Scientific Journal of Impact Factor (SJIF): 4.72

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 4, Issue 4, April -2017

Reverse Image Search

Image retrieval using various techniques

Lad Ami D.[1], Mori Shilpa I.[2], Raulji Urvashi K..[3], Shaikh Saliha H.[4]

[1]*Information Technology, SVM Institute of Technology*
[2]*Information Technology, SVM Institute of Technology*
[3]*Information Technology, SVM Institute of Technology*
[4]*Information Technology, SVM Institute of Technology*

Abstract -*Widely spreaded applications in different kinds of fields and also large amount of database in image here image extraction , retrieval techniques are developed and enhanced. The image retrieval is a very interesting and rapidly growing methodology in all fields. It is an effective and well organized approach for retrieving the image. Image mining is the concept of extracting information from the large collection of image database. It is a process of searching the information and discover the knowledge from that data. This techniques deals with many other emerging techniques such as handling the hidden knowledge using extraction, data association, finding the additional patterns. Here we provide overviews of the Image Mining techniques emerged with various image detection techniques.*

Keywords- *Image Retrieval , Image Extraction, Clustering, Edge Detection, Content Base Image Retrieval, Image Features , Color Model.*

I. INTRODUCTION

Content-based image retrieval is known as **query by image content** and **content-based visual information retrieval**. It is the application of the techniques that are related to computer vision techniques with extraction of image and retrieval. The problem related to image retrieval is searching of image from large amount of databases of image. Here CBIR is traditional Concept – based image indexing or we can say concept –based approach.

The meaning of "Content-based", searching the real content of image, in short Content-base is focuses on original content of image not metadata like keywords, descriptions, tags which associated with image data.

Here "content" key term defines by colors, shape, textures, and many other information that can derived from the image.

Here normally image can be retrieved using the way of entering keywords or metadata from large database can be time consuming not capture the keywords to describe the following image. Image search technique is a subjective technique and has not been well-define.

In the same regard, CBIR systems have similar challenges in defining success.[2]

1 History

2 Technical progress

3 CBIR techniques

3.1 Query

"Content-based image retrieval" originated in 1992 , it was used by T. Kato for experiments into automatic retrieval of images from a database, based on the colors and shapes present.[2] then, the term has been used for describing the process of retrieval of images from a large collection of database on the basis of syntactical image features. The techniques, tools, and algorithms that are used like statistics, pattern recognition, signal processing, and computer vision[1].

The CBIR system was developed by IBM and was called QBIC (Query by Image Content).[3] Recent network and graph based approaches presented as alternative to existing methods.[4]

Textual information of the images can be easily searched using following technology, but for that purpose this technique requires humans for manually describe each image.

Systems based on categorizing images in semantic classes like "apple" as a subclass of "fruits" can avoid the miscategorization problem, but this thing will require more efforts and hardwork by a user to find images that might be "apple", but that only classified as an "fruits". Here for that purpose Many standards have been developed to categorize images, but still face the problems like scaling and miscategorization.[2] CBIR systems were developed to search databases based on image color, texture, and shape properties.

Based on the properties like color , texture and shape CBIR systems were developed to search image from large collection of database.

Day by day many CBIR systems have been developed but main goal of retrieving image on the base of pixel value remains unsolved.[5]

Many Content Based Image Retrieval systems have been developed, but the problem of retrieving images on the basis of their pixel.

Computer technology become worldwide with tremendous database such as numerical data, categorical data various images, sounds, voice, videos which is a part of our daily routine. Here there is huge amount of data is associated with it or we can knowledge is hidden in these data. During 19s image research techniques were developed and then text based image retrieval technology came. After than context web technology were introduce. Using image person can find out the information which is related to him, which he want, but amount valuable knowledge extraction from dataset was not easy task[5].

Image mining concept introduced for the purpose of the extraction of implicit knowledge, patterns stored in image data, relationship with image data.

Image mining is the extension of the data mining. Image Mining is the concept related with computer vision, artificial intelligence, image processing, image understand ability, machine learning.

Image mining is combination of two different approaches 1) Extract images from database or the image collection. 2) Mining or digging out the possible combinations with is related with image here, possible combination of associated alphanumeric data.

Image Mining classification can be done within two direction 1) specified with Domain 2) General application.

These Bothe are used to extract features of image and generate its relevant patterns.

Medical, astronomy, industry, education, sport here, these all kinds of fields are related with associative image data, so we can say that large amount of image data is generated daily.

Medical, astronomy, industry, education, sport here, these all kinds of fields are related with associative image data, so we can say that large amount of image data is generated daily. But there are many issues in the progress of image mining. Here the fundamental is the extract knowledge relating to the image from various web pages.

Image retrieval system is for search through database to find the image that are nearly related to the query image. CBIR is an very important and enhanced technique to the text based image search. CBIR techniques is generally used for extracting low level features such as texture, color, shape of the image, layout of the object.

TBIR is the text based image retrieval, here in TBIR image was retrieved by description or we can say using text.

But this technique is not applicable for the large scale database because these system is task depended queries and manually task is not applicable.

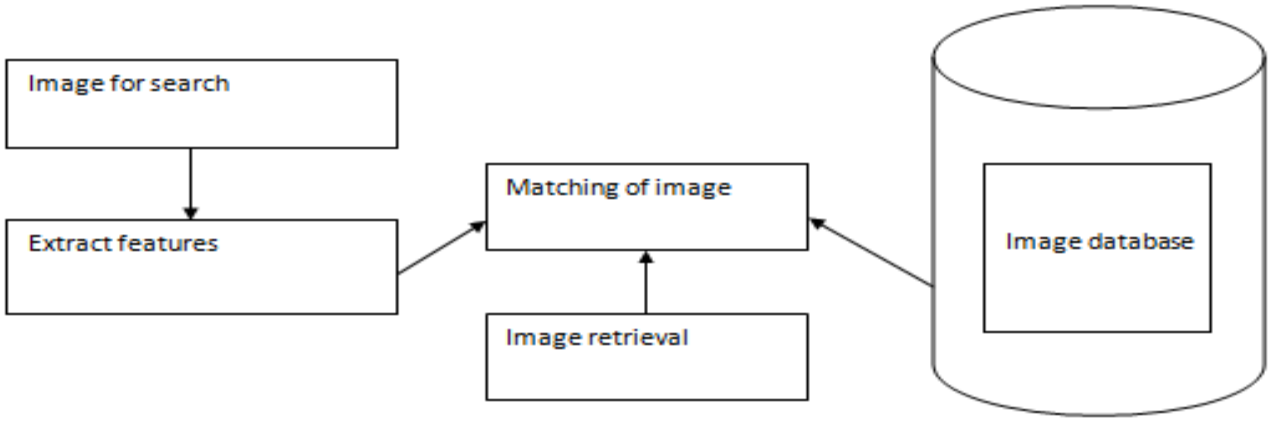

Fig-1 General Image Retrieval Process [5]

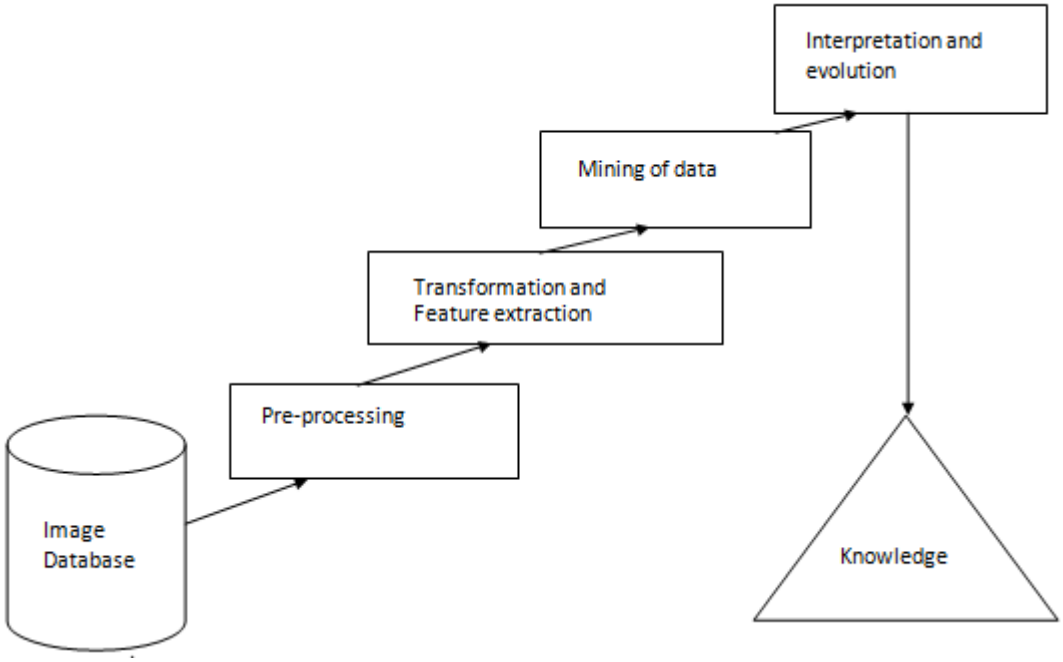

Fig-2 Image Mining Process [4]

This paper is organized as follow: Over view of image mining techniques and works in this area of image retrieval and image extraction.

II. IMAGE-MINING AND ITS MENDETORY REQUIRED FIELDS

This method is utilized method from various fields such as computer vision, image processing, image retrieval, data mining, machine learning, database and artificial intelligence[6]. There are two techniques first one is standalone and second one is integrated collections of the image[7].
Image mining process usually divided into several parts .

A. Data preprocess

In image mining the mining is related with large collection of data with high resolution, dimensions and multiple features. During analysing the data the cost of time and space are relatively very high. Dealing with large set of data there exist some problems. Such problems are dealing with lot of dirty, noisy data. So these kinds problems creates bad results for that we have to preprocess the data cleaning up the data reduce the noise. So that these process can improve the quality and give the efficient quality of data. Thresholding, border tracing, wavelet-based segmentation are used to improve quality of data.

B. Extracting feature

There are key points that will be used to express the data of image, generally features texture, color, edge, shape. For better mining result, it is necessary to meet features. Here some of the main features which is very important for image mining, features are color, edge, textures.[11] Using image processing technologies such as image segmentation, picking up the edge to extract task- related feature vectors, form multi-dimensional feature vectors.

C. Obtaining high-level knowledge

Classifications, clustering, neural networks, association rules are the most commonly used techniques of mining such as data mining , image mining. Object recognition, image indexing and retrieval, image classification and clustering, neural network are oftenly used for feature vectors and mining. Used to evaluate high- level knowledge and explain their knowledge.
1.Image classification

Classification and clustering are two main data-mining techniques which are widely used during mining. Here classification is supervised learning method. Classification technologies are: Bayesian technique, neural network,

decision tree, K-nearest–neighbor-classifier, genetic algorithm etc. Process involves in classification 1) feature extraction 2) model description of established class 3) indexing of the images.

2.Image clustering

Image clustering is the another mining technique as classification technique but clustering is unsupervised learning technique. Here comparison done with similarity of image data without any prior technique which is called as clustering. Image within a cluster have high similarity in comparison to one another but are very dissimilar to images in other clusters. The process is done in following form: 1) Image representation, extraction of features and selection 2)image mapping, set up similarity metrics 3) Image clustering 4) Generate cluster. When clusters are generated ,they going to be examine by field experts where they are labelled with abstracting concept. there are many clustering algorithms such as: partitioning methods, hierarchical methods, grid-based methods, model-based methods, etc.

3.Association rules mining

Association rue mining related with correlation relationships among the large data set of image data. association rule shows the attribute value that frequently occur in given dataset. Association rule mining consists of first finding frequent item sets, from which strong association rules are generated. There are rules which used to satisfied the minimum confidence threshold. A-priori is a typical algorithm for association rule mining, based on which there are a lot of improvement, for example, in [12]. Spatial mining algorithm to the spatial association rules from an image database, proposed more efficient than A-priori algorithm ,2-5 times faster than A-priori.

III. REVERSE IMAGE SEARCH AND RELATED FIELDS

Clustering algorithm:

As above mention clustering means grouping technique in which similar kind of data are grouped in one formation and other unmatched data are grouped into another formation. This is data mining process used to analyze the data. This technique mainly used in many fields such as image analysis, data compression, pattern recognition etc. it is an iterative process of knowledge discovery which is emerged with trail and failure, it is not specific algorithm but common process.

Connectivity-based clustering (hierarchical clustering)

Hierarchical clustering is also process of connectivity based clustering. Main idea for this is object related with another object. Here connectivity done with similar kind of objects. Here cluster defines by maximum distance needed to connect the different parts of the clusters. different clusters define and according to that which cluster is present can be represented using a dendrogram. Dendogram means where the common name "hierarchical clustering"

.

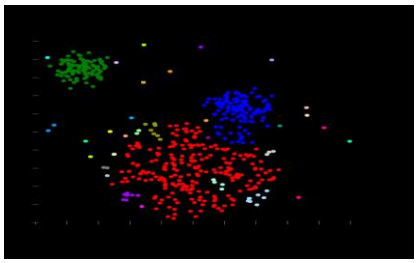

Fig. 1 on Gaussian data

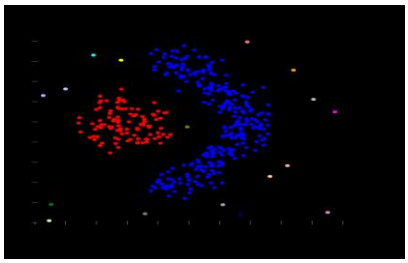

Fig. 2 on density-based cluster

Fig.1 Single-linkage on Gaussian data. At 35 clusters, the biggest cluster starts fragmenting into smaller parts, while before it was still connected to the second largest due to the single-link effect.
Fig .2 Single-linkage on density-based clusters. 20 clusters extracted, most of which contain single elements, since linkage clustering does not have a notion of "noise".
Advantages:
The main advantage of a clustered solution is automatic recovery from failure, that is, recovery without user intervention.
Disadvantages:
Clustering are complex and inability to recover from the corrupted database.

CANNY ADGE DETECTION:

Canny edge detection is a technique to extract useful structural information from different vision objects and dramatically reduce the amount of data to be processed. It has been widely applied in various computer vision systems. Canny has found that the requirements for the application of edge detection on diverse vision systems are relatively similar. Thus, an edge detection solution to address these requirements can be implemented in a wide range of situations. The general criteria for edge detection includes:

1. Detection of edge with low error rate, which means that the detection should accurately catch as many edges shown in the image as possible
2. The edge point detected from the operator should accurately localize on the center of the edge.
3. A given edge in the image should only be marked once, and where possible, image noise should not create false edges.

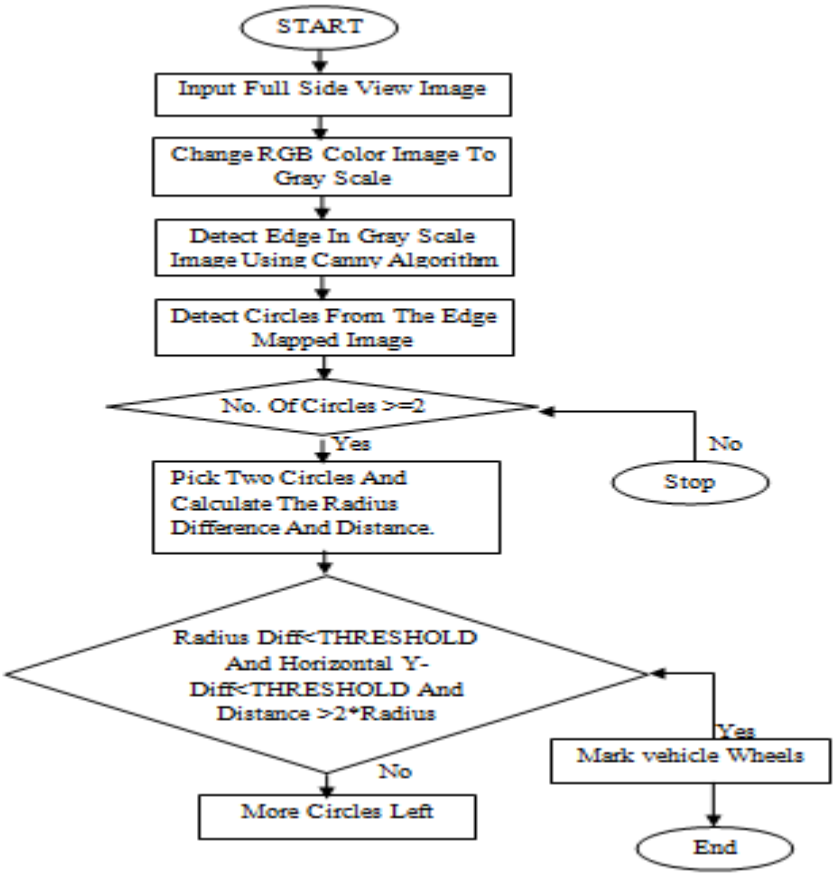

Fig.3 Flowchart of Edge Detection

A. The Process of Canny edge detection algorithm can be broken down to 5 different steps:

1. Apply Gaussian filter to smooth the image in order to remove the noise
2. Find the intensity gradients of the image
3. Apply non-maximum suppression to get rid of spurious response to edge detection
4. Apply double threshold to determine potential edges
5. Track edge by hysteresis: Finalize the detection of edges by suppressing all the other edges that are weak and not connected to strong edges.

Every step will be described in details as following. The introduction of procedure below is developed based on Prof Thomas Moeslund's lecture note for digital image processing in Indian Institute of Technology.

K-means clustering:

K-means algorithm is a method of vector quantization, from signal processing, that is known for analysis of cluster in data mining. The goal of k-means clustering is to divide *n* observations into k clusters in which each observation belongs to the group with the nearest mean, serving as a example of the cluster. An output of this is in a partitioning of the data space into Voronoi cells. These are usually similar to the expectation-maximization algorithm for mixtures of Gaussian distributions via an iterative refinement approach employed by both algorithms. The problem is computationally difficult, there are efficient heuristic algorithms that are commonly employed and converge quickly to a local optimum. They both use cluster centers to model the data. k-means clustering used to identify group of comparable spatial extent, while the expectation-maximization mechanism allows clusters to have different shapes.

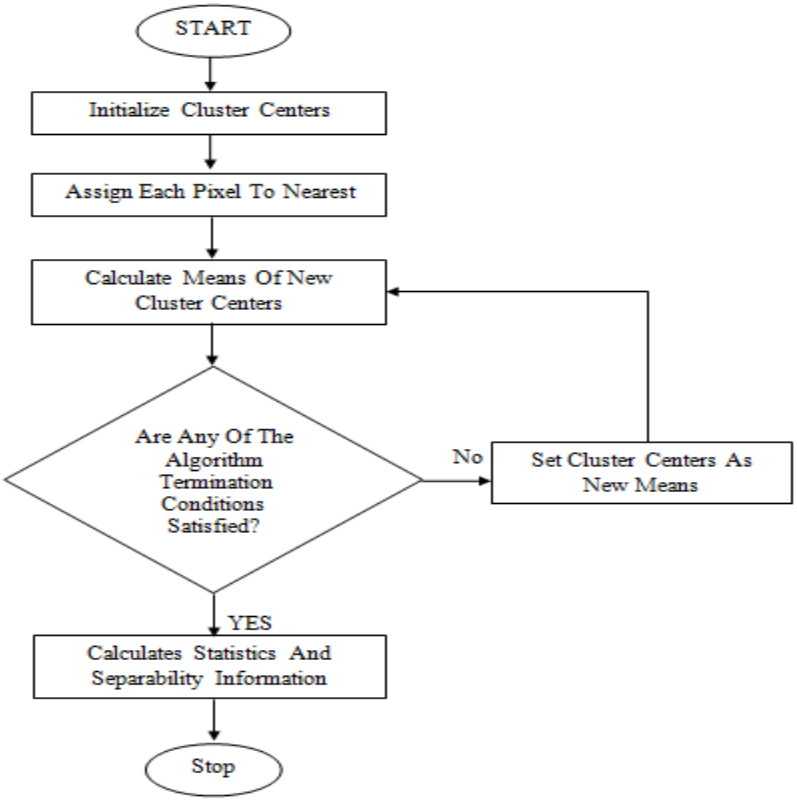

Fig.4 Flowchart of K-means Clustering

The algorithm has a loose relationship to the *k*-nearest neighbor classifier, a popular machine learning technique for classification that is often confused with *k*-means because of the *k* in the name. One can apply the 1-nearest neighbor classifier on the cluster centers obtained by *k*-means to classify new data into the existing clusters. This is known as nearest centroid classifier or Rocchio algorithm.

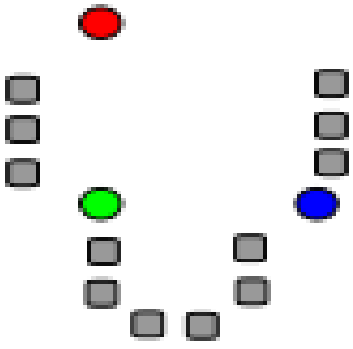

Fig.4(a)

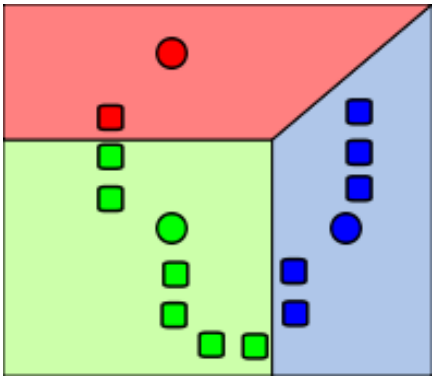

Fig.4(b)

Fig.1. Initial "means" (in this case *k*=3) are randomly generated within the data domain.
Fig.2 Clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.

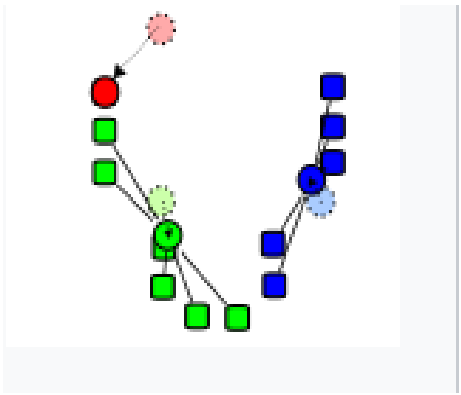

Fig.4(c)

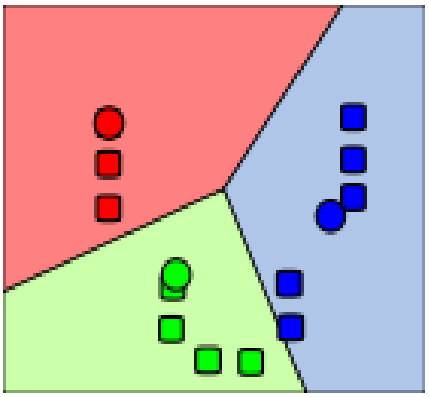

Fig.4(d)

Fig.3 The centroid of each of the *k* clusters becomes the new mean.
Fig.4 Steps 2 and 3 are repeated until convergence has been reached.

Color detection and retrieval:

The following image shows the home page . I which one file up loader and one upload button have added. Two radio buttons are also there for color and shape detection. The image selected from file up loader is displayed I the image box. After selecting image when user hit the upload button then control is redirected to the second page where in the image grid all the retrieved images will be shown. As seen in the above figure all the blue color images are displayed… because the user selected following image as input image.

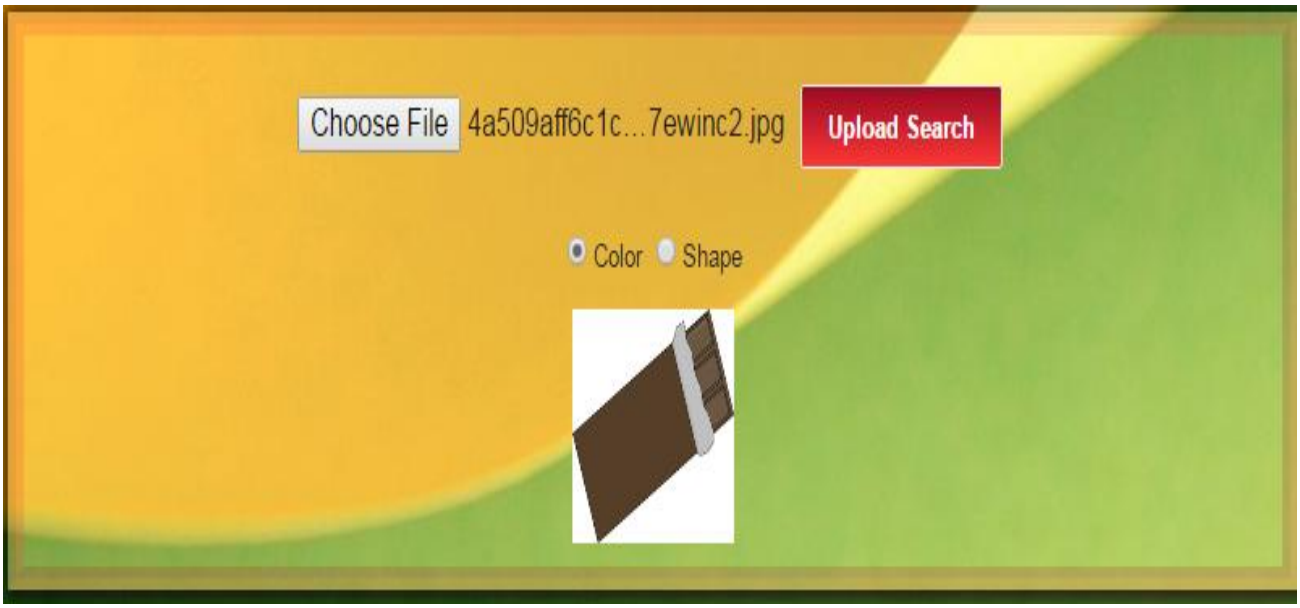

Fig.5 Upload Image

As user has selected brown chocolate image then the program will mined all the brown colored images.

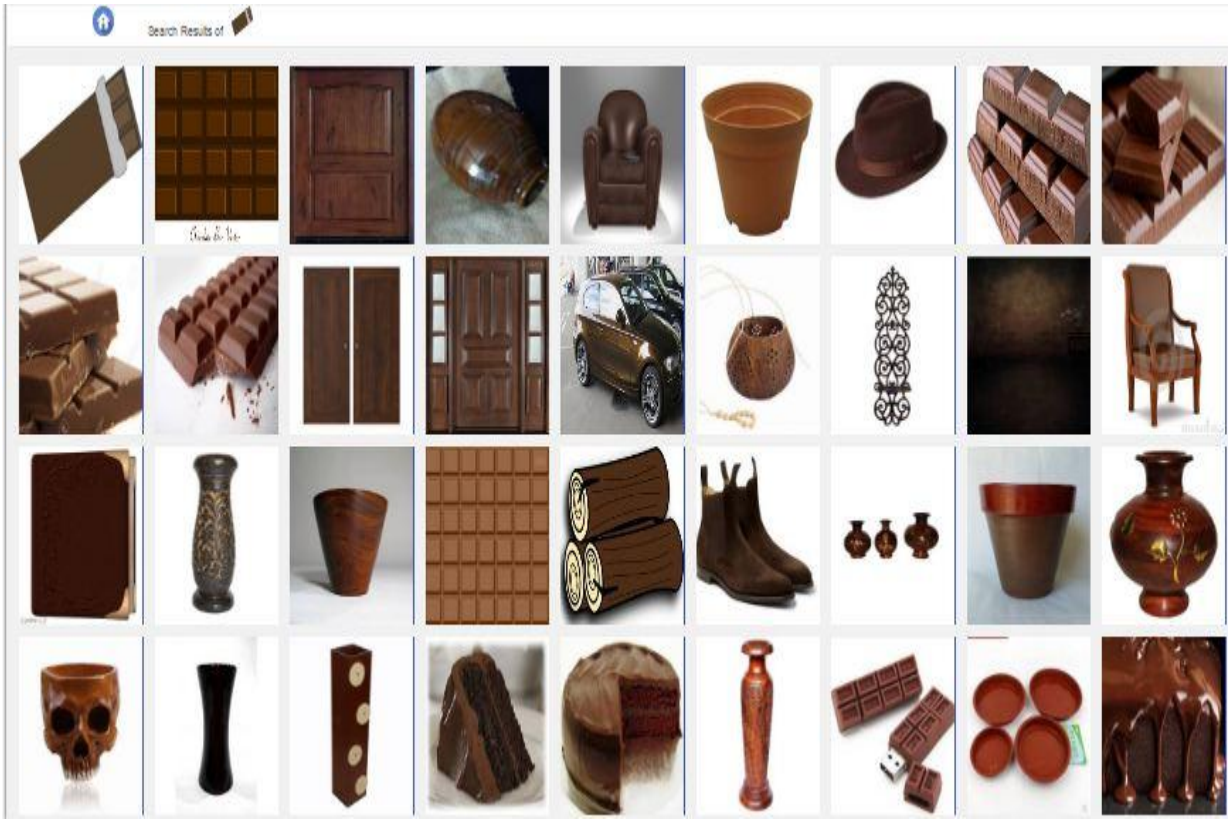

Fig.5(a) Image Retrieval

In the above figure it can be seen that all the brown images are selected to display as an output.
And one home button is also there to go back to the home page.

Edge detection:

Before doing shape detection the important step is to detect edges of an objects. By Edge detection system comes to know the objects of an image and then easily detect the shape available in that particular image.

Fig.6 Edge Detection

The edge detection process can be seen in the figure. Here canny edge detection algorithm is applied. Canny edge detection algorithm is the simple and effective algorithm through which more and more edges can be detected.

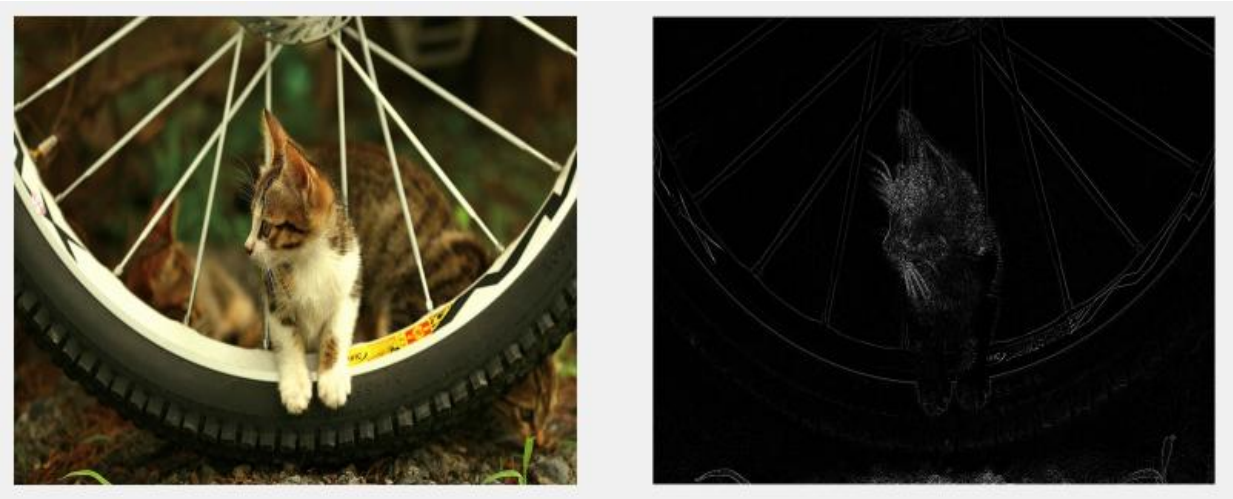

Fig.6 (a) Edge Detection

By applying canny edge detection algorithm on the above cat image all the possible edges are detected and the processed image is displayed as an output image.

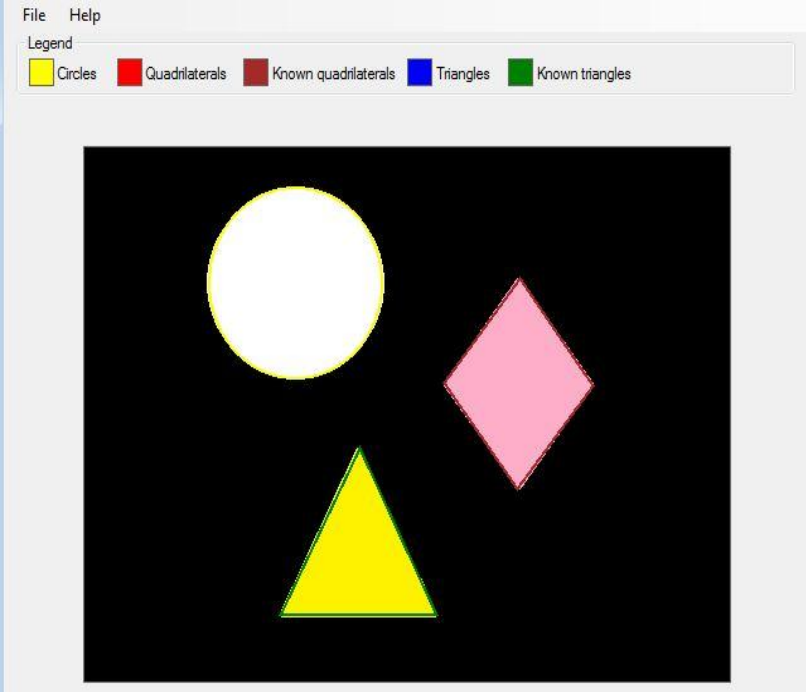

Fig.6 (b) Canny Edge Detection

Canny edge detection algorithm helps to detect shapes in the image. The above module is prepared to show that how the system detect the shapes and even recognize the shape. As seen in the figure system have detected circle, triangle and diamond shapes.

IV. CONCLUSION

Image Mining and Image retrieval techniques are increase day by day with the increasing the demand of various multimedia application over the worldwide. Here these paper provides the approach of image extraction and image retrieval with various techniques such as color detection and edge detection, through which we can extract exact data which fulfill our requirements. In this paper we provide overview of the basic theories and various techniques related with image mining. There is not a single technique that gives in best for all user's requirements, here inventions of new methodologies according to requirements will be increase.

REFERENCES

[1] Hu Min and Yang Shuangyuan."Overview of Image Mining Research",The 5th International Conference on Computer Science Education Hefei, China. August 2427, 2010.
[2] Chang S.K. and Hsu A., Image information systems:where do we go from here? IEEE Trans. On Knowledge and Data Engineering, 5(5), 431-442(1992) .
[3] P. S. Hiremath, JagadeeshPujari, Content Based ImageRetrieval using Color, Texture and Shape features,International Conference on Advanced Computing andCommunications,780-784 (2007).
[4] Zhang Ji, Hsu, Mong and Lee (2001), "Image Mining: Issues, Frameworks and Techniques,"Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), in conjunction with ACM SIGKDD conference, San Francisco, USA, 26th August.
[5] Mussarat Yasmin, Sajjad Mohsin and Muhammad Sharif," Intelligent Image Retrieval Techniques:ASurvey", Journal of Applied Research and Technology.
[6] C. Ordonez and E. Omiecinski (1999), "Discovering association rules based on image content,"Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99).
[7] O. Zaiane and J. Han (1998), "Mining MultiMedia Data," CASCON'98: Meeting of Minds, Toronto,Canada, pp 8396,November. [8] R. Missaoui and R. M. Palenichka (2005), "Effective image and video mining: an overview of modelbased approaches," In MDM '05: Proceedings of the 6th international workshop on Multimedia data mining, pp. 43–52.
[9] Chin-Chuan Han, Hsu-Liang Cheng, et al. Personal authentication using palm-print features. Pattern Recognition 36 (2003) 371 – 381.
[10] Deepak Kolippakkam*, Huan Liu, et al. Feature Extraction for Image Mining.
[11] Anthony J.T. Lee, Ruey-Wen Hong, et al. Mining spatial association rules in image databases. Information Sciences 177 (2007) 1593–1608. H. Mohamadi, A. Shahbahrami, J. Akbari, " Image retrieval using the combination of text-based and contentbased algorithms", Journal of AI and Data Mining, Published online: 20, February-2013.
[12] Pabboju , S. and gopal, R. (2009). A Novel Approach For Content- Based Image Global and Region Indexing and Retrieval System Using Features. International Journal of Computer Science and Network Security. 9(2), 15-21.

