A Performance Perspective: Content Based Image Retrieval System

Senthil Kumar Sundararajan, B ShankaraGomati, D Saravana Priya

ABSTRACT: ***The content-based image retrieval (CBIR) plays a significant aspect in digital image processing. The CBIR is novel and a high-speed process in information recovery. In order to sequentially access the huge set of database the oldest incisive search engines like google, yahoo, bingo was based on textual explanation of images.Hence, the performance of this system was not adequate. Therefore, a new process is required, which could be user friendly in order to access the image data. This paper used different set of various attributes like texture, shape and colorderived from the image extracted from query and the training images for comparing and retrieving the image. Thus, it repeatedly produces descriptions directly from the media data by using digital image processing methods. In this paper we analysis some of the procedural aspects of CBIR method and even specify its advantages and disadvantages.***

Keywords: Image Retrieval, CBIR, Feature Extraction, Classification, Image Mining.

I. INTRODUCTION

A. Background Study

The digital images are uploaded daily in the contemporary period due to the express growth of internet and private networks. The huge collection of images has led to increased challenges of computer systems, further to accumulate and handle data successfully. The consumption of outsized images database for a diversity of applications is possible. Also the databases from satellite and medical imagery are attracted many users in different professional fields [1].It is essential now to efficiently obtain preferred images from huge and different image databases.Thus, the users are not contented with the customary recovery methods based on text. (CBIR) Content Based Image Retrieval is aapproach, which easily access and display the digital images from a huge set of database with the help of image facial appearance. The aim of CBIR is to avoid the use of textual descriptions. CBIR is a kind of method that permits the user to easily access the data. Therefore, in CBIR retrieving of images are established on their contents such as textures, colors, shapes etc. are lesser level facial appearance of images. The term “Content-Based” refers the real information/data or content of an image not the Meta data associated to the query image. Presently the word content specifies texture and color information that could be derived from the various imagesby itself. The CBIR is vital because most of the search engines operating on web is purely depends on the metadata and which leads to lot of malicious results[2,3].

CBIR METHODS

CBIR approachis one of the functions exists in the vision methods related to image retrieval. It is used to identify a specific image required by the user based on the contents fed as input. In general, most of the CBIRis different to traditional Concept-Based Approaches. CBIR is advantageous because exploration depends only on metadata and they are reliant on explanation value and comprehensiveness. Images are interpreted using keywords or meta-data in databases. Most of the time, the images cannot be interpreted and expressed using exact keywords; hence the CBIR is not been well defined. So, the authors in [4] proposed two different kinds of IR systems (Image Retrieval)like text using and content using. In text-using CBIR, the images are labelled or annotated using texts and stored in the database. The demerits of this text-using CBIR is, labelling or annotations is obtained is done by manual. Then, observing an image using text makes factual error. To surmount these, dis- advantagesin text-based retrieval system,to overcome the loss of Content-Based Image Retrieval System(CBIR) was developed and introduced.Author in [5] used features representing image appearance like color, texture and shape in creating CBIR. These features are used to extract the image from a large size database by mapping features of the images extracted from the query with the data base images. Images are regained from the sources by resemblance in features, in order query conditions are evaluated with relevance from the image, which are stored in database to decide which image that matches correspondingly with known features. The CBIR model[6] contains three primaryprocess, a) retrieval system design multidimensional indexing and c) visual features extraction The visual features (content) mining is the basis of CBIR, the features comprise of features based on text (annotation, keyword) and the features based on the visual attributes (faces, color, shape, text). These features are used as additional, unique and domain explicit information for representing the images. Earlier days CBIR includes color, texture and shape, further applications used pattern recognition and image comparison. Figure-1 below specify the image retrieval system architecture.

Revised Manuscript Received on March 25, 2019.

Senthil Kumar Sundararajan Research Scholar, Department of Computer Science, Bharathiar University, Coimbatore,India

Dr B ShankaraGomati Department of Electronics and Instrumentation, National Engineering College, Kovilpatti, India

Dr D Saravana Priya, Dept of Computer Science and Engineering, PA College of Engineering and Technology, Pollachi, India

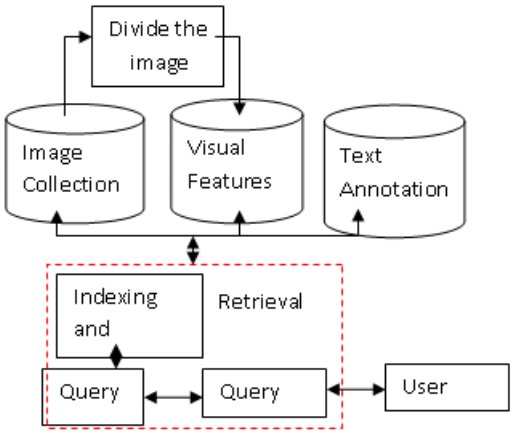

Figure-1: CBIR Architecture

The scientist Kato established the (CBIR) technique in the year 1992. He derived this method when he tested to fetch the images from a tiny database, by means of their visual content.The algorithms, tools, techniques used by CBIR were investigated from machine vision [7], linear systems, signal processing, and pattern recognition.Databases are filled with descriptors, which are originate from theimages having the visual content. Mainly CBIR structure are fretful with the near queries whose goal is to locate images, which are visually related to a particular targeted image. In many cases, the goal of CBIR system is to duplicate individual observation of image[8].There are six phases available in the CBIR like image preprocessing, image acquisition, feature extraction, resultant retrieval, similarity matching.

B. IMAGE ACQUISITION

It is the first phaseof the process to extract a digital image from the image, which are store in the database. The database consists oflargefigure of images that rely onon the user choice and options.

C. IMAGE PROCESSING

The process, which refine the image in its contents and show it in an efficient way, It engages extraction, analysis and recognition of image coding,normalization, segmentation filtering and object identification. Image can be separatedin to several parts by using Image Segmentation methods. The production of this phase is a collection of major regions and objects[8].

D. SIMILARITY MATCHING

The method requires information regarding all image used for calculation which are stored in feature vectors. Thesimilarity are measured in the feature vectors, which are coordinated with images extracted through query, The similarity measure method is also named as the distance method. There are various distance processes like Canberra Distance, City Block Distance, Euclidean Distance, [9].The output would be parallel images containing typical features as like query image [10].

E. SHAPES

In the image retrieval system, shapes refer to exact region that is required. It is achieved, firstly by pertaining segmentation or edge detection to an image. The shape depiction could be classified into two types:
Border Line – which utilize simply the outer border of the shape.
Area – which utilize the complete shape area.

The victorious representative of these two classifieds isby changes in the Fourier and moments. Fourier changes represent the edge and shape-based features of the image. Rui et al. suggested a customized Fourier Descriptor that is Invariant to geometric transformation [11].

F. SEGMENTATION

One of the image processing methods used for separating the image objects from the entire image is image-segmentation. Segmentation process divide the image into several segments [12]. An image segment is aaccumulation of pixels representing an object in the current image. The aim of separatingan image is used to scale down the splitting the images into shorter images or objects where helps to explore the image. It also helps to locate the objects and obtain the boundaries of the objects are present in the image. It also helps to label the pixels in the image, where it leads to obtain the similar label and visual characteristic based objects in the image. The outcome of the image separate process is a collection of image segments where it can be grouped into a complete image or collection of regions wielded from the image. Every pixel in an area is same by means of certain characteristics or calculated property such as intensity, texture, color. Neighboring areas are considerably different by similar characteristics [12].

G. CONTENT LEVELS

Many researchers had agreed thatdifferent levels of content are exists. Ex: color, Luminance are observed as a lowlevel content in addition to substantial objects (Like persons or automobile) are observed as a highlevel content. (Patterns and Texture) which merge several kinds of contents could be observed as mid-level content. Though there is no larger union regarding how many levels of content could be alleged by a human, how many kinds of content are there in every level or how the content of an exact image could be categorized into levels [13]. And types

H. APPLICATIONS OF CBIR

There are several prospective uses for CBIR. They are,

I. Examination: face recognition system, patent on the internet.
II. Shape Recognition: Recognition of imperfection and blunder in industrial automation.
III. Mechanical Analysis: Tumors recognition,advanced MRI and CT scan.

IV. Journalism,Graphic design,Media Advertisement.
V. Remote Sensing: Several Information Systems,Weather Forecast and Satellite images.
VI. Trademark Databases, archaeology,museums and art galleries.
VII. Engineering designs
VIII. Cartography:Map creations from photography,creations of weather maps.
IX. Digital Forensics:Identification of finger prints for crime recognition.
X. Radar Engineering: Assist in recognition and detection of targets.

II. EXISTING APPROACHES

In initial period in 1990's, CBIR System has turned out to be a dynamic study and is followed by the different option mentioned below.

i) Arbitrary Browsing
ii) Explore by Example
iii) Explore by Sketch
iv) Explore by Text(including keyword or speech)
v) Routing by modified image categories.

Nowadays there is a stipulation for affluent collection of search options, although in sensible applications that engage real users still requires organized study to discover the trade-off between various options specified above. Now we would choose few envoy systems and emphasize their different characteristics[14].

III. REAL WORLD REQUIREMENTS

Constructing a real world system entails standard user feedback throughout the development process.For communal usage, very few image retrieval systems are organized.Still there are huge number of proposition in the real world execution. CBIR has been useful in various fields like astronomy, mineralogy, botany, and remote sensing [15-18]. Screen shots could be obtained in figure-2 and figure-3 correspondingly based on theknowledge, by executing real-world data in CBIR system, which are used in public extensively. Few concerns that are raised to be dangerous for real-world deployment are given in Table-1.
The machine algorithm predicts about the semantic concept of the query image. Thus, in its place of ruling the similarities among the query image including the images persisted in the database, it is established among the query image and merely the images that fit into the query image group. Various existing CBIR system [14] are stated below,
QBIC:

IV. IMAGE CONTENT RETRIEVAL BY QUERY

It is the primary profitableCBIR .QBIC method allows the users to provide their queries in a specific manner in accordance to the visual appearance like texture, shape and color.

i) Visual Seek and Web seek are visual element search engine. Web seek is a World Wide Web learning text/image search engine, were these two are urbanized at Columbia University.
ii) In Virage Inc, Virage was developed by content-based image search engine. It maintains color, spatial site matching and also texture matching.
iii) Netra- this system utilizes color, shape, spatial layout, texture matching and even segmentation.
iv) Multimedia analysis and retrieval systems or MARS, this system formulates color, spatial layout, texture and shape matching.
v) Visual Information processing for enhanced retrieval or VIPER-this system recovers the images based on color and texture matching.
vi) The image (Anaktisi) is a web based CBIR system that provides color and texture facial appearances.

There are three different class-fields are used for feature extraction. They are histogram value based, region-based and color-based searching. After retrieval, all the entities are grouped together using a weighted sum matric [17] also called as merging schemes [18].

Table-1: Issues Found in Real World Problems

1	Performance	The most serious issue is the excellence of retrieval and how it is related to the domain-specific user community. Mainly the present effort is determined by improving the presentation condition with their accuracy.
2	Volume of Data	Communal image database has a tendency to produce at huge proportions. The software system should be able to professionally handle indexing and retrieval at such level.
3	Heterogeneity	If the images are created from different sources then the constraints such as quality, resolution and color depth are likely to differ. This in turn causes a difference in color and texture extorted.
4	Concurrent Usage	The on-line image retrieval system has various synchronized users. Where, the majority of system has a high resource supplies for feature extraction indexing. They should be well defined so that they do not tire out the host server resources.
5	Multi-Modal Features	The existence of reliable meta-data like audio or text captions that are related to the image could easily know the enhanced image content.
6	User-Interface	A better attempt is required to design sensitive interface for image retrieval so that people could operate the tool easily.
7	Operating Speed	In on-line system, time is vital because the reply time desires to slow down for excellent interactivity. Execution must be perfectly carried out using proficient algorithms particularly for a larger database.
8	System Evaluation	The image retrieval systems are also essential to be assessed in order to analysis the possibility of empowering in a novel version. The design of a CBIR standard needs cautious design perhaps to detain the inbuilt bias in image retrieval. Such proposals are obtained in [16].

V. GLOBAL FEATURE BASED CBIR SYSTEM

The real time CBIR systems extort the features from the whole image and not from the specific region and the features are termed as Global features.

The Algorithmuses Histogram Search methods [19] will disseminate an image by its color dissemination. There are several range used to describe the resemblance of two-color histogram image, where Euclidean distance is most frequently used.The demerit of the histogram value of the image is, the histogram value is redundant. Also, the color histogram value is response to represent color, intensity variation and color cropping with color bend values.

Image are segmented into various blocks and the normal color of every block is saved in simple color layout indexing[19].Thereforethe color layout isthe low-resolution depiction of the usual image.Images are depicted by a collection of local properties[18].Mostly image retrieval by color feature frequently provides inadequate results, as images with parallel colors do not contain analogous content.D.Zhang [20] projected a technique combining color, texture features to get improved retrieval presentation. At the time of retrieval procedure, a query image will give the images present in database are initially positioned using color and feature. In the next step,the total number of top-level images are preferred and re-leveled based on their texture feature. The user uses any of these two options such as retrieval based color feature and retrieval based combined feature.

REGION BASED CBIR SYSTEMS

Various images are represented at the object-level because this Region Based CBIR Systems try to defeat the insufficiency of universal characteristic based search, The image decompose can be done by segmenting images in the sector based image retrieval system., it will also communicate to object, incase of the decomposition is idyllic [21]. Theregion-based retrieval system comprise of the NatraSystem [22] and the Blob-world System [23].

The Natra and the Blob-world systems evaluate images based on the entity regions.The users semantic perceptive of an image is at upper level than compared to region depiction. Natsev et al. reflects the resemblance model WALLRUS[24], this is a healthy model for scaling and conversion of objects inside the image.Every images are divided into regions.Thecoincidence determined between two images is stated as, the division of the range of two images enclosed by identical regions. Nevertheless, WALRUS spot lights the progress of efficient segmentation technique, as an alternative of an image-to-image similarity compute.The author suggested a gluttonous heuristic approach for evaluating the similar region duo collection by the highest area. In[25] the mean shift algorithm is utilized for segmenting the images and significantly, regions are filed using cluster-based tree to augment the competence of retrieval procedure.Hence,this structure utilizes merely color as image signature that is receptive to cropping, rotation,shifting and scaling.For color, the histogram of image regions are calculated for texture incidence matrix based entropy energy is computed and for edge solidity, (EHD) Edge Histogram Descriptor is discovered.

Li and Wang et all[26], projected the [IRM] Integrated Region Matching algorithm, which permits similar sector of one image to various regions of different image to evaluate the similarity among images. As output, the similarity between two images is denoted as the mass sum of distance in the attribute space, amid every regions from various images.Blobworld and IRM approach reduces the force of imprecise segments by leveling the vagueness in distance. Fuzzy club [27], deals with the problem of valuable and resourceful CBIR by representing an indexing and retrieval system. Fuzzy club highlights on civilizing color feature “imprecision” difficulty in the region based literature. It first partitions the image into regions of 4×4 blocks and removes texture features and color on every block. The K algorithm is utilized in order to group similar pixels as one, in turn to obtain a region.

VI. FEATURE EXTRACTION

It is a form of substantial semantic information from the images and it is utilized as a name for the image.Feature extraction in an image database is usually accomplished off-line inorder calculation difficulty is not a major issue.It contains three features like texture, color, shape.

VII. COLOR

Color is the important feature, which could be visually renowned by individual in images. In CBIR system, color feature is most extensively utilized andpremeditated in literature.It is a dominant descriptor reduces object, which areidentified,The principle of a color space is to assist the requirement of colors. Every color represented in the space is a sole point that are expressed in an organized system. Various color space like HSV, RGB,CIEL*a*b and CIELLL*u*v, has been urbanized for several purposes[28]. A suitable color system is needed for to guarantee perceptual equality.

Hence, RGB is not appropriate for CBIR, since it is aeffective unreliable, device reliant system [29]. Various color descriptors has been explored from several presentation methods, like color edge [32], color texture [33] ,colorcorrelograms [34], color moments [31], color histograms [30],

VIII. COLOR HISTOGRAMS

In an particular image, Histograms are widely used for differentiate the color feature for an image. In the bar graph ,The height indicates the sum of all color in the color space which are used in every image[28],These bars are titled as bins and represents the x-axis and number of pixels in every bin represents y-axis. In color histogram, quantization is a method, where quantity of bins is minimized by considering the colors, which are similar to one another and inserting them in the equivalent bin. Quantization simplifies the information that states the content of images [35]. They are differentiated into two kinds i) Global Color Histogram (GCH) and ii) Local Color Histogram(HCG).GCH is the usual process for retrieving color based images[36]. LCH holds many information regarding images and it is little costlier. In figure6, an example of the color histogram in HSV is shown in detail

Figure-2: Sample set of Images in Database

IX. TEXTURE

Texture is denoted as repetitive model of pixels above a spatial domain. Its characteristic are the visual models in an image that has materials of homogeneity, which does not affect the occurrences of intensity. The various texture properties are alleged by individual for example, reliability and directionality shown in figure (a) to (c). The several brightness intensities, that increase to merge, various individual perception of texture are shown in figure(a) and (c).In image processing, Image texture has computer vision and functional applications.They comprise of identification of image regions utilizing texture properties referred as texture categorization, the identification of texture boundaries utilizing texture properties referred as texture synthesis, texture segmentation. In accordance to Manjunath ND MA [37] the frequently used process for texture description are arithmetical, model-based and transfer-based methods.The texture feature are categorized into,

X. STATISTICAL METHODS

This method examines the spatial distribution of grey value by evaluating limited characteristic at each point in the image and determine a collection of statistics from the allocation of the limited feature. The most frequently used statically method is the Grey-Level-Co-Occurrence Matrix(GLCM) [38]. This is classified as matrix include two dimension. With the join probability among duo of pixels divided by a distance d, in a known direction r.This method is useful for applications that recognize fabric imperfection and in rock texture categorization and retrieval [33].

XI. MODEL BASED APPROACH

This method seeks to confine the process that produces the texture. An evaluation algorithm is utilized to position the criteria of the model to capitulate the robustness [39]. To denote a random field, consider that, the image is modeled as a function f(r,ω) where r is a position vector which denotes the pixel area in the 2-D space and ω is a random parameter. Function f(r,ω) is known as a random field.When a exact texture ω is chosen, f(r,ω) is an image that functions over 2-D grid indexed by r.At present three main model based methods are exists like :Markover Random Fields by Dubes and Jain [40], Fractals by Pentland[41] and the multi-declaration auto regressive features are pioneered by Mao and Jain[42].

XII. TRANSFORM DOMAIN FEATURE

The term transform denotes the arithmetical illustration of an image. There are different texture categorization adopting to change the domain features such as Transform, Discrete Wavelet Discrete Fourier Transform, and Gabor Wavelet. Manjunath and Ma [37] have discovered image retrieval utilizing the Gabor features.

XIII. SHAPE

Shape is one of the best frequently used characteristic in CBIR system.Shape of an object is the feature shell pattern as signified by the sketch. And one of the form is shape recognition from which the individual perceptions of the surroundings are implemented. Shape feature depictions are classified based on the process used. They are region-based and boundary-based [43]. Regular region-based process applies moment descriptor to specify shape [44/51]. Hu [45] stipulate the seven invariant moments, resulting from the second and third classified regularized inner moments. Compound Boundary-Based Descriptors compriseof Grid Descriptors, Fourier Descriptors and ChainNodes [36].

XIV. SHAPE FEATURE

Based on the application, shape information could be 2D or 3D in character. At present three different shape descriptors like ContourShape, Region Shape, and 2D, 3D Shape Descriptors.

XV. REGION SHAPE

The object shape could be a sole region or a collection of regions or either some crack found in object. Based on the moment invariants[46], The Region Shape Descriptor, utilizes all the pixels that constitute the shape within a frame; it could be illustrated in any shape.The Region Shape Descriptor benefitis its capacity to depict various shapes competently.

The feature extraction and match processing are very simple. Since it is having the calculationcomplexities for the tiny order,they are apt for shaping trail for sequences[47] appeared in video.

XVI. CONTOUR SHAPE

It captures, the feature of a shape placed on its outline. It depends on the (CSS) Curvature Scale Space [48], depiction catchesthe shape having significant features. The descriptor importantly characterizes the points of high bend, beside the outline. It is also forceful to non-rigid motion [46, 47].

XVII. SIMILARITY MEASURE

The similarities between two images are classified by a similarity calculation. The metrics for similarity collection has a hysterically impact on the presentation of CBIR,Euclidean distances are most frequently applied to evaluate the rangebetween two points in multi-dimensional space [49].In a multi-dimensional space [49] , Euclidean distances are most frequently applied to evaluate the interval between two points Various kinds of attributes like color,histogram,and Euclidean distance cannot be a perfect similarity metric. Swan and Billard to explore objects within images by color histograms projected histogram intersection.

XVIII. INDEXING STRUCTURES

While operating a huge database a high quality indexing is required.Processing each solitary item in a database is excessively incompetent and sluggish.During image search, this method is much more complicated. A well-known multi-dimensional indexing process comprise of the R*tree algorithms [23] and R-tree. The SOM (Self Organizing Map) is also one of the indexing structures. The SOM is accomplished to proportionate In theactual space the pattern of the dataOnce the query appears, the nearest SOM node is calculated, it is also known as the BMU (Best Matching Unit) and it is used for getting the data in image feature vector. When the query is completed, We are able to match the images present in the database, which are close to the images extracted from the query set.

XIX. SIMILARITY MEASUREMENT

There are many similarities finding methods are available in data mining. In this paper finally the similarity measurement between the patches of both the images are given by

$$SM = \sqrt{\sum |Means\ of\ bins\ for\ template) - (Means\ of\ bins\ for\ patch)}$$

-----[13]

Also the SM uses the Euclidian distance for more accuracy, and it is defined as

$$distance = \sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2)}$$-------[14]

The x1, x2, y1, y2 are the Cartesian co-ordinates of all the next, next pixels in the images. Also, if the SM value is very lesser value, then the matching between the patch is high, else the matching is low. Also for more accuracy of the image comparison we use image information are compared using MATLAB. The complete functionality of the proposed approach is written as a pseudo code and it will be programmed in any language.

Procedure Proposed ( ) {

1. Let G be the group of images clustered manually using features
2. G = { g_1, g_2, g_3, …… g_i, …. , g_n} $\forall\ g_i \in G$
3. Assign T as test image folder, TR is the trained image folder as G
4. Input a test image t_i
5. Pi = DCT(t_i, nb x nb)
6. For I = 1 to length(p_i)
7. Tec1 = d_{mn} (Pi(i, j))
8. Edg1 = Pi(f,k)
9. For j= 1 to length(TR)
10. P2i = DCT(trj, nb x nb)
11. For K = 1 to length(P2i)
12. Tec2 = d_{mn} (Pi(i, j))
13. Edg1 = Pi(f,k)
14. If ((Tec1 == Tec2) && (Edg1 == Edg2))
15. Flag =true;
16. End
17. End
18. If (flag == "true")
19. Resultimg = trj;
20. end
21. End
22. End
23. }

G is the main folder contain all the images, and the g_1, g_2 ,,, g_n are the sub folders where, those folders are manually clustered as animals, flowers, human and cars, etc., The test image ti is patched using DCT, and the trained image tr_j is patched using the DCT. Now corresponding patch's texture, color and the edges are compared using the distance method and display if both the images are very similar.

XX. EXPERIMENTS AND RESULTS

In this paper, an effective software tool MATLAB 2012a is used for feature extractions and comparison. There are a number of images are taken and classified manually by categories wise, color wise, size wise, background wise in separate folders. Here we consider the folders are the databases for manual verification. All these features were derived based on the concern domain, The features are derived from the training image to validate the images, Based on this, A system was developed and the same technique is applied to trained the other images also, The rightness of the features can be evaluated by certain criteria. The next set of images are used for validation and comparision.Since the validation of the images are in same size, they can be termed as blocks, which are used for comparison. The input image is processed and extracts wavelet coefficients, and 1x20 feature vectors containing the first 2 moments of waveletcoefficients in the output.

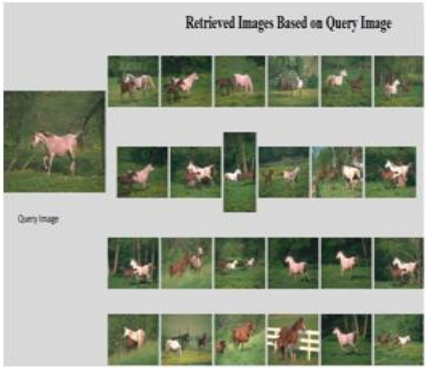

Figure-4a: Image Retrieval by shape based

In this above output, Query Image is the first image, based on the Query image we can get exact output or query based related output from our database. In our database we have stored various types of images .Here we have try to retrieve 10 related images based on the query image. In our database, we might store 10 related images so that we can get required images what we given as input. The retrieved images are most accurately matched with the Query image and no one of the partial image will be return.

Figure-4b: Image Retrieval by color based

In this above output, here also the first image is used as Query image based on that only we can get output. Here we have try to retrieve three images based on that query image. We had store only two images, which are, based on the query images. No other images will be match for original image. However, in our output, we get two images, which are exactly paired with the query image and return another image, which is not exactly matched with that input. Even though we get that image as output because which is partially related to our input.

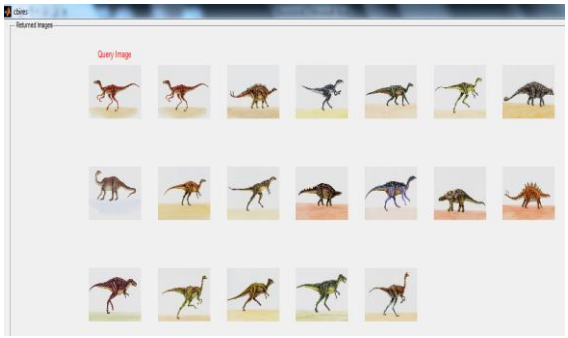

Figure-4c: Image Retrieval by Texture based

The above image shows that the accurate output is the second image located from the query image right side in the dialog box. The other images are the relevant images, which is texture wise similar, color wise similar and shape wise similar. All the other images than the second image is are relevant according to the features considered in this paper. The second image is the accurate image comparing with the query image, which is as same as for the input image is texture, color and shape wise. Also the number of relevancy input by the user is 20, but it retrieves only 18 images, which is more relevant. Also the comparison of the images is based on the color, texture, shape of the image in each patch.

The image separately display above a set of all images in Figure-1, is the input query image. Where the set of images displayed row wise are the output images. The same way we can browse and select the input images from the test folder, this proposed approach retrieves the accurate and the relevant images in the list.

XXI. QUERY IMAGE AND DATABASE

The algorithm is validated with the database having 1000 images includes 50 images with the various categories like Sunset, Flower, Mountain, Horse, Mickey, Barbie, Elephant, Dinosaur, Bus, Building images. By using DCT wavelet and DCT,Feature vectors are extracted by combined with the texture, shape and color. The CBIR is validated with the query by a sample image .When the system collected the query image it will automatically excerpts feature vector as the similar way of extracting database images. By classification of similarity measure, Euclidean distance is calculated based on the results of Ten queries from the Ten classed with the proposed algorithms All the sample images from all the classes are in the in the Figure5 below.

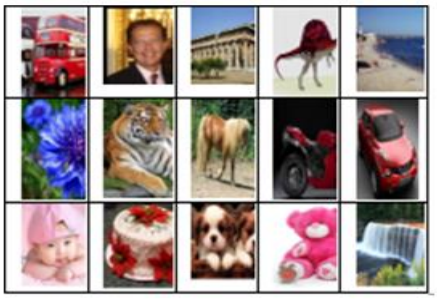

Fig 5: Sample images of all classes

XXII. 1000 SIMILAR IMAGES FETCHED FROM THE DATABASE.

Once the Query is executed it will process based on the algorithm and extract the contents to feature vector of the every patch of the image. As mentioned in the step 1 the DCT patch creation of the each image is segregated in to different plane like R, G, B, The extracted result set is saved in three different types of databases for each plane like R, G, and B.

The query results for the three planes will be compared with the database image respectively, When experiments conducted on the system ,some variation is purpose fully done in the selection of co-efficient in the form of feature vector. We have selected the initial few co-originates having significant information of the feature vector for the effort in transform domain to analyze, Make use of the energy condensation property of them. By changing, the no of co-efficient different size feature vectors can be derived .Initially we selected all the co-efficient and later by lowering the extend of the feature vector. Totally 18 types of set were experimented with the spectrum of feature vector size from 256 to 4 co-efficient for each plane with each way. In addition, the texture, color and shape of the images are compared for image mining. There are two possible outcomes of the image prediction. The outcomes are True Positives (TP), and False Positives (FP) where TP are those retrieved images that are correct, FP are retrieved images that are actually incorrect.

[1]

$$True\ Positive\ rate\ TP = \frac{Number\ of\ Images\ Retrieved\ Correctly}{Total\ Number\ of\ Image} = \frac{99}{100}$$

[2]

$$False\ Positive = \frac{Number\ of\ images\ Retrieved\ incorrectly}{Total\ Number\ correct\ images} = \frac{1}{15}$$

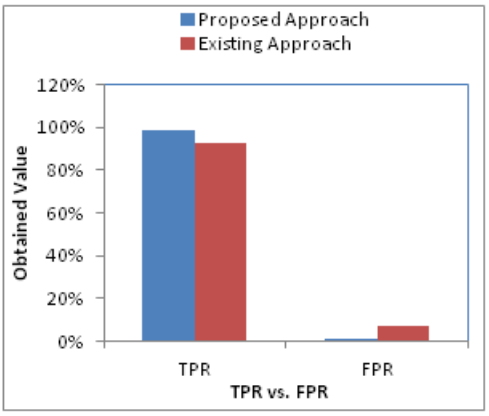

Table 2: Performance Evaluation of Proposed Approach

Method	TPR	FPR
Proposed System	99%	1.00%
Existing System	93%	7.00%

Fig 6: Performance evaluation

The table-1 shows the accomplishment evaluation of the projected approach by TP and FP values of the image mining. The proposed approach gives 98% of the True positive and .3% of the FP; hence, the proposed approach is an effective approach for image mining using feature selection based clustering.

In the above graph, The Validation of the new approach is shown. Where the true positive rate of the projected approach is 98%, the existing system is 93%, and the false positive rate is 3% in proposed system, 7% in the existing system, hence the projected approach is better approach than the existing approaches studied in the literature survey.

XXIII. CONCLUSION

The paper provides a concise review on work, related to the electrifying fields of CBIR and presents a feature analysis of the various works authorized in this field and explains about the different methods used for extorting the relevant low-levelfeatures and several interval measure to locate the resemblance among images in minimizing the semantic gap among high level semantic notations. Low-levelfeatures, several approaches of CBIR and evaluation of different process with reverence to data are also prepared.

UTURE RESEARCH

This paper explained about the relative study of CBIR drift and a different approach to determine a few problems that stumble upon CBIR. As an alternative of utilizing an untainted data driven valuation, an advanced level information regarding regions could be utilized. As the image quintessence presents a compound depiction of a shape and appearance, it is potential to attain an enhanced appraise of homogeneity of the segments. In order to resolve the semantic information quandary, an application-dependent knowledge must be integrated into an assessment process, so that this process knows the desired attributes of segment. Various processes could be used to comprise previous information, regarding desired features.

REFERENCE

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. Remco C. Veltkamp and MirelaTanase, “Content Based Image Retrieval Systems: A Survey,” International Journal of Engineering Science and Technology, Vol. 20, No. 5, pp. 1-62, 2002.
3. Dr. H.B.Kekre, Sudeep D. Thepade and AkshayMaloo, “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.
4. Danish, M., Rawat, R., & Sharma, R. (2013). A Survey: Content Based Image Retrieval Based On Color, Texture, Shape & Neuro Fuzzy. Int. Journal OfEngineering Research And Application, 3(5), 839–844.
5. Stockman, G. C., & Shapiro, L. G. (2001). Compter Vision (2001st ed., pp. 279–325). New Jersey: Prentice-Hall.
6. Jaswal, G., &Kaul, A. (2009). Content Based Image Retrieval – A Literature Review. In Communication and Control. India: National Institute of Technology.

7. Fierro-Radilla, A., Perez-Daniel, K., Nakano-Miyatakea, M., Perez-Meana, H., &Benois-Pineau, J. (2014)," An Effective Visual Descriptor Based on Color and Shape Features for Image Retrieval", MICAI 2014, Springer.
8. Viitaniemi, V. (2002). Image Segmentation in Content-Based Image Retrieval (Master Thesis). Helsini University OF Technology, Department of Electrical and Communications Engineering, Finland.
9. Danish, M., Rawat, R., & Sharma, R. (2013). A Survey: Content Based Image Retrieval Based On Color, Texture, Shape & Neuro Fuzzy. Int. Journal Of Engineering Research And Application, 3(5), 839–844.
10. Kekre, H. (2011), "Survey of CBIR Techniques and Semantics", International Journal of Engineering Science and Technology (IJEST), 3(5).
11. Jaswal, G., &Kaul, A. (2009). Content Based Image Retrieval – A Literature Review. In Communication and Control. India: National Institute of Technology.
12. Stockman, G. C., & Shapiro, L. G. (2001). Compter Vision (2001st ed., pp. 279–325). New Jersey: Prentice-Hall.
13. Black Jr., J. A., Fahmy, G., &Panchanathan, S. (2002). A method for evaluating the performance of CBIR systems. Arizona State University.
14. KanchanSaxena, VineetRicharia, Vijay Trivedi, "A Survey on Content Based Image Retrieval using BDIP, BVLC AND DCD", Journal of Global Research in Computer Science, Vol.3, No. 9, Sep. 2012, ISSN-2229-371X.
15. Z. Wang, Z. Chi, and D. Feng, "Fuzzy integral for leaf image retrieval," Proc. IEEE Intl. Conference on Fuzzy Systems, 2002.
16. H. Muller, W. Muller, D. M. Squire, S. Marchand-Maillet, and T. Pun, "Performance evaluation in content-based image retrieval: Overview and proposals," Pattern Recognition Letters, 22(5):593-601, 2001.
17. F. Long, H. Zhang, H. Dagan, and D. Feng, "Fundamentals of content based image retrieval," in D. Feng, W. Siu, H. Zhang (Eds.): "Multimedia Information Retrieval and Management. Technological Fundamentals and Applications," Multimedia Signal Processing Book, Chapter 1, Springer-Verlag, Berlin Heidelberg New York, 2003, pp.1-26.
18. M. Ghahroudi, M. Sarshar, and R. Sabzevari, "A Novel Content-Based Image Retrieval Technique Using Tree Matching," Proceedings of The World Congress on Engineering, pp1797- 80, May 2008.
19. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, and P. Yanker, "Query by image and video content: The QBIC system," IEEE Computer, vol. 28, no 9, pp.23-32, Sep. 1995.
20. D. Zhang, "Improving Image Retrieval Performance by Using Both Color and Texture Features," In Proc. of IEEE 3rd International Conference on Image and Graphics (ICIG04), Hong Kong, China, Dec.18-20, 2004, pp.172-175.
21. J. Wang, J. Li, G. Wiederhold, "Simplicity: semantics-sensitive integrated matching for picture libraries," IEEE Transaction on Pattern Analysis and Machine Intelligence, vol. 23, no. 9, pp. 947–963, Sep. 2001.
22. W. Ma and B. Manjunath, "Natra: A Toolbox for Navigating Large Image Data bases," Proc. IEEE Int'l Conf. Image Processing, Santa Barbara, 1997, pp. 568- 571.
23. C. Carson, M. Thomas, S. Belongie, J.M. Hellerstein, and J. Malik, "Blob world: A System for Region-Based Image Indexing and Retrieval," Proc. Visual Information Systems, pp.509-516, June 1999.
24. Natsev, R. Rastogi, and K. Shim, "WALRUS: A Similarity Retrieval Algorithm for Image Databases," IEEE Trans. On Knowledge and Data Engineering, vol.16, pp. 301-318, Mar. 2004.
25. M. Sudhamani, and C. Venugopal, "Image Retrieval from Databases: an Approach using Region Color and IndexingTechnique," International Journal of Computer Science and Network Security, vol.8 no.1, pp. 54-64, January 2008.
26. J. Li, J. Wang, and G. Wiederhold, "Integrated Region Matching for Image Retrieval," In Proceedings of the 2000 ACM Multimedia Conference, Los Angeles, October 2000, pp. 147-156.
27. R. Zhang, and Z. Zhang, "A Clustering Based Approach to Efficient Image Retrieval," Proceedings of the 14th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'02), Washington, DC, Nov. 2002, pp. 339-346.
28. R. Gonzales, R. E. Woods, "Digital Image Processing," 2nd Ed., New Jersey Prentice Hall, 2002.
29. T. Gevers and A. Smeulders, "Pictoseek: Combining color and shape invariant features for image retrieval," IEEE Trans. Image Processing, vol. 9, no. 1, pp.102– 119, Nov. 2000
30. Ouyang, and Y. Tan, "A novel multi-scale spatial-color descriptor for content-based image retrieval," Proceedings of the 7th International Conference on Control, Automation, Robotics and Vision, Mexico, August 2002, vol. 3, pp. 1204-1209.
31. H. Yu, M. Li, H. Zhang, and J. Feng, "Color texture moments for content-based image retrieval," Proceedings of the International Conference on Image Processing, Rochester, New York, USA, September 22-25, 2002, vol. 3, pp. 929-932.
32. T. Gevers, and H. Stokman, "Classifying color edges in video into shadow-geometry, highlight, or material transitions," IEEE Transactions on Multimedia, vol. 5, no. 2, pp. 237-243, Sep. 2003.
33. H. Guan, and S. Wada, "Flexible color texture retrieval method using multi-resolution mosaic for image classification," Proceedings of the 6th International Conference on Signal Processing, vol. 1, pp. 612-615, Feb. 2002.
34. H. Moghaddam, T. Khajoie, and A. Rouhi, "A new algorithm for image indexing and retrieval using wavelet correlogram," Proceedings of the International Conference on Image Processing, vol. 3, pp. 497-500, May 2003.
35. M. Lew, N. Sebe, C. Djeraba and R. Jain, "Content-Based Multimedia Information Retrieval: State of the Art and Challenges," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 2, no. 1, pp. 1-19, February 2006.
36. J. Fuertes, M. Lucena, N. Perez, and J. Martinez, "A Scheme of Color Image Retrieval from Databases," Pattern Recognition Letters, vol. 22, pp.323–337, June 2001.
37. B. Manjunath and W. Ma, "Texture features for Browsing and retrieval of image data," IEEE transactions on pattern analysis and machine intelligence, vol. 18. No. 8, pp. 837-842, August 1996
38. R. Haralick, K. Shanmugam, and I. Dinstein. "Texture Features for Image Classification," IEEE Trans. on Systems, Man and Cybernetics, SMC, vol.3, no 6, pp. 610–621, Nov. 1973.
39. R. Datta, J. Li, and J. Wang, "Content-based image retrieval - approaches and trends of the new age," ACM Computing Surveys, vol. 40, no. 2, Article 5, pp. 1-60, April 2008.
40. M. Chantler, and J. Wu, "Rotation Invariant Classification of 3D Surface Textures using Photometric Stereo and Surface Magnitude Spectra," Proceedings of British Machine Vision Conference, Vol.2, pp 486-495, Jan. 2002.
41. R. Dubes, and A. Jain, "Random field models in image analysis," Journal Applied Statistic, vol. 16, no. 2, pp.131-164, Nov. 1989.
42. Pentland, "Fractal-Based Description of Natural Scenes," IEEE Transaction on Pattern Analysis Machine Intelligence, vol. 6, no. 6, pp. 661-674, 1984.
43. J. Mao, and A. Jain, "Texture Classification and Segmentation using Multi-Resolution Simultaneous Autoregressive Models," Pattern Recognition, vol. 25, no. 2, pp. 173-188, 1992.
44. C. Teh and T. Roland, "On image analysis by the methods of moments," IEEE Trans. On Pattern Analysis and Machine Intelligence, vol. 10, no. 4, pp. 496-513, 1988.
45. M. Hu. "Visual Pattern Recognition by Moment Invariants," IEEE Transactions on Information Theory, IT, vol. 8, pp. 179-187, Feb. 1962.
46. M. Bober, "MPEG-7 Visual Shape Descriptors", IEEE Trans. on Circuits and Systems for Video Technology, vol. 11, no. 6, pp. 716-719, June 2001.http://mpeg.telecomitalialab.com/standards/mpeg-7/mpeg-7.htm
47. F. Mokhtarian, and R. Suomela, "Robust Image Corner Detection Through Curvature Scale Space", IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 20, no. 12, pp. 1376-1381, December 1998
48. G. Qian, S. Sural, Y. Gu, and S. Pramanik, "Similarity between Euclidean and cosine angle distance for nearest neighbor queries," Proceedings of ACM Symposium on Applied Computing, vol. 12, no. 22, pp. 1232-1237, 2004.
49. M. Swain, D. Ballard, "Color indexing," International Journal of Computer Vision vol. 7, no 1, pp 11-32, Nov. 1991.
50. J. Laaksonen, M. Koskela, S. Laakso, and E. Oja, "Picsom-content-based image retrieval with self-organizing maps," Pattern Recognition Letters, vol. 21, pp. 1199– 207, Feb. 2000.