



A FOCUS ON IMAGE MINING TECHNIQUES FOR THE RETRIEVAL OF IMAGE BASED KNOWLEDGE

A.C.YAMINI, MCA, M.Phil.,
Guest Lecturer,
Department of Computer Science,
Arignar Anna Govt. Arts College for Women,
Walajapet, Vellore, Tamilnadu, India
ac_yamini1989@yahoo.co.in

Abstract: The Internet is the largest global multimedia data repository. Digital Pictures play an important role in Visual Communication. The access of large volume of multimedia data will be a major transforming theme in the global world. Every centralized and distributed database consist of images in either compressed or raw form. So the knowledge retrieval is a necessity. The current Internet technology needs more advanced data mining technologies to interpret the knowledge from the data distributed all over the world. Thus automatic mining of these data is mandatory in every field. The Image mining deals with the extraction of implicit knowledge about the image. The traditional mining technique includes object recognition, image retrieval, image indexing, image classification and clustering, and association rule mining based on exact matching of keywords. In current trend, the search is done not only through text but also via images. As it deals with spatial information, the mining algorithms need to be unique than in traditional data mining. This paper focus on the Content Based Image Retrieval technique which involves the retrieval of image using the visual features such as color, texture, size and shape.

Keywords: *Data mining, Image mining, Image features, Content based image retrieval*

I. INTRODUCTION

Data Mining plays a prominent role in the Internet. It refers to the knowledge that is discovered from the available data. The data is distributed everywhere in the form of text, images, etc. Every distributed database consists of images in either compressed or raw form. The access of large volume of multimedia data will be a major transforming theme in the global world. So the current Internet technology needs more advanced data mining technologies to interpret the knowledge from multimedia data. The extraction of knowledge from various databases involves an iterative process which includes understanding and analysing the application, Data Selection, Data pre-processing, Data Mining and Data Presentation. Searching information within images represents a special entity of data processing. Images as a unique category of data differ from text data in several aspects as in terms of their nature so in terms of storing and retrieving. Images have visual features, which can be represented in numerical form.

II. IMAGE MINING TECHNIQUES

Image is an important class of multimedia types. The Internet is the largest global multimedia data repository. Digital pictures play an important role in the Internet for Visual communication. Image

mining deals with the extraction of implicit knowledge, image data relationship, or other patterns which is not explicitly stored in the images. As it deals with spatial information, the mining algorithms need to be different than in traditional data mining. The basic techniques involved in data mining are Classification, Association, Clustering and Outlier Detection. [4] proposed a principle in which the division is the urgent stage in iris acknowledgment. We have utilized the worldwide limit an incentive for division. In the above calculation we have not considered the eyelid and eyelashes relics, which corrupt the execution of iris acknowledgment framework. The framework gives sufficient execution likewise the outcomes are attractive. Assist advancement of this technique is under way and the outcomes will be accounted for sooner rather than later. Based on the reasonable peculiarity of the iris designs we can anticipate that iris acknowledgment framework will turn into the main innovation in personality verification. In this paper, iris acknowledgment calculation is depicted. As innovation advances and data and scholarly properties are needed by numerous unapproved work force. Therefore numerous associations have being scanning routes for more secure confirmation strategies for the client get to. The framework steps are catching iris designs; deciding the area of iris limits; changing over the iris limit to the binarized picture; The framework has been actualized and tried utilizing dataset of number of tests of iris information with various complexity quality.

There are two main techniques used for image retrieval.

1. Text Based Image Retrieval
2. Content Based Image Retrieval

In text based image retrieval, the retrieval of image is based on matching the textual information to the annotation of the image. Then the relevant images are retrieved from the database. The images in the database are preprocessed and classified based on the data mining techniques. The steps involved in the retrieval of information in image is depicted in fig.1.

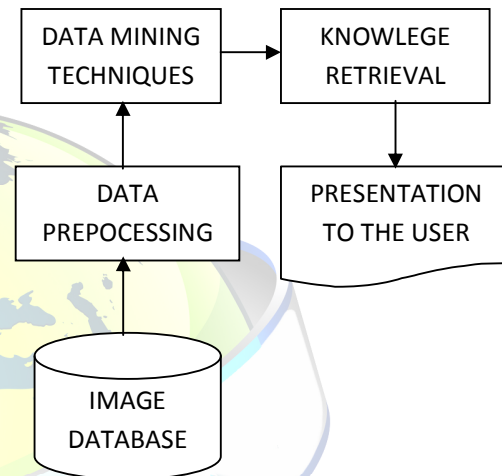


Fig. 1. Steps in Image mining Process

III.CONTENT BASED IMAGE RETRIEVAL TECHNIQUE

The Content Based Image Retrieval technique involves the retrieval of image using the visual features such as color, texture and shape. The Conventional image mining techniques includes Object recognition, image retrieval, image indexing, image classification and clustering and association rule mining. The retrieval problem is solved only by analyzing content of the image based on the available characteristics of its pixels. The steps involved in the CBIR process is presented in fig.2.

The modules used in this system are:

- Features Extraction



- Image Indexing
- Image Retrieval

1. Features Extraction

An object can be identified based on various features such as color, shape, texture, etc.

i. Color

The recognition of image is easily attained by colors. It can be represented by RGB (Red, Green and Blue) model, HSV (Hue, Saturation and Value) or HSB (Hue, Saturation and Brightness). The HSV model is better when compared to RGB model. The vector with three co-ordinates represents three colors.

Hue of a color represents the relative color appearance

Value indicates the darkness of the color

Saturation represents the strength of the color

Brightness refers to the brightness of the color

Color histogram has been found to be very effective in characterizing the global distribution of colors in an image, and it can be used as an important feature for image characterization. To define color histograms, the color space is quantized into a finite number of discrete levels. Each of these levels becomes a bin in the histogram. The color histogram is then computed by counting the number of pixels in each of these discrete levels. Using the color histogram, we can find the images that have similar color distribution based on the vectors.

ii. Texture

The characteristic of texture is the repetition of a pattern or patterns over a region in an image. The elements of

patterns are sometimes called textons. The size, shape, color, and orientation of the textons can vary over the region. The difference between two textures can be in the degree of variation of the textons. It can also be due to spatial statistical distribution of the textons in the image. Texture is an innate property of virtually all surfaces, such as bricks, fabrics, woods, papers, carpets, clouds, trees, lands, skin, etc. It contains important information regarding underlying structural arrangement of the surfaces in an image.

The texture properties are coarseness, contrast, directionality, line-likeness, regularity and roughness. Wavelet transforms have been applied in texture analysis and classification of images, based on multiresolution decomposition of the images and representing textures in different scales.

iii. Shape

Shape may be defined as the surface configuration of an object. Shape features vary as per the orientation of an object such as scaling, translation and rotation. It is determined based on the edge detection.

2. Image Indexing

In traditional database, the indexing refers to the access structure of the database files in terms organization of the records. The indexing mechanism is concerned with the process to assign terms to a document so that the document can be retrieved based on these terms such as phrases or keywords. In Content Based Image Retrieval technique, the descriptors of the multimedia data are extracted based on the feature vector of the data. These content descriptors are then organized into a suitable access structure for retrieval. Tree-based indexing techniques have been designed for traditional database



multimedia database indexing remains a challenge as it uses the image.

3. Image Retrieval

After indexing of images in the image database, it is important to use a proper similarity measure for their retrieval from the database. Similarity measures based on statistical analysis have been dominant in CBIR. Distance measures such as Euclidean distance, Mahalanobis distance, Manhattan distance, and similar techniques have been used for similarity measures.

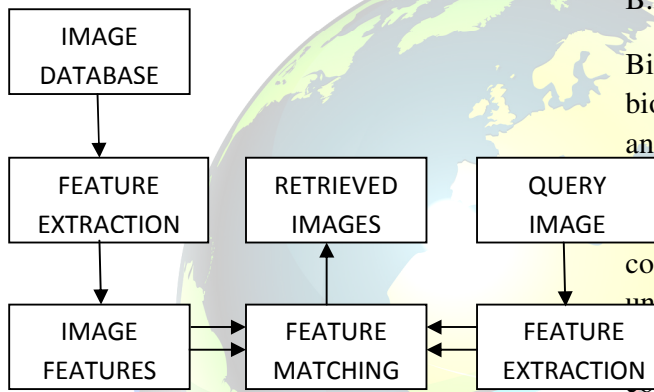


Fig.2. CBIR process

The important issues in CBIR technique are:

1. Selection of image database
2. Feature Extraction of an image
3. Similarity measurement
4. Performance evaluation

IV. APPLICATIONS

CBIR technique is widely used in various fields. Content-Based Image Retrieval has been used in several applications, such as medicine, biometrics, biodiversity information systems, digital libraries,

crime prevention, historical research, etc. Some of the applications mentioned are:

A. Medical Applications

The number of medical images produced by digital devices has increased more and more. The task of handling such data is time-consuming. Thus Content Based Image Retrieval system finds its application.

B. Biodiversity Information Systems

Biologists gather many kinds of data for biodiversity studies, including spatial data, and images of living beings. Ideally, Biodiversity Information Systems (BIS) should help researchers to enhance or complete their knowledge and understanding about species and their habitats by combining textual, image content-based, and geographical queries.

C. Digital Libraries

There are several digital libraries that support services based on image content. This digital library includes a module responsible for content-based image retrieval based on color, texture, and patterns. Geographical images can be provided as an input so that place names associated with retrieved images can be displayed by cross referencing with a Geographical Name Information System (GNIS) gazetter.

V. CONCLUSION

Image mining technique is widely used in the online networking. In online shopping, the search of an image is also done through



the images. The relevant images are retrieved based on the image that is given as an input. This paper presented a novel approach for Content Based Image Retrieval by combining the colour shape, and texture features. Similarity between the images is ascertained by means of a distance function. The retrieval of an image can be done efficiently by combining the textual information with the image content. Some more effective techniques need to be implemented for handling the multimedia data with better performance.

REFERENCES

- [1]. João Augusto da Silva Júnior, Rodiney Elias Marçal, "Image Retrieval: Importance and Applications", WVC 2014
- [2]. Barbora Zahradnikova, Sona Duchovicova and Peter Schreiber, "Image Mining: Review and New Challenges", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 7, 2015
- [3]. K. R. Yasodha, K.S. Yuvaraj," A Study on Image Mining Techniques, Innovation", International Journal of Applied Research ISSN: 2347-9272 (Volume-1, Issue-1) December 2013.
- [4]. Christo Ananth,"Iris Recognition Using Active Contours",International Journal of Advanced Research in Innovative Discoveries in Engineering and Applications[IJARIDEA],Volume 2,Issue 1,February 2017,pp:27-32.
- [5]. Preeti Chouhan , Mukesh Tiwari, "Image Retrieval Using Data Mining and Image Processing Techniques",

International Journal Of Innovative Research In Electrical, Electronics, Instrumentation And Control Engineering
Vol. 3, Issue 12, December 2015.

- [6]. Janani M and Dr. Manicka Chezian. R, "A Survey On Content Based Image Retrieval System", International Journal of Advanced Research in Computer Engineering & Technology, Volume 1, Issue 5, pp 266, July 2012.