International Journal of Enhanced Research in Science, Technology & Engineering
ISSN: 2319-7463, Vol. 5 Issue 4, April-2016

# Image Mining Techniques For Efficient Searching of Images from Internet

Aarigin Hazra[1], Krishnendu Bhattacharjee[2], Arnab Ghosh[3] , Tamasree Biswas[4]

[1,2,3,4] MAKAUT, Department Of Information Technology, Narula Institute of Technology, West Bengal, India

**ABSTRACT**

**Modern developments in the field of technology have made transmitting, storing, retrieval of huge amount of data much more affordable than ever before that too in an efficient way. Data mining techniques is a part of that revolution. Data mining simply refers to the extraction of hidden pattern information from large databases which help organization to focus on their important facts in databases. In the contrast of today's world, where digitalization is expanding in a rapid fashion, one important emerging topic in the context of data mining is image mining. Combining data mining with image processing we get a new field image mining. Image mining is focused on image data relationship which is not explicitly found in the images from databases or collections of images. The application of image mining in today's world is increasing day by day, such as in the fields of biotechnology, medicine, management and environmental control, telecommunications etc. The important function of the mining is to generate all significant patterns without prior information of the patterns. The purpose of this paper is to provide an overview of image mining and emphasizes on past work related in this field.**

**Keywords: Content based image retrieval (CBIR), Data mining, Image Mining, Image Pattern Design, Object Recognition.**

## 1. INTRODUCTION

The Computer Industry has seen a large growth in technology – access, storage and processing fields. This combined with the fact that there are a lot of data to be processed has paved the way for analyzing and mining data to derive potentially useful information. Various fields ranging from Commercial to Military want to analyze data in an efficient and fast manner. Particularly in the area of Multimedia data, images have the stronghold. However there is a general agreement that sufficient tools are not available for analysis of images [ZHL01]. One of the issues is the effective identification of features in the images and the other one is extracting them. One of the difficult tasks is knowing the image domain and obtaining a priori knowledge of what information is required from the image. This is one of the reasons the Image Mining process cannot be completely automated.

This research is concerned with the study and analysis of sonographic images, image mining methods of identifying appendicitis to improve the efficiency and effectiveness of the diagnosis and detection process of acute appendicitis using image mining on sonographic images. Image mining is more than an extension of data mining to image domain. It is an interdisciplinary endeavor that draws upon expertise from computer vision, image processing, image acquisition, image retrieval, data mining, machine learning, database and artificial intelligence. Advances in image acquisition and storage technology have led to tremendous growth in very large and detailed image databases. Analysis of images will reveal useful information to the human users. Image mining deals with the extraction of implicit knowledge, image data relationships or other patterns not explicitly stored in the images.

Image mining has led to tremendous growth significantly to large and detailed image databases. The most important areas belonging to image mining are: image knowledge extraction, content-based image retrieval, video retrieval, video sequence analysis, change detection, model learning and object recognition. Two different types of input data for knowledge extraction from an image collection are original image and symbolic description of the image. Discovering knowledge from data stored in alphanumeric databases, viz relational databases, has been the focal point of much work in data mining. However, with advances in secondary storage capacity, coupled with a relatively low storage cost, more and more non-standard data is being accumulated and one category of non-standard" data is image data. There is currently a very substantial collection of image data that can be mined to discover a new and valuable knowledge.

The central research issue in image mining is how to preprocess image sets so as to represent in a form that supports the application of data mining algorithms. Advances in image acquisition and storage technology have led to great growth

in very large and detailed image databases. A huge amount of image data is generated in our daily life as medical image like CT images, Cardiogram images, MR images, Mammogram images, X-ray images and Ultrasound Images etc. These images involve a great number of useful and implicit information that is difficult for users to discover. Image mining can automatically discover this implicit information and patterns from the high volume of images. Research in image mining can be broadly classified into two main directions. The first direction involves domain-specific applications where the focus is to extract the most relevant image features into a form suitable for data mining. The second direction involves general applications where the focus is to generate image patterns that may be helpful in understanding the interaction between high level human perceptions of images and low level image features.

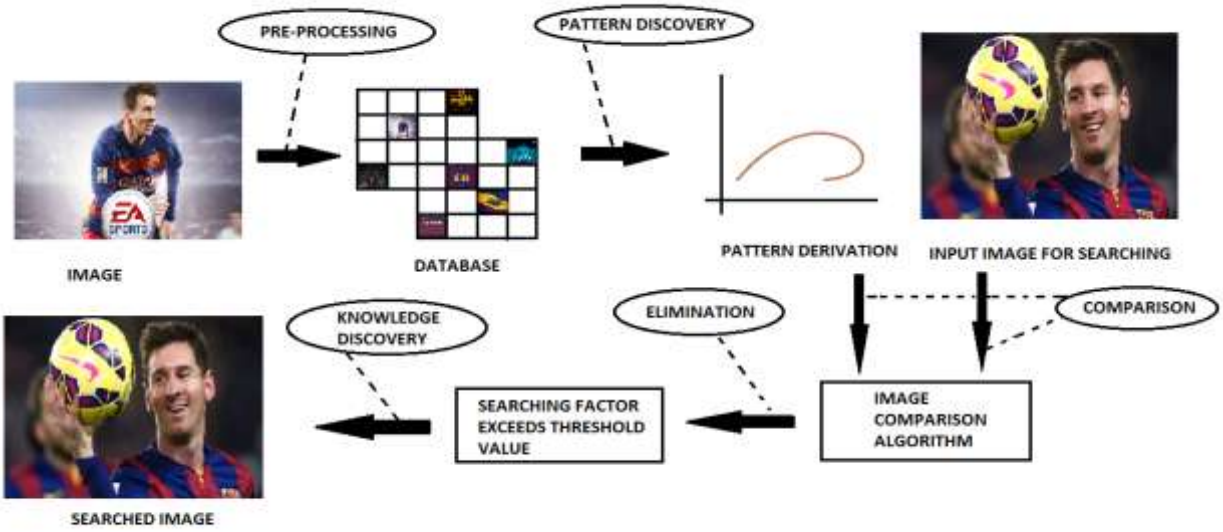


**Fig.1 Image Mining Process**

The fundamental challenge in image mining is to reveal the knowledge relating to the images from the web pages. In imaging science, image processing is processing of images using mathematical operations by using any form of signal processing for which the input is an image, a series of images, or a video, such as a photograph or video frame; the output of image processing may be either an image or a set of characteristics or parameters related to the image. The images stored in the database are used for pattern discovery after a common pattern is derived the input image and the pattern derived are compared using some image comparison algorithms if the searching factor exceeds the threshold level knowledge is discovered if not then eliminated.

## 2. LITERATURE SURVEY

1. There are two types of image mining systems one is function driven image mining frameworks and the other one is information driven image mining frameworks. Majority of the architectures of image mining falls under Function Driven Image Mining Frameworks these are application based and follow module functionality. Mihai Datcu and Klaus Seidel [1] proposed an intelligent satellite mining system that has two modules.

(a) First data is collected, preprocessed and archived which leads to the extraction of images which is further stored in form of raw images and then used for retrieval of images.
(b) An image mining system which gives user the chance to enquire into the image meaning and further leading to relevant events.

2. These has been quite a large advancement in the World Wide Web, large amount of data on many different fields has become available online. In an efficient and effective manner users have successfully retrieved images. Different techniques have been developed and used to solve the image retrieval problem on the basis of image features such as color, texture and shape. The technology used is known as Content Based Image Retrieval. It plays an important role in Image Indexing and Retrieval [2].

3. Object recognition has been an active research focus in field of image processing. Using object models that are derivable, an object recognition system find objects in the real world from an image. This is an important task in image mining. Machine learning and meaningful information extraction can only be realized when some objects have been identified and recognized by machine. The object recognition problem can be referred to as a supervised labeling problem based on models of known objects i.e. given a target image containing one or more interesting objects and a

set of labels corresponding to a set of models known to system, what is object recognition to assign correct labels to regions, or a set of regions, in the image [3].

After extracting objects are encoded as follows:

O1.Circle
O2.Square
O3.Diamond
O4.Star

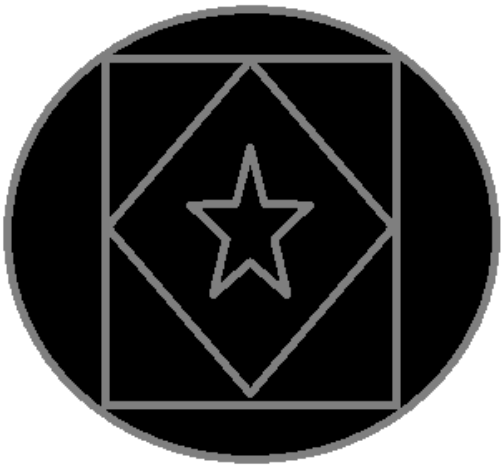

**Figure 2 Example of Image**

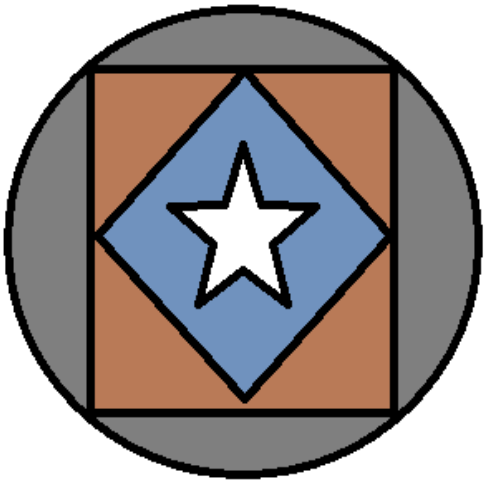

**Figure 3 Object extraction using edge detection**

[4] Yoshitaka A. and Ichikawa T [4] discussed about multimedia database storage system (MMDS). Emphasized on manipulating, retrieval of data through MMDS. Simple conventional way of retrieval is not adequate for managing multimedia data. This paper surveys recent studies on content-based retrieval for multimedia databases from the point of view of three fundamental issues.

[5] J. Borges and M. Lelene [5] Workshop on web usage analysis and user profiling discussed about the application of different data mining techniques and derive the hidden patterns from web content. Work includes text, image, video, hyperlink creation etc.

[6] In 2009,Kun-Cheet al[6] proposed a new technique in the field of image mining based on decision tree concept. Basically the actual idea was to encrypt personal information.

This information can be hidden in images and then we can develop a common framework depending on the decision tree for mining process and processing image data. In this concept, the pixel wise characteristic of the image is being changed into table format saved in a database. The concept is basically pixel wised image processing based on a new set of data created with respect to the images rather than being confined to the exact information.

[7] In recent years there has been a growing interest in developing effective methods for content-based image retrieval (CBIR) [7]. Image clustering and categorization is a means for high-level description of image content.

The goal is to find a mapping of the archive images into classes (clusters) such that the set of classes provide essentially the same information about the image archive as the entire image-set collection. In 2010, Victor and peter put forth a new minimum spanning tree based clustering algorithm for image mining. The clustering is done by creating a weighted graph from an image. It increases the efficiency of image mining even with the low variability image regions.

[8]In 2003, Yanai [8] described an image classification method where the knowledge about the images is being automatically acquired from the web. This comprised of three phases. Firstly, it collects the images related to particular keywords from the Web.

Secondly, information is being extracted from the images which are being gathered in the first phase, besides creating a relationship or a proper mapping among the images. Image attributes are being identified and matched with the respective classes. Lastly, in this phase the classification is done in actual which categorizes an image in an unrelated class proportionate to the keywords of that particular class by creating an association between the image and class attributes.

[9] Silakeri et al. (2009) [9] developed a structure which concentrated on color as the key attribute of an image using the Color Moment and Block Truncation Coding (BTC) to obtain the features of the image and fit them into respective classes of dataset . BTC is a general technique of compressing images by dividing the image into blocks and then using a quantiser lower the grey levels of the image while maintaining the same mean and standard deviation.

[10] Content based image mining approach was explained by Conci & Castro (2002)[10]. 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 (of likeness) 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. 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.

## CONCLUSION

Image mining is an extended section of Data Mining. We know that image provides a huge amount of information so if we can prevent the data loss and extract the vital information from images then it will be very helpful. It is a promising field for future works. Image mining is still in its grass-root level and needs lot of research works to be successfully implemented. Some vital issues that need to be checked are:

(A) Some dominant query language for image databases.
(B) Survey new image mining techniques considering the uniqueness of image data.
(B) Integrate new visualisation techniques that can be used for image pattern visualisation.

In this paper we have discussed about some early works in the field of image mining. We are looking forward to incorporate some new techniques by which we can mine images from web pages efficiently.

## 3. ACKNOWLEDGMENT

It is difficult to acknowledge the precious debt of knowledge and learning. But we can only repay it through our gratitude. We convey our sincere thanks to W. Strelilein , A. Waxman , M. Datcu , K. Seidel , Md. Farooque , Yoshitaka A. , Ichikawa T,J. Borges , M. Lelene,Lu Kun-Che , Yang Don-Lin , S.P.Victor , S.John Peter , Keiji Yanai , Sanjay Silakari , MaheshMotwani , Manish Maheshwari , Aura Conci , Everest Mathias M. M. Castro for their original work..

## REFERENCES

[1]. W. Strelilein and A. Waxman. Fused multi-sensor image mining for future foundation data. **Proceedings of the Third International Conference on Information Fusion (FUSION 2000)** ,Volume:1, 2000.
[2]. M. Datcu and K. Seidel. **Image information mining: exploration of image content in large archives. IEEE Conference on Aerospace, Vol.3, 2000.**
[3]. Md. Farooque, "**Image Indexing and Retrieval", DRTC Workshop on Digital Libraries: Theory and Practice March DRTC, 2003.**

[4]. Yoshitaka A. and Ichikawa T., "A Survey on Content-based Retrieval for Multimedia Databases". IEEE Transaction on Knowledge & Data Engineering, Vol 11, 1999, pp. 81-93
[5]. J. Borges and M. Lelene, "Data Mining of Use Navigation Pattern", in proceeding of WEBKDD 99,
[6]. Lu Kun-Che and Yang Don-Lin(2009),"Image Processing and Image Mining using Decision Trees", Journal of information science and engineering, vol. 25,no. 4,pp.989-1003
[7]. 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
[8]. Keiji Yanai (2003),"Web Image Mining toward Generic Image Recognition",World Wide Web Conference Series.
[9]. Sanjay Silakari,MaheshMotwani and Manish Maheshwari (2009),"Color Image Clustering using Block Truncation Algorithm",IJCSI International journal of Computer Science Issues, Vol. 4,No. 2.
[10]. Aura Conci and Everest Mathias M. M. Castro (2002), "Image mining by content", Expert Systems with Applications, vol. 23, no. 4, pp. 377-383