Implementation of Content Based Image Retrieval Using Improved K-Means Clustering Algorithm

Manisha Sharma[1], Abhinav Vidwans[2]
[1,2] Vikrant Institute of Technology & Management, Gwalior

Abstract- *The content based image retrieval (CBIR) is the well-liked and heart favorite area of research in the field of digital image processing. The key goal of content based image retrieval (CBIR) is to excerpt the visual content of an image directly, like color, texture, or shape. There are several applications of the CBIR technique such as forensic laboratories, crime detection, image searching etc. For the purpose of feature extraction of well-matched images from the database, a universal CBIR system utilizes texture, color and shape based techniques. In this presented work, we have offered an efficient approach for the content based image retrieval, where images are decomposed using the wavelet transform, it means that the image features are converted in the matrix form and a color feature data set is prepared. In order to improve search results we have used k-means algorithm. It is shown by experimental results that, the efficiency of the proposed method is improved in contrast with the existing method.*

Keywords- Image Retrieval, Clustering, Wavelet Transform, Haar Wavelet Transform, Feature Extraction, K-Means Technique.

I. INTRODUCTION

A lot of piece of information from old books, newspapers, journals etc. has been revealed in computer understandable formats in the past few decades. Lots of archives of films, images, music, books, satellite pictures, magazines and newspapers, have been made available for computer users. Human beings can access the enormous amount of information with the help of internet. A lot of information offered about a specified topic on the internet, therefore, one of the utmost challenges of the WWW (World Wide Web) is to discover accurate and appropriate information from this large amount of information. Most of the users recognize what information they want, but they are unconfident where to discover it. Search engines provide the facility for users to discover such related information. Image retrieval is the process of surfing, examining and retrieving images from a huge database of digital images. Due to development of multimedia technology and increasing vogue of the computer network, the conventional information/image retrieval systems are not able to overcome the users' current needs. There are various areas in which digital images are used such as-crime prevention, commerce, finger print recognition, surveillance, hospitals, engineering, architecture, fashion, graphic design, academics, historical research, and government institutions etc. Because of this widespread demand we need to enhance in retrieval precision and minimized retrieval time. The prior methods were only dependent on text based searching instead of its visual feature. Many times just one keyword is redundantly used with more than one images, therefore it leads to erroneous outcomes. Consequently, Content Based Image Retrieval (CBIR) is evolved to defeat the restriction of text based retrieval. Problems which we are identified in the existing image retrieval systems are as follows- How to retrieve the search image accurately. How we can manage a large database of images. How we can make the searching process efficient.
The design and improvement of operative and proficient Content Based Image Retrieval (CBIR) systems are still an investigation problem. The prior image retrieval methods were only dependent on text based searching instead of its visual feature.

Many times just one keyword is redundantly used with more than one images, therefore it leads to erroneous outcomes. Consequently, Content Based Image Retrieval (CBIR) is evolved to defeat the restriction of text based retrieval. "Content-based" means that the search will analyze the actual contents of the image rather than the metadata such as keywords, tags, or descriptions associated with the image. Here the 'content' refers to colors and textures information that can be derived from the image itself.

CBIR is quite desirable because most web based image search engines rely completely on metadata and this produces a lot of false detection in the results. Research objectives of this thesis work are concise as follow:

1. To make a strategy which retrieves the search image accurately.
2. To manage a large database of images, such that it can take less space for storage.

3. To create the process of searching an image from database efficient.

II. LITERATURE SURVEY

Carrying out literature analysis is very major part in any research project as it obviously establishes the need of the work and the background improvement. It makes related queries regarding improvements in the previously performed study and permits unsolved problems to get figured out and thus clearly explain all boundaries concerning the development of the research plan.

A lot of research work has been done in the field of image retrieval in image processing and still a lot need to be done on it. Many researchers have suggested work in this field. Few of the most remarkable works are shown in this section. CBIR is the technique of retrieval of images based on visual features same as shape, color and text.

There are numerous CBIR systems which exist, and are being constantly expanded. Color Selection utilized CBIR system [7] facilitates query-by-color. Such system is standing on eleven color categories. The low frequency DCT coefficients that are transformed from YUV color space as feature vectors, are used for retrieval of images [8].

It allows users to select its dominant feature of query images. But, the technique is sufficient enough to perform effective retrieval by introducing users' opinions on the query images. The Color Selection exploited CBIR system [7] facilitates query-by-color. The scheme is based on 11 color categories used throughout by the people.

Later, the low frequency DCT coefficients that are transformed from YUV color space as feature vectors, are used for retrieval of images [8]. This newly proposed model allows users to select its dominant feature of query images in order to improve the retrieval performance. This technique is useful for performing effective retrieval by letting users' opinions on the query images.

The Universal Model for Content-Based Image Retrieval combines the three featured extraction methods namely color, the feature and the edge histogram descriptor [10]. All the image properties analyzed in this work are done by using computer vision and image processing algorithms. The color the histograms of images are calculated for texture co-occurrence matrix based entropy, energy is calculated. For edge density, it is Edge Histogram Descriptor (EHD) that is found. A novel idea is developed based on the greedy strategy to reduce the computational complexity for the image retrieval. All these existing approaches are require large storage spaces and lots of computation time to calculate the matrix of features. So, authors have proposed the efficient content based image retrieval using the advanced color and texture feature extraction, is now deployed. In such scheme, the color features are extracted using three color moments and texture features are extracted directly from block based DCT coefficients which are in transform domain. Consequently, it does not necessitate any complex computation for texture feature extraction.

Some more techniques proposed in literature are as follows:

1. Ivan Lee, Paisarn Muneesawang and Ling Guan [12]:

Ivan Lee, et.al. (1996) have presented the analysis of the CBIR system with the human controlled and the machine controlled relevance feedback, over various different network topologies including centralized, clustered, and distributed content search.

In their experiment for the interactive relevance feedback using RBF, they saw a higher retrieval precision by showing the semi-supervision to the non-linear Gaussian-shaped RBF relevance feedback. Relevance feedback suffers from few vital problems - User interaction for providing feedback is time challenging and it is a tiring process.

2. E.L. van den Broek, L.G. Vuurpijl, P. Kisters and J.C.M. Von Schmid Nijmegen [7]:

E.L. van den Broek et al have proposed a CBIR system with Color Selection scheme which provides facilities for query-by-color, which is depends on eleven color categories, utilized for color scheme by everyone. Here images are extracted via a low frequency DCT coefficient which is changed from YUV color space.

Thus, ultimately system proposed by authors provides proper and improved retrieval performance to help people choose from the dominant features of query images. This methodology will ultimately improve effective retrieval based upon people preferences on the queried image. In the region of interest, Image Indexing System [9], a person can search images from the database through region of interest (ROI). This will provide results based on region preferences while searching through database. But this technique is only sufficient for performing effective retrieval by introducing users' opinions on the query images; this is the main drawback of this technique.

3. Wei-Ying Ma and B. S. Manjunath [13]:

Wei-Ying Ma et al have proposed a model for image retrieval system which was evolved at the University of California at Santa Barbara. For the feature extraction, this system utilizes a hybrid methodology which integrates texture, color, and shape details from an image using indexing technique.

The most significant characteristic of the NETRA is to utilize segmented local regions for indexing of images inside a database. So, both global and local types characteristics are utilized. The retrieval performance of NETRA is based on low-level feature similarity of images and it was soon realized that the performance of NETRA was limited due to the semantic gap as they were not able to infer the interest of the users. To overcome this, most content-based image retrieval systems typically utilize mouse-clicks and other traditional forms of input to identify the regions or objects of interest.

4. John M. Zachary, Jr. and Sitharama S. Iyengar [14]:

John M. Zachary et al have presented the issues from the perspective of real-world system formation. It also discusses some major feature extraction methods utilized in existing CBIR systems. It reviews numerous CBIR system implementations.

5. Rong Zhao and William I. Grosky [15]:

Rong Zhao and William I. Grosky (2002) view that bridging the semantic gap between the low-level features and the high-level semantics is within the interface between the user and the system.

In this paper, author utilized latent semantic analysis which finds different image features co-occur with similar annotation keywords, and consequently lead to improved techniques of semantic image retrieval, but it is observed that latent semantic analysis does not not correctly and efficiently finds the image features. Another research direction is towards improving aspects of CBIR systems by finding the latent correlation between low-level visual features and high-level semantics and integrating them into a unified vector space model.

6. Peter Stanchev, David Green Jr., and Boyan Dimitrov [16]:

Peter Stanchev, et. al. proposed that several visual descriptors exist for representing the physical content of images, for instance color histograms, textures, shapes, regions, etc. Depending on the specific characteristics of a data set, some features can be more effective than others when performing similarity search. For instance, descriptors based on color representation might be effective with a data set containing mainly black and white images.

All techniques based on the statistical analysis of the data set and queries are useful. Even if it is ain't possible to overcome the semantic gap in image retrieval by feature similarity, it is possible to increase the retrieval effectiveness by properly choosing image features, amongst those in theMPEG-7 standard, depending on the characteristics of the various image data sets (obviously, the more homogeneous the data set is, better results can be obtained).

7. Ryszard S. Chora´s [11]:

Ryszard S. Chora´s (2007) contributes their work for the identification of the problems existing in CBIR and Biometrics systems describing image content and image feature extraction. They have described a possible approach of mapping image content onto low-level features. Their paper investigated the use of a number of different color, texture and shape features for image retrieval in CBIR and Biometrics systems.

8. Gaurav Jaiswal and Amit Kaul [17]:

Gaurav Jaiswal and Amit Kaul concluded that content based image retrieval is not a replacement but rather a complementary component to text based image retrieval. The integration of the two can result in satisfactory retrieval performance. In this paper, they reviewed the main components of a content based image retrieval system, including image feature representation, indexing, and system design, along with highlighting the past and current technical achievement.

9. Chih-Chin Lai and Ying-Chuan Chen [18]:

Chin-Chin Lai et al have tried to decrease the gap among the retrieval outcomes and the users' anticipation by properly demonstrating an interactive genetic algorithm (IGA). Here, authors have utilized the attributes of color like the standard deviation, mean value, and image bitmap. The features of texture for example entropy depend upon the gray level co-occurrence matrix and the edge histogram, which are utilized by the authors. In this paper, authors have proposed IGA which is quite difficult and complex for identification of images, further more entropy feature is used as texture feature in which data distributions leads to lacks of the constraints.

10. Sagar Soman, Mitali Ghorpade, Vrushali Sonone and Satish Chavan [19]:

Sagar Somanet have utilized Content Based Image Retrieval (CBIR) techniques on color and texture of images. Here, author demonstrates two valid and different methods for the feature extraction.

General CBIR system utilizes color, texture and shape as the base criterion for the feature extraction technique to get better search results if we query images from databases. In a proposed CBIR system, author proposes use of color and texture for the feature extraction. In order to retrieve texture features, they are applying block-wise Discrete Cosine Transforms (DCT) on the complete image and to extract color feature they have utilized moments of colors (Deviation, Mean and Skewness) on the set of queried images from database.
For getting better results in the image retrieval, they were comparing feature vectors of the query image with the feature vectors of the images in database. They were computed by the separate and combined vectors by utilizing color and texture features and in comparison they showed that the combined feature vector outcomes were comparatively better.

In this paper, authors have utilized Discrete Cosine Transforms (DCT) (to retrieve a texture feature) which is a quite complex method and does not generates better image features. Its application area is limited to the image compression. This technique utilizes the combination of features so it required more memory to store the image features.

11. Swapnalini Pattanaik and Prof. D. G. Bhalke [20]:

Pattanaik, Bhalke (2012) has worked to prove that the Content Based Image Retrieval has surpassed all the limitations of Text Based Image Retrieval by considering the contents or features of image. A query image can be retrieved efficiently from a large database. A Database consists of different types of images has been implemented on the system. Different features such as histogram, color mean, color structure descriptor texture is taken into consideration for extracting similar images from the database. From the experimental result, it is seen that the combined features can give better performance than a single feature. So, selection of a feature is one of the important issues in the image retrieval. The system is said to be efficient if the semantic gap is minimum.

The proposed technique utilized combination of features so it required more space to store the image features, this is the major drawback of this image. The result can be improved in future by introducing feedbacks and users choice in the system.

12. Pooja Verma and Manish Mahajan [9]:

Pooja Verma and Manish Mahajan(2012)have applied canny and sobel edge detection algorithm for extracting the shape features for the images. After extracting the shape feature, the classified images are indexed as well as labeled for making it easy for applying retrieval algorithm in order to retrieve relevant images from the database. In their work, retrieval of the images from a huge image database as required by the user, can be achieved perfectly by using canny edge detection technique according to results.

In this paper, authors compare the results based on the shape feature of image which is extracted by canny and sobel edge detection algorithm, but the results have not shown any improvement as compared to results produced when consider color feature of an image.

13. Swapnalini Pattanaik and D.G. Bhalke [21]:

Swapnalini Pattanaik and D.G. Bhalke have demonstrated the basic idea of proficient retrieval of images by the most common features of Mpeg-7 (Multimedia Content Description Interface- 7). To present standard techniques for demonstrating that the multimedia content is the foremost goal of Mpeg-7 and it also permitted fast and efficient content identification with supporting a large amount of applications.
Here, a color structure descriptor is utilized for color and edge histogram descriptor is used for texture by the authors. We can also increase the performance of CBIR system by utilizing these two features. Here, authors utilized MPEG-7 descriptors to get better results but the implementation of these descriptors is utterly complex and time consuming.

14. Devyani Soni and K. J. Mathai [2]:

Devyani Soni and K. J. Mathai have showed a technique for image retrieval dependent upon text, color space methodologies by the basis of color correlogram and color histogram. Color space methods utilized global color histogram and local color histogram. Those are putted on both color spaces- RGB color space and HSV color space, and then they were compared. The color correlogram is the second order statistical method to compute spatial correlation.

In this work, firstly the image will be investigated according to annotated text and then the color features will be taken out with the help of a color histogram and color correlogram. In this paper, color correlogram achieves a good

retrieval performance among many others, such as color histograms, color coherence vector, but it does comes along with many infeasibility problems such as its massive memory requirement and computational complexity.

III. PROPOSED METHODOLOGY

Since the processors became increasingly powerful, and memories become increasingly cheaper, the deployment of large image databases for a variety of applications now became more realizable and ascertain.

Databases of art works, satellite and medical imagery have been garnering more and more users in various professional fields. For instance, geography, medicine, architecture, advertising, design, fashion, and publishing. The way of effectively accessing desired images from large and varied image databases is now a compulsion. Due to the development of multimedia technology and increasing vogue of the computer network, the conventional information retrieval systems fail to come in accordance with the users' current need.

There are numerous areas in which digital images are applied such as-crime prevention, commerce, finger print recognition, surveillance, hospitals, engineering, architecture, fashion, graphic design, academics, historical research, and government institutions etc. Because of this prevalent demand, we ought to improve retrieval precision and minimized retrieval time.

The primitive methods were only dependent on text based searching instead of its visual feature. A lot of times just a single keyword is redundantly used with more than one image, therefore it lead to erroneous outcomes. Consequently, Content Based Image Retrieval (CBIR) is evolved to defeat the restriction of text based retrieval.

Problems faced in the existing image retrieval systems are: How to retrieve the search image accurately; how we can access and manage a large database of images; how we can make the searching process efficient.

In this dissertation work, we will propose a unique and efficient technique for content based image retrieval. It will utilize a combination of both wavelet transform and the K indicates clustering algorithm. Figure- 3.1 shows the elucidated system for image retrieval.

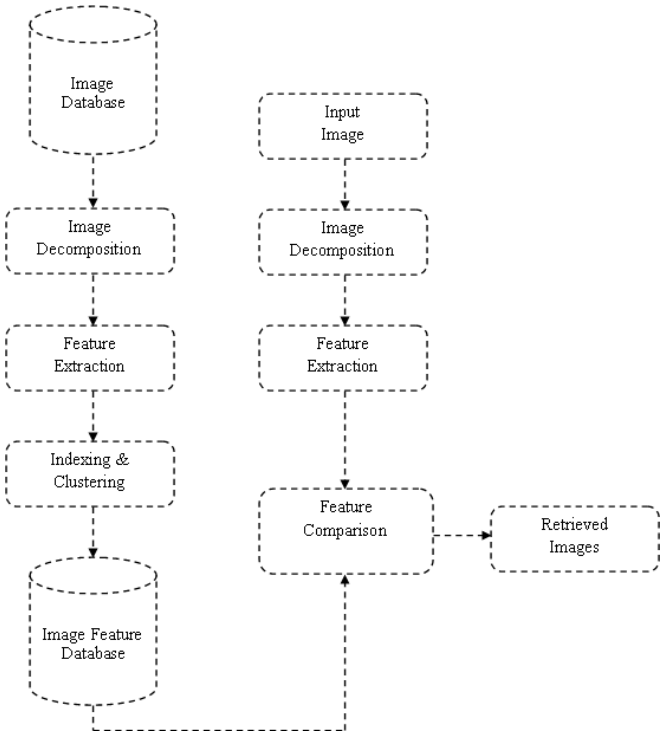

Figure-1: Proposed System for Image Retrieval

As shown in figure-1 there are four important phases of proposed system, image decomposition, feature extraction, indexing & clustering and feature comparison.

- **Image Decomposition& Feature Extraction:**

Image will be decomposed with the assistance of Haar wavelet transform. It tells us that the firstly image will be decomposed and then the color feature database will be prepared for all the images. In decomposition, each image is firstly converted in matrix form, which means that the images get converted in a matrix form and then by utilizing RGB scheme color feature, data set is prepared. In RGB scheme, we get the color i.e., intensity values of the pixels are generated by Haar wavelet transform. We will perform the process of intensity picking in programming with the assistance of getRGB(x, y) function.

- **Image indexing & Image clustering:**

After the image decomposition, we get all the image features in the form of a matrix. Then the images indexed and images with similar features are kept in a cluster using the k-means algorithm. This process is known as indexing & clustering. It will help in improving searching results and ultimately makes image retrieval efficient.

- **Image Search & Image Retrieval:**

Sample image is provided as an input and the elucidated system returns all the similar images as output. The sample image is decomposed and the features are extracted in the form of a matrix, and a color dataset is prepared.
All the features of the input image are compared with the features of the images resides in the database and on the basis of Euclidian similarity measure the relevant images are returned as output.

Haar Wavelet Transform:

- **Wavelet Transform:**

Wavelet transform utilizes both the frequency and spatial correlation of data by means of enlargements (or shrinkages) and transformations of the mother wavelet proceeding the input data. It cares of multi-resolution investigation of data i.e. it can be employed to diverse scales according to the information needed, which permits progressive communication and zooming of the digital image without the necessity of further storage. Another promising characteristic of wavelet transform is its symmetric behavior that is forward as well as inverse transform both has the equivalent complexity, constructing fast decompression and compression procedures.

Its properties well appropriate for digital image compression comprises the capability to take into interpretation of Human Visual System's (HVS) properties, very decent energy compaction skills, great compression ratio, strength under transmission etc. [23].

Haar Wavelet Transform:

The Haar wavelet transform is one of the easiest and elementary transformations from the time and space area to a local frequency area, which exposes the time and space variant spectrum.

An interesting feature of the Haar wavelet transform, comprising fast for implementation and capable to analyze the local feature, build it a probable candidate in modern computer engineering and electrical applications, for instance signal and digital image compression [22]. To get the working of wavelets transform, we take a simple example.

Suppose, we have a one dimensional digital image with resolution of 4 pixels andthe values of pixels are- (9, 7, 3, 5) respectively. We can utilize the Haar wavelet to illustrate this digital image by means of calculating a wavelet transform. For achieving this purpose, we first compute the pairwise average of all given four pixels together and fromthis computed pairwise average wecreate the new lesser resolution digital image with the pixel values (8, 4). Obviously, in this averaging procedure certain information/data is lost.

Now, we ought to store certain detail coefficients to get well the original values of given four pixel from the obtained two average values. For our example, we choose 1 for the first detailed coefficient, for the reason that the computed average is 1 fewer than 9 and 1 additional than 7. We utilized this single number to get well the values of first two pixels from the original four pixel digital image.

In the same manner, we select the second detail coefficient as -1, because3 = 4 + (-1) and 5 = 4 - (-1). Hence we have decomposed the initial image into a set of detail coefficients and a lesser resolution (here into two pixels) description[23]. On repeating this procedure recursively on the obtained averages, we derive the complete decomposition, this is depicted in Table- 1

Table-1: Decomposition of given digital image pixels into lesser resolution

Resolution	Averages	Detail Coefficients
4	[9735]	
2	[84]	[1-1]
1	[6]	[2]

Therefore, by employing Haar wavelet transform on a given four pixel (9, 7, 3, 5) one dimensional digital image the wavelet transform is given by (6, 2, 1, -1).

The decomposition of the two dimensional digital image is a two dimensional generalization of the one dimensional wavelet transformed. It employs the one dimensional wavelet transform to every row of values of pixel. By performing this step, we get an average value of pixels together with detailed coefficients for every row.

Then, these renovated rows are accepted and employ the one dimensional transform to every column. The resultant pixel values are absolutely detail coefficients excluding a particular complete average coefficient. With the purpose to accomplish the transformation, this procedure is recurring recursively merely on the quadrant comprising averages [23].

Now, we will see that how the two dimensional Haar wavelet transform is achieved. The digital image is consist of pixels described by numbers is revealed by G. Beylkin, R. Coifman, and V. Rokhlin in [24]. Now, we will consider a digital image of 8×8 taken from a particular section of a classic image exposed in figure-3.2. The two dimensional matrix representation of this digital image is shown by figure-3.3.

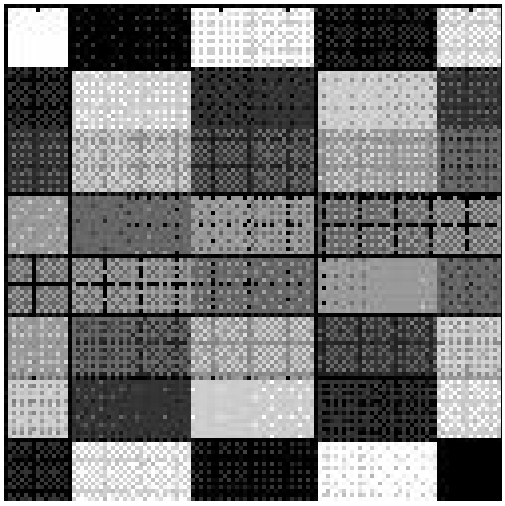

Figure-2: A digital image of 8×8

64	2	3	61	60	6	7	57
9	55	54	12	13	51	50	16
17	47	46	20	21	43	42	24
40	26	27	37	36	30	31	33
32	34	35	29	28	38	39	25
41	23	22	44	45	19	18	48
49	15	14	52	53	11	10	56
8	58	59	5	4	62	63	1

Figure- 3: Two dimensional matrix representation of this image shown in figure-2

Now, we will accomplish the task of averaging andthen differencing to attaina new array/matrix which represents the equivalent digital image in a more succinct manner. Firstly, we consider the 1strow of the figure-3. Averaging: (64+2)/2=33, (3+61)/2=32, (60+6)/2=33, and (7+57)/2=32. Differencing: 64–33 =31, 3–32=–29, 60–33=27, and 7–32= –25. Therefore, the resultant transformed row grows into (33, 32, 33, 32, 31, –29, 27,–25).

Now, we will go through the same process on the average values-(33, 32, 33,32). After that we will perform the similar process on the averages that is on the first two elements of the newly converted row. Averaging: 33+32/2=32.5, and 33+32/2=32.5. Differencing:33-32.5=0.5, and 33-32.5=0.5. Therefore, the new converted row is- (32.5, 32.5, 0.5, 0.5, 31, -29, 27,-25).

Now, the similar process is applied on the average values-(32.5, 32.5). Averaging:32.5+32.5/2 = 32.5 Differencing: 32.5-32.5=0 Ultimately, the converted row grows into- (32.5, 0, 0.5, 0.5, 31, –29, 27,–25). We derive a new matrix shown in figure-4after applying the same process on each row of the whole matrix of figure-3. Now, perform the same process on every column of the matrix in shown in figure-4, we derive the concluding transformed matrix as exposed in figure-5 [23].

32.5	0	0.5	0.5	31	–29	27	–25
32.5	0	–0.5	–0.5	–23	21	–19	17
32.5	0	–0.5	–0.5	–15	13	–11	9
32.5	0	0.5	0.5	7	–5	3	–1
32.5	0	0.5	0.5	–1	3	–5	7
32.5	0	–0.5	–0.5	9	–11	13	–15
32.5	0	–0.5	–0.5	17	–19	21	–23
32.5	0	0.5	0.5	–25	27	–29	31

Figure-4: Transformed matrix after performing the process of averaging and differencing on each row

32.5	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0
0	0	0	0	4	–4	4	–4
0	0	0	0	4	–4	4	–4
0	0	0.5	0.5	27	–25	23	–21
0	0	–0.5	–0.5	–11	9	–7	5
0	0	0.5	0.5	–5	7	–9	11
0	0	–0.5	–0.5	21	–23	25	–27

Figure-5: Final transformed matrix after one step (performing the process of averaging and differencing on each column)

Now, we look the result of process of one step averaging and differencing on adigital image. Figure- 6 (a) shows the original imageof Lena and figure- 6 (b) shows the converted image after performing the process of one step

averaging and differencing. We can perform more steps to get more level of decompositions. Figure- 6 (c) shows the converted image after performing the second step and figure- 6 (d) shows the converted image after performing the third step.

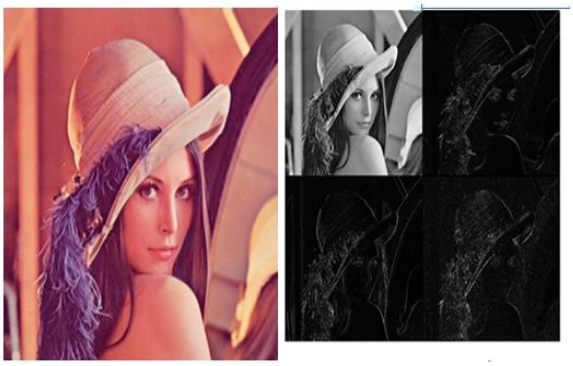

Figure 6(a): Original Lena Image
Figure 6(b): level decomposition after performing the one step averaging and differencing

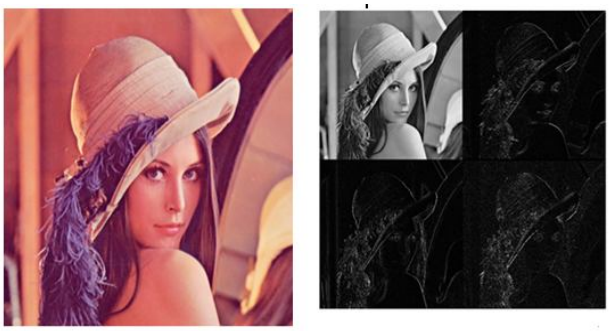

Figure 6(d): Level-2 Decomposition
Figure 6(d)- Level-3 Decomposition

K- means Clustering Algorithm:

Clustering is the technique of dividing a collection of data points into a lesser number of groups, these created groups is known as clusters. In common, assume that we have 'n' data points- X_i, wherei=1 to n and we want to separatethis data points into 'k' clusters. So, our aim is to allocate a cluster to each data point.

Suppose, we are having some object of different colors (for example- objects of 3 colors) and we are given some clusters (for example- number of clusters is 3), so our aim is to allocate a cluster toeach object according to their similarity. Figure- 3.7 shows that thesimilar colored objectsare grouped together into clusters.

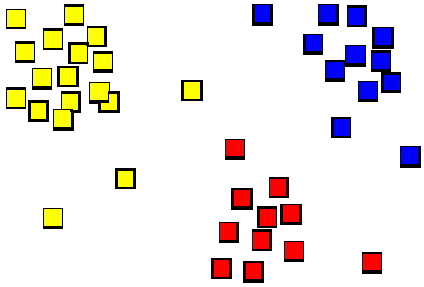

Figure- 7: Similar objects are grouped together into clusters [25]

K-means Clustering:

K-means algorithm is a clustering algorithm that intentions to discover the locations of the clusters that reduce the distance from the objects / data points to the cluster.

The K-Means Clustering Algorithm:

We are given that some data points and some cluster to allot to the datapoints and our aim is to allocate a cluster to each data point. Steps of the k means clustering algorithm are as follows [26]-

- Select the center of clusters randomly.
- Compute the distance from each data point to each cluster center.
- Allot the data points to that clusterswhose distance from the cluster center is lowest.
- Repeat the above step until all the given data points are allotted to one of the cluster.

3.1.1 Euclidian Similarity Measure:

Similarity measure is the measure of similarity in which we compare two given lists of numbers and calculate a single value/number which evaluates their similarity. Euclidian similarity measure utilizes the Euclidian distance formula to calculate the similarity between two data points [27].

Euclidean distance also known as Euclidean metric is the ordinary means straight line distance between two data points in thespace.

The Euclidian similarity measure for finding the distance is given as follows-

$$D = \sqrt{\sum_{i=1}^{n} (X_i - Y_i)^2}$$

Where, X for the input image and Y for the image in the database.

IV. RESULT ANALYSIS

In this dissertation work, we presented a novel and proficient technique for the content based image retrieval. In this technique, we utilized the amalgation of both wavelet transform and k-means algorithm to make the image retrieval process efficient.

Table- 2 indicates the existing efficiency of the DCT &color moments system and proposed system. The comparison graph for the efficiency of existing DCT &color moments system and proposed system for different category of images is shown by figure- 8. The figure- 8 shows that the proposed system offers the greater efficiency in comparison with the existing DCT &color moments system. Table- 2 indicates the efficiencies of existing systems like-Funt's Method, DCT & color moments technique and proposed system.

Table- 2: Comparison of efficiency of existing DCT & color moments system and proposed system

Category	Efficiency of DCT & Color Moments	Efficiency of Proposed System
Dinosaurs	70	75
Horses	70	75
Buses	50	85
Roses	60	70
Mountains	50	65
Average	60	74

Table- 3: Comparison of existing techniques of image retrieval

Techniques	Funt's Method	DCT & Color Moments	Proposed Technique
Efficiency	36%	60%	74%

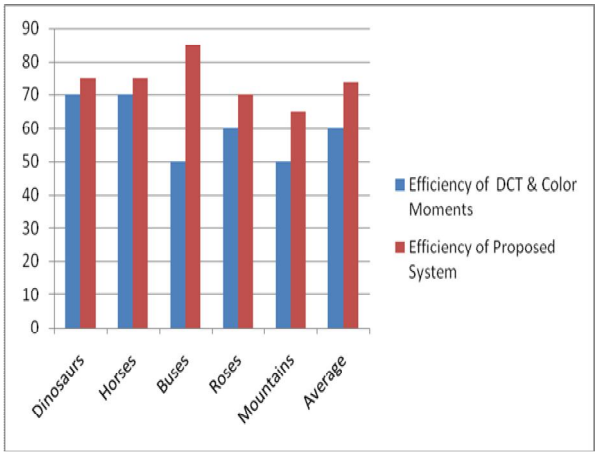

Figure- 8: Comparison graph for efficiency of existing DCT & color moments system and proposed system for different category of images

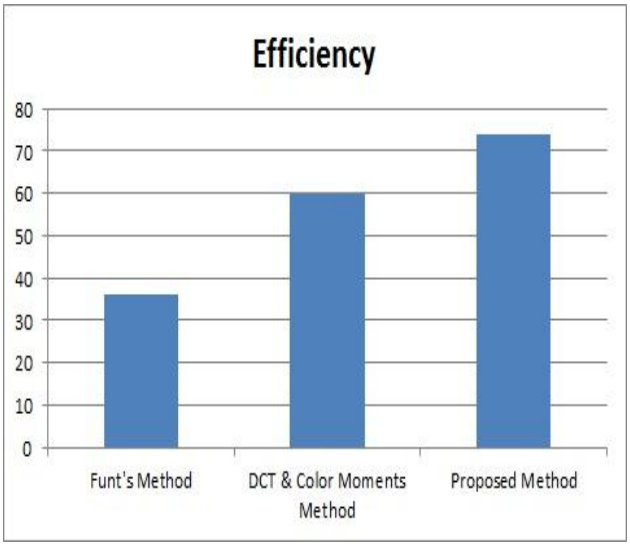

Figure- 9: Comparison graph for efficiency of existing systems like- Funt's Method, DCT & color moments technique and proposed system

Figure- 9 shows that the Comparison graph for efficiency of existing systems like- Funt's Method, DCT & color moments technique and proposed system. It is cleare from figure-9, that the proposed system for content based image retrieval performs better as compared to the existing image retrieval techniques like- Funt's Method and DCT & color moment's technique.

V. CONCLUSION

Due to development of multimedia technology and increasing vogue of the computer network, the conventional information retrieval systems are inept to overbear the users' current need. There are numerous regions where digital images are used, they are crime prevention, commerce, finger

print recognition, surveillance, hospitals, engineering, architecture, fashion, graphic design, academics, historical research, and government institutions etc.

Due to the far-reaching demand, there is a necessity to improve retrieval precision and minimized retrieval time. The primitive methods were dependent on the text based searching instead of its visual feature. A lot of times just a single keyword is redundantly used with more than one image, leading to erroneous outcomes.

Consequently, Content Based Image Retrieval (CBIR) is educed to surmount the restriction of text-based retrieval. Content based image retrieval is really an edifying field to have research upon. In this dissertation work, we presented a novel and proficient technique for the content based image retrieval.

In this technique, we utilized the amalgamation of both wavelet transform and k-means algorithm to make the image retrieval process efficient. For the purpose of similarity computation between input image and other images of dataset, we utilized Euclidian similarity measure. It is shown by experimental results that, the proposed system executes better than the existing CBIR techniques (Funt's Method, Swain's Method, and DCT & color moments technique).

The future works are as follows:

- The future works is mentioned in the following: In this research we consider only color feature of images, in future we will take features like texture and shape and combination of all the features to enhance the retrieval technique.
- The presentation of the retrieval system is based on the efficient decomposition of given image, so by utilizing other methods in future we will improve the process.

REFERENCES

[1] Sang-Mi Lee, Hee-Jung Bae and Sung-Hwan Jung, "Efficient Content based Image Retrieval Methods using Color & Texture", ETRI Journal, Vol. 20, No. 3, pp.272-283, 1998.

[2] Devyani Soni and K. J. Mathai, "An Efficient Content Based Image Retrieval System using Text, Color Space Approach and Color Correlogram", International Journal of Advance Foundation and Research in Computer (IJAFRC), ISSN 2348 – 4853, Volume 1, Issue 8, August2014.

[3] GulfishanFirdose Ahmed and RajuBarskar "A Study on Different Image RetrievalTechniques in Image Processing", International Journal of Soft Computing and Engineering (IJSCE),ISSN: 2231-2307, Volume-1, Issue-4, September 2011.

[4] Dr. H.B.Kekre, Sudeep D. Thepade and Akshay Maloo, "Image Retrieval using Fractional Coefficients of Transformed Image using DCT and Walsh Transform", International Journal of Engineering Science and Technology, Vol. 2, No. 4, pp. 362-371, 2010.

[5] Khan W., Kumar S., Gupta N., and Khan N., "A Proposed Method for Image Retrieval using Histogramvalues and Texture Descriptor Analysis", IJSCE, ISSN: 231-2307, Volume-I Issue-II, May 2011.

[6] Vadivel A., MajumdarA. K., and SuralS., "Characteristics of Weighted Feature Vector in Content-Based Image Retrieval Applications", 2000.

[7] E.L. van den Broek, L.G. Vuurpijl, P. Kisters and J.C.M. Von Schmid Nijmegen, "Content Based Image Retrieval: Color Selection exploited," International Journal of Engineering Science and Technology, Vol. 30, No. 3, pp. 456-462, 1997.

[8] Tienwei Tsai, Taiwan, Yo Ping Huang and Te-Wei Chiang, "Fast Image Retrieval Using Low Frequency DCT Coefficients," International Journal of Engineering Science and Technology, Vol. 6, No. 3, pp. 106-120, 2003.

[9] Pooja Verma, Manish Mahajan, "Retrieval of better results by using shape techniques for content based retrieval",IJCSC ,Vol. 3, No.2, January-June 2012, pp. 254-257, ISSN: 0973-7391.

[10] S. Nandagopalan, Dr. B. S. Adiga, and N. Deepak, "A Universal Model for content Based Image Retrieval," International Journal of Computer Science, Vol. 4, No. 4, pp. 531-538, 2009.

[11] Ryszard S. Chora´s, "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems", International Journal of Biology and Biomedical Engineering Issue 1, Vol. 1, 2007.

[12] Ivan Lee, Paisarn Muneesawang and Ling Guan, "Automatic Relevance Feedback for Distributed Content-Based Image Retrieval", ICGST, ieee.org FLEXChip Signal Processor (MC68175/D), Motorola, 1996.

[13] Wei-Ying Ma and B. S. Manjunath, "NeTra: A toolbox for navigating large image databases", Journals of Multimedia Systems, pp 184–198, 1999.

[14] John M. Zachary, Jr. and Sitharama S. Iyengar, "Content Based Image Retrieval Systems", 2001.

[15] Rong Zhao and William I. Grosky, "Bridging the Semantic Gap in Image Retrieval", Wayne State University, USA, Idea group publishing, 2002.

[16] Peter Stanchev, David Green Jr., and Boyan Dimitrov, "MPEG-7: The Multimedia Content Description

Interface", International Journal of Information Theories & Applications, Vol.11, 2004.
[17] Gaurav Jaiswal and Amit Kaul, "Content Based Image Retrieval", National Conference on Computing, Communication and Control, National Institute of Technology, Hamirpur-177001, Himachal Pradesh (India), 2009.
[18] Chih-Chin Lai, Member, IEEE, and Ying-Chuan Chen," A User-Oriented Image Retrieval System Based on Interactive Genetic Algorithm", IEEE transactions on instrumentation and measurement, vol. 60, no. 10, October 2011.
[19] SagarSoman, MitaliGhorpade, VrushaliSonone andSatishChavan, "Content Based Image Retrieval Using Advanced Color and Texture Features", International Conference in Computational Intelligence (ICCIA) 2011.
[20] Swapnalini Pattanaik and Prof. D. G. Bhalke, "Beginners to Content Based Image Retrieval", International Journal of Scientific Research Engineering &Technology (IJSRET), Volume 1 Issue2 pp 040-044 May 2012.
[21] SwapnaliniPattanaik and Prof. D.G. Bhalke, "Efficient Content based Image Retrieval System using Mpeg-7 Features", International Journal of Computer Applications (0975 – 8887),Volume 53– No.5, September 2012.
[22] Pei-Yu Chao "Haar Transform and Its Applications", Time frequency analysis and wavelet transform tutorial.
[23] Kamrul Hasan TalukderI and Koichi Harada, "Haar Wavelet Based Approach for Image Compression and Quality Assessment of Compressed Image", IAENG International Journal of Applied Mathematics, 36:1, IJAM_36_1_9, February 2007.
[24] G. Beylkin, R. Coifman, and V. Rokhlin, "Fast wavelet transforms and numerical algorithms I", Communications on Pure and AppliedMathematics, 44(2): 141-183, March 1991.
[25] http://www.statisticshowto.com/cluster-analysis/.
[26] A. K. Jain, M.N. Murty and P.J. Flynn, "Data Clustering: A Review", ACM Computing Surveys, Vol. 31, No. 3, September 1999.
[27] Anna Huang, "Similarity Measures for Text Document Clustering", NZCSRSC 2008, Christchurch, New Zealand, April 2008.
[28] https://netbeans.org/community/releases/73/.
[29] James Gosling, "Java: an Overview", jag@sun.com, February 1995.
[30] http://www.oracle.com/technetwork/topics/newtojava/overview/index.html.
[31] http://wang.ist.psu.edu/docs/related/.