

# Image Mining: frameworks and techniques

Madhumathi.k<sup>1</sup>, Dr.Antony Selvadoss Thanamani<sup>2</sup>M.Phil, Department of computer science, NGM College, Pollachi, Coimbatore, India<sup>1</sup>HOD Department of Computer Science, NGM College, Pollachi, Coimbatore, India<sup>2</sup>

**ABSTRACT:** Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the image database. 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. Research in image mining can be broadly classified into two main directions. The first direction involves domain-specific applications where the focus is in the process of extracting the most relevant image features into a form suitable for data mining. The second direction involves general applications where the focus is on the process of generating image patterns that maybe helpful in the understanding of the interaction between high-level human perceptions of images and low level image features. The latter may lead to improvements in the accuracy of images retrieved from image databases.

**Keywords:** Image mining, image indexing, and image clustering.

## I. INTRODUCTION

By definition, 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, where as 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.

## II. IMAGE MINING PROCESS

The image mining processes are shown in Figure 1. The images from an image database are first preprocessed to improve their quality. These images then undergo various transformations and feature extraction to generate the important features from the images [2]. With the generated features, mining can be carried out using data mining techniques to discover significant patterns. The resulting patterns are evaluated and interpreted to obtain the final knowledge, which can be applied to applications.

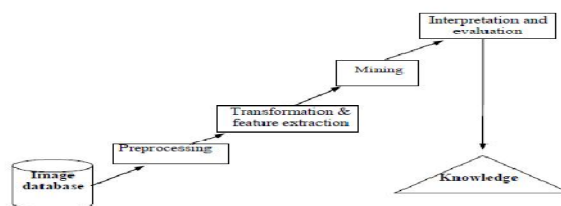


Figure 1: Image Mining Process



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

### Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014

### III. IMAGE MINING FRAMEWORKS

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. So an image mining system is often complicated because it requires the application of an aggregation of techniques ranging from image retrieval and indexing schemes to data mining and pattern recognition. A good image mining system is expected to provide users with an effective access into the image repository and generation of knowledge and patterns underneath the images [4]. Therefore to this end, such a system typically encompasses the following functions: image storage, image processing, feature extraction, image indexing and retrieval, patterns and knowledge discovery.

Two kinds of frameworks used to characterize image mining systems are

- Function-driven frameworks
- Information-driven frameworks

#### Function-driven frameworks

Several image mining systems have been developed for different applications. The majority of existing image mining system architectures fall under the function-driven image mining framework. The function-driven framework spotlighted on the functionalities of different component modules to organize image mining systems, while the latter is a hierarchical structure with an emphasis on the information needs at various levels of hierarchy. This framework serves the purpose of organizing and clarifying the different roles and tasks to be performed in image mining [5]. It is exclusively application oriented and the framework was organized according to the module functionality.

#### Information-driven frameworks

The image database containing raw image data cannot be directly used for mining purposes. Raw image data need to be processed to generate the information that is usable for high-level mining modules. An image mining system is often complicated because it employs various approaches and techniques ranging from image retrieval and indexing schemes to data mining and pattern recognition [3]. Such a system typically encompasses the following functions: image storage, image processing, feature extraction, image indexing and retrieval, patterns and knowledge discovery. Indeed, a number of researchers have described their image mining framework from the functional perspective.

### IV. THE FOUR INFORMATION LEVELS

Four information levels are important that should be noticed in image mining. They are

- Pixel Level
- Color
- Texture
- Edge information



## **International Journal of Innovative Research in Computer and Communication Engineering**

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

### **Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)**

**Organized by**

**Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014**

#### Pixel Level

The Pixel Level is the lowest layer in an image mining system. It consists of raw image information such as image pixels and primitive image features such as color, texture, and edge information.

#### Color

Color is, perhaps, the most widely used visual features in most image management database system. Color is widely represented by its RGB values (three 0 to 255 numbers indicating red, green, and blue). The distribution of color is a global property that does not require knowledge of how an image is composed of component objects. Color histogram is a structure commonly used to store the proportion of pixels of each color within the image. It is invariant to under translation and rotation about the view axis and changes only slowly under change of view angle, change in scale, and occlusion.

#### Texture

Texture is the visual pattern formed by a sizable layout of color or intensity homogeneity. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment.

#### Edge information

Edge information is an important visual cue to the detection and recognition of objects in an image. Typically, edge information is obtained by looking for sharp contrasts in nearby pixels. Once the edges have been identified, these edges can be grouped to form regions.

## **V. IMAGE MINING TECHNIQUES**

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

#### 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 finds objects in the real world from an image. This is one of the major tasks in image mining. Automatic machine learning and meaningful information extraction can only be realized when some objects have been identified and recognized by the machine [7]. The object recognition problem can be referred to as a supervised labeling problem based on models of known objects. That is, given a target image containing one or more interesting objects and a set of labels corresponding to a set of models known to the system, what object recognition does is to assign correct labels to regions, or a set of regions, in the image. An object recognition system typically consists of four components, namely, model database, feature detector, hypothesizer and hypothesis verifier. The model database contains



## International Journal of Innovative Research in Computer and Communication Engineering

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

### Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)

Organized by

Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014

all the models known to the system. These models contain important features that describe the objects. The detected image primitive features in the Pixel Level are used to help the hypothesizer to assign likelihood to the objects in the image [6]. The verifier uses the models to verify the hypothesis and refine the object likelihood. The system finally selects the object with the highest likelihood as the correct object

#### Image clustering

Image clustering groups a given set of unlabeled images into meaningful clusters according to the image content without priority knowledge. Typical clustering techniques include hierarchical clustering algorithms, partitional algorithms, mixture-resolving and mode-seeking algorithms, nearest neighbor clustering, and fuzzy clustering. Once the images have been clustered, a domain expert is needed to examine the images of each cluster to label the abstract concepts denoted by the cluster. Clustering will be more advantage for reducing the searching time of images in the database. Fuzzy C-means (FCM) is one of the clustering methods which allow one piece of data to belong to two or more clusters. In this clustering, each point has a degree of belonging to clusters, as in fuzzy logic, rather than belonging completely to just one cluster. Thus, points on the edge of a cluster may be in the cluster to a lesser degree than points in the centre of cluster. FCM groups data in specific number of clusters

#### Image Indexing

While focusing on the information needs at various levels, it is also important to provide support for the retrieval of image data with a fast and efficient indexing scheme. Typically, the image database to be searched is large and the feature vectors of images are of high dimension (typically in the order of 102), search complexity is high. Two main approaches are: reducing dimensionality or indexing high dimensional data. Reducing the dimensions can be accomplished using two well-known methods: the Singular Value Decomposition (SVD) update algorithm and clustering

## VI. CONCLUSION

The main objective of the image mining is to remove the data loss and extracting the meaningful information to the human expected needs. The images are preprocessed with various techniques and the texture calculation is highly focused. Here, images are clustered based on RGB Components, Texture values and Fuzzy C mean algorithm. Entropy is used to compare the images with some threshold constraints. This application can be used in future to classify the medical images in order to diagnose the right disease verified earlier.

## REFERENCES

- [1]. U. M. Fayyad, S. G. Djorgovski, and N. Weir: Automating the Analysis and cataloging of Sky Surveys. Advances in Knowledge Discovery and Data Mining, 471-493, 1996.
- [2] Babu, G P and Mehtre, B M. Color indexing for efficient image retrieval. Multimedia Tools and applications, 1, 327-348, 1995.
- [3].C. Ordóñez and E. Omiecinski, "Image Mining: A New Approach for Data Mining", 1998.
- [4]. J. R. Bach, C. Fuller, A. Gupta, A. Hampapur, B. Horowitz, R. Humphrey, R. Jain, C-F. Shu. Virage Image Search Engine: An Open Framework for Image Management, in Storage and Retrieval for Image and Video Databases (SPIE), pp 76-87, 1996.



**International Journal of Innovative Research in Computer and Communication Engineering**

(An ISO 3297: 2007 Certified Organization)

Vol.2, Special Issue 1, March 2014

**Proceedings of International Conference On Global Innovations In Computing Technology (ICGICT'14)**

**Organized by**

**Department Of CSE, JayaShriram Group Of Institutions, Tirupur, Tamilnadu, India on 6<sup>th</sup> & 7<sup>th</sup> March 2014**

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

[6] G. A. Carpenter, S. Grossberg and J. H. Mrkuzon. Fuzzy ARTMAP: A Neural Architecture for Incremental Supervised Learning of Analog Multidimensional Maps, IEEE Transactions on Neural Networks, 3(5), 698-713, 1688-1692. IEEE Transactions on Neural Networks, 3(5), 698-713, 1688-1692.

[7]. A. Kitamoto. Data Mining for Typhoon Image Collection. In Proceedings of the Second International Workshop on Multimedia Data Mining (MDM/KDD'2001), San Francisco, CA, USA, August, 2001.