International Journal of Advance Engineering and Research Development (IJAERD)
Volume 1,Issue 5,May 2014, e-ISSN: 2348 - 4470 , print-ISSN:2348-6406

Feature based Image retrieval based on clustering, classification techniques using low level image features

Mit Patel[1], Keyur Bhrahmbhatt[2], Kanu Patel[3]

[1]*PG Scholar,Department of Computer Engineering, BVM Engineering College, ,Gujarat Technology University VallabhVidyanagr,Anand,Gujarat,India„mtpatel1989@gmail.com*
[2]*Assistant Professor,Department of Information Technology, BVM Engineering College, ,Gujarat Technology University VallabhVidyanagr,Anand,Gujarat,India, keyur.brahmbhatt@bvmengineering.ac.in*
[3]*Assistant Professor,Department of Information Technology, BVM Engineering College, ,Gujarat Technology University VallabhVidyanagr,Anand,Gujarat,India,kanu.patel@bvmengineering.ac.in*

Abstract - People are able to take photos using hand held devices and there is a massive increase in the volume of photos digitally stored. However, this tremendous increase in the number of digitally captured and stored images necessitates the development of advanced techniques capable of classifying and effectively retrieving relevant images when needed. Thus, Image retrieval has been popular for several years. There are different system designs for image retrieval system. Image Retrieval is a technique of automatic indexing and retrieving of images from a large data base.

Proposed system uses concept of data mining and image processing . Proposed system which uses a well-known clustering algorithms k-means, fuzzy rule based classification and a database indexing structure to facilitate retrieving relevant images in an efficient and effective way. Colour histogram moments for RGB components and for HSV components are used for colour information extraction. Gabor filter is used for texture feature extraction. Image is clustered by its texture information, it is first level clustering which is done using K-mean clustering and secondly image is classified using its colour information which is done using Fuzzy rule based classification. Use of clustering and classification reduces the search space. The selection of cluster based on texture and colour information is done by various experiments on different images. Fuzzy rule based classification allowed one data point to belong to more than one class so retrieved image share good similarity than normal clustering.

KeyWord:- Feature extraction, Image retrieval, Clustering Algorithm, Rule Based Classification

I.INTRODUCTION

In recent years, very large collections of images and videos have grown rapidly. In parallel with this growth, content-based retrieval and querying the indexed collections are required to access visual information. Now days data management of images is done based on text or tags assigned to these images. So when database is searched the image is retrieved based on the text tags assigned to that image. The content of this search shows ambiguity. There are two disadvantages with this approach. The first is that a considerable level of human labour is required for manual annotation. The second is the annotation inaccuracy due to the subjectivity of human perception. So Content based image retrieval system is used to do operation on image database like image retrieval and image classification.

Image retrieval is a technique, which uses visual contents to search images from large scale image databases according to user's interests. The main goal of the feature based image

retrieval is to find images which are similar to query image visually without using any textual descriptions for the image. There will be 4 major steps to perform image retrieval based on the similarity: 1. Load Query Image 2. Generate Signature using Suitable Techniques 3. For every images in the database generate the signatures and store it in database 4. Calculate Distance for Key Image Signature and Database Image Signature using a Suitable Distance Measure and find the possible matches.[4]

1.1 Related work

The development of automated feature based image retrieval (FBIR) or CBIR systems is very attractive research area because of its wide range of applications in different fields medical imaging, digital photo library, space imaging, bioinformatics, etc.

Efficient retrieval of images is using different Mpeg-7 Features. The main objective of Mpeg-7 is to provide a standardized set of technologies for describing multimedia content. It has allowed quick and efficient content identification, and addressing a large range of applications.[4] Images are to be clustered based on the RGB components. The Grey Level Co-occurrence Matrix (GLCM) is used to extract second order statistics from an image. GLCMs have been used very successfully for texture calculations. From GLCM all the features are calculated and stored into the database. Then, the given query image to be retrieved from the concerned cluster based on the texture features. By which, the time complexity can significantly be reduced at most level [2] Other propose a novel approach which uses a well-known clustering algorithm k-means and a database indexing structure B+-tree to facilitate retrieving relevant images in an efficient and effective way. used Daubechies wavelet transformation for extracting the texture feature vectors of images.[6] Another propose segmentation and grid module, feature extraction module, K-means and k-nearest neighbor clustering algorithms and bring in the neighborhood module to build the system.[3] The significant contributions in their approach include expanding the association model to numerous reference objects, integrating the spatial information into the Bayesian decision rule as spatial priors for background classification, and facilitating dynamic queries by using directional associations as spatial parameters with support for the visibility of image areas that are incompletely enclosed by reference objects. They also demonstrated the efficiency of this technique using quantitative and qualitative results on contextual classification and retrieval of elevated spatial resolution satellite imagery.[7] All the system which are described here uses one level clustering or no clustering for similar image retrieval .

II PROPOSED APPROACH

Feature based Image retrieval is very vast area and it can be modelled as per the application. Here it is divided into following tasks:

1. Texture and Colour Feature generation from image stored in database
2. Clustering algorithm is applied on texture information
3. Classification is applied on texture cluster based on colour information
4. Similarity Search and output the result

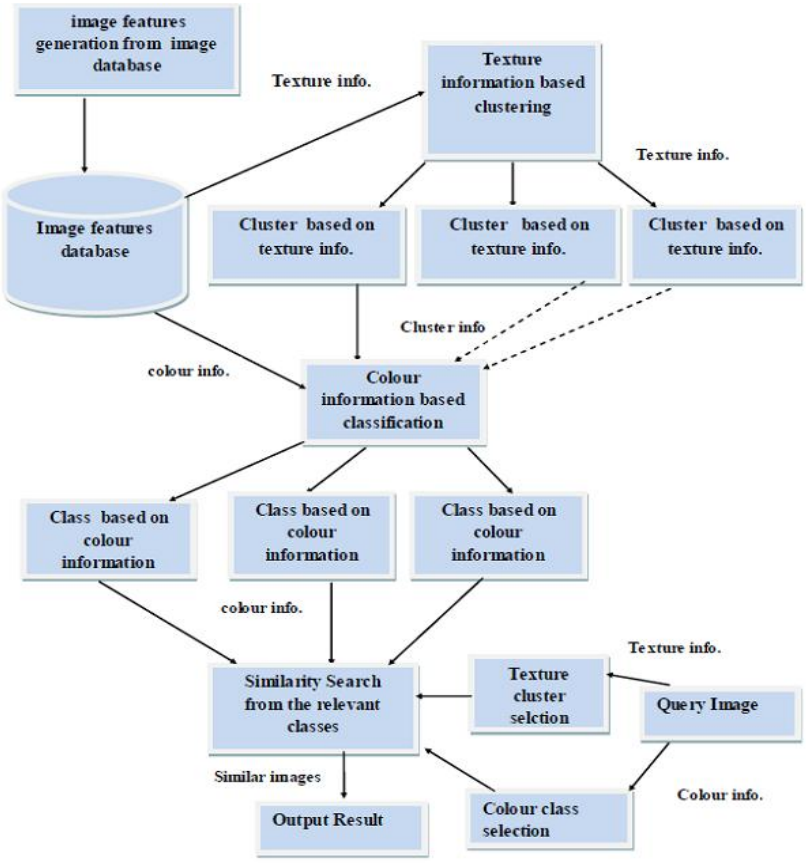

Figure 1 Proposed Model

2.1 Feature information generation

2.1.1 Texture feature generation

For texture feature generation Gabor filter banks is used. Gabor Filter were originally introduced by Dennis Gabor, it is used for edge detection. Scale and orientation representation of Gabor filter is similar to human visual system[1]. They have been used widely in image analysis due to their nature of scalability, orientation selectivity and frequency characteristic. A two-dimensional Gabor function consists of a sinusoidal plane wave of some frequency and orientation, modulated by a two dimensional translated Gaussian envelope.[8]

The 2D Gabor Filter is given by:

Frequency part

```
Gabor{s} = exp((-(log(radius/fo)).^2) / (2 *log(sigmaOf)^2));

Gabor{s} = Gabor{s}.*lp
```

Orientation part

```
ds = sintheta * cos(angl) - costheta * sin(angl);

dc = costheta * cos(angl) + sintheta * sin(angl);

dtheta = abs(atan2(ds,dc));
```


spread{o} = exp((-dtheta.^2) / (2 * thetaSigma^2));

Filter by multiplying frequency and orientation part.

filter = Gabor{s} .* spread{o};

Here s=frequency, o=orientation, lp=lowpassfilter, sigmaOf=Gaussian deviation, angl=orientation's angle, thetaSigma=ration of orientation and frequency part fo=central frequency. Here Gabor filter banks with frequency scale of 3,6,13,27 and orientation of 0,45,90,135,180,225,270 are used. So total 24 filter are applied and mean_applitude and mean_energy is calculated. But by doing this the resultant feature set will be of 24 floating points so it is very hard to apply clustering on that dimension dataset. So mean, variance and skewness is calculated for each feature set and used as texture feature set for image retrieval. So resultant feature set will be of 6 floating point.

2.1.2 Colour feature generation

For colour information extraction RGB colour model is used. Red, Green and blue are primary colours. An RGB colour images is an M*N*3 array of colour pixels, where each colour pixel is a triplet corresponding to the red, green, and blue components of an image at a spatial location[11]. An RGB image can be viewed as the stack of three gray scale images that, when fed into the red, green, blue inputs of a colour monitor, produce the colour image on the screen.

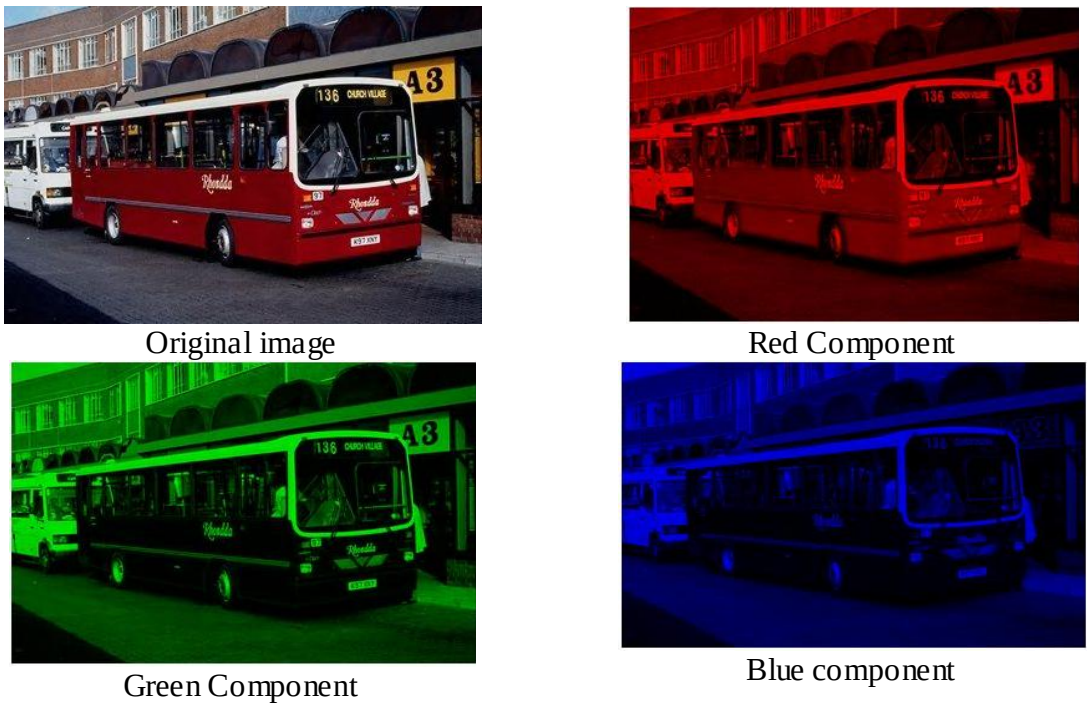

Figure 2. Colour image and its components

Mean and variance is calculated for each component of image that are in database and stored as a colour feature in feature dataset. RGB colour model used because classification is performed on the bases of mean value of R,G and B component.

2.2 Clustering on texture feature set

Centroid based K-mean clustering algorithm is used for clustering. In K mean clustering each cluster is represented by cluster mean[10][9]. So for similarity search we have to only check for mean of the each cluster.

Steps: 1: Texture information is extracted from the feature database

Step 2: Texture data set is applied to k mean clustering algorithm

Loop until maximum iteration or predefined threshold is reached

Assignment step: Assign each data point to the cluster whose mean yields the least within-cluster sum of squares

Updation step: Calculate the new means to be the centroids of the observations in the new clusters

Stecp 3: Centroids of each clusters are stored for searching purpose. [10]

2.3 Classification based on colour information

Here classification is done on mean value of R component, G component, and B Component. This value is normalized by dividing it by total and that normalized value is used as a membership value in four classes which are Red class, Green class, Blue class, Mix classes. Red class contains images which have more red colour then other two, Green class contains image with major green component, Blue class contain images with major blue component, Mix class contains image which contain all three component equally. Here if some image contain yellow part then that image will belong to more than 1 class, in this case yellow image would be in red and green class.

Step 1: colour information is extracted for each texture cluster and stored in a file

Step 2: colour class membership value is calculated from the colour information.

r_m=red membership value;

g_m= green membership value;

b_m= blue membership value;

th1 = 0.1;

th2 = 0.05;

Rule 1:- For selection of images which belongs to only one class

if((r_m-g_m)>th1 && (r_m-b_m)>th1)

image in red class

elseif ((g_m-r_m)>th1 && (g_m-b_m)>th1)

image in green class

elseif ((b_m-r_m)>th1 && (b_m-g_m)>th1)

image in blue class

else

Rule 2:- For selection of images which belongs to two class

color=[r_m,g_m,b_m]; color=sort(color);

if((color(1)==r_m && color(2)==g_m) || color(2)==r_m && color(1)==g_m && abs(r_m-g_m)>0.5)

belongs to red and green class

elseif (color(1)==r_m && color(2)==b_m || color(2)==r_m && color(1)==b_m && abs(r_m-b_m)>0.5)

belongs to blue and red class

elseif (color(1)==g_m && color(2)==b_m || color(2)==g_m && color(1)==b_m && abs(b_m-g_m)>0.5)

belongs to blue and green class

Rule 3: for images which belongs to all three classes

belongs to mix class;

The images which belongs to two classes are added to both classes. If image falls in blue and red class then that image will be added to both blue and red class. Thresholds are selected by experimenting with different images.

2.4 Similarity Search and output

Here Search by Example type of querying is used so image is provided.

Step1: First Image's texture and colour information is extracted

Step 2: From the stored cluster centroid select the cluster centroid which is nearest one.

Step 3: From the colour information select the appropriate colour class for that image and use the similarity search methods to find similar images.

Step 4: Display the top images which are specified by user.

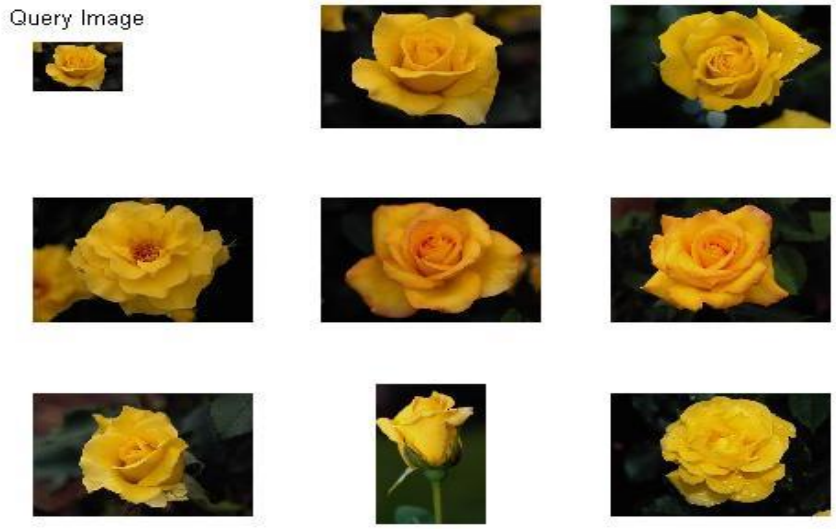

Figure 3 Result of proposed method based image retrieval

Here left image of the query image which is fetched from the database and other most similar images are shown from left to right in all row in first row query image, 1st similar image in second row 2nd image, 3rd similar image, 4th similar image Third row 5th similar image, 6th similar image, 7th similar image.

III. PERFORMANCE EVALUATION OF PROPOSED METHOD

In this paper, The WANG database is a subset of 1,000 images of the Corel stock photo database which have been manually selected and which form 10 classes of 100 images each[12]. is used for Clustering and Classification. The result is taken for different set of images and .On considering the results, the time complexity has been considerably reduced in proposed technique.

Table 1. Time Complexities of Image Retrieval with normal and proposed method

Total images	Proposed Method	Normal Method
200	5.5288	6.1095
500	6.1445	6.6747
600	6.6621	6.4755
700	6.7981	6.9421
800	6.8721	7.1111
900	7.2620	8.3017
1000	8.2991	9.7250

To calculate retrieval time output are taken 10 times and then average time is considered. To calculate retrieval time Matlab's in-built function "tic-toc" is used. From above table we can say that proposed method retrieves images in less time than normal method.

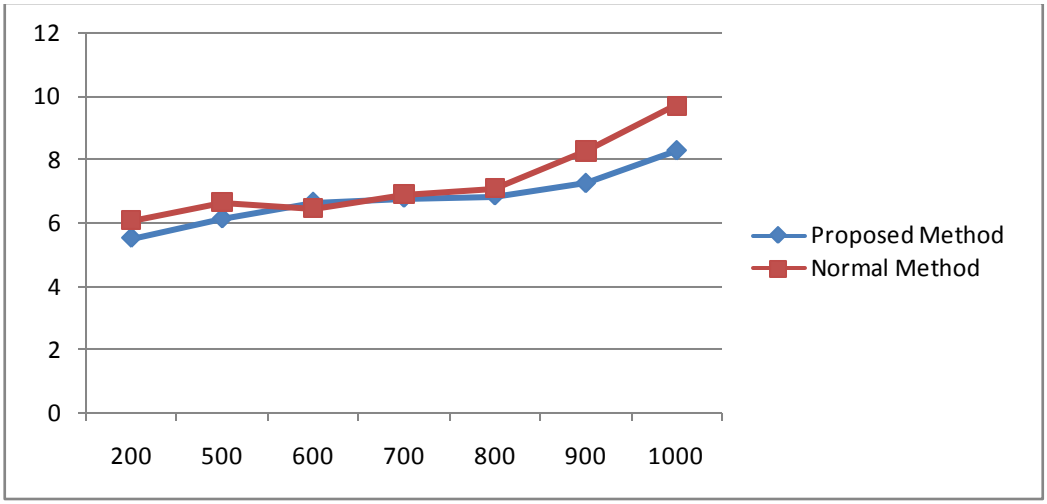

Figure 4 Comparison of both system

Table 2 Accuracy of both System

Total images	Proposed(%)	Normal(%)
200	74	74
500	76	76
600	70	75
700	80	82
800	83	79
900	84	83
1000	79	73

Proposed system provide accuracy of 78% and normal method provides accuracy of 77%. Proposed method is accurate as normal method.

IV.CONCLUSION

When feature dataset is divided into similar image classes using clustering and classification the similar image retrieving time decreases and it also improve the result, because similar images are placed together. Efficiency of System is increased by decreasing the retrieval time. Proposed method's accuracy is better than Normal method because it cluster and classify same images in one place. So that retrieval time decreases and accuracy increases. Proposed method is useful for all type colour image retrieval because it is not dependent on image dataset used.

REFERENCES

[1] R. Datta, D. Joshi, J. Li, and J.Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," ACM Comput. Surv., 40(2), Apr.2008,

[2] A.Kannan1 Dr.V.Mohan2 Dr.N.Anbazhagan3, "Image Retrieval Based on Clustering and Non-Clustering Techniques using Image Mining"-Int J Engg Techsci Vol 1(1) 2010,54-61

[3] Ela Yildizer, Ali Metin Balci, Tamer N. Jarada, Reda Alhajj ,"Integrating wavelets with clustering and indexing for effective content-based image retrieval", (2012)

[4] Zhang D, and Lu G , "Evaluation of Similarity measurement for image retrieval" , in proceedings of IEEE Neural network and signal processing, 2003, pp-928-938.

[5] Ali Mosleh1, Farzad Zargari2, Reza Azizi -"Texture Image Retrieval Using Gabor Transform" (2010)

[6] Markus Koskela, Jorma Laaksonen, and Erkki Oja -"Comparison Of Techniques For Content-based Image Retrieval" Conference on Image Analysis (SCIA 2009). Bergen, Norway, pages 579-586.

[7] Ravishankar , K.C.Prasad, B.G. Gupta, S.K. Biswas, "Dominant colour region based indexing for CBIR", International conference on image analysis and processing,2001

[8] Urszula Marmol - "Use Of Gabor Filters For Texture Classification Of Airborne Images And Lidar Data" - Archives of Photogrammetry, Cartography and Remote Sensing, Vol. 22, 2011, pp. 325-336 ISSN 2083-2214

[9] Ray-I Chang 1, Shu-Yu Lin 1,2, Jan-Ming Ho 2, Chi-Wen Fann 2, and Yu-Chun Wang 2-"A Novel Content Based Image Retrieval System using K-means/KNN with Feature Extraction" - DOI: 10.2298/CSIS120122047C

[10] Jiawei Han, Micheline kamber, Jain pei, "Datamininng concept and techniques" The morgan Kaufmann series in Data management system.

[11] R.C. Gonzalez, R.C. Woods, Digital Image Processing, Addison-Wesley,

[12] http://savvash.blogspot.in/2008/12/benchmark-databases-for-cbir.html

