Review of Different Existing Image Mining Techniques

Prabhjeet Kaur
Student Masters of Technology
Department of CSE
Sri Guru Granth Sahib World University
Fatehgarh Sahib, Punjab, India

Kamaljit Kaur
Assistant Professor
Department of CSE
Sri Guru Granth Sahib World University
Fatehgarh Sahib, Punjab, India

Abstract: In the area of Data Mining ,Image Mining technology has been considered an advanced field for discovering information related to the images.Image Mining is a process of extracting knowledge concerning images.The demand of image mining increases as the need of image data is growing day by day.There are many techniques developed in the earlier researches and eventually these techniques can reveal useful information according to the human requirements,but Image Mining still require more development especially in the area of web images.Now-a-days Image Mining is one of the major technology used in the world of internet.There is a need of effective and efficient methods to be encounter for obtaining the information of web images.
This paper presents a survey on various image mining techniques.It also provides a improvements for future research.

Keywords: Image Mining, Data Mining, Image Retrieval, Image Indexing, Object Recognition, Image Classification and Clustering, Association Rule Mining.

I.   INTRODUCTION: Image Mining

Image Mining is an extended branch of data mining that is concerned with the process of knowledge discovery concerning images**.**Image Mining deals with the extraction of image patterns from a large collection of images. In Image Mining, the goal is the discovery of image patterns that are significant in a given collection of images [4]. Image Mining deals with extraction of knowledge, image data relationship and other required patterns and uses ideas from image processing, image retrieval and machine learning, databases [6].The focus of image mining is on the extraction of knowledge patterns from a large collection of images.While there seems to be some overlap between image mining and content-based retrieval (since both deal with large collections of images), image mining goes beyond the problem of retrieving relevant images. In image mining,the goal is to discover image patterns that are significant in a given collection of images and the related alphanumeric data [2]. The fundamental challenge in image mining is to reveal out the knowledge relating to the images from the web pages.

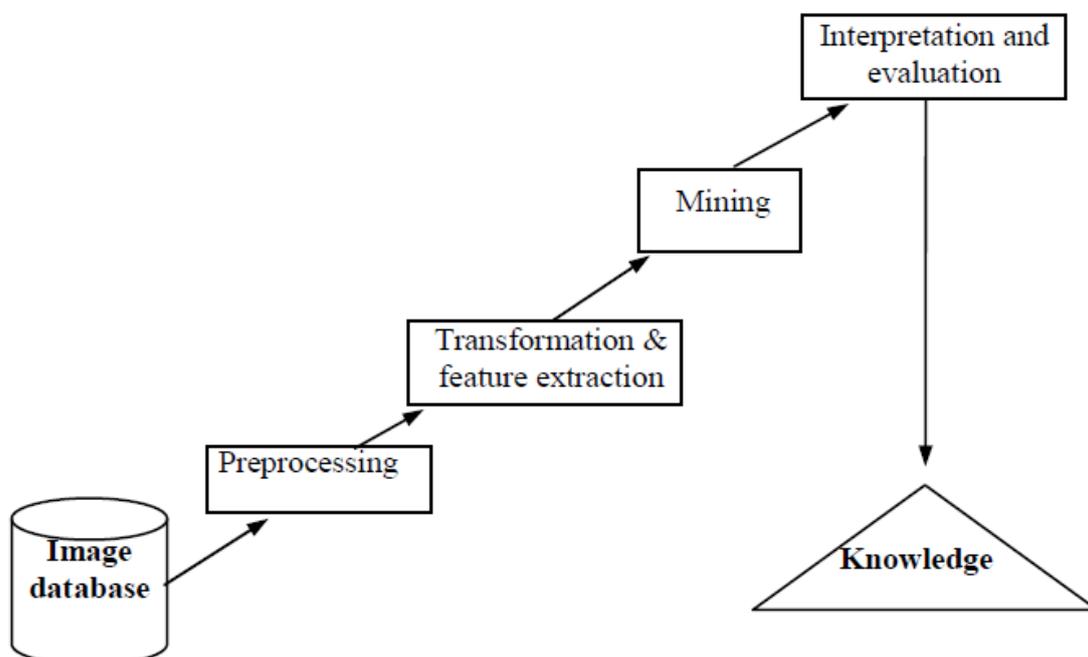Fig. 1 illustrates the typical image mining process.

Fig.1 Image Mining Processs[5]

1.1 Preprocessing:

In image data, the spatial segmentation can be done at region and/or edge level based on the requirements of the application. It can be automatic or manual and should be approximate enough to yield features that can reasonably capture the image content.

1.2 Feature Extraction and Transformation:

Color, edges, shape, and texture are the common image attributes that are used to extract features of mining.Feature extraction based on these attributes may be performed at the global or the local level. There are obvious trade-offs between global and local descriptors.Global descriptors are generally easy to compute, provide a good representation, but they tend to integrate and therefore are often unable to discover subtle patterns or changes in shape. Local descriptors, on the other hand, tend to generate more elaborate representations and can yield useful results even when part of the underlying attribute[5].

II.   Image Mining vs. Data Mining

The most common misconception of image mining is that image mining is nothing more than just applying existing data mining algorithms on images. This is certainly not true because there are important differences between relational databases versus image databases. The following are some of these differences:

(a) Absolute versus relative value. In relational databases, the data value are semantically meaningful. For example, age is 37 is well understood. However, in image databases, the data value themselves may not be significant unless the context supports them. For example, a grey scale value of 48 could appear darker than a grey scale value of 88 if the surrounding context pixels values are all very bright.

(b) Spatial information (Independent versus dependent position):Another important difference between relational databases and image databases is that the implicit spatial information is critical for interpretation of image contents but there is no such requirement in relational databases

(c) Unique versus multiple interpretations: A third important difference deals with image characteristics of having multiple interpretations for the same visual patterns. A new class of discovery algorithms is require to consider to the special needs in mining useful patterns from images [15].

III.   Existing Techniques of Image Mining

Besides investigating suitable frameworks for image mining, early image miners attempted to use existing techniques to mine image information. The image mining techniques include object recognition,,image retrieval, image classification ,image clustering, association rules mining, and neural network. We will explain these techniques and how they have been applied to image mining in the following subsections:

To perform Image Mining techniques select a collection of images belonging to the same field(E.g. geographical images)

Figure 2. Example of image

3.1 Object Recognition

Object recognition has been an active research focus in field of image processing. Using object models that are known a priori, an object recognition system find objects in the real world from an image. This is a major 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 referr 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 [15].
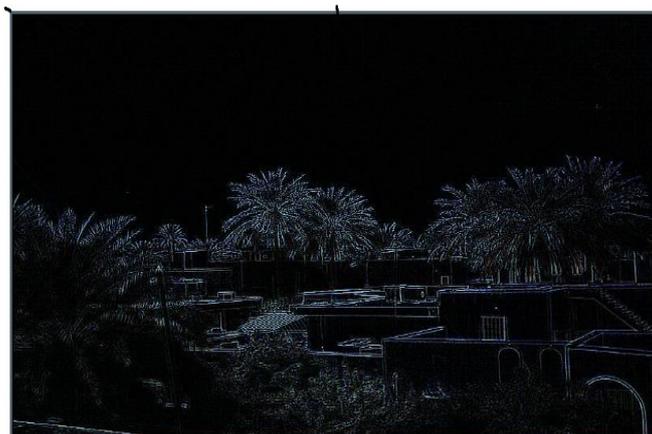
Fig.3 Object extraction using edge detection

After extracting object are encoded it as follows:
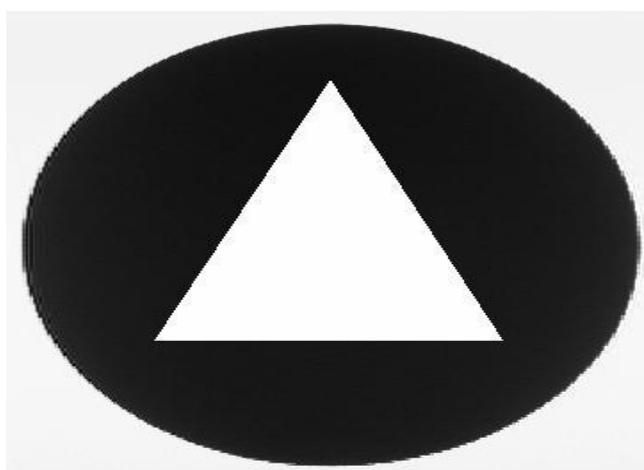O1: circle.
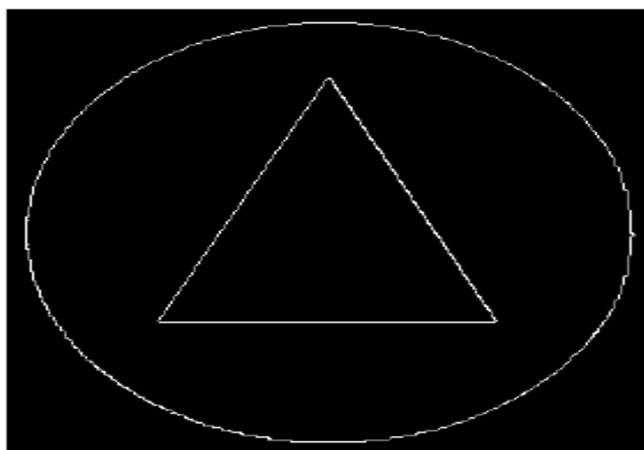O2: triangle.
O3: square.

Fig.4 Example of an image

Fig.5.Object extraction using edge detection

3.2 Image Retrieval
Image mining requires that images be retrieved according to some requirement description. The requirement description can be classified into three levels of increasing complexity [12]:
(a) Level 1 comprises image retrieval by primary features such as color, texture and shape .
(b) Level 2 comprises image retrieval by derived or logical features like objects of a given type or individual objects or persons.
(c) Level 3 comprises image retrieval by obtaining attributes, involving a significant amount of high level reasoning about the meaning or purpose of the objects or scenes depicted.

Text based Image Retrieval

The images are indexed and retrieved based on the descriptions such as their size, type, date and time of capture, identify of owner, keywords description of the image. This is often called description based or text based image retrieval process. The text based descriptions of the images are typed manually for each image by human descriptors, because the automatic generation of keywords for the images is difficult without visual information and feature extraction. Based on these contents, desirable image features can be mined and used as index. The keyword from dictionary of the images and on the basis of these keywords searches the requested images. Image came in the form of text stores in the data dictionary and here finds the keywords are available in the data dictionary then it shows the number of specific images,if data not find then it shows a message the word does not match. The user's has given input text the images are displayed based on their relevance feedback[8].

Query Based Image Retrieval

The query image can be extracting the visual features and can be compared to find matches with the indices of the images stored in the database and these features are used to retrieve the similar images from the image database. In the comparing of two images, resemblance of the visual features of the query image is measured with the features of each image. The similarities of two images are measured by computing the distance between the feature vectors of two images the retrieval system returns the k images. Several image characteristics have been used to index images for content based image retrieval system. Query gives to this system with an image and then search is based upon search algorithms which vary according to its application; result images must all share elements.Image which is already existed may be supplied by user or taken from stored Database[8].

3.3 Image Indexing

The objective of image indexing is to retrieve similar images from an image database for a given query image (i.e., a pattern image). Each image has its unique feature. Hence image indexing can be implemented by comparing their features, which are derived from the images. The basis of similarity among images may be based on the features such as color, intensity, shape, position and texture, and above mentioned other image attributes.[9]
 Image indexing techniques are of two types:
1. Textual (manual)
2. Content- based (automated)

1. Textual

   It is very simple techniques; keeping in mind the user approach keywords are given for a specific image. These includes
- Caption indexing
- Keyword additions
- Standard content titles, Classification, etc.

The difficulty with this indexing is that it is
- Labor intensive
- More subjected to inter indexer consistency difficulties than indexing of text
- Of-ness, thing-ness, about-ness ambiguities

2. Content-based

In this technique images are indexed based on their content like color, shape, direction and texture etc. This kind of indexing is taken care by software itself, algorithms are arised which can distinguished the color, shape, textures etc. The image retrieved through this technique is known as Content Based Image Retrieval(CBIR).
Research and development issues in CBIR cover a range of topics, many shared with ordinary image processing and information retrival. Some of the most important are:
- understanding image users requirements and information-finding behavior
- recognition of suitable ways of explaining image content
- retrieving such features from unprocessed images
- provide storage area for large image databases
- combine query and stored images in a way that reflects human similarity discrimination
- efficiently access stored images by content
- providing usable human interfaces to CBIR systems[9]
-

3.4 Image Classification and Image Clustering

Image classification are the supervised and unsupervised classification of images into groups.
In supervised classification,give a collection of labeled (pre classified) images, and the problem is to label a newly encountered, yet unlabeled images.
In unsupervised classification (or image clustering), the problem is to group a given collection of unlabeled images into meaningful clusters according to the image content without a priori knowledge [2].

Image Texture Classification:
The texture represents the energy content of the images. If an image contains high textures, then the energy will be high as estimated to that of average and low texture images.So when combining the energy values specified for a local patch of an image the values will be high for highly textured areas and will be low for simple areas. Also the local areas of same kind of textured areas will approximate same energy level. So it can be known as "Texture Activity Index". If it is capable to fit the power values into any distribution then the classification of images into High, Average and Low description of images can be easily and effectively done because the statistical parameters of the respective distribution will be different for all the three categories as because they possess different energy levels[13].

Image Clustering:
Clustering will be more advantage for reducing the searching time of images in the database.There are a variety of clustering methods:hierarichal,partitioning,density-based,grid based and fuzzy clustering methods.Fuzzy C-means (FCM) is one of the clustering methods used frequently in image mining which allow one piece of data to connect to two or more than two clusters. In this clustering, each point has a degree of connecting to clusters, as in fuzzy logic, rather than connecting completely to just one cluster. Thus, points on the edge of a cluster can be in the cluster to a lesser degree than points in the centre of cluster. FCM groups data in particular number of clusters[13].

3.5 Association Rule Mining:
Association rule mining is frequently used in data mining to uncover interesting trends, patterns and rules in large datasets. Recently, association rule mining has been applied to large image databases [4]. Although the current image association rule mining approach is far from mature and perfection compared its application in data mining field, there opens up a very promising research direction and vast room for image association rule mining. There are two main methods:The first method is to mine from large collections of images alone, and the second method is to mine from a combined collection of images and associated alphanumeric data . A typical example of the first kind of association mining of image is to find if there is some pattern existing for an individual city or between different cities by studying a collection of satellite imagery of various cities of the United States. An example of the second case may involve medical imagery and patient records. Image data and patient records can be viewed together to find interesting associations[4].

The association rule mining is an appropriate tool for pattern detection in knowledge discovery and data mining. Its objective is to remove useful information from very large databases. By using rules extracted from images, the content of images can be suitably analyze, and the information require for image classification can be obtained.

The association rule can be represented by an expression $X \Rightarrow Y$, where X and Y can be any discrete entity. As we discuss image database, X and Y can be some feature elements mined from images. The meaning of $X \Rightarrow Y$ is: Given an image database D, for each image $I \in D$, $X \Rightarrow Y$ expresses that whenever an image I contains X then I will probably holds Y. The support of association rule is defined as the probability $p(X \subseteq I, Y \subseteq I)$ and the confidence of association rule is defined as the conditional probability $p(X \subseteq I \mid Y \subseteq I)$. A rule with support bigger than a specified minimum support and with confidence bigger than a specified minimum confidence is considered as a significant association rule[16].

3.6 Neural Networks
Neural Networks is one of the technique used in image processing and image retrieval. Neural Networks are computational systems made up of simple processing units called neurons which are usually organized into layers with fully or partially connections. The main task associated with a neuron is to receive the activation values from its neighbors (the output of other neurons), compute an output based on its weighted input parameters and send that output to its neighbours. Artificial neural network models have been studied for many years in the hope of achieving humanlike performance in several fields such as speech and image understanding.A neural network, by definition, is a massively parallel distributed processor made up of simple processing units, each of which has a natural tendency for storing experiential knowledge and making the knowledge available for use. Neural networks are fault tolerant and are good at pattern recognition and trend prediction. In the case of limited knowledge, artificial neutral network algorithms are frequently used to construct a model of the data. There are some key differences in the way conventional programs and neural networks work. The former require programming whereas neural networks are train by using of train data. The conventional programming uses serial processing, while neural networks use parallel processing[18].

IV. Literature Survey

Nishchol Mishra et al.(2012)[13]The emergence and proliferation of social network sites such as Facebook and Linkedin and other multimedia networks such as Flickr has been one of the major events of this century.The networks have acquired immense popularity and have become a part of the daily lives of millions of people. Many of these network sites are thus extremely rich in content, and contain a tremendous amount of multimedia content waiting to be mined and analyzed. Analyzing this huge amount of multimedia data to discover useful knowledge is a challenging task. It has opened up opportunities for research in Multimedia Data Mining (MDM). Multimedia Data Mining can be defined as the process of finding interesting patterns from media data such as audio, video and text that are not ordinarily accessible by basic queries and associated results. This paper mainly focused on Image Mining techniques and how Content-based Image Retrieval can be helpful for Image Mining.

Mahip M.Bartere et al.(2012)[8] Image mining presents special characteristics due to the richness of the data that an image can show. Effective results of image mining by content requires that the user point of view is used on the performance parameters. Comparison between different mining by similarity systems is particularly challenging owing to the great variety of methods implemented to represent likeness and the dependence that the results present of the used image set. In this paper we proposed an evaluation framework for comparing the influence of the distance function on image mining by color and also a way to mine an image from its name. 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 .

Shaikh Nikhat Fatma(2012)[20] The goal of this paper is to discover the image patterns that are significant in agiven collection of images.The basic challenge is to find low level pixel representation contained in a unprocessed image can be efficiently and effectively processed to identify high level features.Overview of image mining frameworks is also discussed briefly.The framework models represent the information presents in the image data and figure out the issues and challenges of finding useful patterns/knowledge belonging to the data.

Ji Zhang et al.(2012)[5] The paper highlights the need of image mining as the demands of image data is increasing day by day.The analysed images give useful information according to the user needs.Image Mining deals with the extraction of knowledge, relationship between image and data and other required information stored in the images.Image Mining is advanced field of data mining.Many algorithms are implemented in research of image mining but it is still required further development in this area.This paper examined the research issue and current developments in Image Mining and also identifies future research directions in Image Mining.

Md. Farooque(2003)[9] With the advancement in the world wide web,a large amount of data on many different felds has become available online.User retrieved images by an efficient and effective manner.Many techniques have been developed to solve the image retrieval problem on the basis of image features such as color,texture and shape.These technology called as Content Based Image Retrieval.It plays an important role in Image Indexing and Retrieval.

V. CONCLUSION

Image Mining is the advanced field of Data Mining technique .The main objective of the Image Mining is to remove the data loss and extracting the meaningful information to the human expected needs.

In summary, Image Mining is a promising field for research. Image Mining research is still in its infancy and many issues remain solved. Specifically, we believe that for Image Mining research to progress to a new height, the following issues need to be investigated:

(a) Design semantically powerful query languages for image databases.
(b) Explore new discovery techniques that take into account the unique characteristics of image data.
(c) Incorporate new visualization techniques for the visualization of image patterns.

In this paper, we have discussed various techniques that are frequently used in the early works in image mining.Our goal will be to make a packing plan such that images must be mined in regular and efficient manner from web pages.

REFERENCES

[1]   A.Kannan, Dr.V.Mohan and Dr.N.Anbazhagan,"An Effective Method of Image Retrieval using Image Mining Techniques", The International journal of Multimedia & Its Applications (IJMA) Vol.2, No.4, November 2010.
[2.]   A. K. Jain, M. N. Murty and P. J. Flynn."Data Clustering: A Review, ACM computing survey",Vol.31, No.3, September 1999.
[3]   B. C. Ooi, K. L. Tan, C. Yu and S. Bressan. Indexing the Edges - A Simple and Yet Efficient Approach to High-Dimensional Indexing, 19th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, pp. 166-174, Dallas, Texas, May 2000.
[4]   C. Ordonez and E. Omiecinski,"Discovering Association Rules Based on Image Content",01Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99), 1999.
[5]   Ji Zhang Wynne Hsu Mong Li Lee," Image Mining: Trends and Developments",School of Computing,Singapore,2012.
[6]   J. Salton, and M. J. McGill,"Introduction to Modern Information Retrieval",McGraw-Hill Book Company,1983.
[7]   K. Lin, H. V. Jagadish and C. Faloutsos. The TV-tree: An Index Structure for High- Dimensional Data. The VLDB Journal, 3 (4): 517-542,1994.
[8]   Mahip M.Bartere and Dr.Prashant R.Deshmukh"An Efficient Technique using Text & Content Base Image Mining Technique for Image Retrieval",International Journal of Engineering Research and Applications (IJERA) Vol. 2, Issue 1, pp.734-739,Jan-Feb 2012.
[9]   Md. Farooque,"Image Indexing and Retrieval", DRTC Workshop on Digital Libraries: Theory and Practice March DRTC,2003.
[10]   Michael Gabay,Sofia and Zaourer,"Variable Size Vector Bin Packing Heuristics",springer,2013.
[11]   M. Antonie, O.R. Zaiane, A. Coman. Application of Data Mining Techniques for Medical Image Classification. In Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), San Francisco, CA, USA, August, 2001.

[12]   M. C. Burl et al. Mining for Image Content. In Systemics, Cybernetics, and Informatics /Information Systems: Analysis and Synthesis, (Orlando, FL), July 1999.

[13]   Nishchol Mishra and Dr. Sanjay Silakari "A Perspective Image Mining in the Context of Content Based Image Retrieval" IJCSI International Journal of Computer Science Issues, Vol. 9, Issue 4, No 3, July 2012.

[14]   N. Katayama and S. Satoh. The SR-tree: An Index Structure for High-dimensional Nearest Neighbor Queries. In proceedings of the 1997 ACM SIGMOD Conference, pages 369-380,Tucson, Arizona, May 1997.

[15]   Prof. Dr.Hilal M. Yousif, *Dr.*Abdul- Rahman Al-Hussaini, Mohammad A. Al-Hamami," Using Image Mining to Discover Association Rules between Image Objects".

[16]   P. Rajendran , M. Madheswaran," Novel Fuzzy Association Rule Image Mining Algorithm for Medical Decision Support System", International Journal of Computer Applications ,pp.0975 − 8887,Volume 1 − No. 20,2010.

[17]   S. Gibson Intelligent Mining in Image Databases, with Applications to Satellite Imaging and Web Search, Data Mining and Computational Intelligence, Springer-Verlag, Berlin, 2001.

[18]   S. Berchtold, D. A. Keim and H. P. Kriegel. The X-tree: An Index Structure for Highdimensional Data. In Proceedings of the 22nd VLDB Conference, pages 28-39, Mumbai, India, September 1996.

[19]   S. Haykin. Neural Networks: A Comprehensive Foundation. Prentice Hall, 1999.

[20]   Shaikh Nikhat Fatma ," Image Mining Method and Frameworks",International Journal of Computational Engineering Research (ijceronline.com)Vol. Issue. 8,2012.

[21]   T.Y.Gajjar,N.C.Chauhan ,"A review on image mining frameworks and techniques" International journal of computer science and information technologies,vol3(3),2012.