ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

CBIR Systems: Results of Feature Extraction with Color Feature Comparison with Standard Database

Bandna Kumari*

Department of Computer Science and Technology, Central University of Punjab, Bathinda, India

E-mail: bandnaaf@gmail.com

Abstract: Image retrieval is a distinguished field in digital image processing. Images can be extracted from a big collection of images on the basis of text, color and structure. In systems, using features which users are connected to like color get many similar and relevant images. In a typical Content based image retrieval (CBIR) system, the visual content of the images in the database are extracted and displayed by multi-dimensional feature vectors. The feature vector of the images in the database is a feature database. Most sought out systems represent images with color feature most related to user and shape feature to relate image to exact object. In this paper we present introduction to CBIR systems, research work on Color scheme with HSV color space making use of color histogram. A comparison with Corel dataset and evaluation with parameters precision, recall and f-measure are computed.

Keywords: Content based image retrieval; Image retrieval; HSV histogram

I. INTRODUCTION

Images play a very crucial role in any visual information Systems and Multimedia. Picture is a visual representation of an object or scene or person or abstraction which is produced on a surface. Image collection is increasing rapidly with the enhancement in various image capturing devices like phone cameras, scanners, digital cameras, mobile cameras and increased use of multimedia data on internet. To deal with this huge collection of images an efficient and effective browsing, searching and retrieval tools are required. An image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital images. This area of research is very active research since the 1970s. The purpose of an image database is to store and retrieve an image or image sequences that are relevant to a query. In a variety of domains such as information retrieval, computer graphics, database management and user behaviour which have evolved separately but are interrelated and provide a valuable contribution to this research subject. Various image retrieval systems are developed which are text based image retrieval and content based image retrieval. In image retrieval color feature is most widely used feature. With color content based image retrieval is to be employed for fast search of image in large scale image collection and to get perfect results.

II. LITERATURE REVIEW

The research in content based retrieval system is done initially using inherent features. It make use of representation of the features that are extracted from the images themselves [1,2]. A content Based Image retrieval (CBIR) system first extract the features, then index those features by using some appropriate structure and then efficiently provide the result to the users. Color histogram is most commonly used approach with global colors [3,4]. Swain [5] has used color histogram approach for indexing .The main idea was to count the number of times a color occurs in image array and then color histogram intersection is used to match two images. Histogram intersection algorithm used in this approach matches the query's image color histogram with the histograms of the each image model in the database. If the match value is higher better it is. The technique of Histogram intersection is good to match color histograms. The biggest weakness in this was it loses space information of an image color entirely. As images are characterized using color histograms so there may be case images with little bit different appearance have similar histogram. So these images need to be compared before retrieval.

Pass [6] put forward color vector aggregation as a means of image index. He used histogram refinement technique for further put condition on histogram based matching Color indexing is based on the consideration that often color is used

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

to encode a functionality [7]. For example: roads are black, forests are green, and sky is blue. An objects color is not major criteria to find its identity. This noise can be removed by using Filters [8]. Plataniotis worked for noise removing filters for images [9]. Aboulmagd [10] purposed new approach to reduce this gap. HSV color space is used in this. Letters H S and V stands for Hue, Saturation and value respectively. Here Hue represents color tone that is red or blue, saturation is the amount of color like pale red or bright red. Image retrieval using color feature alone often does not meet user's requirements and provide disappointing results [11]. Texture provides surface characteristics for images. It can be identified by smoothness, coarseness and regularity [12]. Feature image detail serve purpose for image retrieval of desired image and coarse feature for discarding the undesired images [13].

Gnanaraja [14] used a reference image with known texture and a model for classifier is trained. In general only color and spatial information are not sufficient in superior segmentation [15], so including texture gives good results. Shape of an object in image is configuration which is represented by an outline [16]. To get the approximate information about the shape of an object, she used automatic segmentation process in her shape retrieval system. In the system, depending on the brightness, image is segmented into five classes. Quoc worked on improving retrieval of images using image region matching [17]. In images for the classification of objects, images can be classified into two types- foreground and background regions [18]. Guoyong combined all three features in his research and showed their advantages. He made use of an algorithm which can reduce the influence of the background. Then to acquire ideal segmentation around the hard partition it use border matting [19]. Improvements are made in the areas like to use Gaussian Mixture Model to replace the histogram and extending grey level images to color images. For color feature extraction he has used HSV by converting values of RGB to it. For shape feature at first color image is converted to gray level image. He proposed seven moments for translation, scaling and rotation. Then eigenvector of shape is formed which is automatically stored in feature library. His experiments has shown increased precision which means number of images associated with the example in the result to the number of images to the number of images returned for the query. The problem with this approach is it has long response time.

Heller and Zoubin specified a simple Bayesian network for content based image retrieval in which distribution of texture and color features are modelled [20]. As a user first specify query ,set of images are extracted by system. Then a Bayesian score is computed for each image by distribution of color and texture features in a large unlabelled collection of images. Shereena did experiments by classifying images using neural networks. And results showed that methods which are based on hybrid combination of features such as color and texture give higher accuracy than the methods which are based on retrieval using single feature extraction. It is difficult to say one feature is superior to other [21]. She showed Texture feature combined with color histogram give better results. Neural network classification improved recall rate and retrieval time.

III. FEATURE EXTRACTION

Image Features are broadly classified as low level features and high level features. Where feature like texture, color or shape is obtained from images and a feature vector database is created. Features can be broadly classified into low level and high level features. These FV databases are used for measuring the similarity between querying image and database images. The various steps in CBIR Systems are:

1. Image Acquisition: Image is acquired using digital devices and image database is created.
2. Image Pre-processing: Images are pre-processed before the retrieval process to enhance the images so that more similar images could be retrieved. It involves enhancement, noise removal, segmentation etc.
3. Feature Extraction: In this feature like texture, color or shape is obtained from images and a feature vector database is created. Features can be broadly classified into low level and high level features. These FV databases are used for measuring the similarity between querying image and database images.
4. Similarity Matching: Similarity matching techniques are used for acquiring percentage of similarity between the queried image and images existing in the database.
5. Output/ Retrieved Images: These are the final outcomes after the whole process is done by matching.
6. User Interaction/Feedback: User can communicate with the system for classifying the images whether they are relevant or not. This process goes on until the user gets satisfied.

3.1 Extraction Using Color Histogram

It is one of the basic features considered so far for the extraction purpose for the easy extraction of images. For color and shape feature extraction many approaches are described earlier. Techniques used in [1] give an overview of CBIR techniques. Color histogram represent values through which images are classified for further retrieval (Figure 1).

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

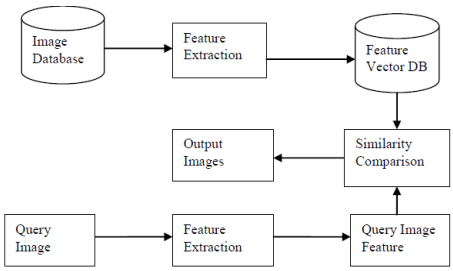


Figure 1: Block diagram for CBIR technique.

A color image represents three values at each pixel for red, green and blue color. Color histogram based approach is done for extracting features of images. This is achieved by calculating RGB values of each image in database at first. Then these values are converted to HSV and 3d histogram was computed for all images and stored in a mat lab's .mat file, this was termed as trained dataset. When query image was given its features are extracted with the same procedure and similarity was computed between query image and database images using similarity measure. Our work uses RGB images as input. At first we found out RGB values for image and then those were converted to HSV. To convert RGB values to HSV values we have made use of following formula:

$$Cos^{1/2[(R-G)+(R-B)]}_{\sqrt{(R-G)2+(R-B)(R-C)}}$$

$$S = 1 - \frac{3}{R+G+B}\left(\min(R,G,B)\right)$$

$$V = \frac{1}{3}\left(R+G+B\right)$$

In a histogram based extraction technique, an image is characterized by its color distribution or histogram. A histogram basically represents colors of image and occurrence of these colors at different levels irrespective of the type of image is under consideration in the form of a graph. Histogram is used to obtain basic features of an image. To form histogram of an image in a system we first need to obtain the array values. Now when retrieval is done on the basis of histogram in a content based image retrieval system, color histogram for each image is being generated and similarity is being computed based on values stored for each pixel of image. For query image and database images color histogram is being computed in matlab. At first input image is read in MATLAB workspace by using matlab function imread ('imagename'). Then histogram is generated for that image using imhist ('imagename') function. Figure 2 explains how a histogram looks like:

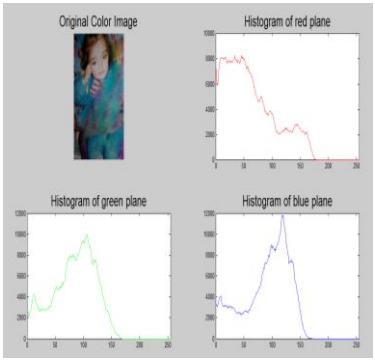


Figure 2: Histogram for red, green and blue pane.

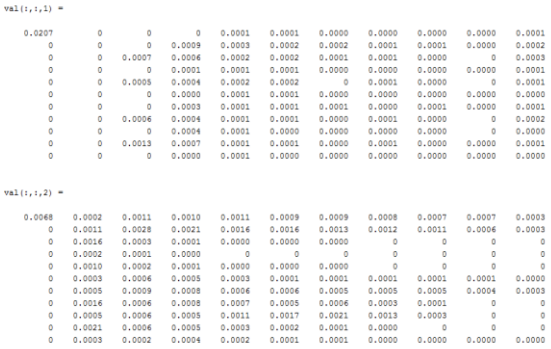
ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

Original color image shows the input Image and for this image red, green and blue histogram generated are shown in the same figure in each plane. The RGB values are converted to HSV and histogram is generated are stored in different bins. Few of the values for above figure are shown below in Figure 3.



```
val(:,:,1) =

    0.0207         0         0         0    0.0001    0.0001    0.0000    0.0000    0.0000    0.0000    0.0001
         0         0         0    0.0009    0.0003    0.0002    0.0002    0.0001    0.0001    0.0000    0.0002
         0         0    0.0007    0.0006    0.0002    0.0002    0.0001    0.0001    0.0000         0    0.0003
         0         0         0    0.0001    0.0001    0.0001    0.0000    0.0000    0.0000    0.0000    0.0001
         0         0    0.0005    0.0004    0.0002    0.0002         0    0.0001    0.0000         0    0.0001
         0         0         0    0.0000    0.0001    0.0001    0.0000    0.0000    0.0000    0.0000    0.0000
         0         0         0    0.0003    0.0001    0.0001    0.0001    0.0000    0.0001    0.0000    0.0001
         0         0    0.0006    0.0004    0.0001    0.0001    0.0000    0.0001    0.0000         0    0.0002
         0         0         0    0.0004    0.0001    0.0000    0.0000    0.0000    0.0000         0    0.0000
         0         0    0.0013    0.0007    0.0001    0.0001    0.0000    0.0001    0.0000    0.0000    0.0001
         0         0         0    0.0000    0.0001    0.0000    0.0000    0.0000    0.0000    0.0000    0.0000

val(:,:,2) =

    0.0068    0.0002    0.0011    0.0010    0.0011    0.0009    0.0009    0.0008    0.0007    0.0007    0.0003
         0    0.0011    0.0028    0.0021    0.0016    0.0016    0.0013    0.0012    0.0011    0.0006    0.0003
         0    0.0016    0.0003    0.0001    0.0000    0.0000    0.0000         0         0         0         0
         0    0.0002    0.0001    0.0000         0         0         0         0         0         0         0
         0    0.0010    0.0002    0.0001    0.0000    0.0000    0.0000         0         0         0         0
         0    0.0003    0.0006    0.0005    0.0003    0.0001    0.0001    0.0001    0.0001    0.0001    0.0000
         0    0.0005    0.0009    0.0008    0.0006    0.0006    0.0005    0.0005    0.0005    0.0004    0.0003
         0    0.0016    0.0006    0.0008    0.0007    0.0005    0.0006    0.0003    0.0001         0         0
         0    0.0005    0.0006    0.0005    0.0011    0.0017    0.0021    0.0013    0.0003         0         0
         0    0.0021    0.0006    0.0005    0.0003    0.0002    0.0001    0.0000         0         0         0
         0    0.0003    0.0002    0.0004    0.0002    0.0001    0.0001    0.0000    0.0000    0.0000    0.0000
```

Figure 3: Histogram values for HSV image.

Then values obtained are stored in number of bins. The values so obtained are stored in .mat file. Each time we will have to find out results we will load this .mat file called model name. Now comes the execution part for query image. For query image too we will generate histogram and their values are quantized into number of bins. As now we have extracted values for both query image as well as database images so we are going to measure similarity. Now Euclidean distance is calculated. After that images will be sorted and matching results are displayed for specific query image. We have used three datasets for calculating results using color feature based extraction as explained earlier. For all these datasets we have executed our code and output was analysed for each. The methodology of work carried out for feature extraction is explained with the help of following flow diagram and algorithm is written further Figure 4.

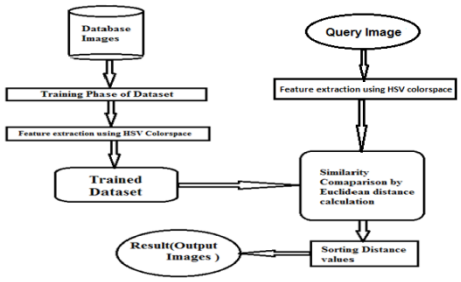


Figure 4: Methodology for color feature extraction.

Following steps explains the algorithm followed for retrieval of images from dataset.

1. First step is loading of database images in the MATLAB workspace by specifying directory where dataset is kept.
2. For all the images in database convert RGB values to HSV values by above specified formula.
3. Then there is generation of histogram for hue, saturation & value and values are quantized into bins.
4. After this, Store values obtained of database images into the .mat file.
5. Now, Load the Query image.
6. Apply the procedure 2 and 3 to find HSV values of Query image.
7. As values are obtained, determine the Euclidean distance of Values calculated of Query image with trained database.
8. Sort the distance values to find out most similar images to query image.
9. Display the results (all the output images) on GUI.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

For both query image as well as database images we have to extract features that what we have done in step 3 and 4.At step 3, histogram is generated for three values separately for hue, saturation and value. Histogram generation was explained earlier in this section. At step 4 the values for each of then needs to be stored, which we have done using MATLAB's .mat file. After these values are computed there is need to calculate difference between them. Step 7 explains this. After calculating distances they are sorted in ascending order so that most similar and relevant images to user's query will be displayed. Step 8 tells about this. Last and final step is to display results to user in a graphical user environment.

IV. EVALUATION OF RETRIEVAL SYSTEM

Most common criteria discussed previously [4] includes precision and recall. Precision and recall values for proposed dataset containing 500 images of 10 classes are shown below:

4.1 Performance Evaluation using Color Feature

For proposed dataset and Wang Dataset algorithm explained in methodology part is applied. For query image and all images of database features are extracted using techniques explained earlier, and while evaluating results asked for different outputs to check accuracy of retrieval. We checked relevant images result for five, ten, fifteen and twenty images as output. We randomly selected one image from each test class of ten classes and used it for checking retrieval. This procedure we have applied for both datasets dataset1 and Wang Dataset. Following table shows the output images (5, 10, 15 and 20 number of images retrieved) for queries of test classes and relevant images out of the output for Dataset 1.

S.no	**Classes**	**Precision values for images retrieved**			
		Five	Ten	Fifteen	Twenty
1	Airplanes	0.60	0.70	0.73	0.70
2	Apple	0.80	0.60	0.67	0.60
3	Basketball	0.60	0.50	0.53	0.55
4	Beach	1	0.70	0.73	0.75
5	Building	0.80	0.70	0.80	0.80
6	Car	0.40	0.40	0.53	0.50
7	Grass	0.80	0.70	0.80	0.85
8	Guitar	0.60	0.50	0.60	0.60
9	Man	0.60	0.60	0.73	0.70
10	Flower	1	0.80	0.80	0.65
Average Precision		0.72	0.62	0.69	0.67

Table 1: Precision values obtained for proposed dataset.

The precision values of the retrieval results for top 5, 10, 15, and 20 retrieved images in response to each of the ten queries are computed for each class as shown in Tables 1 and 2. It can be noticed from this table that the precision values are high when small numbers of images are retrieved, and these values get decreased as the number of retrieved images increase. This indicates that the system gives a good ranking of retrieved images, for an instance in Flowers query when the first top 5 is retrieved ,all were found to be relevant and for top 10 output images the retrieved images are relevant too, whereas in the top 15 and top 20 retrieved images we found some irrelevant too.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

S.no	Classes	Precision values for images retrieved			
		Five	Ten	Fifteen	Twenty
1	African	1	0.90	0.80	0.75
2	Beaches	0.80	0.80	0.87	0.85
3	Architecture	1	0.90	0.80	0.80
4	Buses	0.80	0.60	0.73	0.80
5	Dinosaurs	1	1	1	1
6	Elephants	1	0.80	0.80	0.85
7	Flowers	1	0.70	0.60	0.70
8	Horses	0.60	0.70	0.73	0.80
9	Mountains	0.80	0.60	0.60	0.65
10	Foods	0.80	0.80	0.80	0.80
Average Precision		0.88	0.78	0.77	0.80

Table 2: Precision values for Wang dataset.

Above results of table shows Performance evaluation of the designed system for color feature extraction using Precision and Recall values. As shown average values recorded for precision for proposed dataset (dataset 1) is 0.68 means 68% efficiency and for Wang Dataset is 0.81 or 81% efficiency. Recall values for proposed dataset is 0.42 and for Wang dataset is 0.44.

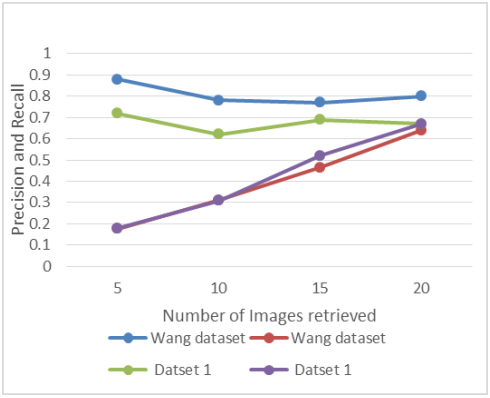


Figure 5: Precision and recall comparison for proposed dataset and Wang dataset using color feature.

For dataset 1 F-measure is calculated and following Figure 5 describes it for top 20 images retrieval for each class. Precision and recall values comparison using each class for both datasets is explained in Figure 5. The parameters precision and recall calculate the accuracy of image retrieval in correspondence to the query and database images but always two values need to be computed to show the effectiveness of the image retrieval of system. But these two measurements cannot be considered as complete accuracy for effective image retrieval [22]. F-measure for all classes of Dataset1 when 20 images are retrieved as output is shown below:

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

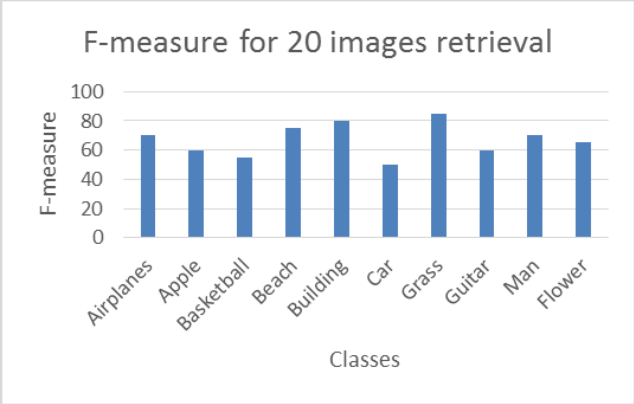


Figure 6: F-measure for dataset 1.

These can be combined to provide a single value which describes the accuracy of image retrieval and this is called F-Score or F-measure for measuring accuracy in Figures 6 and 7. Both parameters precision and recall are combined to calculate the score and it is also termed as a weighted average or harmonic mean of the precision and recall. F-measure for all classes of Wang dataset when 20 images are retrieved as output is shown below [23].

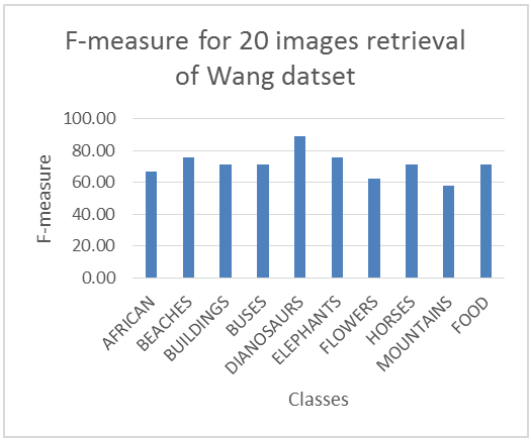


Figure 7: F-measure for Wang dataset

V. CONCLUSION

With time multimedia technology is becoming more popular and users are not satisfied with the traditional information retrieval technique. Content based image retrieval is becoming source of fast and exact retrieval. Color feature when employed for retrieval give images most connected with users as generally human perceive images based on its color taken as example when one says grass its green that comes to mind first. So color histogram technique as described by many researchers was providing best results. Wang dataset a standard dataset for images is used for comparison purpose with proposed dataset. Result evaluation with precision, recall and F-measure describes Wang dataset generates best retrieval efficiency i.e. 81% as compared to proposed dataset 1, 68% with HSV technique. The performance of the proposed method can be further improved by computing moment feature for constructing feature vector to make overall system self-adaptive. Relevance feedback concept can be added to give user more satisfying results.

ISSN(Online): 2320-9801
ISSN (Print): 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Vol. 5, Issue 11, November 2017

VI. REFERENCES

1. PS Meenakshi, GK Sunil. Image Retrieval: A Literature Review. International Journal of Advanced Research in Computer Engineering and Technology (IJARCET) 2013; 2: 2077-2080.
2. G. Kaur, A Research Paper on Content Based Image Retrieval. International Journal of Advanced Research 2015.
3. M Priyanka, SV Gumaste, A Review Paper on Content Based Image Retrieval. International Research Journal of Engineering and Technology (IRJET) 2015; 2: 883-885.
4. Y Hechao, Z Xuemei, Research of Content Based Image Retrieval Technology. Proceedings of the Third International Symposium on Electronic Commerce and Security Workshops (ISECS) 2010: 314-316.
5. JS Michael, BH Dana, Indexing via color histogram. Active Perception and Robot Vision 1992: 261-273.
6. P Greg, Z Ramin, Histogram refinement for content based image retrieval. IEEE Computer Society 1996.
7. S Markus, O Markus, Similarity of Color Images. 1995
8. SH Syed, K Noor ul Amin, et al. Content-Based Image Retrieval Using Texture Color Shape and Region (IJACSA) International Journal of Advanced Computer Science and Applications 2016; 418-426.
9. KN Plataniotis, D Androutsos, et al. Adaptive fuzzy systems for multichannel signal processing. IEEE 1999: 1601-1622.
10. A Heba, El G Neamat, et al. A new approach in content based image retrieval using fuzzy. Telecommunication Systems 2008.
11. T Satish, D Jayadevappa, et al. A Comparative Study of Content Based Image Retrieval Trends and Approaches. International Journal of image processing 2015: 127-155.
12. PW Huang, SK Dai, Image retrieval by texture similarity. Pattern Recognition 2001: 666-679.
13. M Wei-Ying, BS Manjunath, NeTra: A toolbox for navigating large image databases. Multimedia Systems 1999; 7: 184-198.
14. R.Gnanaraja, B. Jagadishkumar, et al. Content Based Image Retrieval (CBIR) using segmentation process. International Conference on Engineering Technology and Science-(ICETS) 2014; 384-389.
15. E. P. Mustafa Ozden, "A color image segmentation approach for content Based Image Retrieval," Pattern Recognition, p. 1318 – 1325, 2007.
16. C Reshma, AM Patil, Content Based Image Retrieval Using Color and shape features. International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering 2012; 1: 386-392.
17. H Quynh Nguyen, TT Ha Nguyen, et al. An efficient content based image retrieval method for retrieving images. International Journal of Innovative computing, information and control 2012; 8: 2823-2836.
18. L Fai Hang, Finding a small number of regions in an image using low level features. Pattern Recognition 2002: 2324-2339.
19. D Guoyong, Content-Based Image Retrieval Research. International Conference on Physics Science and Technology (ICPST) 2011.
20. H Katherine, G Zoubin, A simple A Simple Bayesian Framework for Content based Image Retrieval. IEEE Computer Society 2006; 2: 2110-2117.
21. VB Shereena, MD Julie, Content Based Image Retrieval: Classification Using Neural Networks. The International Journal of Multimedia and Its Applications (IJMA) 2014; 6: 31-44.
22. KV Aher, A Survey on Feature based Image Retrieval. International Journal of Advanced Research in Computer Science and Software Engineering 2014.
23. PP Mane, NG Bawane, Image Retrieval by Utilizing Structural. International Journal of Image, Graphics and Signal Processing 2016; 8: 68-74.