International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 01 , January 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

RESEARCH AND PERFORMANCE EVALUATION OF PROPOSED CBIR SYSTEM OF CONTENT BASED IMAGE RETRIEVAL USING COLOR, TEXTURE AND IMAGE PROCESSING TECHNIQUES

Mr. Kommu Naveen[1], Dr. S. Sudarshan[2], Mr. Raghu Kumar Lingamallu[3]

[1]M.Tech, Ph.D Scholar Department of Electronics and Communication Engineering, Sathyabama University, Jeppiaar Nagar, Chennai, Tamilnadu (India)

[2]Department of Electronics and Communication Engineering Sathyabama University, Jeppiaar Nagar, Chennai, Tamilnadu (India)

[3]Ph.D Research Scholar Dept. of Computer Science Engineering Engineering, O.P.J.S University, Churu, Rajasthan,(India)

ABSTRACT

This paper has a look for the in the lands ruled over of image processing, image mining is move-forward in the field of facts mining. image mining is the extraction of put out of the way knowledge for computers, association of image data and added good example which are quite not clearly able to be seen in field that has to do with, image processing, data mining, Machine Learning, not natural quick brains and knowledge-base. The lucrative point of image Mining is that without any before information of the designs it can produce all the important designs. This is the writing for a make observations done on the mixed, of different sort's image mining and knowledge for computers mining expert ways of art and so on. Facts mining says something about to the getting from of knowledge /information from a very great knowledge-base which is stored in further number times another heterogeneous knowledge-bases. Knowledge/ information is making an exchange of note through straight to or roundabout way of doing. These techniques join neural network, coding into groups, connection and association. This writing gives a first paper on the application fields of facts mining which is full of changes into telecommunication, making, Fraud discovery, and marketing and education part. In this way of doing we use size, feeling of a material and chief colour factors of an image. Grey Level Co-occurrence matrix (GLCM) point is used to come to a decision about the feeling of a material of an image. Points such as feeling of a material and colour are normalized. The image acts to get back point will be very sharp using the feeling of a material and colour point of image gave with the form point. For similar types of image form and feeling of a material point, weighted Euclidean distance of color point is put to use for getting back points.

Index Terms: Data Mining, Feature Extraction, Image recovery, Clustering, database, ray Level Co-occurrence Matrix, centroid, Weighted Euclidean Distance.

International Journal of Advance Research in Science and Engineering
Vol. No.6, Issue No. 01 , January 2017
www.ijarse.com
IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

I. INTRODUCTION

BY this present scenario, image plays full of force part in put in order as low-level and high-level points. users every point of view of business such as business images, can question example images based on these features satellite images, medical images and so on. If we such as feeling of a material, colour, form, field, range and others. By observations these knowledge for computers, which can give knowledge of useful similarity comparison the Target image from the information to the man-like users. But, unhappily image repository is put in good order again. Meanwhile, the next there are certain difficulties to get the idea those knowledge for computers in an important phase today is put at point at which rays come together on coding into group's right way. Needing payment to not complete knowledge for computers, the expert ways of art and so on. Coding into groups algorithms can offer higher information gathered is not processed further for any organization of multidimensional facts for working well reasoned opinion. Acts to get back. Coding into groups algorithms let a nearest- person living near look for to be with small amount of money did. In another end, image acts to get back is the tightly growing and hard make observations area with in connection with to both still for this reason, the image mining is rapidly getting more and moving images. Many content based image attention among the persons making observations in the field of facts acts to get back (CBIR) system prototypes have been mining, information acts to get back and sound and view offered and few are used as business, trading systems. Knowledge-bases. spatial knowledge-bases is the one of the CBIR aims at looking for image knowledge-bases for special ideas of a quality common to a group which plays a chief part in sound and view images that are similar to a given question image. It also System. makes observations can clear substance semantically gives one's mind to an idea at undergoing growth new techniques that support purposeful information from image data are working well looking for and taking grass for food of greatly sized by numbers, electronic increasingly in request. Image libraries based on automatically formed from picturing points. It is a rapidly getting wider (greater) make observations

1.1 Comparison of image Mining with other area placed at the point going across of knowledge-bases, expert ways of art and so on information acts to get back, and computer uncommonly beautiful. Although CBIR is still not full grown, there has been abundance of image mining normally low price offers with the extraction of before work. If true, then some other is necessarily true knowledge, image data relation, or other

1.1 Comparison of Image Mining With Other Techniques

The makes observations in image mining can be put in order into 2 kinds. The image processing is one in which, it has to do with a lands ruled over special application where the chief place is in the process of getting from the most on the point image features into a right form and the image mining is one in which, it has to do with general application where the chief place is on the process of producing image designs that may be able to help in the getting rightly of the effect on one another between high-level man-like power being conscious of images and low-level points. So, the latter may be the best one to lead the getting better in the having no error of images got back from image knowledge-bases. image mining normally low price offers with the extraction of if true,

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

then some other is necessarily true knowledge, image data relation, or other patters not clearly, with detail stored from the low-level computer act or power of seeing and image processing expert ways of art and so on. i.e.) the chief place of image mining is the in the extraction of designs from a greatlysized getting together of images, the chief place of computer act or power of seeing and image processing techniques is in getting rightly or getting from special features from a single image Number in sign 1.1 shows the image mining process. The images from an image knowledge-base are first pre-processed to get better their quality. These images then undergo different great changes and point extraction to produce the important features from the images. With the produced points, mining can be does using facts mining techniques to discover important designs. The coming out designs are valued and took as having a certain cause to come to be the last knowledge, which can be sent in name for to applications. Current techniques in image acts to get back and order (2 of the chief tasks in image mining) get, come together at one point on content-based expert ways of art and so on. Different systems like the QBIC, Retrieval Ware and Photo Book and soon have a range of points, but are still used in one fields (of knowledge). Jain et Al use colour features has at need with form for order. Ma et Al use colour and feeling of a material for acts to get back. Smith and Chang use colour and the spatial arrangements of these colour fields, ranges. Since power of being conscious of is subjective, there is no single point which is enough and, in addition, a single Representation of a point is also not enough. For this reason number times another Representations and a mix of features are necessary

II. PROBLEM DEFINITION

In the colour based image acts to get back the RGB colour design to be copied is used. Colour images normally are in three regular sizes. RGB colour parts are taken from each and every image. Then the mean value of R, g, and b values for both question image and Target images are worked out. These three mean values for each image are stored and thought out as points. By using these stored features the Target image from the repository is got back with respect to the question image. Then the top position on scale images are re-grouped according to their feeling of a material points. In the feeling of a material- based move near the parameters gathered are on the base of statistical move near. Statistical features of grey levels were one of the good at producing an effect methods to put in order feeling of a material. The grey Level Co-taking place matrix (GLCM) is used to get out second order statistics from an image. GLCMs have been used very successfully for feeling of a material answers by mathematics. The different feeling of a material parameters like entropy, contrast, unlikeness, being made up of parts of the same sort, standard amount gone away from straight, mean, and authority to change of both question image and Target images are worked out. From the worked out values the needed image from the repository is got from. Then, the pre-processed images in the knowledge-base are put in order as low-texture, average-texture and high-texture detailed images separately on the base of some cause like MLE (greatest point chance rough statement) rough statement. The put in order images are then subject to color point extraction. The got back outcome is pre-clustered by Fuzzy-C means expert way of art and so on. This is moved after by GLCM feeling of a material parameter extraction where the feeling of a material factors like comparison, connection, mean, authority to change and quality example authority to change are mined. The resulted values of both the question image and Target images are made a comparison by Euclidean distance careful way.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

2.1 Proposed Solution

In this, a new method for image classification is formulated in order to reduce the searching time of images from the image database. The coarse content of image is grouped under three Categories as

(i) High-texture detailed Image

(ii) Average-texture detailed Image

(iii) Low-texture detailed Image

In that way, we can get changed to other form the look for space by one third of what was earlier. If we go more number of groups or less number of groups, they may give knowledge of unnecessary partly covering overhead problems or may produce rough outcomes. So, the main chief place on this order is by making use of textures present in an image. This is because this texture-based order is simple, simple, and not hard and good at producing an effect for true time applications as made a comparison to system of ordering based on entropy careful way as well as segmentation based expert ways of art and so on. The first end of this work is to undergo growth algorithms in order to make come into existence true knowledge-bases from collections of sound and view facts (specially images) by mining What is in from the knowledge for computers off- line in order to with small amount of money support complex questions at run-time.

2.2 Image Retrieval

Image Retrieval from the image collections involved with the following steps

1. Pre-processing
2. Image Classification based on some true factor
3. RGB processing
4. Preclustering
5. Texture feature extraction
6. Similarity comparison

a. Target image selection

2.3 Preprocessing

Pre-processing is the name used for operations on images at the lowest level of idea, not fact. The purpose of the pre-processing is a getting better of the image that keeps secret unwilling twisting or gives greater value to some image points, which is important for future processing of the images.

2.4 RGB Components Processing

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. 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. By convention the three images form an RGB images are called as red, green and blue components.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

The average values for the RGB components are calculated for all images after calculating the mean values of Red, Blue and Green components, the values are to be compared with each other in order to find the maximum value of the components. For eg., if the value of Red component is High than the rest of the two, then we can conclude that the respective image is Red Intensity oriented image and which can be clustered into Red Group of Images. Whenever the query image is given calculate the RGB components average values. Then compare this with the stored values.

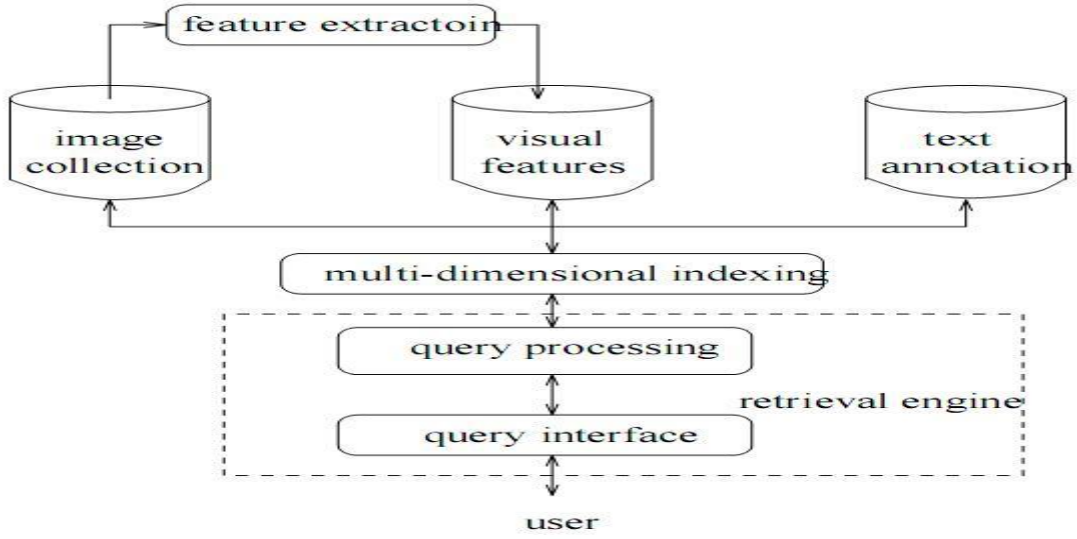

Fig: Work Flow Diagram

This step gives one's mind to an idea on image point processing.

2.5 Image Texture Classification

Texture Classification is the problem of distringuishing between textures, a classic problem in pattern recognition. Since many very sophisticated classifiers exist, the key challenge here is the development of effective features to extract from a given textured image. Many approaches have been defined to extract features, such as Gabor filters or wavelets, however a great deal of recent work has focused on patch-based methods, whereby a texture is classified strictly based on a set of small patches of pixels extracted from a given textured image.Recent work by Liu proposed using random linear functions of patches. The number of such random features needed turns out to be relatively modest, therefore it is suddenly feasible to do texture classification using large image patches, then with some number of random features. Astronishingly, such an approach outperforms the texture classification of finely-designed, state-of-the-art texture filter

Texture classification assigns a given texture to some texture classes Two main classification methods are supervised and unsupervised classification. Supervised classification is provided examples of each texture class as a training set. A supervised classifier is trained using the set to learn a characterisation for each texture class. Unsupervised classification does not require prior knowledge, which is able to automatically discover different classes from input textures. Another class is semi-supervised with only partial prior knowledge being

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

available.The majority of classification methods involve a two-stage process. The first stage is feature extraction, which yields a characterisation of each texture class in terms of feature measures. It is important to identify and select distinguishing features that are invariant to irrelevant transformation of the image, such as translation, rotation, and scaling. Ideally, the quantitative measures of selected features should be very close for similar textures. However, it is a difficult problem to design a universally applicable feature extractor, and most present ones are problem dependent and require more or less domain knowledge.

The second stage is classification, in which classifiers are trained to determine the classification for each input texture based on obtained measures of selected features. In this case, a classifier is a function which takes the selected features as inputs and texture classes as outputs.

In the case of supervised classification, a knn nearest neighbour knn classifier is usually applied which determines the classification of a texture by computing distances to the $ k$ nearest training cases. The distances are computed in a multi-dimensional feature space constructed by selected texture features. Euclidean distance, Chi-square distance, and Kullback-Leibler distance are mostly used as distance metrics for distributions and thus similarity metrics for textures. A Bayesian classifier that performs classification via probabilistic inference is also frequently used. A general two-class histogram calculation as well as Bayesian classifier can be specified by the following Bayes formula [27],

$$Pr(w_j|\mathbf{F}) = \frac{Pr(\mathbf{F}|w_j)Pr(w_j)}{Pr(\mathbf{F})}$$

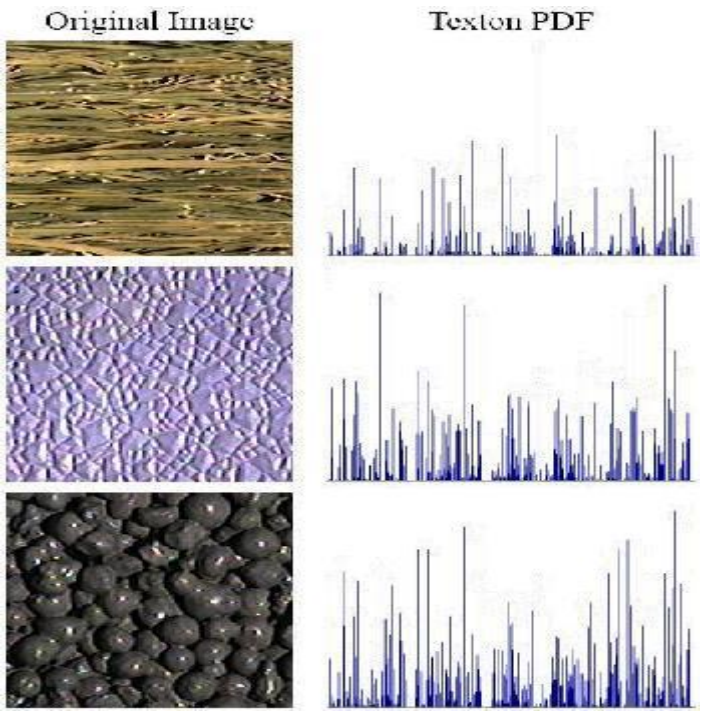

Fig: Histogram Feature in a Image

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

2.8 Image Clustering

Feature selection: Since the inclusion of superfluous features during classifier training can lead to degradation in prediction accuracy [9], we wished to select a subset of the 116 initial features that were most representative of our data for use in cluster analysis. We screened the entire feature set using a backward elimination process based on the Lagrangian Support Vector

Machine classifier [10,11], which was used because it generalizes well. The process starts from the full feature set. In each iteration, one feature is eliminated from the remaining set by evaluating all the possible subsets (n subsets, each containing n-1 features, need to be evaluated for an n-element feature set) and selecting the subset that achieves the smallest training error rate as our next feature set. We use a low training error as an approximation of the importance of that feature. All the features can thus be ranked according to when they are eliminated in the backward elimination process. We repeat this process for all 8 mutant types in a pairwise fashion and generate 28 sequences of ranked features. We proceed with the features that appear at least five times among the top 10 features of each of 28 sequences and use these 18 features as our subset to represent the feature data. Parallel analyses were conducted using three scaling methods. We also compared the classification error of the first few principal components with feature subsets using these scaling methods. We concluded that the data were well represented using a subset of 18 features, with less than 4% cross-validation error rate. These features included several measurements of speed averaged over different time periods, as well as the features described in

Section 2

K-means clustering and stopping rule: To further investigate the clustering of the data points, we applied the k-means clustering algorithm to find the natural clusters in the behavioral data. For this analysis, each data point was treated individually without regard to mutant type. We generated sufficiently many (10,000) random initializations for each k and tracked the error at the convergence to be reasonably confident that the global minimum was found. Figure 4 shows the cluster centers identified by the k-means algorithm; for each case, thecenters are marked by black squares. Although the actual k-means clustering was done using all 18 selected features, the data were visualized by showing the first two principal components. Table 1 shows the Euclidean distance between prototype centers (cluster centers).

2.9 Similarity Comparison

The retrieval process starts with feature extraction for a query image. The features for target images (images in the database) are usually precomputed and stored as feature files. Using these features together with an image similarity measure, the resemblance between the query image and target images are evaluated and sorted. Similarity measure quantifies the resemblance in contents between a pair of images. Depending on the type of features, the formulation of the similarity measure varies greatly. The Mahalanobis distance and intersection distance are commonly used to compute the difference between two

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

histograms with the same number of bins. When the number of bins is different, the Earthmover's distance (EMD) is applied. Here the Euclidean distance is used for similarity comparison.

III. PERFORMANCE EVALUATION OF PROPOSED CBIR SYSTEM

Evaluation of retrieval performance is a crucial problem in Content-Based Image Retrieval (CBIR). Many different methods for measuring the performance of a system have been created and used by researchers. We have used the most common evaluation methods namely, Precision and Recall usually presented as a Precision vs Recall graph. Precision and recall alone contain Insufficient information. We can always make recall value 1 just by retrieving all images. In a Similar way precision value can be kept in a higher value by retrieving only few images or Precision and recall should either be used together or the number of images retrieved should be specified.

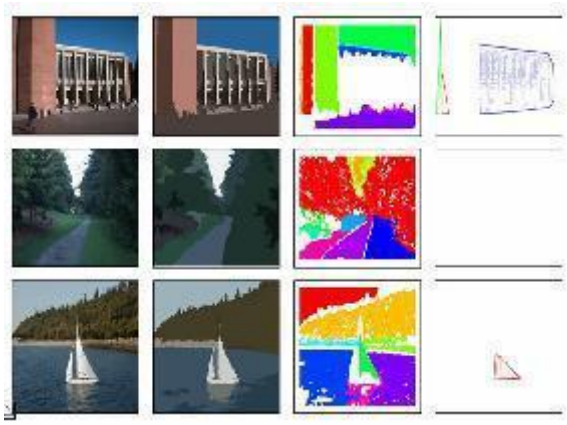

Fig: Feature Based Image Retrieval System

With this, the following formulae are used for finding Precision and Recall values.

Retrieve dimage srelevant of No. ecisionPr =(retrieve dimagesrelevant of No/ retrieve dimages of no Total)

IV. CONCLUSION

The main objective of the image mining is to remove the data loss and extracting the meaningful potential information to the human expected needs. There are several Content Based Image Retrieval Systems existing in this present scenario. However, this particular system will be applied in Medical transcription in an effective manner not only based on the contents of the image but based on the given query image too for comparing certain frequent diseases affected earlier in human bodies. In this system, a new image retrieval technique based on clusters is also introduced in order to reduce the searching time space. Moreover, the RGB components o the colour images are classified in different dimension in order to create Red, Blue and Green image clusters.

IJARSE
ISSN (O) 2319 - 8354
ISSN (P) 2319 - 8346

REFERENCES

[1] Image Mining: Trends and Developments, Ji Zhang Wynne Hsu Mong Li Lee

[2] U. M. Fayyad, S. G. Djorgovski, and N. Weir: Automating the Analysis and ataloging of Sky Surveys. Advances in Knowledge Discovery and Data Mining, 471-493, 1996.

[3] W. Hsu, M. L. Lee and K. G. Goh. Image Mining in IRIS: Integrated Retinal Information System (Demo), in Proceedings of ACM SIGMOD International Conference on theManagement of Data, Dallas, Texas, May 2000

[4] Kitamoto. Data Mining for Typhoon Image Collection. In Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), San Francisco, CA,USA, August, 2001.

[5] C. Ordonez and E. Omiecinski. Discovering Association Rules Based on Image Content. Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99), 1999.

[6] P. Stanchev, A. Smeulders and F. Groen, An Approach to Image Indexing of Documents,In E. Knuth and L. Wegner (Eds.), Visual Database Systems II, North Holland, 1992, pp.63-77.

[7] P. Stanchev, Medical Knowledge Acquisition for Expert Consultation System, in M.Chytel and R. Engelbrecht (eds.), Medical Expert Systems, Sigma press, London 1988, pp.153-163.

[8] R. Agrawal, C. Faloutsos and A. Swami, Efficient Similarity Search in Sequence Databases, FODO Conferense , Evanston, Illinois, 1993.

[9] P. Stanchev and V. Vutov, A Model-Based Technique with a New Indexing Mechanism for Industrial Object Recognition, IEEE/CHMT, IEMT, 1990, pp. 56-60.