses features with spatial information to analyze the visual content of the image. Then, the feature extraction process is followed by applying the fuzzy c-means (FCM) clustering algorithm to reduce the search space and speed up the retrieval process. The experimental results show that using the spatial features increases the system accuracy and that the clustering algorithm speeds up the image retrieval process. This shows that the proposed system works with texture and non-texture images.

Keywords: CBIR, HSV 3D Histogram, GLCM, Fuzzy c-means clustering.

استرداد الصور باستخدام تقنية تنقيب البيانات

سرمد ثائرعبد الصمد* ، سوسن كمال
قسم علوم الحاسوب ، كلية العلوم ، جامعة النهرين ، بغداد ، العراق

الخلاصة
على الرغم من أن استرداد الصور يعتبر أحد أهم مجالات البحث في العقدين الماضيين ، إلا أنه لا يزال هنالك مجال للتطوير. لذلك ، لا يزال غير مرضي لكثير من المستخدمين. تتمثل مشكلات استرداد الصور في الحاجة الى تحسين دقة النظام بالاضافة الى زيادة سرعته من أجل تحقيق رضا المستخدم وأيضًا جعل نظام استرداد الصور مناسبًا لجميع المنصات. في هذا البحث ، يستخدم نظام الاسترداد المقترح ميزات مع معلومات مكانية لتحليل المحتوى المرئي للصورة. بعد ذلك ،يتبع عملية استخراج الميزات تطبق خوارزمية التجميع الضبابية (FCM) لغرض التقليل من المساحة التي يتم البحث فيها و زايدة سرعة عملية الصورة. تُظهر النتائج التجريبية أن استخدام الميزات المكانية يزيد من دقة النظام بالإضافة إلى أن خوارزمية التجميع تزيد من سرعة عملية استرداد الصورة. ويوضح أيضًا أن النظام المقترح يعمل مع صور النسيجية وغير نسيجية.

Introduction
In recent years, the development of visual and multimedia applications led to the widespread of digital images. Also, the revolution of image messaging applications and the usage of photos in social media platforms make thousands of images shared each second, leading to millions of images being accumulated each day. However, managing and organizing these digital images presents a problem.

*Email: Sarmad.thaer991@rdd.edu.iq.

Thus, the concept of indexing and retrieval was introduced to overcome this issue. Indexing relates to "how to store images in database to retrieve them (through querying) more efficient", whereas Retrieval relates to "how to retrieve images that are most relevant to the query image from images in database" [1].

Two retrieval methods are used to retrieve digital images from the database. The first method is known as Text-Based Image Retrieval (TBIR) that depends on metadata associated with each image and uses traditional query techniques to retrieve images from the database by a keyword. This method works well with small digital image databases and it is not good with a huge database. This is because it is very difficult and time consuming when performing keyword generation for such databases. The most important problem in TBIR is that different users use different words to describe the same image (subjectivity of the human). This problem will adversely affect the efficiency of the text-based image search [2]. Hence, a need for a more effective image retrieval system has appeared. This needed system must perform an automatic indexing and retrieving. Hence, the second method depends on image content for indexing and retrieving. Therefore, this method is generally known as Content-Based Image Retrieval (CBIR). CBIR also, known as Query By Image Content (QBIC) and Content-Based Visual Information Retrieval (CBVIR) was introduced in the 1990s. It depends on analysis of the visual content of the digital image which can be analyzed by extracting image features such as color, texture and shape, which are called low level features [3].

The data should provide knowledge and information for decision making. Hence, data mining is the concept of data analysis and the process of finding an interesting pattern from a large amount of data. The data is stored in different databases such as data warehouse, World Wide Web, and external data sources. The goals of data mining are fast retrieving of data or information, knowledge discovery from the databases, identification of hidden patterns and other patterns that are not explored before, reducing the level of complexity, saving processing time, and many other goals which are all useful in CBIR. Data mining is occasionally called Knowledge Discovery from Database (KDD) [4].

In order to design and implement generic CBIR applications, both advanced algorithms in image understanding field and advances in computer hardware are needed. Therefore, most efforts are directed to specific CBIR applications [5]. A wide range of CBIR applications varied from personal to medical diagnoses, crime prevention, education, military and many others [6].

This study proposed an approach of CBIR system for both texture and non-texture images. Color and texture features with spatial information are used to analyze the image visual content. The proposed system used a segmentation technique followed by feature extraction. Also, the system employed clustering method on the images in the database. The need of using segmentation technique and clustering algorithm is to speed up the image retrieval process as well as to increase the system accuracy.

Materials and Methods

The next sections show how to design and implement the proposed system and describe in detail every material and algorithm needed for this work.

Image database

In this work, INRIA Holidays database is used to evaluate the proposed system. This database contains some personal holiday photos. It includes a very large variety of scenes such as pyramids, forests, sunset, boats, etc. The dataset contains 1000 high resolution JPEG images [7].

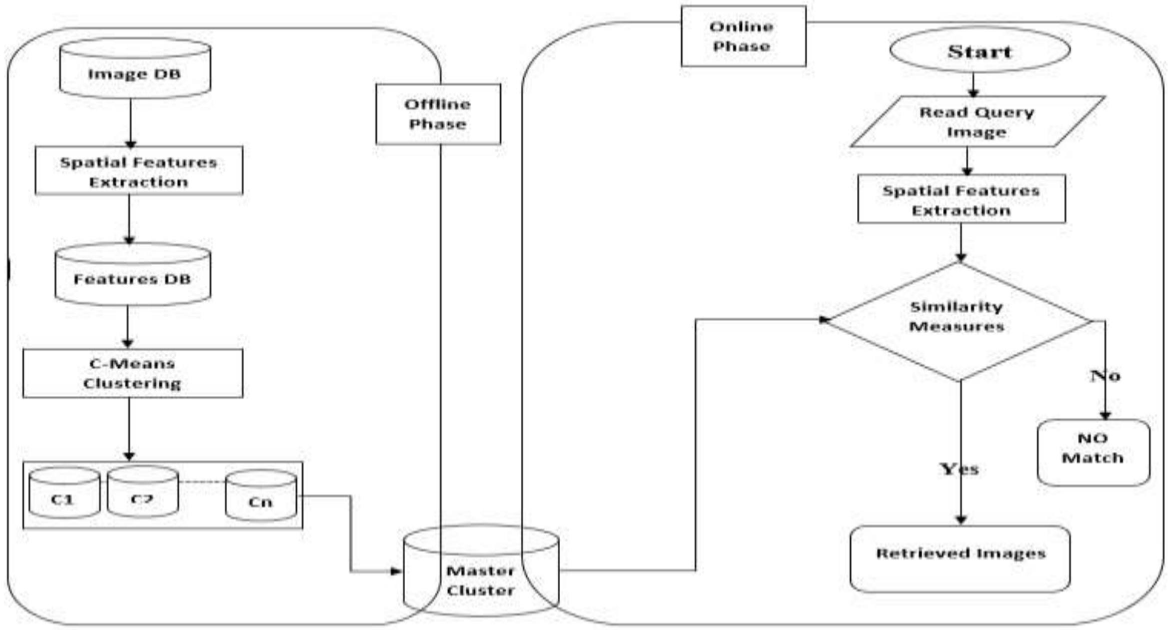
System Implementation Environment

The proposed image retrieval system is implemented using C# .Net 2015 programming language which is used under Microsoft Visual Studio IDE (Integrated Development Environment). This IDE is installed on Windows 10, 64-bits OS produced by Microsoft company. CSV files (Comma-Separated Values) are used to store the features vectors and produce a feature database. The hardware platform (physical component) that the proposed image retrieval system works on is as follows:

- CPU (Intel core i7 2.7 GHz chipset).
- Memory (RAM 8GB).
- Graphics card (VGA 1GB AMD Radeon).
- STORAGE (SSD 240 GB).

The proposed image retrieval system

An overview of the proposed retrieval system architecture is presented in Figure-1. The suggested system consists of two main stages: *offline stage* and *online stage.* In the offline stage, the features of images are extracted and saved in features database.



Figuer 1- Flow chart of the proposed system

The extracted feature vectors are clustered into many clusters that have similar values for each cluster. While, in the online stage (retrieval stage), the features of the query image are extracted and compared with the clusters' centroids in the master cluster. Then, the images in the cluster that have similar values are retrieved.

HSV color space

The color is an important descriptor. Using the information of color helps to extract the details of the image, such as the object of interest. HSV color space is widely used in computer graphics and is a more intuitive way of describing color. HSV stands for Hue, Saturation and Value. Hue represents a mixture of two primary colors, one of which is at full intensity. Saturation indicates the mean of the color purity, which is how much of pure spectrum is diluted by mixing with white light/color in it. Value indicates the Chroma notation of intensity that is called as brightness. In a short description, the chromatic information of the color will be given by HS, while V will give the intensity information for that color [8].

Three-dimensional HSV color histogram

The color histogram plays an important role in image analysis. 3D HSV histogram is widely used in image retrieval field since HSV color space is close to the human perception. In this paper, a region-based histogram is used. Image is segmented into 9 equally sized segments to increase the gained spatial information, and a separated histogram is calculated for each segment. Color histogram is easy to compute and it is one of the most effective descriptors in characterizing the distribution of colors in an image [9]. Increasing the number of bins of the color histogram will increase its power of discrimination. However, the large number of bins will increase the computational time cost and will be inappropriate for building efficient indices for image databases and vice versa. Thus, for better feature extraction, a proper selection of bins is required. The number of bins can be selected directly proportional to the size of the dataset. Less number of bins will be taken if the images' database is small [10]. In this paper, 9 bins are taken for Hue, 5 for Saturation, and 4 for Value. As mentioned before, the chromatic information of the color will be given by HS and V will give the intensity information. For that reason, the bins for HS are selected to be more than those for V. Thus, the total number of features extracted is $9\times5\times4=180$ for each segment and $180\times9=1620$ for each image.

Color Correlogram

The spatial information of the extracted feature is the main drawback of the color histogram. For example, all the images shown in Figure-2 have the same color proportion, but different spatial distribution.

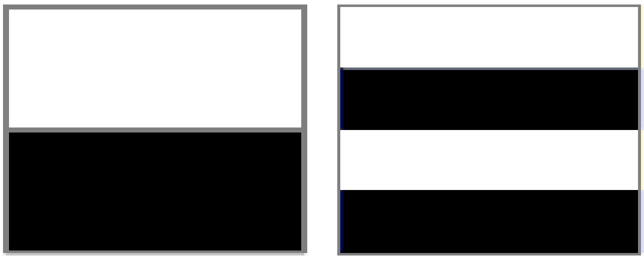


Figure 2- Images having the same color proportions but different spatial distribution.

Correlation histogram (correlogram) tries to fix this histogram's drawback by taking the spatial correlation of color distribution into account. It shows how the spatial correlation between pairs of colors is changing with distance [11].

A color correlogram can be represented as a table indexed by color pairs (i,j), where the d^{th} entry specifies the probability of finding a pixel with I color at a distains d form the pixel with j color in the image.

Let [D] denote the set of distances $\{d_1,\ldots,d_D\}$. Then the color correlogram for the image I for color pair (c_i,c_j) at a distance d can be denoted as [12]:

$$C^{d}_{c_{i},c_{j}}(I) = probability_{\,P1\in I_{c(i)},P2\in I}\left[P_2 \in I_{c(j)} \mid |P_1 - P_2| = d\right] \quad (1)$$

where:

P_1, P_2 are the probabilities of the color occurrence

Auto-correlogram shows the spatial correlation between only identical colors in an image. The auto-correlogram of the image I for color Ci at a distance d can be denoted as:

$$AC^{d}_{c_i}(I) = C^{d}_{c_{i},c_{i}}(I) \quad (2)$$

The experiment shows that both correlogram and auto-correlogram are computationally expensive. Hence, correlogram with a small number of color and distance values still gives a very good result without increasing the computational cost [12]. Thus, it is used in this work.

Row sum and Column sum

Row and column sums are image features that have very crucial information about the image. For two similar images, they are nearly the same. In this work, these features are calculated in RGB color space. At the beginning, the image is resized to 256×256 in order to have the same numbers of rows and columns for each image, and then the red channel, green channel and blue channel are extracted to have a full indication about each color combination in each row and column. Then, the sum for each row and column of each channel is calculated. These features are another type of the spatial features that give an indication about the spatial relationship between pixels [13].

Texture Feature

The spatial arrangement of the pixel within an image is defined as image texture. The GLCM statical approach is used to extract texture features in the current paper. In this approach, gray level spatial dependence of texture is explored. A co-occurrence matrix, $I_{d,\theta}(i, j)$, is a matrix in which the $(i, j)^{th}$ element represents the frequency of occurrence of two pixels separated with d distance, and in the direction θ with grey levels i and j. The variations of texture in a region can be captured through the co-occurrence matrix by various θ and d. That is, the co-occurrence matrix characterizes the spatial interrelationships of the grey levels in a textured pattern and it is invariant under monotonic grey-level transformations [14]. The extracted texture features are given in Table-1 [15].

Table 1-GLCM features used for extraction.

Feature	Formula
$Energy$	$\sum_{i,j} P(i,j)^2$
$Entropy$	$\sum_{i,j} P(i,j) \log P(i,j)$
$Homogeneity$	$\sum_{i,j} \frac{P(i,j)}{1+(i-j)^2}$
$Contrast$	$\sum_{i,j} (i-j)^2 P(i,j)$
$Correlation$	$\sum_{i,j} \frac{(i-\mu_i)(j-\mu_j)P(i,j)}{\sigma_i \sigma_j}$

Feature vectors clustering by fuzzy C-means

The clustering of data is a process of grouping data in classes or clusters. Each object in a cluster has high similarity in the extracted features, but objects' features in other clusters are dissimilar. In this work, the fuzzy c-means clustering algorithm is applied to the feature vectors stored in the features database. After extracting the features for each image in database and saving the extracted feature vectors in feature database, the latter is portioned into a number of groups each with similar features. Then, each cluster center is stored into a master cluster. The distance is calculated between the query image feature vector and the clusters' centers in the database. If the distance is minimum compared to all cluster centers in the master cluster, then the query image is similar to the images in that cluster.

The proposed system steps

A. Dataset Collection

- INRIA holidays Dataset is used
- Convert the image from RGB color space to HSV color space

B. Feature Extraction

- Feature extraction is carried out using spatial features such as region-based histogram, auto-correlogram, row sum and column sum, and texture features using GLCM.
- The extracted features for each image are registered with the corresponding feature vector in a CSV file.
- All the feature vectors are forwarded to fuzzy c-means clustering algorithm.
- The centroid for each cluster is put into a Master-cluster.

Similarity Measure

The similarity between the query image and the images in the database is found out using Euclidean distance. The distance between query image and database image is calculated by the following formula

$$\mathrm{Euclidean}(\mathrm{x},\mathrm{y}) = \sqrt{\sum_{i=0}^{n}(x_i - y_i)^2} \quad (3)$$

Experimental results

16 clusters selected to form the image database were used to test the system, which represents 20% of the image database. Many experiments with different combinations of spatial features were conducted on these clusters to reach the acceptable accuracy and speed. Based on these experiments, the weakness and strength points of the system will be explained. The precision, recall and accuracy of each cluster were calculated. The next subsections will show the results of the implemented experiments

Experiment using Histogram and GLCM features combination

In this experiment, a combination of three dimensional HSV color histogram for five segments and texture features extracted using GLCM was experimented. The results are shown in Table-2.

Table 2- The experimental results of 3D Histogram and GLCM features

No.	Cluster ID	Precision	Recall	Accuracy%
1	Under water	1	0.2	96.52
2	Egyptian pyramids	1	0.071428571	89.51
3	Mountains	1	0.166666667	95.68
4	Boating Excursions	1	0.142857143	94.87
5	Sunsets cross the ocean	1	0.125	94.06
6	Forest	1	0.125	94.06
7	Flowers	1	0.166666667	95.68
8	Coral reefs	1	0.142857143	94.87
9	Food plate	1	0.25	97.36
10	Child	1	0.833333333	99
11	Algae	1	0.166666667	95.68
12	Fish	1	0.142857143	94.87
13	Blue sky and cloud	1	0.166666667	95.68
14	Traffic signs	0.423076923	1	86.48
15	Garden	0.059701493	1	43.24
16	Sunset	1	0.2	96.52
Average		0.9	0.3	91.5

Experiment using Histogram, row sum and column sum, and GLCM features combination.

In this experiment, a combination of three-dimensional HSV color histogram for five segments and texture features extracted using GLCM were experimented as in the first experiment. A gray level row sum and column sum were also added to the tested features combination. The results of the current experiment are shown in Table-3.

Table 3- The experimental results of 3D Histogram, GLCM, and gray level row and column features.

No.	Cluster ID	Precision	Recall	Accuracy%
1	Under water	1	0.2	96.52
2	Egyptian pyramids	1	0.071428571	89.51
3	Mountains	0.454545455	0.833333333	93.75
4	Boating Excursions	1	0.142857143	94.87
5	Sunsets cross the ocean	0.666666667	0.25	94
6	Forest	1	0.125	94.06
7	Flowers	1	0.166666667	95.68
8	Coral reefs	1	0.142857143	94.87
9	Food plate	1	0.25	97.36
10	Child	1	0.166666667	95.68
11	Algae	1	0.166666667	95.68
12	Fish	1	0.142857143	94.87
13	Blue sky and cloud	1	0.166666667	95.68
14	Traffic signs	0.130952381	1	34.23
15	Garden	1	0.25	97.36
16	Sunset	1	0.2	96.52
Average		0.89	0.26	91.29

Experiment using Histogram, row sum and column sum, GLCM, and correlogram features combination

The current experiment tested the combination of features that were used in the previous experiment. In addition to the previously extracted features, a correlogram was added to the combination. The correlogram was computed for quantized HSV color space. The results of the current experiment are shown in Table-4.

Table 4- The experimental results of 3D Histogram, GLCM, gray level row sum and column, and HSV correlogram features.

No.	Cluster ID	Precision	Recall	Accuracy%
1	Under water	1	0.2	96.52
2	Egyptian pyramids	0.237288136	1	59.45
3	Mountains	1	0.666666667	98.23
4	Boating Excursions	1	0.142857143	94.87
5	Sunsets cross the ocean	0.777777778	0.875	97.32
6	Forest	1	0.375	95.68
7	Flowers	1	0.166666667	95.68
8	Coral reefs	1	0.142857143	94.87
9	Food plate	1	0.75	99.1
10	Child	1	0.166666667	95.68
11	Algae	1	0.166666667	95.68
12	Fish	1	0.142857143	94.87
13	Blue sky and cloud	1	0.333333333	96.52
14	Traffic signs	0.714285714	0.454545455	93.16
15	Garden	1	0.5	98.23
16	Sunset	0.133333333	0.4	85.96
Average		0.86	0.4	93.24

Experiment using Histogram, row sum and column sum, GLCM, and auto-correlogram features combination

In this test, a combination of a gray-level histogram for nine segments, gray-level row and column sum, gray-level auto-correlation, and GLCM, was used to extract the texture features. The gray-level color space was used in order to decrease the feature vector size. The results of the current experiment are shown in Table-5.

Table 5- The experimental results of gray-level histogram, GLCM, gray level row and column, and gray level auto-correlation features.

No.	Cluster ID	Precision	Recall	Accuracy%
1	Under water	0.25	0.2	93.91
2	Egyptian pyramids	0.297297297	0.785714286	74.56
3	Mountains	0.238095238	0.833333333	84.82
4	Boating Excursions	1	0.142857143	94.87
5	Sunsets cross the ocean	1	0.75	98.23
6	Forest	1	0.375	95.68
7	Flowers	0.8	0.666666667	97.34
8	Coral reefs	0.6	0.857142857	95.53
9	Food plate	0.5	0.75	96.42
10	Child	1	0.166666667	95.68
11	Algae	1	0.5	97.36
12	Fish	0.5	0.142857143	94
13	Blue sky and cloud	1	0.333333333	96.52
14	Traffic signs	1	0.090909091	91.73
15	Garden	0.6	0.75	97.32
16	Sunset	1	0.4	97.36
Average		0.73	0.48	93.83

Experiment using Histogram, row and column sum, and GLCM features combination

The current experiment tested a three-dimensional HSV color histogram and GLCM for texture feature extraction. Also, the row sum and column sum features were extracted, but with an RGB color space. For each channel, a row sum and column sum features were calculated. The results of the current experiment are shown in Table-6.

Table 6 - The experimental results of 3D level histogram, GLCM and RGB row, and column sum features.

No.	Cluster ID	Precision	Recall	Accuracy%
1	Under water	0.454545455	1	94.59
2	Egyptian pyramids	0.4	0.857142857	82.3
3	Mountains	1	0.166666667	95.68
4	Boating Excursions	1	0.142857143	94.87
5	Sunsets cross the ocean	1	0.125	94.06
6	Forest	0.833333333	0.625	96.49
7	Flowers	1	0.833333333	99.1
8	Coral reefs	0.2	0.142857143	91.45
9	Food plate	1	1	100
10	Child	0.75	0.5	96.49
11	Algae	1	0.833333333	99.1
12	Fish	0.208333333	0.714285714	81.41
13	Blue sky and cloud	1	0.833333333	99.1
14	Traffic signs	1	0.272727273	93.27
15	Garden	1	0.75	99.1
16	Sunset	1	0.8	99.1
Average		0.8	0.59	94.76

Experiment using Histogram, row and column sum, GLCM, and auto-correlogram features combination

The current experiment is the one selected for implementing our proposed system. In this experiment, a combination of three-dimensional HSV histogram for nine segments, row sum and column sum for RGB color space, HSV auto-correlogram, and GLCM for texture features are extracted. The results of this experiment are better than those of the previous experiments. The results are shown in Table-7 below.

Table 7- The experimental results of 3D level histogram, GLCM, RGB row, and column, HSV auto-correlogram features.

No.	Cluster ID	Precision	Recall	Accuracy%
1	Under water	0.833333333	1	99
2	Egyptian pyramids	0.875	1	98.19
3	Mountains	0.857142857	1	99
4	Boating Excursions	0.857142857	0.857142857	98.21
5	Sunsets cross the ocean	0.857142857	0.75	97.34
6	Forest	1	0.5	96.52
7	Flowers	1	0.666666667	98.23
8	Coral reefs	0.875	1	99
9	Food plate	0.8	1	99
10	Child	1	1	100
11	Algae	1	1	100
12	Fish	0.875	1	99
13	Blue sky and cloud	1	1	100
14	Traffic signs	0.916666667	1	99
15	Garden	1	1	100
16	Sunset	1	1	100
Average		0.921	0.923	98.94

Result Analysis

In the previous section, a spatial feature descriptor (3D HSV histogram, row sum and column sum, auto-correlogram, correlogram, and GLCM) was extracted with different color spaces and different quantization schemas. Six experiments with different combinations of the extracted features were conducted until reaching the desired result. At first, the time needed to extract the features for one image in each of the six experiments is shown in Figure-3 below.

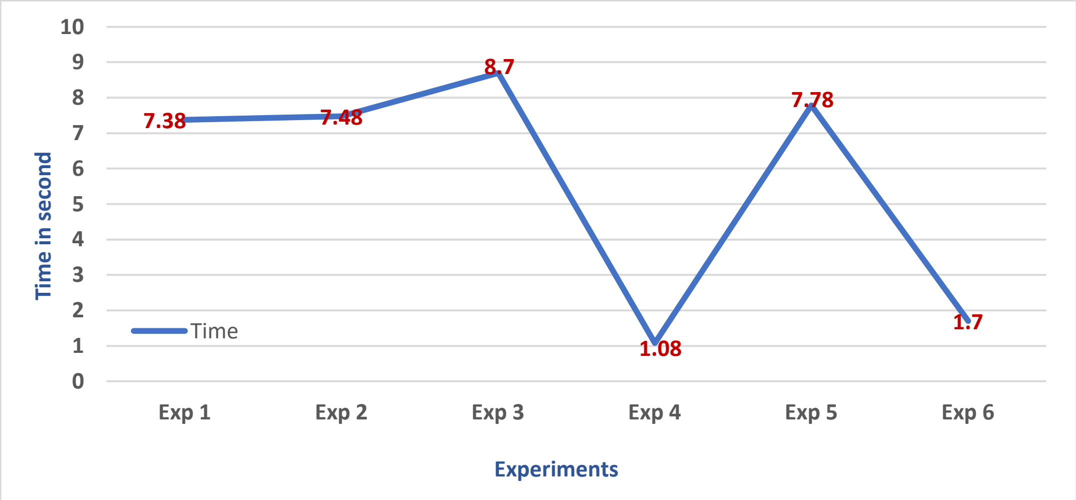


Figure 3- Changing in time for each experiment feature extraction.

The change in time can be very clear in experiments 4 and 6. This change is attributable to image segmentation methods. In experiments 1, 2, 3 and 5, five segments were used to well extract the histogram, whereas in experiments4 and 6, nine fixed size segments were used. The time needed to segment the image into five regions was seven seconds. Thus, using fixed size segmentation reduced features extraction time.

The best retrieval result was obtained when using 3D HSV histogram from nine fixed segments, RGB color, row sum and column sum, HSV auto-correlogram, and GLCM. While the worst performance was given by the first experiment when using 3D HSV histogram and GLCM. From the experiments, it is clear that using spatial features will give a good result, since spatial features give information about the relationship between pixels. It is clear that adding row sum and column sum increases the accuracy of the results. In this proposed system, twelve GLCM were extracted to analyze the texture feature using four orientations (0^o,45^o,90^o and135^o) and three distances (1,4 and 8) between pixels in each direction. These GLCMs will describe the texture of the image as much as possible. Also, this will increase the efficiency of the system.

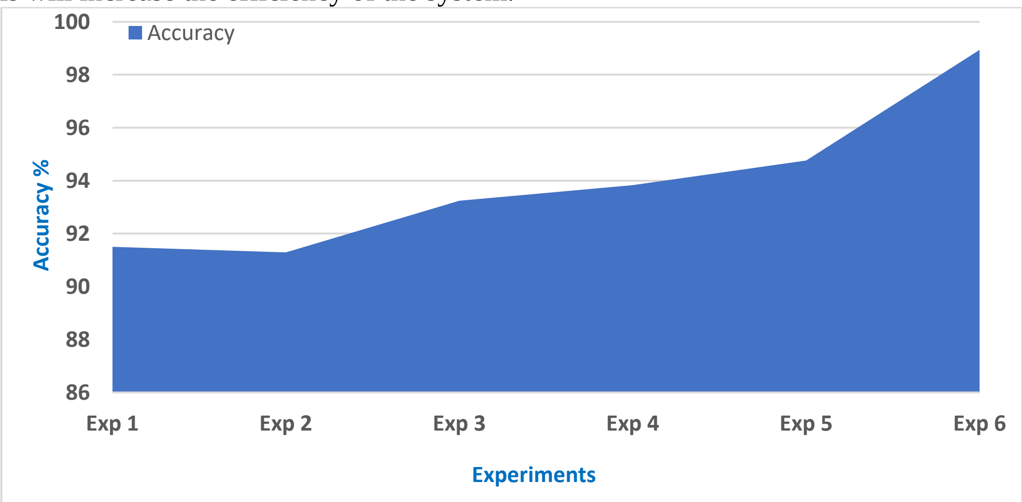


Figure 4- Experiments accuracies.

The size of the feature vector was increased by using the correlogram, requiring more processing time. Thus, we experienced the auto-correlogram, and this gave us similar results and effects on the performance of the system, but with lower feature vector size.

Normalizing the data in the feature vector also will increase the performance of the system. At first, the features' vectors data were tested without normalization. It was found that the results sometimes depend on big values in feature vector such as the values of the histogram that represent the count of specific colors in the image. Normalization limited the values between 0-1, so that no data will be having big values that will affect the computations.

Comparing the results with other works

In this section, the results of the current proposed system are compared with previous related work. The comparison is illustrated in Table-8.

Table 8 - Compression with other related works.

No.	The System	Features	Accuracy
1	**The proposed system**	**Spatial features (region-based histogram, row sum, column sum auto-correlogram and GLCM)**	**98.94%**
2	**Vishal Lonarkar, Ashwath Rao B**	**Histogram, color moment and GLCM**	**97.56%**

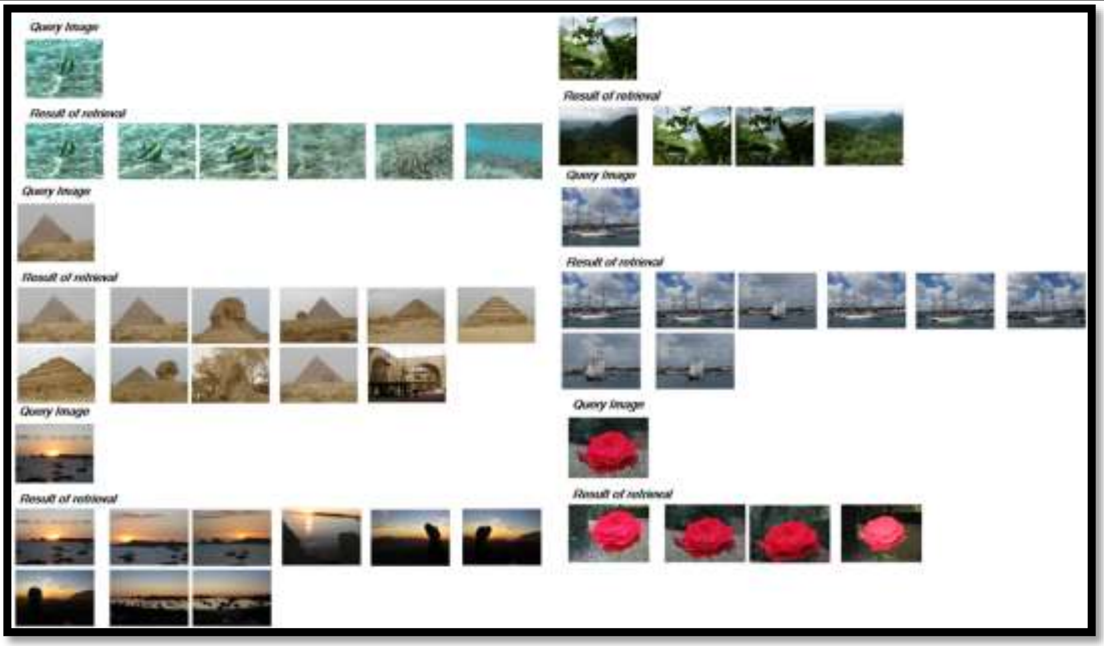


Figure 5- Sample of the proposed system results

Conclusion

In this paper, a CBIR system was developed using spatial features and clustering approach. The system was designed to search both texture and non-texture images. Using 3D HSV histogram showed better results than the other histograms since this type of histograms has many features. The first feature is that it uses the HSV color space that is the closest color space to human perception. Second, it computes the histogram for a combination of three channels H, S and V, which gives more information about color. In Addition to the previous features, the histogram was calculated for nine image segments. This will give spatial information about color. The GLCM calculated the texture features depending on the spatial relationship between the image pixels. The results showed that the spatial features make the system's result more accurate and that the color and texture features with spatial information are sufficient, with no need to use more classifiers. Using FCM clustering will reduce the searching space and speed up the system processing. The experiment finds that the clustering algorithm also increases the retrieval system accuracy. The returned images may be the exact match to the query image, since the searching space was limited in image groups having similar features. The FCM clustering algorithm reduces the number of iterations needed to clustering the images in database by choosing good initial centroids, which will speed up the offline stage. From the experimental results, we concluded that it is better to use spatial features to describe the texture and non-texture images.

References

1. Narasimha, Y. R., Pavithra, L. K. and Sree, S. T. **2018**. Analysis of Supervised and Unsupervised Learning in Content Based Multimedia Retrieval. International Conference on Computer, Communication, and Signal Processing (ICCCSP) pp. 1-5. IEEE.
2. Qi, Y., Song, Y. Z., Zhang, H. and Liu, J. **2016**. Sketch-based image retrieval via siamese convolutional neural network. In 2016 IEEE International Conference on Image Processing (ICIP) pp. 2460-2464. IEEE.
3. Lonarkar, Vishal, and B. Ashwath Rao **2017**. Content-based image retrieval by segmentation and clustering. International Conference on Inventive Computing and Informatics (ICICI), pp. 771-776. IEEE, 2017.
4. Chouhan, P. and Tiwari, M. **2015**. Image Retrieval Using Data Mining and Image Processing Techniques. *Image (IN)*, **3**(12).
5. O. A., Valle, E. and Torres, R. D. S. **2012.** Comparative study of global color and texture descriptors for web image retrieval. *Journal of Visual Communication and Image Representation*, **23**(2): 359-380.
6. Singh, B. and Aggarwal, M.M. **2018**. A Review on Content Based Image Retrieval. *International Journal of Computers & Technology*, **17**(2): 7226-7235.
7. INRIA Holidays dataset: http://lear.inrialpes.fr/people/jegou/data.php.
8. Doimari, D.C., Khusro, S. and Anand, S. **2018**. Content Based Image Retrieval using RGB to HSV conversion. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology*, **3**(3).
9. Jain, R. and Johari, P.K. **2016**. An improved approach of CBIR using Color based HSV quantization and shape based edge detection algorithm. In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* pp. 1970-1975. IEEE.
10. Niranjanan, S. and Gopalan, S.R. **2012**. Performance efficiency of quantization using HSV colour space and intersection distance in CBIR. *International Journal of Computer Applications*, **42**(21), pp.48-55.
11. Hazra, D. **2011**. Retrieval of color image using color correlogram and wavelet filters. *In International Conference on advances in computer engineering*, pp. 151-154, 2011.
12. Mistry, Y., Ingole, D.T. and Ingole, M.D. **2017**. Content based image retrieval using hybrid features and various distance metric. *Journal of Electrical Systems and Information Technology*, **5**(3): 874-888
13. kshatri, M. and Ratore, Y. **2012**. Row Mean and Column Mean Based Cbir System for Gray Scale Image. *International Journal of Engineering Research & Technology (IJERT)*, **1**(6).
14. Löfstedt, T., Brynolfsson, P., Asklund, T., Nyholm, T. and Garpebring, A**. 2019**. Gray-level invariant Haralick texture features. *PloS one*, **14**(2): e0212110.
15. Chandana, P., Rao, P.S., Satyanarayana, C.H., Srinivas, Y. and Latha, A.G. **2017**. An Efficient Content-Based Image Retrieval (CBIR) Using GLCM for Feature Extraction. In *Recent Developments in Intelligent Computing, Communication and Devices* pp. 21-30. Springer, Singapore.