Cuckoo Search Algorithm Based Feature Selection in Image Retrieval System

Manar Abdulkareem Al-Abaji

Computer Sciences Dept. , College of Education for Pure Science , University of Mosul, Iraq

Abstract

Efficiency an d retrieval time are very important issues in any content-based image retrieval system. In this study, an efficient image retrieval system was introduced depending on several features extracted from the database images, namely color moment (Mean, Standard Deviation), GLCM, and DWT(only LL-sub band). To increase the retrieval speed, Cuckoo search algorithm was used to select the important positions that contain full power features from the (LL-sub band). On using the Cuckoo search algorithm, only (50) important positions were chosen out of the total (24576) positions within (LL- sub band). These positions were stored for later use when entering a query image. Thus, the time taken to retrieve images was greatly reduced and this process also increased the efficiency of the system due to the fact that the selected positions gave the lowest distance measures between the query images and the similar images when evaluated using Manhattan distance measure. Two effectual performance measures (precision & recall) were used to calculate the accuracy of the system. The findings proved the system efficiency when compared to other previous works.

Keywords: CBIR, Color Moment, GLCM, DWT, Cuckoo Search algorithm, Manhattan measure

DOI: 10.7176/JEP/10-15-08

Publication date:May 31st 2019

1. Introduction

Due to the abundance of the images provided by the devices of image capture side by side with the recent introduction of the internet, the need for an effective system to retrieve the images has emerged. Hence, the content-based image retrieval (CBIR) has been one of the most important and difficult research topics related to computer vision (Fazal & Baharum 2013).

CBIR systems, also known as content-based visual information retrieval systems, retrieve images similar to the query images from the database and based on the visual components of the images, namely color, texture and shape (N. Pavithra & K.Sivaranjani 2016).

CBIR system includes two stages, viz. the features selection and extraction stage and the searching stage. In the feature extraction stage, many techniques are used to extract features related to the color, texture, and shape from the database images. To increase the efficiency and speed of these systems, just the most important features are chosen to create features vector for each image and store them in a special database. In the searching stage, features vector is configured of the query image and is compared with all the vector features of the database images for the purpose of retrieval of the most similar images (Mutasem 2018) (Fazal & Baharum 2012).

In this paper, the focus was on proposing a fast and efficient image retrieval system by using robust techniques to extract color features from images that include color moment (Mean, Standard Deviation) and texture features such as GLCM, and 1- level DWT. Only LL-sub band from the transformed images is used because it contains most features of the images. Cuckoo search is a meta-heuristic algorithm used to optimize the results. It is applied to an LL-sub band to select solely the positions that contain relevant and robust features. This process not only increases the efficiency of image retrieval but also reduces the retrieval time. The features vector for each image is created and stored. All the above-mentioned techniques are applied to the query image to extract features vector, viz. Mean, Standard Deviation, GLCM, and 1- level DWT from it. Only The features in the positions selected by cuckoo search are the only ones taken from the LL-sub band. Finally, Manhattan measure is used to calculate the distance between features vector of query image and features vectors of all database images and to retrieve the images with the shortest distance.

The remaining of the current paper is organized as follow: Section 2 includes the research methodology wherein all techniques used in the research are clarified. Description of the proposed image retrieval scheme and the steps followed are elaborated in Section 3. Section 4 presents the experimental results including materials, methods of evaluation, and comparison with previous works. Finally section 5 subsumes the conclusion.

2. Methodology

In this research, we have used images from the COREL database and extracted its features. The features have been reduced by selecting the optimal ones using meta-heuristic algorithm and storing them in a database. The stored features will be compared with the selected features of the query image for the purpose of retrieve of all related images by using one of the similarity measures. Performance measures have also been used to measure the efficiency of the retrieval system and compare it with results of the previous researches.

The retrieval system in this research subsumes the following steps:
- Features extraction using two techniques: color and texture.
- Features locations selection using Cuckoo search algorithm.
- Similarity measure.

2.1. Features Extraction
Feature extraction forms a crucial step in any CBIR system. The increase in data size, represented by the number of image pixels, increases the size of the search space, which in turn makes it difficult to perform additional data processing. As such, an appropriate representation of image pixels is required. Two kinds of features are used to represent image: A Low-level feature which includes color, texture, and shape features and A High-level feature which is extracted by using machine learning techniques (Atif *et al.* 2018) (Mohammad & Zahra 2017) . Here, the two techniques have been used to extract low-level features:
A. Color Feature Extraction
Color Moment is one of the most famous and best methods used to derive the color features of the image (Mean, Standard Deviation, Skewness). It extracts the low-level visual features that are used to differentiate images by representing their color distribution. In this research, we have relied on two color features only, viz. Mean and Standard Deviation, which have been calculated by using the following equations (K. P. Rajeswari *et al.* 2018) (Preeti & R. R. Welekar 2016) :

$$\mu_i = \frac{1}{N}\sum_{j=1}^{N} h_{ij} \qquad (1)$$

$$\sigma_i = \sqrt{\left(\frac{1}{N}\sum_{j=1}^{N}\left(h_{ij} - \mu_i\right)^2\right)} \qquad (2)$$

Where h_{ij} is i^{th} color channel & j^{th} pixel .
B. Texture Feature Extraction
Texture is an important feature of the image retrieval system. It is also a low-level visual feature used to distinguish images with high efficiency (Lakshmi *et al.* 2014).
- GLCM

One of the most prominent statistical methods used to derive texture feature is the Gray level co-occurrence matrix of which four characteristics can be identified by using the following equations   (K. P. Rajeswari *et al.* 2018) (Lakshmi *et al.* 2014):

Energy = $\sum_i \sum_j x^2 (i,j)$ 　　　　(3)
Entropy = $\sum_i \sum_j x(i,j) \log x(i,j)$ 　　　(4)
Contrast = $\sum_i (i-j)^2 x(i,j)$ 　　　　(5)
Homogeneity = $\sum_i \sum_j \frac{x(i,j)}{1+|i-j|}$ 　　　(6)

Where i is no. of rows , j is no. of columns of the image x .
- Wavelet Transform

Wavelet forms a useful tool to analyze signals and images. It is used to extract features from an image by decomposing it into the following 4-sub images, each of which is ¼ size of the original image: (LL) smoothing image which looks like the original image as it contains important information about the original image, (LH) which contains horizontal edge details, (HL) vertical edge details, and (HH) diagonal details of the image (Atif *et al.* 2018) (Raj *et al.* 2018).

2.2.Feature Selection Using Cuckoo Search Algorithm.
In the feature selection process, a subset of the original features is selected. This results in a reduction of the data, speed and efficiency in data mining (Lakshmi *et al.* 2014). Here Cuckoo search is used to select the features.
CSA is one of the most recent meta-heuristic algorithms inspired by nature. It was developed by Yang and Deb in (2009). What distinguishes cuckoo birds as well as their beautiful sounds is the strategy they possess for aggressive reproduction where some kinds of them put eggs in host nests and remove the host's eggs so as to increase the likelihood of hatching their eggs.
The following three rules are used to simplify the description of cuckoo search strategy:
- Each bird puts one egg at a time, and casts its egg in the nest that has been chosen at random.
- The best nests that contain high quality eggs will move on to the next generation.
- The number of available host nests is fixed and the egg placed by the cuckoo is discovered by the host bird with a probability between [0,1]. (Ramandeep & Ashok 2014)(Yang & Deb 2009)

Cuckoo search based feature selection algorithm
Begin
Divide dataset (vectors 1*24576 of LL-Sub bands) into a training set and testing set
Initialize the population of (25) host nest xi (i=1,2,…,25), each nest of (50) eggs
While (t < max Generation or stop Criterion) do
Get a cuckoo randomly by Levy Flights
Evaluate its fitness Fi on training set
Randomly choose nest among n available nests (say j)
If (F i > F j)
Replace j by the new solution
End if
Abandon a fraction (pa) of worse nests and new ones are built
Keep the best solution
Rank the solutions and find the current best
End while
Calculate the classification accuracy of the selected feature position subset on the test set
Return the training and test classification accuracies
end

2.3. Similarity Measure

In the current research, Manhattan is the distance measure that has been used for comparing between query image feature vector and database image feature vector, Manhattan equation (Atif *et al.* 2018) (K. Seetharaman & S. Sathiamoorthy 2014) :

$$d\left(V_{query}, V_{dbase}\right) = \sum_{i=1}^{N} \left| v_{query}(i) - v_{dbase}(i) \right| \qquad (7)$$

Were, N represent no. of features in each vector, $v_{query}(i)$ is query image feature vector , and $v_{dbase}(i)$ is database image feature vector

3. The Proposed Image Retrieval System

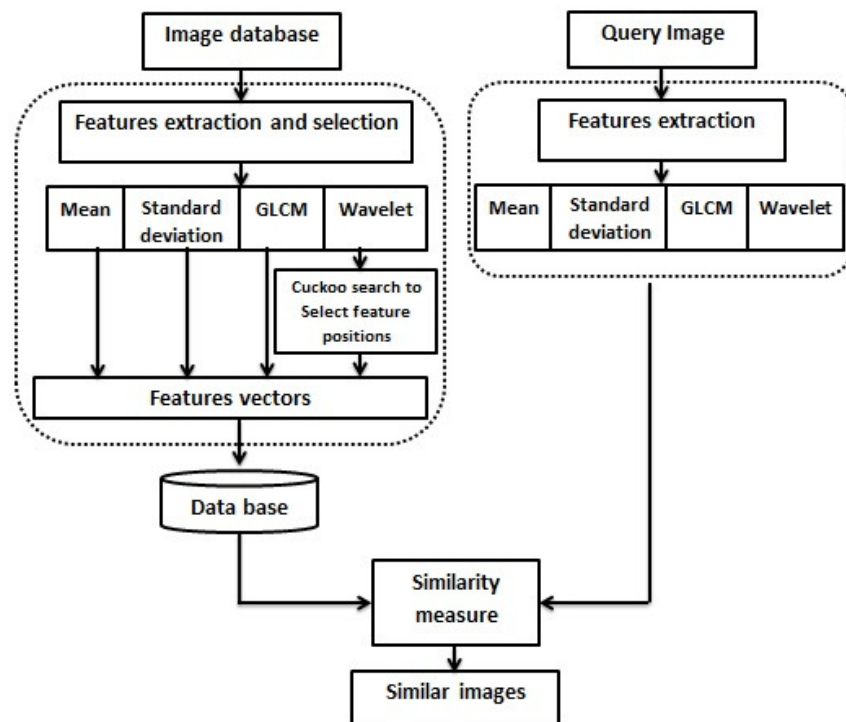To understand the overall steps in the proposed image retrieval system see figure 1:

Figure 1. The Proposed System

Step1: Load all images database; each of size (256×384)
Step 2: Compute the Mean, Standard Deviation, and GLCM of all images and store them (18 features for each image).

Step 3: Apply DWT to all images and take only (LL) sub-band. Size of LL is (128×192) = 24576 DWT features
Step 4: Create one dimension vector of size 24576 that contains numbers (from 1 to 24576) in order to represent the positions of DWT features
Step 5: Apply cuckoo search to the vector of features positions to obtain the most significant (50) features positions.
Step 6: Take the features in the positions selected in previous steps from (LL) sub band, (50) features
Step 7: Create features vector of size (68) for each image and use them for comparison with query image. [(1-feature) Mean,(1-feature) Standard Deviation, (16-features)GLCM, (50-features) DWT]
Step 8: Read query image, extract its features using Mean, Standard Deviation, GLCM, and DWT.
Step 9: Take only (50) features from (LL) sub band depending on the positions that have been previously selected.
Step10: Create features vector for query image.
Step11: Use Manhattan distance measure to compare between query image vector and all database images vectors.
Step12: Retrieve similar images from database.

4. Experimental Results
This section attends to the description of the materials used and the presentation of the methods adopted to evaluate the findings of the current work, and compare them with those of previous works.

4.1. Materials
This work was performed using Matlab R2017b language. A collection of Coral database images was also taken represented by 4 classes (Bus, Food, Elephant and Dinosaur), with 25 images taken from each class. Table (1) illustrates the parameter values used in Cuckoo search:

Table (1): Parameter Values Used in Cuckoo Search

parameter	value
No. of nests (population)	25
No. of eggs in each nest (no. of features)	50
No. of generation	100
Probability pa	0.25

See Figure 2 , 3 and 4 that illustrate the images retrieved after applying Cuckoo search to select features:
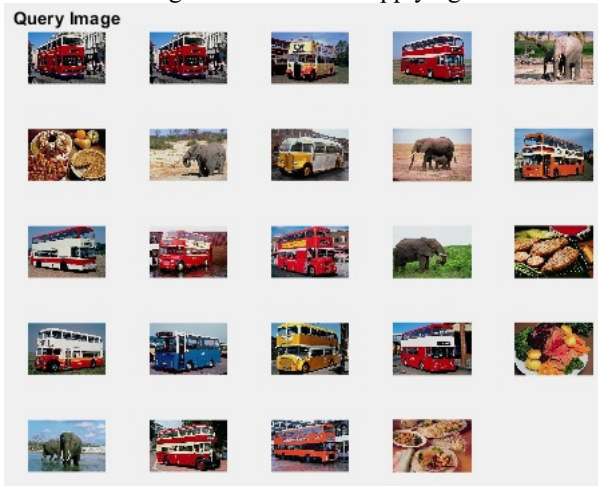

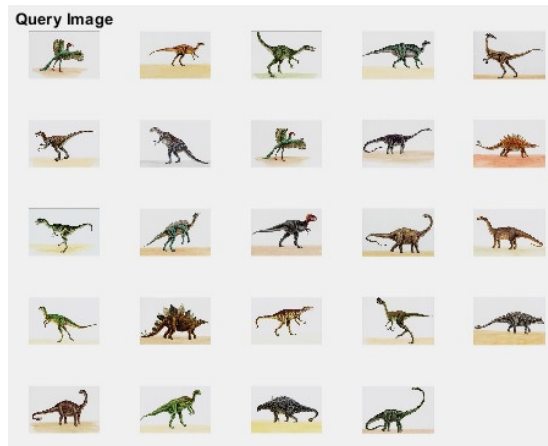Figure 2. The Query Image & the Retrieved Bus Images

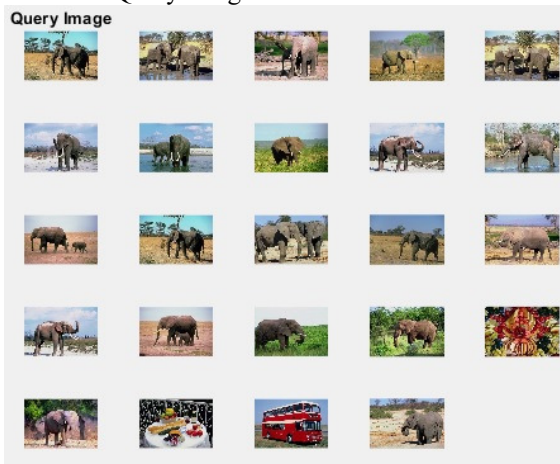Figure 3. The Query Image & the Retrieved Dinosaur Images

Figure 4. The Query Image & the Retrieved Elephant Images

4.2. Methods of Evaluation

To evaluate the accuracy of the proposed image retrieval system, two effectual performance measures have been used, namely precision & recall; the mathematical equations of which are as follows:

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

$$Recall = \frac{number\ of\ relevant\ images\ retrieved}{Total\ number\ of\ relevant\ images\ in\ database} \qquad (9)$$

And to calculate the execution time before and after using the cuckoo search, tic and toc Matlab function has been used. The time is calculated from reading the query image until the completion of the comparison between its features vector and those of the database images.   Table (2) illustrates the results of precision, recall and execution time before and after applying cuckoo search.

Table (2): The Results with and without Cuckoo Search

Class	Mean, Standard Deviation, GLCM, DWT			Mean, Standard Deviation, GLCM, DWT+ cuckoo search		
	Time (s)	Precision (%)	Recall (%)	Time (s)	Precision (%)	Recall (%)
Bus	15.8	47.82	44	0.3831	60.86	56
Dinosaurs	15.9	86.95	80	0.3735	100	92
Elephant	15.9	69.56	64	0.3715	82.60	76
Food	15.4	52.17	48	0.3685	56.52	52

Table(2) shows a noticeable difference between the time it takes to retrieve images before and after the Cuckoo search is used. This difference is due to the reduction in the number of features used from (24594) to (68) , which

in turn reduces the time required for image comparison operations. The results further show an improvement in precision & recall values; this indicates the strength of the features selected by the Cuckoo search algorithm. The precision values in all classes increased from range (47.82-86.95%) to (56.52-100%) on the application of Cuckoo search to extract features. Added to that, the recall values increased from range (44-80%) to (52-92%). Figure 5, 6 and 7 illustrate the precision, recall and execution time before and after applying Cuckoo search:
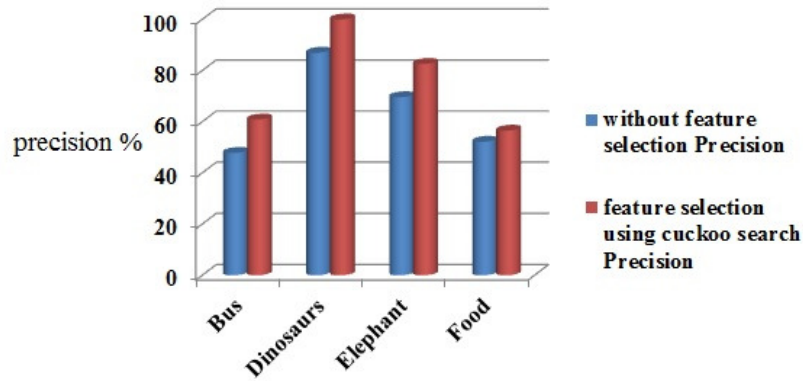

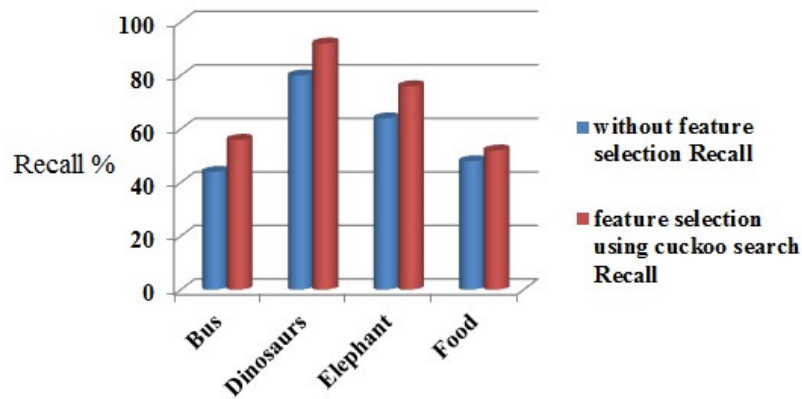Figure 5. Precision (%) with and without Cuckoo Search

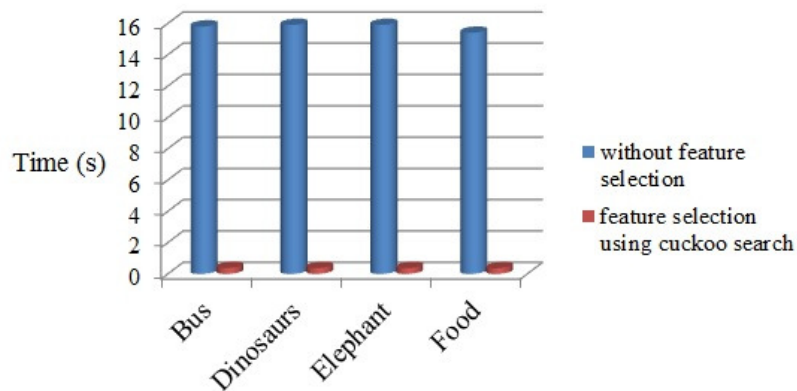Figure 6. Recall (%) with and without Cuckoo Search

Figure 7. Time(s) with and without cuckoo search

Figure 7 shows the difference in the retrieval time for all classes before selecting the features (24594 features) and after using the Cuckoo search in selecting the features (68 features).

Table (3): Comparison in Precision between the Proposed Work and Previous Works

Class	Previous Work [2014]		Previous Work [2016]		proposed Work	
	Feature extraction	Feature selection	Feature extraction	Feature selection	Feature extraction	Feature selection
	Precision (%)	Precision (%)	Precision (%)	Precision (%)	Precision (%)	Precision (%)
Bus	18.75	58.33	46.2	59.4	47.82	60.86
Dinosaurs	58.22	100	85	100	86.95	100
Elephant	41.66	51.66	43.2	58.9	69.56	82.60
Food	41.66	50	50.6	57.2	52.17	56.52

Table (4): Comparison in Recall between the Proposed Work and Previous Works

Class	Previous Work [2014]		Previous Work [2016]		proposed Work	
	Feature extraction	Feature selection	Feature extraction	Feature selection	Feature extraction	Feature selection
	Recall (%)	Recall (%)	Recall (%)	Recall (%)	Recall (%)	Recall (%)
Bus	12	28	27.6	35.8	44	56
Dinosaurs	28	48	82.4	93.36	80	92
Elephant	15.3	20	16.8	23	64	76
Food	20	23	30.2	50.89	48	52

To evaluate the present proposed system, its obtained results in terms of (precision, recall) have been compared with the results of two previous studies(table 3 & table 4): The first study by Lakshmi P.S and others (2014) used GLCM to extract texture features from images, and applied Genetic Algorithm to select optimal features, and used Euclidean distance in similarity measure process. The second study by Preeti K. and R. R. Welekar (2016) used color moment, GLCM and edge histogram descriptor to extract features, while the features selection process was done by using Genetic Algorithm and k-means algorithm for clustering. In the current study, the features have been extracted by using color moment ,GLCM and DWT(LL-sub band); features selection has been applied to (LL only) using Cuckoo search algorithm and Manhattan has been used for similarity measure. The comparison has been done on the same image classes from Corel image database. The results (precision & recall) in tables 3 &4 illustrate the efficiency of our proposed system in most image classes. Figure 8 and 9 show the comparison between our proposed system and previous studies in relation to precision and recall.
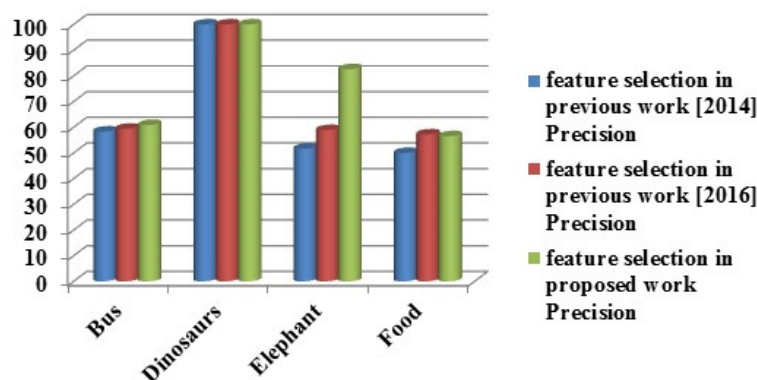

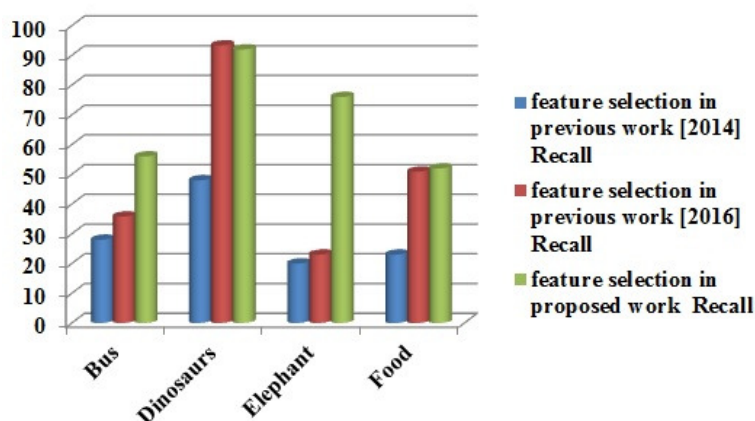Figure 8. Precision (%) between the proposed work and previous works

Figure 9. Recall (%) between the proposed work and previous works

5. Conclusion

In the current study, an efficient CBIR system based on extracting multi low-level features like color moment (Mean, Standard Deviation), GLCM, and DWT(only LL-sub band) has been proposed. Cuckoo search algorithm was used to select the important positions that contained power full features from the (LL-sub band). Once a query image was entered, its low-level features were extracted in the same way. Manhattan similarity measure was used to retrieve images from the database related to query mage. It is concluded that the use of the Cuckoo search algorithm to select the features in the proposed system significantly reduced the time required to retrieve images similar to the query image and increased the efficiency of retrieval in terms of precision& recall, compared to the system that was not   using the cuckoo search algorithm. Our proposed system was compared with previous works and proved to give good results in terms of precision and recall.

References

Atif N., Rehan A., Talha H., Nouman A.,(2018), "Content Based Image Retrieval System by using HSV Color Histogram, Discrete Wavelet Transform and Edge Histogram Descriptor", *international Conference on Computing ,Mathematics and Engineering Technologies*- iCoMET .

Fazal M., Baharum B.,(2012), "Analysis of distance metrics in content-based image retrieval using statistical quantized histogram texture features in the DCT domain", *Journal of King Saud University-Computer and Information Sciences*.

Fazal M., Baharum B.,(2013), "The Statistical Quantized Histogram Texture Features Analysis for Image Retrieval Based on Median and Laplacian Filters in the DCT Domain*", The International Arab Journal of Information Technology*, 10(6).

K. P. Rajeswari, A. B. Roshini, R. Shwetha, S. Ravi Sankar, (2018), "Feature Selection For Image Retrieval Using Genetic Algorithm Deep Convolutional Neural Network*", International Journal of Electronics and Communication Engineering,* 76-80.

K. Seetharaman , S. Sathiamoorthy, (2014), "A unified learning framework for content based medical image retrieval using a statistical model", *Journal of King Saud University-Computer and Information Sciences.*

Lakshmi P. S., Thusnavis B. M., A. Vasuki, Ansa Saju, Kolcharam Y.,(2014), "Image Retrieval Based on Feature Selection Method", *International Journal of Research in Engineering and Technology*, 03(07), 34-38.

Mutasem K. Alsmadi,(2018), "Query-sensitive similarity measure for content-based image retrieval using meta-heuristic algorithm", *Journal of King Saud University-Computer and Information Sciences*, 30, 373-381.

Mohammad S. N., Zahra A. B.,(2017), "Content-Based Image Retrieval through Combined Data of Color Moment and Texture", *International Journal of Computer Science and Network Security*, 17(2), 94-97.

N. Pavithra, K.Sivaranjani,(2016), "Content Based Image Retrieval System Data Mining Using Classification Technique", *International Journal of Computer Science and Mobile Computing*, 5(7), 519-522.

Preeti Kushwaha, R. R. Welekar,(2016), "Feature Selection for Image Retrieval based on Genetic Algorithm*", International Journal of Interactive Multimedia and Artificial Intelligence*, 4(2), 16-21.

Raj K. S., Moni K. C., Pratiksha G., Asif U.,(2018), "DWT Features Based Image Retrieval Using Neural Network*", International Journal of Computer Science and Mobile Applications*, 6(6), 1-9.

Ramandeep K., Ashok K. B.,(2014), "Enhanced Content-Based Image Retrieval Using Cuckoo search Algorithm", *International Journal of Advanced Research in Computer Science and Software Engineering*, 4(9), 233-241.

Yang X. S., Deb S.,(2009), "Cuckoo Search Via Levy Flights", *IEEE 2009,* 210-214.