

A SURVEY ON CONTENT BASED IMAGE RETRIEVAL SYSTEM

M.Janani

Research Scholar, Department of Computer Science,
NGM College, Pollachi, India.
Email: m.jananict@gmail.com

Dr.R.Manicka Chezian

Assistant Professor, Department of Computer Science,
NGM College, Pollachi, India.

Abstract – Image mining is a vital technique which is used to mine knowledge from image. This is simply an expansion of data mining. The function of image mining is to retrieve similar image from huge database. The development of the Image Mining technique is based on the Content Based Image Retrieval system. Color, texture, pattern, shape of objects and their layouts and locations within the image, etc are the basis of the Visual Content of the Image and they are indexed. This paper presents a survey on various image mining techniques that are proposed earlier.

Keywords- Data Mining, Image Mining, Feature Extraction, Image Retrieval.

I. INTRODUCTION

Image mining are growing explosively and more frequently in our daily life. Image mining is a technique commonly used to extract knowledge directly from image. Image mining is a process to search significant images based on user input. This technique is mainly used to retrieve the similar image. A typical image mining process first extracts the image feature. The query image must be given as input, it compares with the image from database. Similar images are displayed as an output. This is mainly used for fast matching. The image retrieval system is categorized based on the type of searches, using either description of an image or its visual content.

II. RELATED WORK

Conci & Castro proposed a Content based image mining approach [1]. Image mining presents unique distinctiveness suitable to the richness of the data that an image can show. Successful assessment of the results of image mining by content requires

that the user point of view is used on the performance parameters. Comparison among different mining by resemblance systems is particularly challenging owing to the great variety of methods implemented to represent resemblance and the dependence that the results present of the used image set. Other obstacle is the lag of parameters for comparing experimental performance. In their paper they describe about an evaluation framework for comparing the influence of the distance functions on image mining by color. Experiments with color similarity mining by quantization on color space and measures of likeness between a sample and the image results have been carried out to illustrate the proposed scheme. Important aspects of this type of mining are also described.

Danzhou Liu, A. Hua, Khanh Vu, And Ning Yu [2] discussed a Fast query point movement technique for large content based image retrieval systems. This system refers to finding a specific image such as a particular registered logo or a specific historical photograph. This approach is able to reach any given target image with fewer iterations in the worst and average cases. In this system they propose a new index structure and query processing technique to improve retrieval effectiveness and efficiency. They also consider strategies to minimize the effects of users' inaccurate relevance feedback. Extensive experiments in simulated and realistic environments show that this approach significantly reduces the number of required iterations and improves overall retrieval performance. This approach can always retrieve intended targets even with poor selection of initial query points which is the main advantage of this approach.

Efficient relevance feedback for content based image retrieval by mining user navigation pattern is proposed by huang wei-jyun, su ja-hwung

and Vincent s [8]. This paper gives a detailed study about image retrieval and also specifies that approach used in it can be divided into two major operations namely offline knowledge discovery and online image retrieval. For online operation, once a query image is submitted to this system, the system first finds the most similar images without considering any search strategy, and then returns a set of the most similar images. Then, by using the navigation patterns, three search strategies, with respect to Query Point Movement (QPM), Query Reweighting (QR), and Query Expansion QEX, are hybridized to find the desired images. Overall, at each feedback, the results are presented to the user and the connected browsing information is stored in the log database.

Sanjay Silakari et al. [3] developed a structure which concentrates on color as characteristic using Color Moment and Block Truncation Coding (BTC) to obtain the features for image dataset. Subsequently K-Means clustering algorithm is conducted to group the image dataset into various clusters. In this paper an image is a spatial representation of an object and represented by a matrix of intensity value. It is sampled at points known as pixels and represented by color intensity in RGB color model. A basic color image could be described as three layered image with each layer as Red, Green and Blue.

Content based image retrieval using hierarchical and K-Means clustering technique is proposed by Murthy et al. [4]. In this paper image is given as input query and retrieves images based on image content. Content based image retrieval is an approach for retrieving semantically-relevant images from an image database based on automatically-derived image features. The unique aspect of the system is the utilization of hierarchical and k-means clustering techniques. This technique consists of two clustering algorithms, the hierarchical and the K-means clustering algorithms to group the images into clusters based on the color content. Both these clustering algorithms have been frequently used in the pattern recognition literature. Here we are going to filter most of the images in the hierarchical clustering and then apply the clustered images to K-Means, so that we can get better favored image results.

Rajendran & Madheswaran [5] discussed an improved image mining technique. An enhanced image mining technique for brain tumor classification using pruned association rule with MARI algorithm is presented in their paper. The method proposed makes use of association rule mining technique to

classify the CT scan brain images into three categories namely normal, benign and malign. Normal ones are those characterizing a healthy patient, benign cases represents CT scan brain images showing a tumor that are not formed by cancerous cells, and Malign cases are those brain images that are taken from patients with cancerous tumors. It combines the low-level features extracted from images and high level knowledge from specialists. The developed algorithm can lend a hand to the physicians for well-organized classification with multiple keywords per image to get better the accuracy. The experimental result on pre-diagnosed database of brain images showed 96% and 93% sensitivity and accuracy respectively.

Image mining technique using wavelet transform is projected by sanjay et al. [6]. It uses common pattern identical, pattern identification and data mining models with the intention that a real life scene/image can be associated to a particular category, support in different prediction and forecasting mechanisms. It is a three-step procedure i.e. image gathering, learning and classification. Since wavelet transform uses time frequency association, it can be utilized for image mining as a substitute of Fourier transform. Wavelet transform is utilized to decompose an image into dissimilar frequency sub bands and a small frequency sub band is used for Principal Component Analysis (PCA). Classification assists in recognizing the category to which an image relates with. They have constructed a prototype system for identification using DWT + PCA system. The conception of image mining as a consequence can be competently used for weather forecasting so that one can know the natural disasters that may occur in advance.

Rajshree [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.

III. IMAGE MINING USING CONTENT BASED IMAGE RETRIEVAL SYSTEM

Image mining is based on the color histogram and texture of image. The image is taken as input query and then the Color Histogram and Texture is taken and based on this the resultant Image is output.

A. FEATURE EXTRACTION

Several features are used in the Image Retrieval system. The popular amongst them are color features and texture features.

B. COLOR FEATURES

One of the features that are used is the color of the Image, since it has very strong co-relation with the underlying objects of the Image. Some of the techniques used are Average color in Gray scale, Average color in RGB format and Average color in YCBCR (Y is the luminance and CB, CR are the Chrominance components). The average color value is computed by

$$\text{Average color} = \frac{\sum (\text{intensity of all pixels in Current block})}{(\text{Total pixels in the block})}$$

The average values for the RGB components are calculated for all images

$$\text{Red average} = \frac{\text{Sum of all the red pixel in the image R (P)}}{\text{No. Of pixels in the image P}}$$

$$\text{Green average} = \frac{\text{Sum of all the Green Pixels in the image G (P)}}{\text{No. Of pixels in the image P}}$$

$$\text{Blue average} = \frac{\text{sum of all the Blue Pixels in the image B (P)}}{\text{No. Of pixels in the image P}}$$

Where R (P) = RED component pixels,

G (P) = GREEN component pixels,

B (P) = BLUE component pixels,

P =No. of pixels in the image.

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. Whenever the query image is given calculate the RGB components average values. Then compare this with the stored values.

C. SHAPE FEATURE

Shape also is an important feature for perceptual object recognition and classification of images. It has been used in conjunction with color and other features for indexing and retrieval. Shape description or representation is an important issue both in object recognition and classification. Many techniques including chain code, polygonal approximations, curvature, Fourier descriptors and moment descriptors have been proposed and used in various applications.

D. TEXTURE FEATURES

Texture is a very interesting image feature that has been used for characterization of images, with application in content based image retrieval. A major characteristic of texture is the repetition of a pattern or patterns over a region in an image. The Size, shape, color, and orientation of the textures can vary over the region. The difference between two textures can be in the degree of variation of the textures. Since basic texture patterns are governed by periodic occurrence of certain Gray levels.

IV. IMAGE RETRIEVAL

Instead of exact matching, content-based image retrieval calculates visual similarities between a query image and images in a database. Accordingly, the retrieval result is not a single image but a list of images ranked by their similarities with the query image. Many similarity measures have been developed for image retrieval based on empirical estimates of the distribution of features in recent years. Different similarity measures will affect retrieval performances of an image retrieval system significantly. Fig 1 represents the image retrieval process

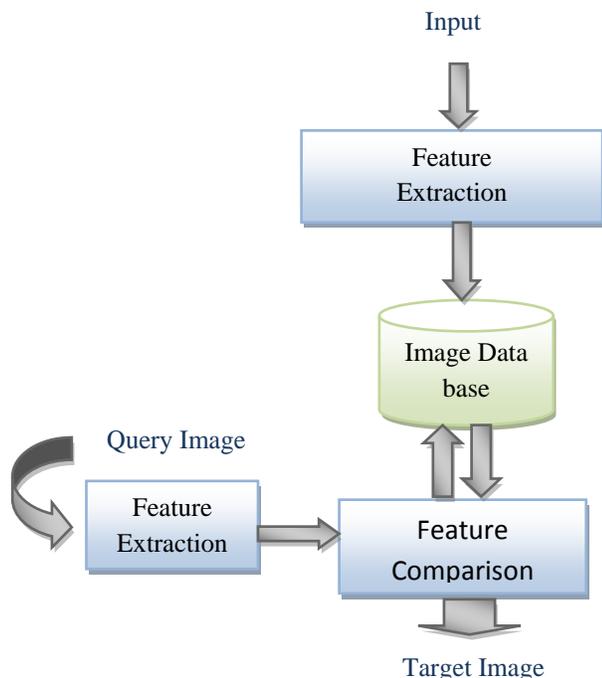


Fig 1: Image retrieval process

V. CONCLUSION

This paper presents survey on various image mining technique. The large number of research publication in the field of image mining especially in recent years shows that its very active and that it is starting to get more attention. This will hopefully advanced the field as new tools and technologies will be developed and performance will increase.

REFERENCES

- [1]. Aura Conci and Everest Mathias M. M. Castro (2002), "Image mining by content", Expert Systems with Applications, vol. 23, no. 4, pp. 377-383.
- [2]. Danzhou Liu, A. Hua, Khanh Vu, And Ning Yu (2010) , "Fast Query Point Movement Technique in large content based image retrieval system" , IEEE Transaction on knowledge and data engineering ,
- [3]. Sanjay Silakari, Mahesh Motwani and Manish Maheshwari (2009), "Color Image Clustering using Block Truncation Algorithm", IJCSI International Journal of Computer Science Issues, Vol. 4, No. 2.
- [4]. V.S.V.S. Murthy, E.Vamsidhar, J.N.V.R. swarup Kumar, P.Sankara Rao (2010), "Content Based Image Retrieval using Hierarchical and K-Means Clustering Techniques", International Journal of Engineering Science and Technology Vol. 2(3),, 209-212
- [5]. P. Rajendran and M. Madheswaran (2009), "An Improved Image Mining Technique For Brain Tumour Classification Using Efficient classifier", International Journal of Computer Science and Information Security, vol. 6, no. 3, pp. 107-116
- [6]. Sanjay T. Gandhe, K. T. Talele and Avinash G. Keskar (2007), "Image Mining Using Wavelet Transform", Knowledge-Based Intelligent Information and Engineering Systems, Springer link book chapter, pp. 797-803.
- [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.
- [8]. Huang wei-jyun, suja-hwung and Vincent s,(2011) "Efficient relevance feedback for content based image retrieval by mining user navigation pattern"IEEE Transaction on knowledge and data engineering ,