HYBRID APPROACH FOR EFFECTIVE FEATURE EXTRACTION TECHNIQUE IN CONTENT BASED IMAGE RETRIEVAL

[1] Ruchi Kapadia, [2]Swarndeep Saket J

[1]*Student of Gujarat Technological University, Department of Computer Engineering, L. J. Institute of Engineering and Technology, Gujarat Technological University, Ahmedabad, Gujarat, India*
[2]*Assistant Professor at L. J. Institute Of Engineering & Technology, Ahmedabad, Gujarat, India*
[1]*ruchikapadia1@gmail.com*, [2]*nandasaket1990@gmail.com*

ABSTRACT

Content Based Image Retrieval (CBIR) is a very important and increasingly popular approach that helps in the retrieval of image data from a huge collection of Image. Image representation based on certain features helps in retrieval process. Content-based image retrieval technique is in huge demand for few domains like Weather forecasting, data mining, remote sensing, medical imaging, education, crime prevention and management of earth resources. Content Based Image Retrieval (CBIR) is process to find similar image data in the large amount of image database when a query image is given by user. A user gives input to the system in the form of specified query image and system return set of relevant images related to query image. In this report we analyze different Content Based Image Retrieval techniques and their comparative study. Image retrieval process and improve visual similarity search in content-based image retrieval many studies have been conducted and many methods developed in recent years, but there are a few issues that need to be addressed. Color, shape and texture feature are extracted from the image. To improve accuracy in terms of Precision and Recall, Ant colony Optimization (ACO) algorithm and Support Vector Machine (SVM) method is used in proposed model.

Keywords: *CBIR, Support Vector Machine, Feature Extraction, Precision, Recall, Similarity measure, Ant colony Optimization.*

I. INTRODUCTION

With the Internet developing rapidly, image databases become bigger and more diverse. The way to effectively retrieve digital images in a large library is still highly challenging. Professions such as law enforcement, graphic design, fashion design, medicine, publishing, advertising, crime prevention, engineering and architectural design now-a-days make extensive use of digital image databases to maintain a record and use them when necessary. This lead to a demand for a system that can quickly and effectively retrieve images which are not only similar but are also relevant[9]. An image retrieval system provides a way to access, browse and retrieve images efficiently, from the large databases . Image retrieval systems came out in light in 1979 when a conference on Database Techniques for Pictorial application was held in Florence. Retrieval of image from large databases are exponentially increasing and becoming a challenging task. For image retrieval, two methods are used namely Content Based Image Retrieval System (CBIR) and Text Based Image Retrieval System (TBIR)[3].

In the text-based system, text descriptors are used to manually annotate the images and then database management systems are used for retrieval of images. But, there are two limitations of using this approach. First, considering the large image databases, it is not feasible to manually annotate all the images and second describing the features present in an image is highly subjective. these limitations This limitations are remove by CBIR. In CBIR images are retrieval on the basis of content. In CBIR systems, image processing techniques are used to extract visual features such as color, texture and shape from images [8]. CBIR has diverse applications in internet, multimedia, medical image archives, crime prevention, entertainment, and digital libraries and it is an important field in image processing [2].

In this paper, Section 2 describes basic working of content based image retrieval .Section 3 describes related work for various method of content based image retrieval . Section 4 describes comparative study of techniques which have been used for content based image retrieval. Proposed work and Conclusion is presented in Section 5 and 6 respectively.

II. CONTENT BASED IMAGE RETRIEVAL

Content Based Image Retrieval (CBIR) is defined as a process to find similar image in the image database when a query image is given. The generalized CBIR system shown in figure 1.

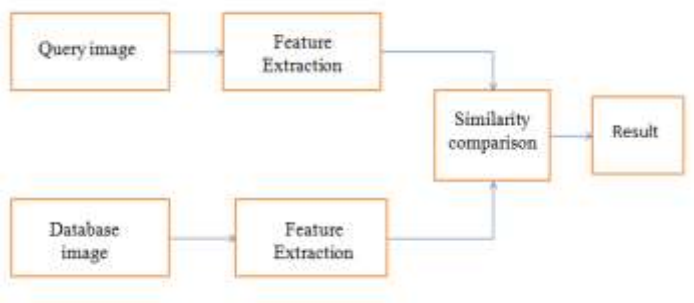

Fig -1: Working of CBIR

Features are extracted for both query image and images in the database. The distance (i.e., similarities) between the features vectors of the query image and database are then computed and ranked. The database images that have highest similarity to the query image are retrieved. Then the performance analysis is carried out using precision and recall. Content based image retrieval is working with different types of image database. The Content based image retrieval involves two phases [2]. First is Feature extraction process to extract the features in terms of information of an image based on their visual contents for comparing images. And second is Matching process in which images are compared with some distance metrics, how much they are similar to each other.

III. RELATED WORK

There have been many researchers who have attempted in Content Based image retrieval System. Some reviews are presented in Table I.

Table- 1: Comparison of Implemented Techniques

Sr. No	Title	Method Used	Advantages	Disadvantages
1	Combination of Global and Local Features using DWT with SVM for CBIR	HDWT , SVM , GLCM , Color correlogram and Color Histogram is used	Improve exactness and execution recovery	Need to implement more classification technique

2	Content Based Image Retrieval using Color Edge Detection and Discrete Wavelet Transform	Canny edge detector Manhattan distance and DWT is used	Robustness is high against query image alteration	Searching speed is Less.
3	Complementary Feature Extraction approach in CBIR	LBP,RGB Color descriptor , Euclidean distance	Better image quality on human vision.	Texture analysis of LBP operator Is less depend on security
4	Content Based Image Retrieval Using Interactive Genetic Algorithm	EHD , IGA , Mean , Standard deviation, Image bitmap ,GLCM is used	Less no of irrelevant images are retrieval	Take only low level feature
5	A Method For Content-Based Image Retrieval Using Visual Attention Model	HOG ,saliency detection algorithm, SIFT ,earth mover's distance	More robust on variation in rotate and scale	Only Work on visual variation
6	A Revised Averaging Algorithm for an Effective feature extraction in CBIR system	K-mean clustering algorithm, Sobel edge detection algorithm	Reduce the complexity of feature extraction model .	Not respond properly for image that contain only Sceneries
7	SIFT implemented Efficient Content Based Image Retrieval System Using Neural Network	Neural network , SIFT ,color moment, Canny edge detection algorithm , GLCM is used	Good for understanding-g complex Semantic	Time consuming
8	Optimization of image retrieval by using HSV color space , Zernike moment and DWT technique	Color Histogram , Color Correlogram ,Zernike moment , DWT ,Gabor filter is used	Robustness is high	Add Extra low level Feature such as shape to improve efficiency

Here Table 1 provides the information about the method used and also include about the advantages and disadvantages of each and every methods.

IV. TECHNIQUE STUDY OF PROPOSED METHOD

In existing system ACO and SVM is added for improving retrieving process. In proposed system applied ACO is applied on color, shape and texture feature. And SVM is used for classify the images.

A.Support Vector Machine

The main element of Support Vector Machine is to create hyper planes or a collection of hyper planes with the help of support vectors in a higher dimension space. SVM used for classification. It divides the space into two half spaces. A 'good separation' is reached through hyper planes that have the major closest data distance to the points. Here decent separation means superior the division between two hyperplanes gives lesser generalization error. That's by it is known as a maximum margin classifier. If geometric gap between hyper planes more elevated than classification error is low [1].

A SVM is a one type of linear separator. If we want to separate the black circle from white circle by drawing a line .this task will achieve by an infinite number of lines.SVM help in that case.SVM find the "maximum-

margin" line. “maximum-margin" line is the line “in the middle”. We draw the straight line in middle to separate data in to two class .But what happened if we don’t use the straight line but a curved one? . In this case "lifting" the features that we observe into higher dimensions. If we draw the line in this higher dimension down to our original dimension, it looks like a curve. This line in higher dimension is known as “hyperplan” .This task efficiently done by Kernel trick. By using Kernel we can easily draw the hyperplan in in very high dimensions, and even infinite dimensions.

B.Ant Colony Optimization (ACO)

ACO can be used for feature selection task by the way that nodes represents features and edges show the selection of next feature. So, the full-connected graph is considered in feature selection. In ACO-based feature selection the objective is find a path with minimum cost in graph. Ant Colony Optimization (ACO) is a population based met heuristic approach to find approximate solutions to difficult optimization problems.
The inspiring source of ACO is the pheromone trail laying behaviour of real ants, which use pheromone as a communication medium. In analogy to the biological example, ACO is modelled based on the indirect communication of a colony of simple agents, called artificial ants, mediated by artificial pheromone trails.

These pheromone trail values are modified at runtime based on a problem-dependent heuristic function and the amount of pheromone deposited by the ants while they traverse between their colonies and a food source. The problem-dependent heuristic function, in the case of famous ACO algorithms for travelling salesman problem, is set to be the inverse of the distance between one city and another city. In ACO, pheromone trail values serve as distributed, numerical information, which the ants use to construct solutions probabilistically. There is one solution per ant.

The higher the pheromone value (initial edge), the higher the probability of an ant choosing that particular trail will be. The pheromone values on lower quality trails which are not reinforced often enough will progressively evaporate. it avoids the algorithm from converging too rapidly toward a suboptimal region (final edge map), therefore, as mentioned above, it is repeatedly applied until a termination condition is satisfied.

C.Speeded Up Robust Feature (SURF)

SURF is an algorithm in computer vision to detect and describe local feature in image. As the name suggest in computer vision, SURF is shown as one of the most fastest algorithm for detect local Feature from image. It is invariant to image rotation, change in scale and illumination. The Speeded-Up Robust Features (SURF) detector-descriptor scheme developed by Bay is designed as an efficient alternative to SIFT. SIFT use the 128 dimensional feature vector to describe the local feature while SURF use the 64 dimensional feature vector to describe the local feature. It is much faster, and more robust than SIFT (Scale Invariant Feature Transform) detector-descriptor. For detection of interest point instead of using Gaussian derivative, the computation is based on simple 2D box filters; where, it uses a scale invariant blob detector based on the determinant of Hessian matrix for both scale selection and locations.

D.HSV Color Histogram

The HSV color histogram decouple the luminance component of a pixel color form its chrominance components (Hue and saturation). HSV is the important color model and it’s design reflect the way Human see the color. For example one can not refer to the color of a any dress material by giving the percentage of RGB component. When we see the color we describe it in terms of its Hue, Saturation and Value. In the HSV color space model, H stands for Hue. Hue describes the property of Pure color. Hue is what artist refer to as “Pigment” example like Yellow, Orange etc.

In the HSV color space model, S stands for Saturation. Saturation measures the degree to which pure color is diluted by white light. Saturation allows adjusting of the ‘Strength’ of color. In the HSV color space model, V stands for Value (Brightness).Main purpose of V is simply the average of R, G and B component. Hue and saturation component provide the color information can be illustrated by the color circle. As a result, this is an ideal tool for developing image Processing algorithm. For HSV color Histogram First convert the RGB model to HSV color space and then histogram is calculate. Histogram provide the graphical representation of color over the image.

E.Canny Edge Detection Algorithm

The Canny edge detection algorithm is known as the optimal edge detector. The detector ensures only one response to a single edge and it provides shape that is optimal at any scale. Canny edge detector is also able to cope up with noise in the image[2].

Basic steps of canny edge detector:

Smoothing: In this step Gaussian filter is used. With the help of Gaussian filter first Smooth the image to eliminate noise from image.

Compute Derivation: After filter the image finds derivation of gradient of filtered image.

Find Magnitude and Orientation: In third phase ,after computing the derivation find magnitude and orientation of gradient image.

Apply Non Maximum Suppression: The algorithm then tracks along these regions and suppresses any pixel that is not at the maximum. It only allows edge whose magnitude is large.

Apply Hysteresis thresholding : In this phase scan the image from left to right, top to bottom and provide track along the remaining pixels that have not been suppressed. If the gradient at pixel at above high declares it is an edge pixel. if the gradient at pixel at below low declare it is an non edge pixel. And if the gradient at pixel at below high and above low and it is connected to edge pixel then declare it is an edge pixel.

F.Color moment

Color moment provide the color distribution information .For Color indexing purpose mostly Color moment method used. It is invariant to scale and rotation. It provide good color feature under the lighting condition .In this purposed model second order color moment (mean and standard derivation) is used. Mean provide the average color in the image. By taking the square root of the variance of the color distribution we obtain the Standard derivation. Standard derivation provide the contras value and spread the data.

G.Color Auto correlogram

Color auto-Correlogram is the feature to represent color information in image. Color Auto correlogram mainly capture the each color distribution in image It includes the spatial relationship between colors. Auto correlogram is provide the overall distribution of local spatial relationship between colors. The information in color correlogram require $O(m^2d)$ while The information in color Auto correlogram require O(md).where m is color and d stands for distance between two pixel in image. It require less space for storing the color feature

V. PROPOSED WORK

As seen all the methods that are invented for CBIR , most of the method suffers through about the less precision, less Recall, Time consuming ,Searching speed, Accuracy, Semantic gap , Less efficient , Extract more feature etc . The proposed model main focus is to improve precision and recall. In order to find more relevant image, a new Content based image retrieval technique has been proposed and following are the steps.

Step 1: Take the image from the database and query image.
Step 2: Applying pre processing technique on image.
Step 3: Extract the color feature using HSV color Histogram, calculate mean and variance and color Auto Correlogram.
Step 4: Extract the Shape feature using Canny method and For texture feature extraction SURF method is applied.
Step 5:Store this feature in feature vector.
Step 6: Apply ANT colony optimization algorithm are used to finding the edges from the image and Select the most important feature among color, shape and texture. ACO provide the optimum result.
Step 7:Apply the Support vector Machine for Classification.SVM provides the Result image which are same as the query image. Use the Manhattan distance for similarity Comparison.
.

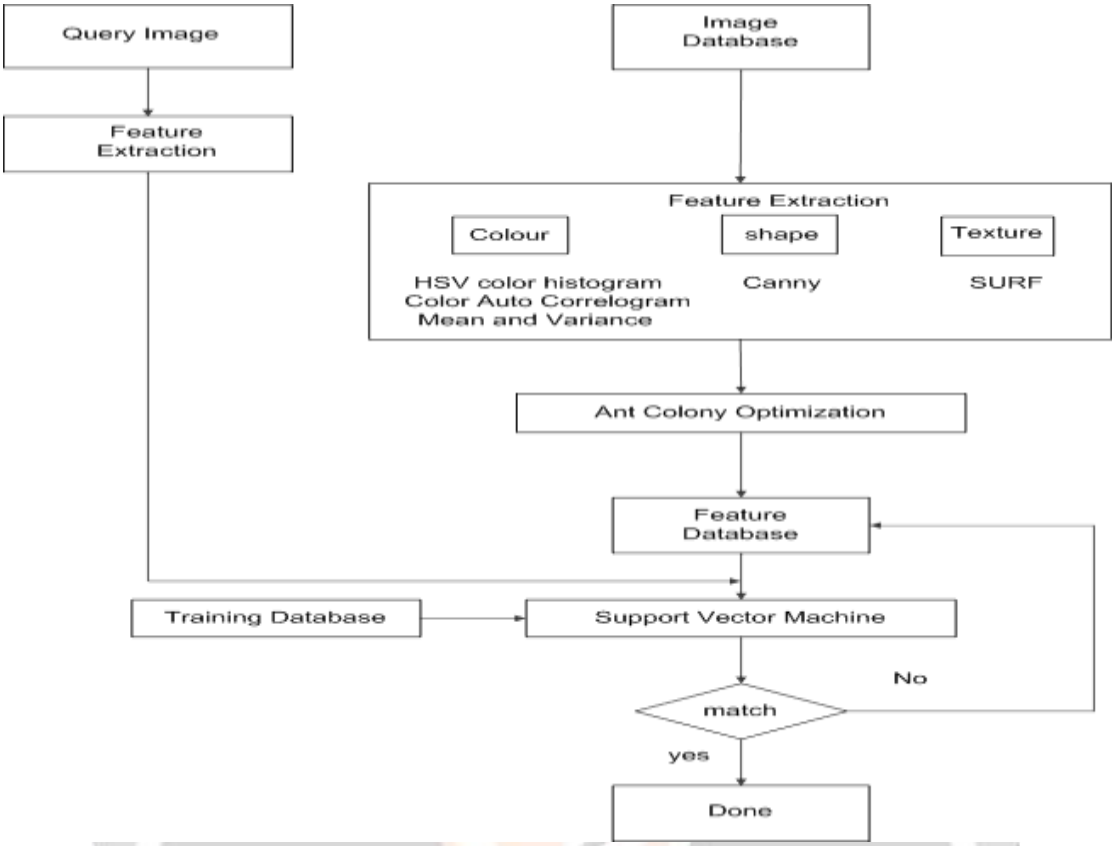

Fig -2: Proposed diagram of CBIR

VI. RESULT ANALYSIS

For implementation of Content based image retrieval system has been experimented through Matrix Laboratory (MATLAB) software which is running on laptop with a 2 GHz Core2duo with 2GB RAM and windows 8 operating system. CBIR performance is analyzed by evaluate the values of precision and recall. Those measures are the basic initial measures of the image retrieval process. Proposed framework is implemented in MATLAB software using Wang Database which contains 1000 images. Wang Database contains 10 class of different Category of Images like: Flower, Bus, Elephant, Dinosaur, Horse, Food, Beach, Building, African, and Mountain.

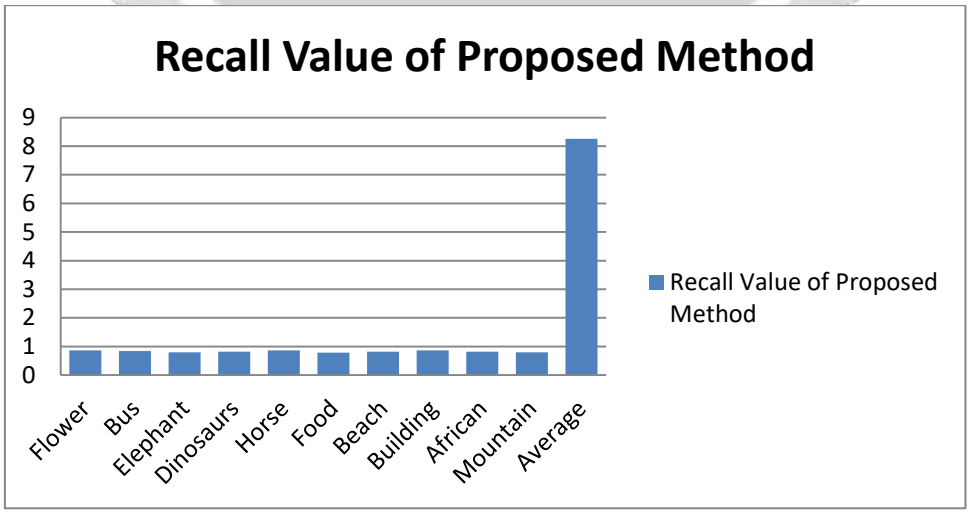

Fig -3: Recall Comparison Graph

VII. CONCLUSION

Content based Image Retrieval System is a method to retrieve the image from large database when query image is given. According to literature review various author have research based on content based image retrieval technique .In proposed method used color histogram, color auto correlogram, mean and variance for extract the meaning full color content. Canny method provide optimum Shape feature and SURF feature provide the texture feature. In proposed work try to get the maximum accuracy of retrieved image using Ant colony optimization and Support Vector Machine algorithm also prove result with different parameter like Precision and recall.

REFERENCES

[1] Gupta, Ekta, and Rajendra Singh Kushwah. "Combination of Global and Local Features Using DWT with SVM for CBIR." 2015 4TH INTERNATIONAL CONFERENCE ON RELIABILITY, INFOCOM TECHNOLOGIES AND OPTIMIZATION (ICRITO) IEEE 2015, doi:10.1109/ICRITO.2015.7359320.

[2] Agarwal, Swati, A. K. Verma, and Nitin Dixit. "Content Based Image Retrieval Using Color Edge Detection and Discrete Wavelet Transform." 2014 INTERNATIONAL CONFERENCE ON ISSUES AND CHALLENGES IN INTELLIGENT COMPUTING TECHNIQUES (ICICT) IEEE 2014, pp.368-372, doi: 10.1109/ICICICT.2014.6781310.

[3] Kumar, Kamlesh, Jian-Ping Li, and Zain-Ul-Abidin. "Complementary Feature Extraction Approach in CBIR." 2015 12TH INTERNATIONAL COMPUTER CONFERENCE ON WAVELET ACTIVE MEDIA TECHNOLOGY AND INFORMATION PROCESSING (ICCWAMTIP) IEEE 2015,pp. 192-197,doi:10.1109/ICCWAMTIP.2015.7493973.

[4] Abdolreza Rashno,Hossein SadeghianNejad and Saeed sadri "An Efficient content-based image retrieval with ant colony optimization feature selection schema based on wavelet and color features" 2015 The International Symposium on Artificial Intelligence and Signal Processing (AISP) IEEE 2015,pp: 59 - 64, DOI: 10.1109/AISP.2015.7123522

[5] Mohammad pour, Mostafa, and Saeed Mozaffari. "A Method for Content-Based Image Retrieval Using Visual Attention Model."2015 7TH CONFERENCE ON INFORMATION AND KNOWLEDGE TECHNOLOGY (IKT)IEEE 2015 ,doi:10.1109/IKT.2015.7288764.

[6] adhav, Pratima, and Rashmi Phalnikar. "SIFT Implemented Efficient Content Based Image Retrieval System Using Neural Network." 2015 INTERNATIONAL CONFERENCE ON INFORMATION PROCESSING (ICIP) IEEE 2015, pp. 772-777,doi: 10.1109/INFOP.2015.7489486.

[7] Bhoomika Gupta , Shilky Shrivastava and Manish Gupta "Optimization of Image Retrieval by using HSV Color Space ,Zernike Moment &DWT Technique" 2015 IEEE INTERNATIONAL CONFERENCE ON COMPUTATIONAL INTELLIGENCE AND COMPUTING RESEARCH (ICCIC) IEEE 2015, doi:10.1109/ICCIC.2015.7435783 .

[8] K.B.A.B.Chathurika and P.K.S.C.Jayasinghe" A Revised Averaging Algorithm for an Effective Feature Extraction in Component- based Image Retrieval System" 2015 IEEE INTERNATIONAL ADVANCE COMPUTING CONFERENCE (IACC) - IEEE 2015, pp.1153-1157, doi:10.1109/ICCIC.2015.7435783.

[9] Sushant Shrikant Hiwale and Dhanraj Dhotre "Content-Based Image Retrieval: Concept and Current Practices" 2015 INTERNATIONAL CONFERENCE ON ELECTRICAL, ELECTRONICS, SIGNALS, COMMUNICATION AND OPTIMIZATION (EESCO) – IEEE 2015, doi: 10.1109/EESCO.2015.7254040

[10] Ashwini Vinayak Bhad and Komal Ramteke "Content Based Image Retrieval a Comparative Based Analysis for Feature Extraction Approach" 2015 International Conference on Advances in Computer Engineering and Applications ,IEEE 2015, pp: 260 - 266, DOI: 10.1109/ICACEA.2015.7164712.

[11] Kamlesh Kumar,Muhammad, Jawad Hussain;,Jian-Ping Li,Asif Khan;,Saeed Ahmed Khan and Riaz Ahmed Shaikh"Complementary semantic model for contentbased image retrieval"2014 11th International Computer Conference on Wavelet Actiev Media Technology and Information Processing(ICCWAMTIP),IEEE 2014,pp:266-270,DOI: 10.1109/ICCWAMTIP.2014.7073405.

[12] Madhavi, K.V.,Tamikodi, R.,&Sudha, K.J.(2016). ‘’An innovative method for retrieving relevant images by getting the top-ranked images first using interactive genetic algorithm”,procedia computer science, 79,ELSEVIER 2016 ,pp:254-261, doi:10.1016/j.procs.2016.03.033.

.