Learning the Comparison of Image Mining Technique and Data Mining Technique

Hlaing Htake Khaung Tin

Faculty of Information Science, University of Computer Studies, Yangon, Myanmar

***Abstract*—Data mining is the computing process of discovering patterns in large data sets involving methods at the intersection of machine learning, statics, and database systems. Image data represents a keystone of many research areas including medicine, forensic criminology, robotics and industrial automation, meteorology and geography as well as education. 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 learning the comparison of image mining technique and data mining technique.**

***Keywords*—data mining; image mining; clustering; classification; image retrieval**

I. INTRODUCTION

In the real world, huge amount of data are available in education, medical, industry and many other areas. Such data may provide knowledge and information for decision making. Data mining is a type of sorting technique which is actually used to extract hidden patterns from large databases.

Image mining deals with the extraction of image patterns from a large collection of images. Clearly, image mining is different from low-level computer vision and image processing techniques because the focus of image mining is in extraction of patterns from *large* collection of images, whereas the focus of computer vision and image processing techniques is in understanding and/or extracting specific features from a *single* image. While there seems to be some overlaps between image mining and content-based retrieval (both are dealing with large collection of images), image mining goes beyond the problem of retrieving relevant images. In image mining, the goal is the discovery of image patterns that are significant in a given collection of images.[1]

Analyzing image data forms a keystone of many research areas including medicine (evaluating MRI, interpreting XRays/CT scans), forensic criminology (fingerprint identification, face recognition), robotics and industrial automation (robotic vision), meteorology and geography (satellite imagery) as well as education (computer-aided visualization) and many other fields. Image data plays vital role in every aspect of the systems like business, hospitals, engineering and so on. Image mining normally deals with the study and development of new technologies that allow easy analysis and interpretation of the images. Image mining is not only the simple fact of recovering relevant images but is the innovation of image patterns that are noteworthy in a given collection of images.

The rest of the paper is organized as follows. Section 2 will discuss research issues that are unique to image mining and goal of data mining. Section 3 discusses analysis and classification of image in image mining technique. Section 4 gives data mining technique. Section 5 discusses data management and section 6 gives the comparison of image mining and data mining. Finally, section 7 concludes with some future research directions for image mining and data mining.

II. RSEARCH ISSUE

Data mining, which is defined as the process of extracting previously unknown knowledge, and detecting intersecting patterns from a massive set of data, has been a very active research. Image mining is more than just an extension of data mining to image domain. It is an interdisciplinary endeavor that draws upon expertise in computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence. Despite the development of many applications and algorithms in the individual research fields cited above, research in image mining is still in its infancy [1].

Image mining is a very important technique which is used to mine knowledge easily from image. Image mining handles with the hidden knowledge extraction, image data association and additional patterns which are not clearly accumulated in the images. The most important function of the mining is to generate all important patterns without previous information of the patterns. Rule mining has been adapting to huge image databases. Numerous researches have been carried on this image mining [13].

The World Wide Web is regarded as the largest global mage repository. An extremely large number of image data such as satellite images, medical images, and digital photographs are generated every day. These images, if analyzed, can reveal useful information to the human users. Unfortunately, there is a lack of effective tools for searching and finding useful patterns from these images. Image mining systems that can automatically extract semantically meaningful information (knowledge) from image data are increasingly in demand. The fundamental challenge in image mining is to determine how low-level, pixel representation contained in a raw image or image sequence can be efficiently

 ISBN: 978-93-5288-448-3

and effectively processed to identify high-level spatial objects and relationships. In other words, image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the image databases. It is an interdisciplinary endeavor that essentially draws upon expertise in computer vision, image processing, image retrieval, data mining, machine learning, database, and artificial intelligence [14].

Image mining is the process of searching and discovering valuable information and knowledge in large volumes of data. Figure 1 shows the Typical Image Mining Process. Some of the methods used to gather knowledge are, Image Retrieval, Data Mining, Image Processing and Artificial Intelligence. These methods allow Image Mining to have two different approaches. One is to extract from databases or collections of images and the other is to mine a combination of associated alphanumeric data and collections of images. In pattern recognition and in image processing, feature extraction is a special form of dimensionality reduction. When the input data is too large to be processed and it is suspected to be notoriously redundant, then the input data will be transformed into a reduced representation set of features. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Several features are used in the Image Retrieval system. The popular amongst them are Color features, Texture features and Shape features.[2]

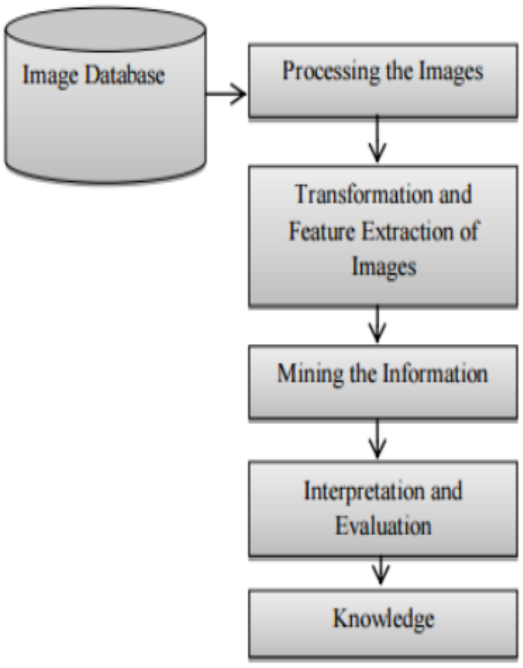

Figure 1. Image Mining Process

The goals of data mining are fast retrieval of data or information, knowledge Discovery from the databases, to identify hidden patterns and those patterns which are previously not explored, to reduce the level of complexity, time saving, etc. Various field adapted data mining technologies because of fast access of data and valuable information from a large amount of data. Data mining application area includes marketing, telecommunication, fraud detection, finance, and education sector, medical and so on. They are:

- Data Mining in Education Sector
- Data Mining in Banking and Finance
- Data Mining in Market Basket Analysis
- Data Mining in Earthquake Prediction
- Data Mining in Bioinformatics
- Data Mining in Telecommunication
- Data Mining in Agriculture
- Data Mining in Cloud Computing

III. Image Analysis and Classification in Image Mining Technique

Image analysis is an inevitable step of image Mining. The analysis is often said to be a pre-processing stage of the image mining. The objective of analyzing an image is to find and extract all relevant features required to represent an image [11].

A. Image Preprocessing

Image preprocessing is an initial step of processing images.
It is utilized for improving the quality of an image before object detection algorithms are applied. Normalizing images is usually performed in order to reduce noise and/or enhance resolution of an image. Different pre-processing procedures might by employed including average, median and wiener filtering for lowering the impact of noise and interpolation based Discrete Wavelet Transform (DWT) and Multiresolution
Image Fusion to enhance the resolution.

B. Object Recognition

Object recognition is a step resulting in segmentation of an image. It focuses on identifying objects in an image and dividing an image into several regions accordingly. It is a task which until recently was considered the main objective of image processing. Visual objects are to be detected from an image according to a model. The model represents certain patterns obtained as an outcome of applying a training algorithm on the training sample. For this purposes, supervised machine learning needs to be deployed.

C. Feature Extraction

Extracting features stands for a process of compressing the information derived from identified objects into a set of attributes. Both local and global descriptors may be used for representing the image. Global descriptors are easier to compute and do not tend to segmentation errors. In comparison, local descriptors provide much precise representation and might discover even subtle patterns. Features are usually represented numerically and provide complex mathematical representation of an image. They describe objects in terms of shape, texture and/or color, etc.

 ISBN: 978-93-5288-448-3

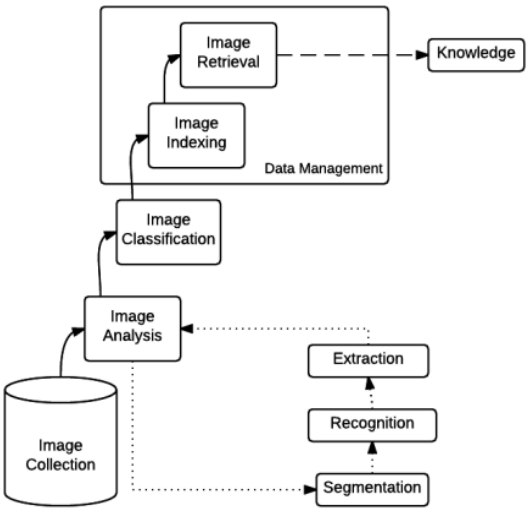

Figure 2. Traditional Image Mining Procedure

Image classification is to categories objects detected in an image. Currently, classification objects are an extensively researched domain. Different approaches have been proposed and tested. However, the field remains in its infancy and categorizing into non-pre-defined classes is still an issue to be solved. The researched methods of categorizing objects as described by[11]are:

D. Supervised Classification

Supervised classification is the original approach of categorizing images. The objective is to divide the detected objects into pre-defined categories. Methods of machine learning (decision tree, rule-based classification, support vector machines, neural networks) are applied on training the system based on the labeled (pre-classified) samples and flowingly, on labeling new images using the obtained (trained) classifiers.

E. Image Clustering

In contrast to standard classification methods, clustering represents unsupervised categorization of objects. The objects are grouped into clusters based on the similarity, not on the basis of predefined labels. Cluster analysis aims at searching for common characteristics without knowing the exact data types. It is oriented on decomposing images into groups of objects similar to each other and different from the other objects as much as possible. The similarity is evaluated based on the calculated features (texture, shape, color,...). Hierarchical clustering, partition based clustering, mixture resolving, nearest neighbor clustering, fuzzy clustering, evolutionary clustering are some of approaches used for unsupervised categorization. After accomplishing the clustering process (dividing the objects into clusters), an expert form the particular field is needed to identify the individual categories (clusters).

IV. Data Mining Technique

Data mining means collecting relevant information from unstructured data. So it is able to help achieve specific objectives. The purpose of a data mining effort is normally either to create a descriptive model or a predictive model .A descriptive model presents, in concise form, the main characteristics of the data set. The purpose of a predictive model is to allow the data miner to predict an unknown (often future) value of a specific variable; the target variable. The goal of predictive and descriptive model can be achieved using a variety of data mining techniques as shown in figure 3[2].

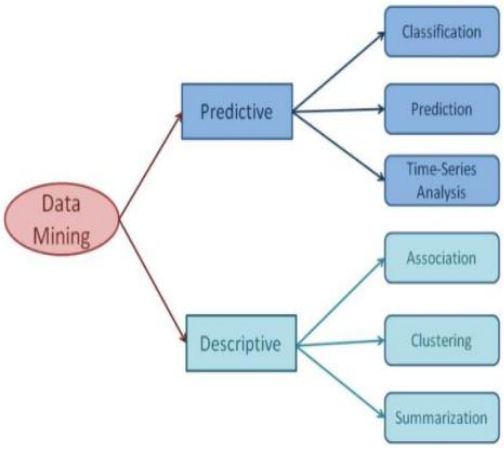

Figure 3: Data Mining Models

A. Classification

Classification based on categorical (i.e. discrete, unordered). This technique based on the supervised learning (i.e. desired output for a given input is known). It can be classifying the data based on the training set and values (class label). These goals are achieve using a decision tree, neural network and classification rule (IFThen). for example, we can apply the classification rule on the past record of the student who left for university and evaluate them. Using these techniques, we can easily identify the performance of the student.

B. Regression

Regression is used to map a data item to a real valued prediction variable [7]. In other words, regression can be adapted for prediction. In the regression techniques target value are known. For example, you can predict the child behavior based on family history.

C. Time Series Analysis

Time series analysis is the process of using statistical techniques to model and explain a time-dependent series of data points. Time series forecasting is a method of using a model to generate predictions (forecasts) for future events based on known past events [8]. For example stock market.

 ISBN: 978-93-5288-448-3

D. Prediction

It is one of a data mining techniques that discover the relationship between independent variables and the relationship between dependent and independent variables [9]. Prediction model based on continuous or ordered value.

E. Clustering

Clustering is a collection of similar data object. Dissimilar object is another cluster. It is way finding similarities between data according to their characteristic. This technique based on the unsupervised learning (i.e. desired output for a given input is not known). For example, image processing, pattern recognition, city planning.

F. Summarization

Summarization is abstraction of data. It is set of relevant task and gives an overview of data. For example, long distance race can be summarized total minutes, seconds and height. Association Rule: Association is the most popular data mining techniques and fined most frequent item set. Association strives to discover patterns in data which are based upon relationships between items in the same transaction. Because of its nature, association is sometimes referred to as "relation technique". This method of data mining is utilized within the market based analysis in order to identify a set, or sets of products that consumers often purchase at the same time [10].

G. Sequence Discovery

Uncovers relationships among data [7]. It is set of object each associated with its own timeline of events. For example, scientific experiment, natural disaster and analysis of DNA sequence.

V. DATA MANAGEMENT

Images cover a huge amount of information. Depending on the way of storing and indexing images, various knowledge might be searched and retrieved from an image database [11].

A. Storing Images

Zhang et al. identified several differences between image databases and relational databases pointing put the misusing and misunderstanding the term of Image Mining. IM cannot be understood barely as applying data mining techniques on images, as compared to relational databases, there are important differences in handling images:

• Relativity of values - Images can be numerically represented, however, in contrast to relational databases, the values are only significant in a certain context.

• Dependency on the spatial information - When working with image databases, the position of individual pixels is an inevitable factor for correct interpretation of image content.

• Multiple interpretations - In comparison with relational databases, image databases are more difficult to handle, as the same patterns derived from images might have multiple interpretations depending on the context and position.

There are different ways of storing images. Several compression formats (JPEG, MPEG 7, DICOM) store the meta
data in one file with an image. According to [12]), this approach might result in difficulties with analyzing images. Databases of such images are not the most suitable candidates for data mining, as they are not optimized for time and memory consumption when performing image retrieval.

VI. 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:

(1) Absolute versus relative values. In relational databases, the data values are semantically meaningful. For example, age is 35 is well understood. However, in image databases, the data values themselves may not be significant unless the context supports them. For example, a grey scale value of 46 could appear darker than a grey scale value of 87 if the surrounding context pixels values are all very bright.

(2) 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. As a result, image miners try to overcome this problem by extracting position independent features from images first before attempting to mine useful patterns from the images.

(3) Unique versus multiple interpretations. A third important difference deals with image characteristics of having multiple interpretations for the same visual patterns. The traditional data mining algorithm of associating a pattern to a class (interpretation) will not work well here. A new class of discovery algorithms is needed to cater to the special needs in mining useful patterns from images [3,4,5,6].

VII. CONCLUSTIONS

Data mining is a powerful concept for data analysis and process of discovery interesting pattern from the huge amount of data, data stored in various databases such as data warehouse, World Wide Web, external sources. Interesting pattern that is easy to understand, unknown, valid, potential useful. This paper is described what the goals of image mining, data mining are and also learning the comparison of mining technique.

ACKNOWLEDGMENT

I would like to thank all of my current and previous colleagues at Computer University for their support and encouragement during my research work. A big thanks you to

my husband Oo Tun Shwe and daughter San ThitSar Nyi for their help and support.

REFERENCES

[1]. Hilal M. Yousif, "Using Image Mining to Discover Association Rules between Image Objects".

[2]. Preeti Chouhan, Mukesh Tiwari, "Feature Extraction Techniques for Image Retrieval Using Data Mining and Image Processing Techniques", International Journal of Advanced Research in Computer and Communication Engineering, volume 5, Issue 5, May 2016.

[3]. Jiawei Han, Micheline Kamber, "Data Mining: Concepts and Techniques", Simon Fraser University.

[4]. Margraet H. Dunham, "Data mining: Introductionary and and Advanced Topics", Southern Methodist University.

[5]. Osmar R. Zaiane, Jiawei Han, Ze-Nian Li, Hean Hou, "Mining Multimedia data", Simon Fraser University, Canada.

[6]. Chabane Djeraba, "Relationship Extraction from Large Image Databases", University of Nautes, France.

[7]. M. Flickner, H Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafne, D. Lee, D. Petkovic, D. Steele and P. Yanker, "Query by Image and Video Content The QBIC System" IEEE Computer, pp-23-32, 1995

[8]. Anil K. Jain and Aditya Vailaya, "Image Retrieval using color and shape", In Second Asian Conference on Computer Vision, pp 5-8. 1995.

[9]. 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.

[10]. Y. Liu, D. Zang, G. Lu and W. Y. Ma, "A survey of content-based image retrieval with high level semantics", Pattern Recognition, Vol-40, pp-262-282, 2007.

[11]. Barbora Zahradnikova, Sona Duchovicova and Peter Schreiber, "Image Mining: Review and New Challenges", International Journal of Advanced Computer Science and Applications, Vol. 6, No. 7, 2015.

[12]. T. Berlage, "Analyzing and mining image databases," *Drug discovery today*, vol. 10, no. 11, pp. 795–802, 2005.

[13]. K. R. Yasodha, K.S. Yuvaraj, "A Study on Image Mining Techniques", International Journal of Applied Research, volume 1, issue 1, December 2013.

[14]. 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.

I received the B.C.Sc and B.C.Sc (Hons:) degree from Government Computer College, Lashio, Northern Shan State, Myanmar in 2003, the Master of Computer Science (M.C.Sc) degree from University of Computer Studies, Mandalay, Myanmar in 2006 and the Ph.D(IT) degree from University of Computer Studies, Yangon, Myanmar in 2014 January. My research interests mainly include face aging modelling, face recognition, perception of human faces and information technology (IT).

 ISBN: 978-93-5288-448-3