A Review: Content Base Image Mining Technique for Image Retrieval Using Hybrid Clustering

Gurpreet Kaur

M-Tech Student, Department of Computer Engineering, Yadawindra College of Engineering, Talwandi Sabo, India

ABSTRACT:

Valuable information can be hidden in images. The need for image mining is high in view of the fast growing amounts of image data. Image mining deals with the extraction of image patterns from a large collection of images in database. Clearly, image mining is different from low-level computer vision and image processing techniques because the focus of image mining is in extraction of patterns from large collection of images according to user queries, whereas the focus of computer vision and image processing techniques is in understanding and / or extracting specific features from a single image. While there seems to be some overlaps between image mining and content-based retrieval, image mining goes beyond the problem of retrieving relevant images. In image mining, the goal is the discovery of image patterns that are significant in a given collection of images as per user queries.

Keywords: - image mining, CBIR, Hierarchical clustering, K-Means clustering.

I. INTRODUCTION

The image mining was introduced to extract implicit knowledge, image and data relationship. Image mining is an extension of data mining. In text based image retrieval system only find out the images those are just concerned with the accurate text that is described by human or relevant query, instead without looking into the content of related images. Images have many more duplications and user is not aware about it.WWW having largest global image repository. So remove this drawback with the help of Content Based Image retrieval. Content based image

Retrieval (CBIR) is widely used for retrieved the images from a very large data collection of images. CBIR extract the Images as per their features. Because it is a huge problem to retrieve the required images from the image database very frequently . The users are always not satisfied with the given technologies they used in present time they always look forward for further enhancement. The CBIR focuses on image features. The Features are further classified as low-level and high –level features. User simply put the query regarding that features such as color, shape, region etc and retrieve the required images. After that we have to focuses on clustering to combine the related images into one cluster and other images into another cluster for fast retrieval.
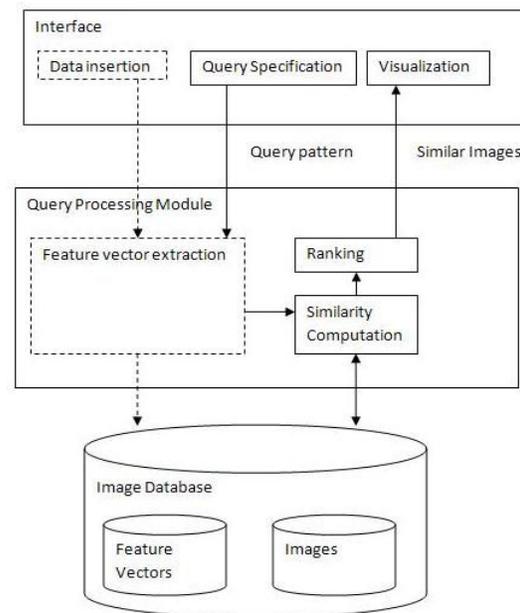

Fig.1 Typical architecture of Content based Image Retrieval System.

II. OVERALL PROCESS OF IMAGE MINING

Image mining is just at its infancy, however, observing from some of the existing image mining systems, overall process can be divided into the following parts:-

1. Data preprocess
A lot of dirty and noisy data exist in large image databases, for instance, images that are extremely unclear. Those data often cause chaos in mining process and give birth to worse mining results, so it is Necessary to preprocess data, clean up the noisy, dirty data to highlight the features if that image

2. Extracting multi-dimensional feature vectors
Using image processing technologies such as image Segmentation, picking up the edge to extract task-related feature vectors, form multi-dimensional feature vectors.

3. Mining on vectors and acquire high-level knowledge
Various methods such as object recognition, image indexing and retrieval, image classification and clustering, neural network are used on feature vectors for mining and acquiring hidden and valuable high-level knowledge, then evaluate and explain that exact query related knowledge.

III. LITERATURE REVIEW

Numerous researches have been carried in Image mining .in this section of the paper present a survey on various images mining techniques those were proposed earlier.

Dingyuan Xia et.al in (2009) IEEE [1] In a word, the above-mentioned researches actually try to solve the technical problems faced in the process of image Retrieval under network environment from the following two aspects:
1. The uniformity of image features description. Namely, to obtain features by extracting for

expressing the image contents widely and effectively. Nowadays, we haven't yet found such a method.
2. The universality of image retrieval method. Namely, it ought not to lead to system performance fluctuation due to the query or man-made factors when querying. It has been difficult for the existing methods to meet with the demand as yet.
we may develop a new and common image retrieval way, analyze the inherent correlation expression of low-level visual Features and high-level semantic features, explore the Feature-level fusion of multi-modes features and decision-making level information fusion among them, and construct the image retrieval system based on the unified description of multi-modes features and interactive relevance feedback between human and computer to adapt to users' needs on the Internet.

Mahip M.Bartere in (2012) IJERA [2] in this paper analyzes the performance of the two different colors quantization and the five distance functions. Each combination is evaluated with respect to the percentage of Retrieval Robustness using the same image set and the similar color groups. Below figures and the others considering all image color groups show that the results are related with the images color composition more than with the number of bins on the histogram. They depend also of the number of images the user wants to find. On average, better results concerning retrieval robustness were obtained using Euclidian distance metric, which is also an easy computed value. Great difference of metric performance can be seen if few numbers of images are asked. On increasing the number of wanted images almost all metric presents quite the same performance.

Jan-Ming Ho et al. in **(2012) IEEE [3]** in this paper a novel system Architecture for CBIR system which combines techniques includes color analysis, with content-based image retrieval as well as data mining techniques. In this first time segmentation and grid module, feature extraction module, K-means clustering and bring in the neighborhood module to build the CBIR system. Concept of neighborhood color analysis module recognizes the side of every grids of image is first contributed in this paper. In this

paperwork not only for Image retrieval it improves quality or fact of being exact and accurate image.

In this method CBIR having two parts: - learning and querying.

1. In learning is a training process in this a sample of images put as an input and k-means algorithm for feature extraction. Finally classification set as a learning code book.
2. In querying part image searching process explained and matched the image with output image and the output matched with user query.
3. The main motive to use segmentation and grid computing used to decrease the computing time. in this image divide into F*F grid , where F describe the Feature Module and neighbor module and again sub divide Into S*S, where S describe the color feature
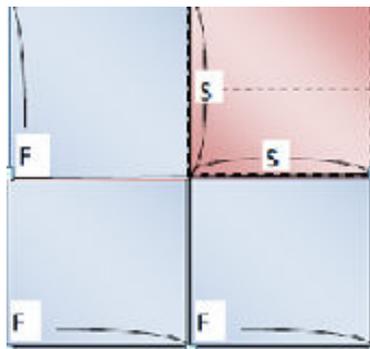

Figure 2. Grid module and segmentation

Shaikh Nikhat Fatma, madhu Nashipudimath in **(2011) IEEE** [4] presented association rule for Image mining. Association rule is deals with the extraction of image pattern from a large database of images. This method help us for Prediction We will discuss this with an example if sky contains black clouds so there are 64% chances it will rain.
The method is as follows:-

1) It segmented the images into blobs (region descriptor) where blob is equal to an object.
2) Compare blob with all other blobs with an id. This works as a pre-processing algorithm

3) After That Create Auxiliary images with identified objects.
4) Apply data mining techniques to produce object association rule.
Basically this technique used for selecting images for a particular field (eg. Weather, medical images)

A.Kannan et al. In (2010) IEEE [5] this paper author describes that image mining is the main concept which can extract potential information from the collection. For color based image extraction RGB model is used, RGB component taken from each and every image. Images are stored by mean values of Red, Green, blue components of target images. The top ranked images are further regrouped according to texture features. The gray level co-occurrence matrix (GLCM) used texture calculations (contrast, dissimilarity, homogeneity).The images are classified into clusters with the help of GLCM based on Low texture, average texture and high texture. Texture based classification is simply easy and efficient for real time applications as compared to Entropy method. The authors also evaluate the performance with the help of precision v/s recall graph. Recall value 1 just by retrieving all images and precision value kept in a higher value by retrieving only few images.

$$\text{Precision} = \frac{\text{No of relevant images Retrieved}}{\text{Total Number of Images Retrieved}}$$

$$\text{Recall} = \frac{\text{No of relevant images Retrieved}}{\text{Total Number of Images in the database}}$$

This method has proved to be highly encouraged performance in terms of simplicity, robustness and efficient. This method removes the data loss and extracting meaningful information as per user requirement. In this texture classification is focused. Entropy is used to compare images with some

threshold constraints. This application helpful for medical images to find out the right disease.

S. P. Victor and S. John Peter (2010) EJSR [6] put forth a new minimum spanning tree based clustering algorithm for image mining. The minimum spanning tree clustering algorithm is proficient of detecting clusters with irregular boundaries. The author presented a minimum spanning tree depending on the clustering technique using weighted Euclidean distance for edges, which is vital constituent in constructing the graph from image. The technique constructs 'k' clusters with segments. This approach is very much capable of protecting detail in low variability image regions while not considering detail in high variability regions which is the main advantage of this approach. This approach has handled the problems of undesired clustering structure and redundant huge number of clusters. Effective research in the field of image retrieval and mining has turned out to be a significant research area because of significant applications in digital image databases.

Rajshree S. Dubey in (2010) IJCSE [7] illustrated about an Image mining methods which is dependent on the Color Histogram, texture of that Image. The query image is considered, then the Color Histogram and Texture is created and in accordance with this the resultant Image is found. They have examined a histogram-based search techniques and color texture techniques in two different color spaces, RGB and HSV. Histogram search distinguish an image through its color distribution. It is revealed that images retrieved by using the global color histogram possibly will not be semantically related although they share comparable color distribution in some results.

IV. FUTURE ENHANCEMENT

Image mining is an extension of data mining technique. Most of the image processing algorithms include image mining. Therefore, image mining is always an emerging field and it has attracted a lot of researchers to investigate its applications in recent years.

- Some possible future studies that may be conducted in the area of image mining include the experimentations on other image elements to improve accuracy such as textures, shape, and so forth.
- We can use hierarchical method with k-means clustering algorithm to Preprocessing is done by applying hierarchical clustering algorithm based on color feature. subjective

V. CONCLUSION

This paper presents a survey on various image mining techniques that was proposed earlier by Researchers for the better development in the field of content based image retrieval. The purpose of this survey is to provide an overview of the functionality of content based image retrieval systems. We found during this survey HC and K-Means strategy can help us in both efficiency and quality. Clustering is very efficient and powerful technology to handle large data sets. It assists faster image retrieval and also allows the search for most relevant images in large image database. K-means is a clustering method based on the optimization of an overall measure of clustering quality is known for its efficiency in producing accurate results in image retrieval. By using k-means user can select the closer group of image so that they gate fast result. K-means is sensitive to noise and thus to get better result we will try k-Centroid clustering and for faster we use hierarchical clustering same time

VI. REFERENCES

[1] Dingyuan Xia et.al, "Trend of content –based images retrieval on the internet"2009 IEEE 5th international conference on image and graphics. Pp733-738.

[2] Mahip M.Bertere et. Al, "An efficient technique using Text & content base image mining techniques for image retrieval" 2012 International Journal of Engineering Research and Applications. Vol. 2, Issue 1, Jan-Feb 2012, pp.734-739

[3] S.Nancy Emymal,C.Yasubai Rubavathi , "spatially related Image Mining on Very Large image

Collections",Proceedings of ICETECT 2011 IEEE. Pp864-869.

[4] Shaikh Nikhat Fatma, Madhu nashipudimath, "Image mining using Association Rule," 2011 World Congress on Information and Communication Technologies, IEEE. pp 587-593.

[5] Jan-Ming Ho, Shu-Yu Lin, Chi-Wen Fann, Yu-Chun wang, Ray-I Chang "A Novel Content Based Image Retrieval System using K-means with Feature Extraction," 2012 International Conference on Systems and Informatics IEEE, Pp 785-790.

[6]S. P. Victor and S. John Peter (2010), "A Novel Minimum Spanning Tree Based Clustering algorithm for Image Mining," European Journal of Scientific Research, vol. 40, no. 4, pp. 540-546.

[7]Rajshree S. Dubey (2010), "Image Mining using Content Based Image Retrieval System", (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, pp. 2353-2356.

VII. BIBLIOGRAPHY

Gurpreet Kaur received her B.Tech degree in Computer Science & Engineering from the Institute of Engineering & Technology (Bhaddal)) under Punjab Technical University in 2010 and pursuing M.Tech (Regular) degree in Computer Engineering from Yadawindra College of Engineering Punjabi University Guru Kashi Campus Talwandi Sabo (Bathinda), batch 2011-2013. Her research interests include in improvement of Content Base Image Mining Technique for Image Retrieval Using Hybrid Clustering Mining in Image mining.