International Journal of Scientific Research & Engineering Trends
Volume 4, Issue 6, Nov-Dec-2018, ISSN (Online): 2395-566X

A Study on Image Mining Tools Techniques and Framework

Assistant Professor C. Rukmani
Department of Computer Science
Adthiyaman Arts and Science College for Women
Uthangarai, Krishnagiri, India
rukmanicb@gmail.com

Abstract- **Image mining is a vital method which is used to mine knowledge from image. Image segmentation is the primary phase in image mining. Image mining handles with lot of unknown information extraction, image data association and additional patterns which are not clearly accumulated in the images. Image data represents a keystone of many research areas including medicine, forensic criminology, robotics and industrial automation, meteorology and topography as well as education. Therefore, obtaining specific information from image databases has become of great importance. Images as a special kind of data differ from text data as in terms of their nature so in terms of storing and retrieving. Image mining as a research field is an interdisciplinary area combining methodologies and knowledge of many branches including data mining, computer vision, image processing, image retrieval, statistics, recognition, machine learning, artificial intelligence etc. This paper aims at reviewing the current state of the IM as well as at describing challenges and identifying directions of the future research in the field.**

Keywords- **image mining, image classification, indexing, image retrieval etc.**

I. INTRODUCTION

Due to the massive research and development of the recent years, the lack of information has not been an issue in the most fields of human movement. On the contrary, besides new technology, there is a huge volume of data available for people. Therefore, sorting the data and obtaining specific information from databases has become of great significance. In the last decade, data mining as a research field has expanded and progress in data processing is getting both more accurate and convenient. Besides text data mining; novel data mining algorithm web mining and social network analysis, image mining belongs to the spheres of interest. Analyzing image data forms a keystone of many research areas including medicine (evaluating MRI, interpreting X-rays/CT scans), forensic criminology (fingerprint identification, face recognition), robotics and industrial automation (robotic vision), meteorology and topography (satellite imagery) as well as education (computer-aided visualization) and many other fields.

Image mining is a technique which handles the mining of information, image data association, or additional patterns not unambiguously stored in the images (Zhang Ji et al. 2001). It utilizes methods from computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence. Rule mining has been implemented to huge image databases (Ordonez & Omiecinski 1999). There are two most significant techniques. The first technique is to mine from huge amount of images alone and the second technique is to mine from the integrated collections of images and related alphanumeric data. Rule mining technique is exploited (Megalooikonomou et al. 1999) to determine relations between structures and functions of human brain. An image mining algorithm using blob required to be carry out the mining of relations within the context of images is provided by Zaiane & Han (1998).

The main intention of image mining is to produce all considerable patterns without any information of the image content, the patterns types are different. They could be classification patterns, description patterns, correlation patterns, temporal patterns and spatial patterns. Image mining handles with all features of huge image databases which comprises of indexing methods, image storages, and image retrieval, all regarding in an image mining system (Missaoui & Palenichka 2005).

The establishment of an image mining system is frequently an intricate process because it implies joining diverse techniques ranging from image retrieval and indexing schemes up to data mining and pattern recognition. Further, it is anticipated that a good quality image mining system provides users with a

useful access into the image storage area at the same time it recognizes data patterns and generates knowledge beneath image representation. Such system basically be supposed to bring together the following functions: image storage, image processing, feature extraction, image indexing and retrieval and, pattern and knowledge discovery.

II. IMAGE MINING FRAMEWORK

Early work in image mining has focused on developing a suitable framework to perform the task of image mining. Normally an image database containing raw image data cannot be directly used for mining purposes. Therefore raw image data has to be first processed to generate the information usable for high-level mining modules. Image mining provides a framework that uses the raw format images stored in the database which cannot be used directly. To use them in high-level modeling they must be processed first. An image mining technique is considered as a good technique if it supports fully user interaction during retrieving the patterns and knowledge from the collection of huge image (Bach et al., 1996) database. The following functions are performed in image mining, they are: image storage, image processing, feature extraction, image indexing and retrieval, patterns and knowledge discovery.

The two kinds of frameworks of image mining are

- Function driven framework: which focused on different modules component and their functionalities?
- Information driven framework: that provided a hierarchical structure of levels and the data needed into all the levels.

III. REVIEW ON IMAGE MINING

Fay et al. (2003) developed a system for multisensory image fusion and interactive mining. This system was dependent on neural models of colour vision processing, learning and pattern recognition. They also had add-on modules which performed image conditioning, image fusion, extraction of context features and interactive image mining. All of these modules were combined together to create a work flow which enabled a user to create vector products of foundation features (e.g., roads, rivers, and forests).

They also highlighted the target detections from raw multisensory or multispectral imagery. In this image mining technique, multispectral imagery modified by virtual ecological conditions was not addressed. In order to process large amounts of remote sensing image data, Daschiel et al. (2005) developed the prototype model of information mining system. It consisted of both an online interface as well as an offline part. The offline part dealt with the generation of features relating to image mining like data reduction, compression, unsupervised content index and the absorption of catalogue entry.

Users can collect information from a vast amount of data which is present in the WWW on demand by using various image mining techniques. The digitalized image which is obtained from the web is relevant in the real world. Sometimes real world images can differ from the obtained results due to the various classification/recognition techniques used. Morsillo et al. (2008) proposed a technique which provided more accurate visualization of objects by reducing the noisy search. This model united both generative and discriminative elements to perform an efficient retrieval of web images. It successfully worked on semi supervised machine learning technique. Zhan et al. (2009) devised the relation between the two main characteristics of web image i.e., visual feature cluster and keyword, using multi-mode association rule.

1. Image mining techniques

The techniques which were used by early image miners prior to the invention of suitable framework include pattern recognition, image indexing and retrieval, image classification, image clustering, association rule mining, and neural network. In the following, is a survey on these techniques? The techniques are classified on five levels of information and the associated image or data mining operations. These levels (from top to bottom) are:

- Knowledge extraction level
- patterns and inter-image relations level
- semantic concept level
- region, objects, or visual patterns level
- E .pixel level.

2. Object recognition

One of the key areas of image mining is object recognition, which operates data on patterns and inter-image relations level. It finds the object relevant to the real world, from the image by processing the provided object models. It is also known as supervised labeling method. The system has four parts, they are:

- feature detector
- model database
- hypothesizer
- Hypothesis verifier.

3. Image retrieval

Image retrieval refers to the process of retrieving a particular image from a large database using data mining. Retrieval of images in image mining (Tahoun et al., 2005) is done based on some requirement specification. There are three levels of requirement specifications and the complexity also increase with the levels. a level 1 retrieve the image based on some basic features of images such as texture, colour, shape

or image elements' spatial location b level 2 is based on image retrieval which derives the logical features such as individual objects or persons from images c level 3 is based on image retrieval by abstract attributes which involves a high level reasoning in order to obtain the meaning of the objects or scenes illustrated.
Kazman and Kominek (1993) introduced three query schemas to retrieve image information. They were

- Query by description.
- Query by associate attributes.
- Query by image content.

Query by associate attributes refers to the technique of taking the conventional table structure to tailor which fulfils the purpose of image needs. Query by description means the method that uses description along with each image, through which the user can locate the images interested. The image description is often referred as label or keyword. With the emergence of large-scale image repositories, the problems of vocabulary and non-scalability caused by manual operation have become more pronounced. Hence, content-based image retrieval (CBIR) was proposed to overcome these difficulties. IBM's QBIC system (Flickner et al., 1995) could retrieve image description by any combination of colour, texture and shape as well as text keyword. This system may be one of the popular systems amongst all other image content retrieval frameworks. It uses R*-tree indexes to improve efficiency.

Image indexing Apart from focusing on the information requirements at various levels, it is also important to provide support for the retrieval of image data with a fast and efficient indexing scheme. On the contrary, the image database to be searched is too large and the feature vectors of images are of high dimension which increases the search complexity. Image indexing handles data and images in region, objects and visual patterns level. Reducing the dimensions can be accomplished using two well-known methods.

- The singular value decomposition (SVD) update algorithm.
- Clustering.

Although, the best way to reduce complexity is to perform appropriate multi-dimensional indexing after performing dimension reduction, which provides non-Euclidean (Rui and Huang, 1997) similarity measures. Lin et al. (1994) introduced an efficient technique of colour indexing for retrieving similar type data. In this work, they increased the search time as the size of the database increased. In 2001, Tan et al. (2001) proposed a multi-level nested R-tree index which retrieved the structure efficiently and effectively.

It helped to select appropriate technique and also helped to design new technique by prolife the retrieval process. This process helped to evaluate the performance of colour-spatial retrieval techniques, which led to the selection of a suitable new technique.

IV. IMAGE CLASSIFICATION AND CLUSTERING

Image classification and clustering refers to the method of arranging the images into clusters which may be done in a supervised or unsupervised way. In supervised classification, the problem is to classify a newly encountered image from a collection of given pre-classified images. Whereas, in unsupervised classification (or image clustering), without any previous knowledge the unlabeled similar type of images are grouped together which leads to cluster generation. Clustering the images based on their content is an important and equally challenging task to infer information from the huge collection of images. This technique is more focused on the levels of inter-image relations, semantic in an image, and regions. However, this technique may operate on the large raw data.

Association rule mining Ordonez and Omiecinski (1998) discussed an algorithm for image mining association rules. This algorithm reduced I/O and CPU overhead and operated data or images on region, objects and visual pattern level. They also built the data mining system on the top of CBIR system. This algorithm first segmented images into blobs. Then identified and labeled objects present in the images. Later, similarity measurement was done on those images.

The value of similarity measurement being one indicated perfect match on all desired features, whereas zero similarity measurement value referred to the worst match possible on those desired features. To interpret the association rules, this process also provided the auxiliary images with identified objects.

1. Image mining framework

There are two different frameworks of image mining (Datcu and Seidel, 2000).

- Function driven framework
- Information driven framework. Most of the existing image mining system architectures fall under the function driven framework.

However, function driven framework is not a generalized framework. It can be application oriented or organization oriented. Datcu and Seidel (2000) introduced function driven framework for intelligent satellite mining system.

The function driven framework for the multimedia miner was proposed by Zaiane and Han (1998). The advantage of this framework was it could organize and

clarify the different tasks to be performed in image mining, but on the contrary, it was unable to differentiate levels of vital information representation to perform meaningful image mining. This drawback was fixed in the information driven framework.

V.CONCLUSION

The goal of this paper was to emphasize the fact that nowadays, users (including doctors, meteorologists, investigators, teachers and students, etc.) need to face and utilize an incredible amount of pictures stemming from the Internet or various private and commercial databases. This paper discussed and compared different image mining techniques and also discussed about various image mining frameworks. The discussion and overview of all such techniques and frameworks, helped to establish a comparative study among the existing image mining methods.

The review aims at stressing out the need of automating their processing and classification with the purpose of obtaining particular information/knowledge from an image collection. Following the objectives, Image Mining was described as an interdisciplinary research area, the particular steps needed for IM were reviewed and the commonly exploited IM techniques were summarized. The tasks of introducing automated detection of unknown patterns in image sets and deriving contextual information based on these patterns were defined as the main purposes of Image Mining. Accordingly, at each IM level, the benefits and bottlenecks of individual techniques pointing out the future focusing were identified. Besides, the final part of the paper outlines the challenges to be faced within the future research.

REFERENCES

[1]. J.S.D. Bonet, Image Preprocessing For Rapid Selection In “Pay Attention Mode” MIT,2000.
[2]. M.C.Burl et al. Mining for image content: In Systemic, Cybernetics and Informatics/ Information Systems: Analysis and Synthesis (Orlando,Fl), July 1999
[3]. G.G.Gargner and D.Keating Automatic Detection Of Diabetic Retinopathy Using An Artificial Neural Network: A Screening Tool. British Journal of ophthalmology 1996.
[4]. Prof Dr.Hilal M.Yousif Using Image Mining To Discover Association Rules Between Image Objects AI.Rafidian University college.
[5]. R.Kazman and J.Kominek , Information Organization In Multimedia Resources, Proceedings of the 11 th annual international conference on system documentation1993 Pg 149-162
[6]. C.Lakshmi Devasena et al , An Experiential Survey On Image Mining Tools, Techniques And Applications, journal on computer science and engineering (IJCSE) 2011Vol :3 No:3 ISSN :0975-3397
[7]. C.Ordonez and E.Omiecinski, Discovering association rules based on image content. Proceedings of the IEEE Advances in Digital Libraries Conference(ADL‟99), 1999.
[8]. P.Perner, Image mining: issues, framework, a generic tool and its application to medical image diagnosis. Journal engineering applications of artificial intelligence Volume:15,No:3.
[9]. Peter Stancher Using image mining for image retrieval IASTED Conference of computer science and technology 2003.
[10]. Ramadass Sudhir A Survey On Image Mining Techniques: Theory And Application 2011 journal of computer engineering and intelligent systems Vol: 2 No:6.
[11]. Rajshree S. Dubey et.al Image Mining Using Content Based Image Retrieval System International Journal on computer science and Engineering Vol:2, No.7 2010.
[12]. P.Rajendran, M.Madheswaran hybrid medical image classification using association rule mining with decision tree algorithm journal of computing Vol: 2, issue 1, January 2010.
[13]. O.R.Zaiane, J.W. Han et al. Mining multimedia Data. CASCON „98‟: Meeting of minds , PP 83-96 Toronto, Canada, November 1998
[14]. J.Zhang, W.Hsu and M.L.Lee. An information driven Framework for image mining, in Proceedings of 12th International Conference on Database and Expert Systems Applications(DEXA),Munich, Germany, September2001.
[15]. JZhang Ji, Hsu, Mong, Lee, Image Mining:Trends and Development, Proceedings of the second international workshop on multimedia data mining (MDM/KDD‟2001), in conjunction with ACM SIGKDD conference. San Francisco, USA, August26, 2001
[16]. Zhang Ji, Hsu, Mong and Lee (2001), “Image Mining: Issues, Frameworks And Techniques ,”Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD‟2001), in conjunction with ACM SIGKDD conference, San Francisco, USA, 26th August.
[17]. N. Katayama and S. Satoh, “The sr-tree: An index structure for high-dimensional nearest neighbor queries,” in ACM SIGMOD Record, vol. 26, no. 2. ACM, 1997, pp. 369–380.
[18]. N. Beckmann, H.-P. Kriegel, R. Schneider, and B. Seeger, The R*-tree: an efficient and robust

access method for points and rectangles. ACM, 1990, vol. 19, no. 2.

[19]. S. Berchtold, D. A. Keim, and H.-P. Kriegel, "The x-tree: An index structure for high-dimensional data," Readings in multimedia computing and networking, vol. 451, 2001.

[20]. B. C. Ooi and K.-L. Tan, "B-trees: bearing fruits of all kinds," in Australian Computer Science Communications, vol. 24, no. 2. Australian Computer Society, Inc., 2002, pp. 13–20.

[21]. R. Kazman and J. Kominek, "Information organization in multimedia resources," in Proceedings of the 11th annual international conference on Systems