Content Based Image Processing Approach on Colored Images using KNN Classifier

Garvita[#1], Priyanka Kamboj[#2]

[#1]*M.tech Computer Science Department, Kurukshetra University*
[#2] *Assistant Professor in Computer Science Department, Kurukshetra University*
JMIT Radaur, Haryana, India

Abstract- *Content Based Image Retrieval (CBIR) is also called as Query By Image Content (QBIC). It has been an active research field since last decades. In contrast with traditional systems, where images are retrieved on the basis of keywords but in CBIR system images are retrieved on the basis of visual content. Content Based Image Retrieval is a methodology that allows a user to extract an image based on a query from the database. Here, the term "content" refers to color, shape, texture, or any other data that can be derived from image itself. Different types of classification we can use Neural network, Support Vector Machine (SVM), KNN, Bayesian etc. In this paper, we are using K Nearest Neighbor (KNN) classifier to find out the relevant images and after that we use Spearman's Rank Correlation Function to reduce the time complexity and improve F-measure.*

Keywords- *KNN, F-measure, Recall, Precision, Spearman's Rank Correlation Function, Test Dataset, Train Dataset, color strings*

I. INTRODUCTION

With the advancement in computer technologies and the advent of World-Wide Web, multimedia and communication there has been an explosion in amount and complexity of digital data being generated, stored, transmitted, analyzed, and accessed. In order to make use of vast amount of data, efficient and effective techniques to retrieve database on its content need to be developed. Content-based Image Retrieval (CBIR) technology, also known as query by image content (QBIC) overcomes the traditional text-based image retrieval technology [10] where search is based on automatic or manual explanation of images. A conventional TBIR searches database for the similar text surrounding the image as given in the query string. To overcome limitations forced by TBIR systems more natural and easy to understand content based image retrieval systems were developed [1].

Content-based Image Retrieval is a method which utilizes visual contents for searching the images from databases. In CBIR, images in database are represented using low level image features such as color, texture, and shape. Among these features, color features are most widely used features for image retrieval because color is most perceptive feature and can be extracted from images easily [9]. The aim of any image retrieval system is to retrieve as many images as possible from a dataset so that retrieved images meet user's requirements [11].

The continuously increasing demand for the multimedia storing has led to need of large image database systems. For example, various applications like medical imagery, satellite imagery and graphic art require fast and sufficient data access. However, we cannot use this information unless the database is organized in such a way that allows efficient storage, browsing and retrieval. Every CBIRS consists of three main stages: pre-processing, feature extraction and similarity measurement (classification).

Pre-processing stage becomes necessary when we have images that are corrupted by some type of distortion. Images with noise, bad illumination, blurred are few examples when we need preprocessing step. Pre-processing usually involves filtering, normalization color transformation and segmentation.

One of the most important issues in image retrieval system is **feature extraction** process, where visual content of images is mapped into a new space, the feature space. The key to get to a successful retrieval system is to choice the right features that represent images as strong as possible. Feature extraction of images may include color [8], texture and shape [5] information.

The final stage of image retrieval system is the **classification** stage where using the extracted feature values, a similarity checking is performed between the query image and database images. Basically, this procedure is clustering process where the query image is compared with stored images and the most similar images are thus obtained. For this purpose,

computation of feature-based distance similarity can be obtained for every image in the database that tells us how similar query image is to each one of the stored images. There are many similarity- checking techniques, but the most usually used is the Euclidean distance [8], which is simple and suitable in most of cases, and especially when we have a large number of images [15].

II. RELATED WORK

Priyadarshini Patil et al. [2] proposed and implemented an efficient image retrieval technique using both color and texture features of an image. Here they compare and analyze performance of an image retrieval using both these features. And we see that CBIR using color features gives high precision whereas CBIR using texture feature features give high recall. **Manoharan Subramanian et al. [3]** proposed a new mechanism for CBIR systems which is based on two works. The first work is based on filtering technology which includes anisotropic morphological filters, hierarchical Kaman filters and particle filters. The second work is based on feature extraction which includes color and gray level features and after this the results were normalized. The experimental results shows that this proposed technique of CBIR using advanced filter approaches are much better than the existing system GLCM and color feature extraction for CBIR process. **R. Malini et al. [4]** was desired to enhance effectiveness of retrieving images by Color Averaging technique. Firstly, an average mean based technique is proposed. Secondly, a feature extraction technique based on central tendency is proposed. The proposed CBIR techniques are tested on the Wang image database and indexed image database. In last, results are compared with existing technique and we see proposed technique gives better performance with less computational complexity. **Sushant Shrikant Hiwale et al. [12]** proposed a CBrn system which extracts features of digital image to retrieve similar images from huge databases. They used Color Histogram, Color Moment, Color Auto-Correlogram, Gabor Wavelet and Discrete Wavelet transform to extract image features. The images are classified using Support Vector Machine (SVM). Finally we have better precision and recall rate. **Y H Sharath Kumara et al. [5]** presents a model for representation and indexing of flower images for retrieving flowers based on query sketch. Here, they swot correctness of Kdtree indexing scheme for flower retrieval system based on shape descriptors like Histogram of Gradients (HOG) ,Scale Invariant Feature Transform (SIFT), and Edge Orientation Histograms (EOH). This proposed method provides better efficiency and achieves good accuracy with indexing approach. **Roshi Choudhary et al. [7]** proposed an approach used to extract color and texture feature from images. To extract the color feature, higher order of color moment is used which is the descriptor of color. To extract texture, LBP is used which is descriptor of texture. **Siddarth Ladhake et al. [6]** provides a system for large scale database is designed and implemented. Here, proposed system exploits semantic binary code generation techniques, fine and coarse similarity measure technique, which improves accuracy, image retrieval speed. Finally,the performance of image retrieval is improved in terms of accuracy, retrieval time and efficiency. **Kommineni Jenni et al. [8]** presented a CBIR approach based on database classification using Support Vector Machine (SVM) and color string coding feature selection. Through this computational complexity decreases and obviously increased the accuracy for image retrieval. **A. A. Khodaskar et al. [13]** proposed innovative framework for effective, intelligent and efficient CBIR based on three soft computing techniques such as Fuzzy Logic, Artificial Neural Network and support vector machine. In this, the SVM based relevance feedback is introduced, that whether the retrieved image is relevant or not. The main purpose is to avoid non-relevant images. CBIR based on soft computing techniques improves retrieval performance in term of accuracy, precision and efficiency. **Devyani Soni et al. [14]** proposed an efficient color space Based Approach for Image Retrieval Using fusion of Color Histogram and color correlogram. During experimentation, both HSV color model as well as RGB color model was used for the same process of retrieval and it was observed that HSV color space gives more accurate result as compared to RGB color space.

III. PURPOSE OF CBIR

Content Based Image Retrieval is a technique that enables a user to extract an image based on a query from the database containing huge amount of images. Here, we have to test a query image from our own built Dataset and provides the accurate result to the user.

IV. PROPOSED WORK

In this paper, we are going to propose Content based Image Retrieval System on the colored images. Here, we built our own dataset. The Dataset is divided into two categories: Train Data and Test Data. The steps involved are:

- The Query image is given by the user corresponding to which user want the results.
- Read that particular Query image.

- Extract features from the Query image on the basis of color and relatively find the Prediction class using **K Nearest Neighbor (KNN)** classifier. Prediction class is used to find out the relevant images.
- Now to order these relevant images we use **Spearman's Rank Correlation Function** to calculate the distances of each relevant image with Query image**.** It will sort all these relevant images and fetch top-n images (top-n < size of Dataset) and print these images.

K Nearest Neighbor: K nearest neighbors is a simple algorithm that stores all available cases and classifies new cases based on a similarity measure (e.g., distance functions). KNN has been used in statistical estimation and pattern recognition.

Spearman's rank correlation coefficient or **Spearman's rho**: is a nonparametric measure of rank correlation (statistical dependence between the ranking of two variables). It assesses how well the relationship between two variables can be described using a monotonic function.

Mathematical notation of Spearman Rank formula is:

$$(R) = 1 - \frac{6\sum d^2}{n^3 - n}$$

The result will always be between 1 and minus 1.

Here,

d= Distances

n= No. of cases

V. EXPERIMENT AND RESULTS SYSTEM DETAILS

A. Hardware details

We have validated our results on machine with the configuration of installed memory (RAM 3GB), 64-bit Operating System, having processor Intel(R) Core(TM) i3-2310M CPU @ 2.10GHz. Here, we have our own created Dataset e.g., images of Flags of different countries.

B. Software details

MATLAB 7.0

WINDOW 7

The Experimental work is done on the MATLAB. MATLAB is a software package for high performance numerical computation and visualization. It provides an interactive environment with hundreds of built-in functions for technical computation, graphics and animation. The name MATLAB stands for **MATrix LABoratory**. MATLAB is an efficient program for vector and matrix data processing. It contains ready functions for matrix manipulations and image visualization. MATLAB provides a suitable environment for image processing. Although MATLAB is slower than some languages (such as C), its built in functions and syntax makes it a more versatile and faster programming environment for image processing.

In this paper, proposed work is done on MATLAB as it contains ready-made functions so this tool is easy to use. We are going to compare existing and proposed approach.

Table: Comparison of Existing and Proposed approach

Approach	Precision	Recall	F-measure	Time complexity
Existing	0.625	0.625	0.625	1.62
Proposed	1	1	1	1.34

From this table it is clear that proposed work reduces the time complexity and improves accuracy as work is done only on the Trained Dataset.

A **precision rate** can be defined as the number of relevant images retrieved by a search divided by the total number of images retrieved by that search. The equation is as follows:

Precision = Relevant Correctly Retrieved / All Retrieved

= A/A+B

Where A is relevant correctly retrieved and B is falsely retrieved.

A **recall rate** is defined as the number of relevant images retrieved by a search divided by the total number of existing relevant images (which should have been retrieved). The equation is as follows:

Recall = Relevant Correctly Retrieved / All Relevant

= A/A+C

Where A is relevant correctly retrieved and *C* is relevant but not retrieved.

As per as the MATLAB concern, we are going to show our results on the MATLAB tool as well. As we have Test Dataset in which we place those images to which we want to test or we can say Query images are placed there. This test data is compared with the rest according to the particular category. Here, we compare existing and proposed approaches and we see that proposed gives 100% result to the user.

Firstly, the results of Existing approach are shown below from figure 1-2:

Figure 1(a): Query image of Austria to be searched

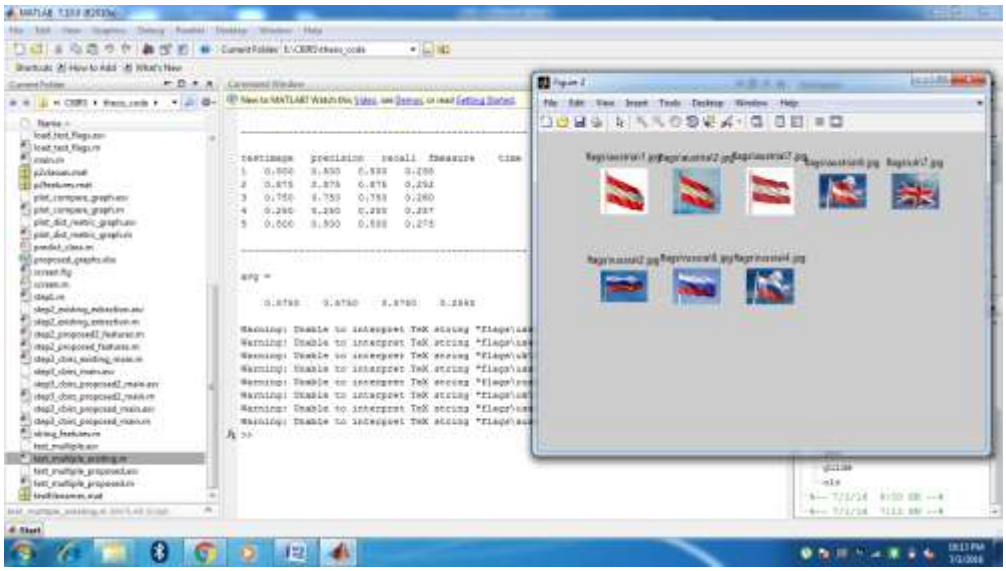


Figure 1(b): Results corresponding to the Query Image Austria

Figure 2(a): Query Image of India to be searched

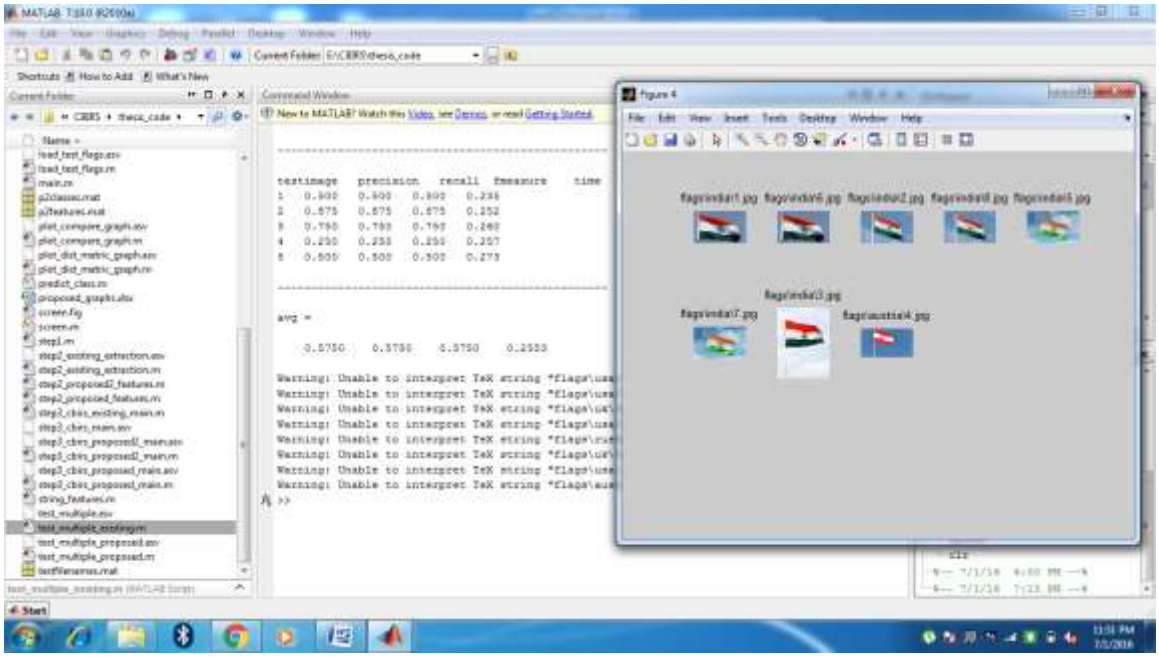


Figure 2(b): Results corresponding to the Query Image of India

From these screenshots we conclude that existing result is not 100%. For example, we have the Query image of Austria country flag and the user wants the similar images to that. But we can see that in results we have some of the images of flags of Austria country and some are of the other countries. So we can say that Existing approach does not give the accurate results. To overcome the limitations of existing approach, we proposed a model that will give 100% results to the user.

The validated proposed results are shown below from figure 3-4:

Figure 3(a): Query Image of Austria Country

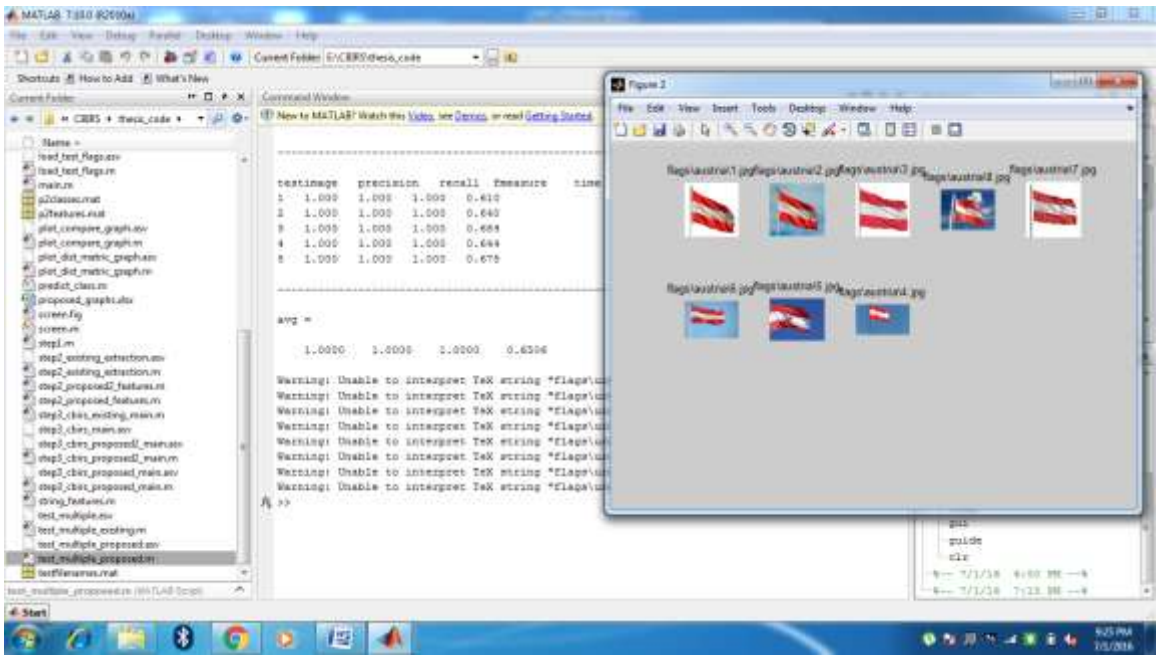


Figure 3(b): Results corresponding to the Query Image of Austria

Figure 4(a): Query image of India Country

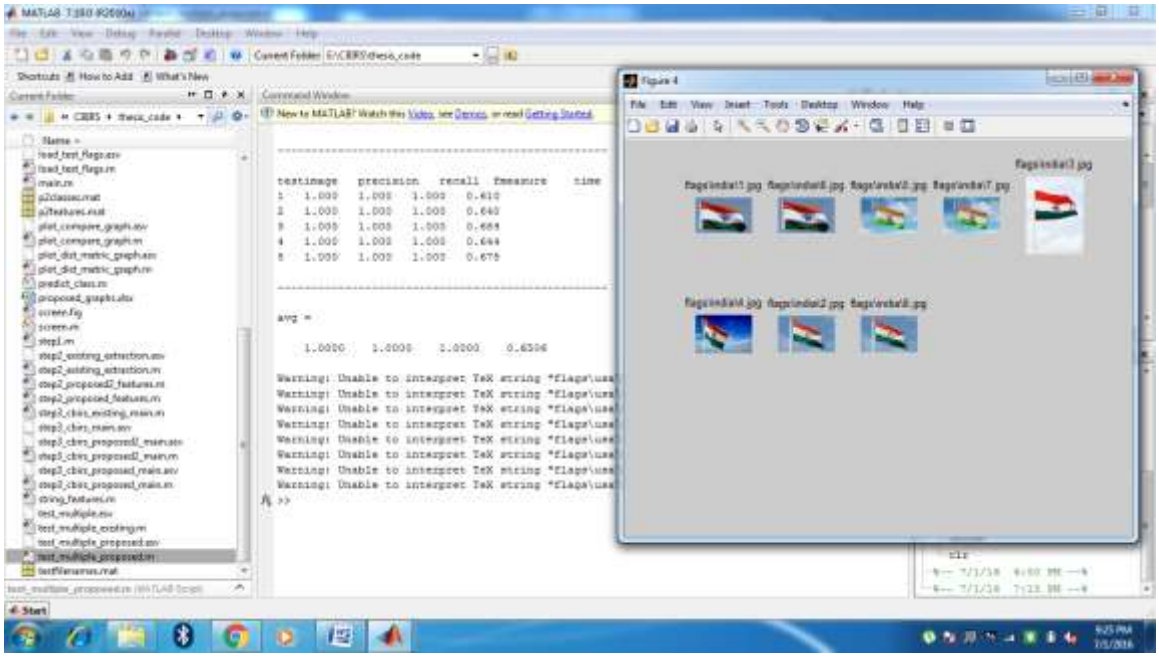
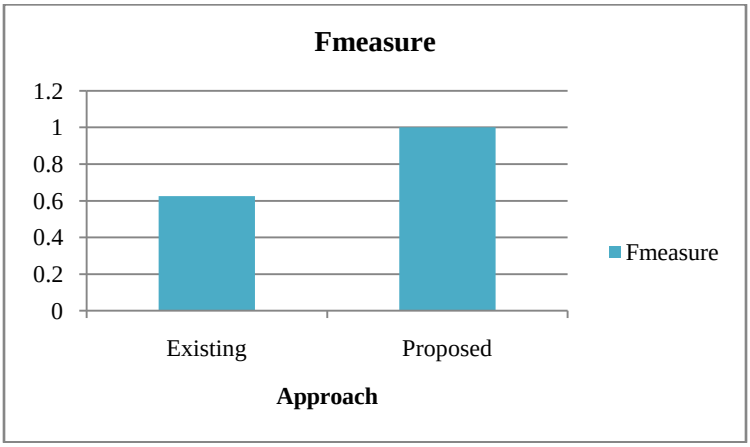
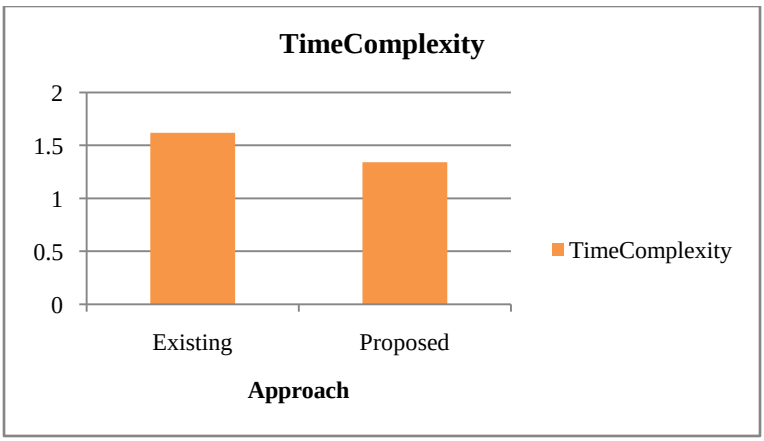


Figure 4(b): Results corresponding to the Query Image of India

These results can also be explained with the help of graphs:



Graph 1: Comparison of Existing and Proposed Approaches on the basis of F-measure.



Graph2 : Comparison of Existing and Proposed Approaches in terms of Time Complexity.

From the above graphs, we see that proposed approach provide better results as compare to existing one as Time Complexity reduces and improves the F-measure value.

VI. CONCLUSION AND FUTURE SCOPE

From above discussion we conclude that the proposed approach perform better than existing approach with reduced time complexity and improves F-measure value. Also we use Nearest Neighbor (KNN) Classification to calculate the relevant images from Dataset and Spearman's Rank Correlation Function for calculating the distances. As here we have own Dataset which contain limited number of images. In future, we can use this concept for the huge database as well and use some other classifiers to enhance the results.

REFRENCES

[1] Garvita and Priyanka Kamboj, Content Based Image Retrieval System: Review, International Journal of Computer Trends and Technology (IJCTT) – Volume 34 Number 3 - April 2016
[2] Priyadarshini Patil and Bhagya Sunag, Analysis of Image Retrieval Techniques Based On Content, IEEE International Advance Computing Conference (IACC)
[3] Manoharan Subramanian and Sathappan Sathappan, An Efficient Content Based Image Retrieval using Advanced Filter Approaches, The International Arab Journal of Information Technology, Vol. 12, No. 3, May 2015
[4] R.Malini and C.Vasanthanayaki, An Enhanced Content Based Image Retrieval System using Color Features, International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 2 Issue 12 Dec, 2013 Page No. 3465-3471
[5] Y H Sharath Kumara, D S Gurub, Retrieval of Flower Based on Sketches, International Conference on Information and Communication Technologies (ICICT 2014)

[6] Anuja khodaskar, Siddarth Ladhake, Promising Large Scale Image Retrieval by using Intelligent Semantic Binary Code Generation Technique, International Conference on Intelligent Computing Communication & Covergence (ICCC-2014) Conference Organized by Interscience Institute of Management and Technology, Bhubaneswar, Odisha, India

[7] Roshi Choudhary, Nikita Raina, Neeshu Chaudhary, Rashmi Chauhan, Dr. R H Goudar, An Integrated Approach to Content Based Image Retrieval, International Conference on Advances in Computing, Communications and informatics (lCACC1) 2014

[8] Kommineni Jenni, Satria Mandala, Mohd Shahrizal Sunar, Content Based Image Retrieval Using Colour Strings Comparison, 2nd International Symposium on Big Data and Cloud Computing (ISBCC'15)

[9] Krishnan Nallaperumal, M.Sheerin Banu and C.Callins Christiyana, Content Based Image Indexing and Retrieval using Color Descriptor in Wavelet Domain, International Conference on Computational Intelligence and Multimedia Applications 2007

[10] B.Jyothi, Y.Madhavee Latha and V.S.K.Reddy, Relvance Feed Back Content Based Image Retrieval Using Multiple Features, IEEE 2010

[11] H B Kekre, V A Bharadi and S D Thepade, Content Based Image Retreival using Fusion of Gabor Magnitude and Modified Block Truncation Coding, Third International Conference on Emerging Trends in Engineering and Technology

[12] Sushant Shrikant Hiwale, Dhanraj Dhotre , Dr. G.R.Bamnote, Quick Interactive Image Search in Huge Databases Using Content-Based Image Retrieval, IEEE Sponsored 2nd International Coriference on Innovations in Iriformation Embedded and Communication Systems ICIIECS'15

[13] A. A. Khodaskar and S. A. Ladhake, A Novel Approach for Content Based Image Retrieval in Context of Combination S C Techniques, International Conference on Computer Communication and Informatics (ICCCI -2015), Jan. 08 – 10, 2015, Coimbatore, INDIA

[14] Devyani Soni , K. J. Mathai, An Efficient Content Based Image Retrieval System based on Color Space Approach Using Color Histogram and Color Correlogram, Fifth International Conference on Communication Systems and Network Technologies 2015

[15] Aly S. Abdelrahim, Mostafa A. Abdelrahman, Ali Mahmoud and Aly A. Farag, Image Retrieval Based on Content and Image Compression, IEEE 2011