
A Study on Image Mining; Its Importance and Challenges

Mohammad Hadi Yousofi^{1,*}, Mahdi Esmaeili², Majide Sadat Sharifian³

¹Young Researchers and Elite Club, Kashan Branch, Islamic Azad University, Kashan, Iran

²Department of Computer, Kashan Branch, Islamic Azad University, Kashan, Iran

³Department of Mechatronic, Kashan Branch, Islamic Azad University, Kashan, Iran

Email address:

Mhu320@yahoo.com (M. H. Yousofi), msxp12@yahoo.com (M. Esmaeili), majidehsadatsharifian@gmail.com (M. S. Sharifian)

To cite this article:

Mohammad Hadi Yousofi, Mahdi Esmaeili, Majide Sadat Sharifian. A Study on Image Mining; Its Importance and Challenges. *American Journal of Software Engineering and Applications*. Special Issue: Academic Research for Multidisciplinary. Vol. 5, No. 3-1, 2016, pp. 5-9. doi: 10.11648/j.ajsea.s.2016050301.12

Received: January 6, 2016; **Accepted:** January 7, 2016; **Published:** June 24, 2016

Abstract: Image mining is an interdisciplinary field that is based on specialties such as machine vision, image processing, image retrieval, data mining, machine learning, databases and artificial intelligence. Although many studies have been conducted in each of these areas, research on image mining and emerging issues is in its infancy. For instance, data mining techniques can not automatically extract useful information from the large amount of data set like images. In this paper, by presenting the unique features of image mining, we discussed about the general procedure of the analysis and the main techniques of image analysis. Finally we explored different image mining systems, and knowledge extraction from images to achieve progress and development in this area.

Keywords: Image Mining, Image Classification, Image Clustering, Data Mining

1. Introduction

Data mining concept is combined with large databases such as Data repository and Data warehouse [1] and its aim is to extract useful unknown information from raw data [2,3]. Although like other concepts of information technology, it evokes several meanings such a data mining, information technology for different people; if it is applied accurately it can be a complex analytical tool for discovering useful patterns automatically among the data of a data repository. In fact, data mining is the advanced form of decision support that contrary to passive query tools generates templates, trends, and planned rules without requiring the user to generate questions [1].

In other words, the ability of data mining is to disclose the patterns not being considered in the user's search, and to answer questions never asked before [4]. Therefore, the ultimate goal of data mining is useful information extraction and knowledge discovery [2,5]. That is why some people call it knowledge discovery from data (KDD) rather than data mining but some others consider data mining as a core of the process of knowledge discovery [6,7,8] and as one of the most important step of knowledge management [9].

Image mining in large set of image is a new approach in the field of research on the one hand, and image database and data mining researches on the other hand [10]. Although, recently this discussion has caused the precise concept of image mining remain a challenge [11], researchers, particularly in recent years, have proposed different definitions of image mining, as well as various methods under this topic. Image mining focuses on the extraction of patterns from large collections of images while the emphasis of image processing and machine vision is on the understanding of certain characteristics of a specific image. A high volume of images, such as satellite images, medical images and digital photos produced on a daily basis. In case of the analysis of these images, a lot of useful information can be gained. The pixels shown in a raw image or series of images in order to detect objects and the relationship among them is the most fundamental challenge in the mining picture [12].

One of the main obstacles in rapid development of image mining is the lack of understanding the topics and research results about image mining.

Many researchers have this wrong presupposition that image mining is a simple extension of data mining applications, while some others consider image mining as an another term for pattern recognition and differ them in terms

of different nature of relational databases and image databases, In other words, image mining is not just utilizing data mining algorithms in images [12].

Image mining is a technique that explores information, images' data dependence and unambiguous patterns stored in the images. There are two basic techniques in this field, the first technique do the exploration in an extensive range of independent pictures. The second technique explores a series of integrated and linked images [13].

The main objective of image analysis is obtaining all significant patterns of images, without knowing the details of the content of the images; this means that without having a basic knowledge of the content of the images you can extract important patterns out of a series of images as an input.

2. Content-Based Image Retrieval (CBIR)

Image mining can be done manually by cutting and fragmenting data to achieve a specific pattern or that can be performed by using programs that analyze the data automatically.

Color, texture and existing shapes in the image, are the primary descriptors in context-based image retrieval system.

Primary descriptors are used to identify and retrieve similar images from a database of images; it is very difficult to extract images from a data set manually, because this is a very large data base [14].

Moreover, CBIR is well known as a Query by Image Content (QBIC) and content-based visual information retrieval (CBVIR) and consists of using machine vision for retrieving digital images of large databases of images [14]. It is confirmed that the previous methods of image retrieval, such as indexing, is very time consuming and inefficient. In these methods an

indexed image is stored in the database and it is connected to a keyword or a number related to the classified descriptions. These old methods were not based on CBIR content.

In CBIR any image which is stored in the database has its own characteristics, which is extracted and compared with the features of the query image. This method is a combination of knowledge in different fields such as pattern recognition, matching objects, machine learning, and microwave filtering and so on. CBIR is intended to receive and discover visual properties of images without having any descriptive text about them.

CBIR plans to look at the database images that are similar to the query image. It also focuses on the development of techniques that would effect on digital libraries of images based on the feature; the image is automatically extracted from the query. CBIR also focuses on the features of images; these features can be classified as low-level features or characteristics of a high level. CBIR images from the database images based on attributes such as color, texture, edge and shape their recovery [16]. In a text-based image retrieval system (TBIR) images based on descriptions, indexing and retrieval, such as size, type, date, time capture, identify the owner of the image, keywords or some other explanatory text on the image [16].

In Figure 1 a general CBIR system is shown. In such a system, concepts of visual images extracted from databases and features are described as multi-dimensional vectors. Feature vector features are going to be in the form of a database. To restore an image, users provide a sample image as input. The application form its own internal system that turns the feature vector. The similarity between the input image and the images in the database search and indexing is performed is calculated, and retrieved with the help of patterns [15].

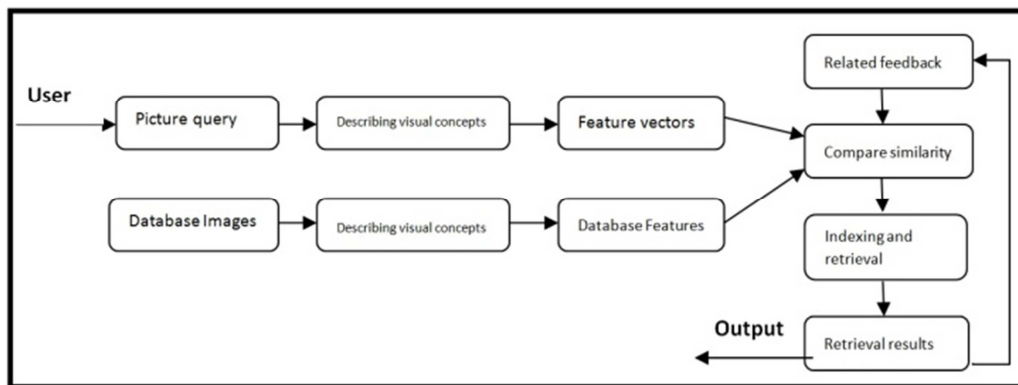


Figure 1. An example system architecture Content-Based Image Retrieval CBIR.

3. Image Mining

In a system of image mining different activities will be done in order to reach the desired images. Many of these activities are based on image processing techniques and pattern recognition. This section introduces some of the processes that occur during the process of image mining and some of the

techniques that refer in any process used to express planned. It should be noted that some of these processes precedence depends on the model which we designed for image mining.

3.1. Pre-processing and De-noising

It is necessary to improve the quality of the images before any processing to make characteristics extraction phase

easier and more reliable. Pre-processing images are done to create high-quality images for more transparent categorization. The main objective is the improvement of preprocessing of images that have been exposed to the undesirable distortion data and improve some characteristics of the image that is in the processing of future importance. This stage focuses on the properties of the image. Filtering is one of the techniques used to change or enhance an image. When we want to highlight some of the features of an image we use filtering. The existing noises in an image are eliminated using linear or nonlinear filtering methods. Low pass filters, high pass and Band pass are some of the methods used to remove noise from images [17].

3.2. Classification

Classification is a supervised method of data grouping. In supervised methods, classification of a set of labeled images is provided, which is called learning set [12]. Classification is usually a two-phase process. Learning phase and test phase. In the first phase, profile images are distinct and learning is made on the basis of class. In the second phase, parts of the specifications are used to classify images [19,18]. The most popular classification methods are decision trees, Bayesian classifier, SVM-based classification rule, neural networks, and fuzzy logic techniques mentioned [19]. One of the methods which are very important in the process of classification is using decision tree. Decision trees, divide decision space to smaller areas as a return based on the whole sample. In this way, decision trees break down the complex decision as a throwback which has a uniform result and naturally reflects the recognition strategy that can be used in human decision-making process [20].

3.3. Color Processing

One of the methods of color image processing is using color histogram. Color histogram of an image may be at the level of the whole picture or for each range, a histogram as a feature in the image used to represent the color distribution [19]. A color image of RGB, is an $M * N * 3$ array of color pixels, the color pixels of which is a triple specifying the amount of red, green, blue part of the image in a space. A color image can be considered as a stack of three black and white images when color display with entries in a red, green and blue are combined to make a color image, which can average each color component in the image as calculated (Formula 1).

$$\text{Average pixels red} = R(P) / P$$

$$\text{Average green pixels} = (G(P)) / P$$

$$\text{Average blue pixels} = (B(P)) / P$$

Formula1: Calculation formula

Where P is the total number of image pixels. R (P) is the number of red pixels. G (P) is the number of green pixels and B (P) is number of blue pixels.

3.4. Clustering

Clustering, a branch of learning, is an unsupervised method and is an automated process in which samples are divided into groups, whose members are similar to the categories called cluster. Therefore, cluster is a collection of objects where objects are similar with each other and with objects in other clusters are dissimilar. Similarly, the various criteria to be taken into account for example, the criteria are to be used for clustering contract and objects that are closer together as a cluster consider that this type of clustering, also called distance-based clustering.

Clustering, divided into a number of subsets or clusters of heterogeneous population is said to be homogeneous. What distinguishes clustering categories is that clustering does not rely on pre-determined categories. In categorization based on model, each data is allocated to a pre-determined category. These categories (such as gender, skin color, etc.) have been determined thorough the finding of previous studies. There is no set of predetermined clustering and data on the basis of similarity are grouped and titles of each group be determined by the user. For example, clusters of symptoms may indicate a variety of diseases and clusters of features customers may be indicative of different market segments. Clustering is usually as a prelude to the use of other data mining analysis or modeling is used [21].

3.5. Feature Extraction

Measuring features of an image is a basis factor to distinguish and categorize an image. The machine vision research is providing modals of objects and scenes of an image to extract image properties for developing decision rules, and then analyze and describe observed image. We use the image processing methods, clustering and measuring image properties for this purpose.

Developing imaging techniques according to image revival system is based on content. Color, texture, style, object shape, arrangement and their situations inside image and etc. are all bases of visual contents of an image and an image is indexed based on these properties [22]. If properties and characteristics are selected correctly, they can express much useful information about an image. Features extraction methods analyze properties, objects and images to extract significant features indicating different classes of objects. Properties are given to categorization as an input to distinguish a class to which the object is related. texture is one of the most important features that can be extracted from images. Texture is referred to informational patterns or structural arrangement observed in an image. Texture may include some initial information and also it may express structural arrangement in an area and it's relation with other limited areas surrounding it. Texture is kind of vision features that it does not depend on color, severity and reflections in natural phenomenon in images. Texture is a collection of all natural features in a surface and for this reason we use from this feature widely in image processing. Many objects are distinguished via only texture and without any additional data. First, texture analysis

was based on first order statistics or second order statistics. There are different methods to measure images textural features such as co – occurrence matrix, fractals, Gabor filters, and microwave converter socializations. Also many techniques were developed to describe local patterns via textural spectrum. We can use co-occurrence matrix and edges data to describe a texture [14].

In a texture-based method, the parameters are collected base on statistical methods. Gray surface statistical features are one of the most efficient ways to categorize texture. Gray – Level Co – occurrence Matrix (GLCM) is one of methods that are used to extract second- order statistics from image. Every element (I, J) in this matrix indicate occurrence count in a relation between pixel I and pixel J in input image. Parameters related to image texture that we can extract are entropy, contrast, dissimilarity, homogeneity, standard deviation, correlation, average and variance [18] [22].

3.6. Selecting Properties

To select properties, we can use measuring methods based on entropy, Gain – ratio, Gini- index, chi square, etc. To discretization of properties, we apply chi- merge discretization cut point, discretization base on MDLP or LVQ. If we use decision tree to categorize, this discretization methods create one or several interval during making decision tree that depend on which ways is used for discretization. Gained tree can be binary or n- number that led to produce more correct and compact trees. To evaluate them, we can use n-fold lateral evaluating methods or test and train method [20].

Selecting features cause to reduce problem dimension and as a result cause to improve prediction and decrease time calculations. This, problem can remove via deleting unrelated, additional and noisily features. Therefore, we always try to select a subset of features. Usually, these features select via search ways. Different search ways were developed to reach this purpose. Of popular algorithms which are used including sequential forward selection, sequential backward selection, genetics algorithm, particle swarm optimization, branch and bound feature optimization [18].

3.7. Histogram Equalization

Histogram equalization is a method that use for contrast setting in image processing. Contrast amount distribute better on histogram via this setting. This matter let limits which has less local contrast to reach better contrast. Histogram equalization performs this operation via developing the most amount contrast. This method is very useful for images that their background and foreground is black and white such as radiology images. One of the other histogram methods in image processing is providing severity histogram. In this kind of histogram, we consider some feature such as average, variance, skewness, elongation, entropy and energy [18].

4. Discussion and Conclusions

Valuable bits of information from sources like satellite,

space, medical and digital images, are produced daily, in such a way that their high magnitude and size has made it impossible for human to analyze them for extracting information or useful and appropriate patterns in decision making processes.

Image mining is a new and promising area for knowledge extraction from images, however is still in the beginning and more studies need to be done for future development to improve techniques such as image processing, feature extraction, image segmentation and identifying objects.

In this paper, we presented the unique features of image mining, proceeded with the general process of analyzing and discussed the main image mining techniques. Furthermore, we introduced the concept of image mining as one of newest research axis in imaging database. Then we accounted for different methods and techniques for image mining proposed by researchers.

References

- [1] G. Eason, B. Noble, and I. N. Sneddon, "On certain integrals of Lipschitz-Hankel type involving products of Bessel functions," *Phil. Trans. Roy. Soc. London*, vol. A247, pp. 529–551, April 1955. (references).
- [2] Tan J. *Medical Informatics: Concepts, Methodologies, Tools, and Applications*. Hershey: IGI Global snippet; 2008.
- [3] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in *Magnetism*, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 1963, pp. 271–350.
- [4] LaTour KM, Eichenwald S. *Health Information Management: Concepts, Principles, and Practice*. Chicago: AHIMA; 2002. p. 478-80.
- [5] Chakrabarti S, Cox E. *Data Mining: Know It All*. Amsterdam: Morgan Kaufmann p. 7; 2009.
- [6] Fayyad U, Shapiro G, Smyth P. *Knowledge Discovery and Data Mining* [Online]. 2011 [Cited2011Aug8]; Available from: URL: [http://www. Aaa.org/](http://www.Aaa.org/).
- [7] Han J, Kamber M, Pei J. *Data Mining: Concepts and Techniques*. Philadelphia: Elsevier; 2011.
- [8] Maimon OZ, Rokach L. *Data Mining And Knowledge Discovery Handbook*. New York: Springer Science & Business; 2010. p. 1.
- [9] Chen H, Fuller SS, Friedman C, Hersh W. *Medical Informatics: Knowledge Management and Data Mining in Biomedicine*. New York: Springer; 2005.
- [10] C. Ordonez, E. Omiecinski, *Image Mining: A new approach for data mining*, Thechnical Report GIT-CC-98-12, Georgia Institute of Technology, College of Computer, 1998.
- [11] J. Zhang, W. Hsu, M. Lee, *Image Mining: Issues ,Frameworks And Techniques*, In *Proc. Of the second International workshop on Multimedia Data Mining*, San Francisco, USA, August 2001.
- [12] Ji Zhang, Wynne Hsu, Mong Li Lee, "An Information. Driven Framwork For Image Mining", *Computer Science*, School of Computer, National University of Singapore, IEEE, August 2001.

- [13] RamadassSudhir, "A Survey on Image Mining Techniques: Theory and Applications", Computer Engineering and Intelligent Systems, Vol2, No, 6, 2011.
- [14] Monika sahu, madhup shrivastava, dr. m a rizvi, "image mining: a new approach for data mining based on texture", IEEE, 2012.
- [15] Nishchol mishra1, Dr. sanjay Silakari, "Image Mining the Context of content based Image Retrieval: A perspective", IJCSI, Vol. 9, Issue4, No3, July 2012.
- [16] Tomas Berlage, "Analyzing and mining image database", DRUG DISCOVERY TODAY: BIOSILICO, DDT, Vol 10, Number 11, June 2005.
- [17] A. Kannan, DR. V. Mohan, Dr. N. Anbazhagan, "Image Clustering and Retrieval using Image Mining Techniques", IEEE, 2010.
- [18] Aswini Kumar Mohanty, Manas Ranjan Senapati, Saroj Kumar Lenka, " A novel image mining technique for classifaction of mammograms using hybrid feature selection, "Springer, 23 February 2012.
- [19] Chidansh Amitkumar Bhatt, Mohan S. Kankanhalli, "Multimedia data mining: state of the art and challenges", springer Science+Business Media, LLC 2010.
- [20] Petra Perner, "Image mining: issue, framework, a generic tool and its application to medical-image diagnosis", Elsevier, 2002.
- [21] Sanjay T. Gandhe, K. T. Talele, and Avinash G. Keskar. "Image Mining Using Wavelet Transform". Springer-Verlag Berlin Heidelberg 2007.
- [22] A. Hema, E. Annasaro, "a survey in need of image mining techniques", International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue2, february 2013.