ISSN: 2277-9655
Impact Factor: 1.852

IJESRT

INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

Image Mining Techniques and Applications

Deepika Kishor Nagthane
Assistant Professor, MVJ College OF Engineering Bangalore, India
dnagthane@gmail.com

Abstract

Digitization in every sector leads to the growth of digital data in a tremendous amount. Digital data are not only available in the form of text but it is also available in the form of images, audio and video. Decision making people in every field like business, public sector, hospital, etc. are trying to get useful and implicit information from the already existing digital data bases. Image mining is the concept used to extract implicit and useful data from images stored in the large data bases. Image mining is used in variety of fields like medical diagnosis, space research, remote sensing, agriculture, industries and even in the educational field.

Valuable information can be hidden in images. The need for image mining is high in view of the fast growing amounts of image data. In this paper, we first point out unique characteristics of image mining, then analyze the overall process and discuss the main technology of image mining, namely, image classification and clustering, association rule mining. Some applications in various areas are introduced. Finally, some future research directions and problems of image mining are presented.

Keywords: Image mining, preprocessing, image classification, image clustering Association rules mining

Introduction

Discovering knowledge from data stored in typical alphanumeric databases, such as relational databases, has been the focal point of most of the work in database mining. However, with advances in secondary and tertiary storage capacity, coupled with a relatively low storage cost, more and more non standard data(e.g., in the form of images) is being accumulated. This vast collection of image data can also be mined to discover new and valuable knowledge. The problem of image mining combines the areas of content-based image retrieval, image understanding, data mining and databases. To our knowledge, no other work has been done with regard to mining knowledge from a collection of images from a database perspective. An initial step towards tapping into the undiscovered wealth of knowledge from mining image-bases is the focus of this paper. This work can also be seen as a starting point for an as yet, unexplored area that that can provide enormous benefits.

Image mining has two main themes. The first is mining large collections of images and the second is the combined data mining of large collections of image and associated alphanumeric data. The data mining objective might be to find if there is some pattern that exists for an individual city (over time) or if there is some pattern that exists between different cities. An example of the second case might involve medical imagery and patient (alphanumeric data) records. To develop an accurate diagnosis or prognosis both image data (such as Xrays, SPECT, etc.) and patient data (such as weight, prior health conditions, family history, etc.) can be examined together to find interesting associations.

In present scenario, image plays vital role in every aspect of business such as business images, satellite images, medical images and so on. If we analysis these data, which can reveal useful information to the human users. But, unfortunately there are certain difficulties to gather those data in a right way [1]. Due to incomplete data, the information gathered is not processed further for any conclusion. In another end, Image retrieval is the fast growing and challenging research area with regard to both still and moving images. Many Content Based Image Retrieval (CBIR) system prototypes have been proposed and few are used as commercial systems. CBIR aims at searching image databases for specific images that are similar to a given query image. It also focuses at developing new techniques that support effective searching and browsing of large digital image libraries based on automatically derived imagery features. It is a rapidly expanding The CBIR focuses on Image ‘features’ to enable the query and have been the recent focus of studies of image databases. The features further can be classified as low-level and high-level features. Users can query example

ISSN: 2277-9655
Impact Factor: 1.852

images based on these features such as texture, colour, shape, region and others. By similarity comparison the target image from the image repository is retrieved. Meanwhile, the next important phase today is focused on clustering techniques. Clustering algorithms can offer superior organization of multidimensional data for effective retrieval. Clustering algorithms allow a nearest-neighbour search to be efficiently performed. Hence, the image mining is rapidly gaining more attention among the researchers in the field of data mining, information retrieval and multimedia databases. Image mining presents special characteristics due to the richness of the data that an image can show. Effective evaluation of the results of image mining by content requires that the user point of view (of likeness) is used on the performance parameters .

Typical CBIR Technique Applied for Medical Images

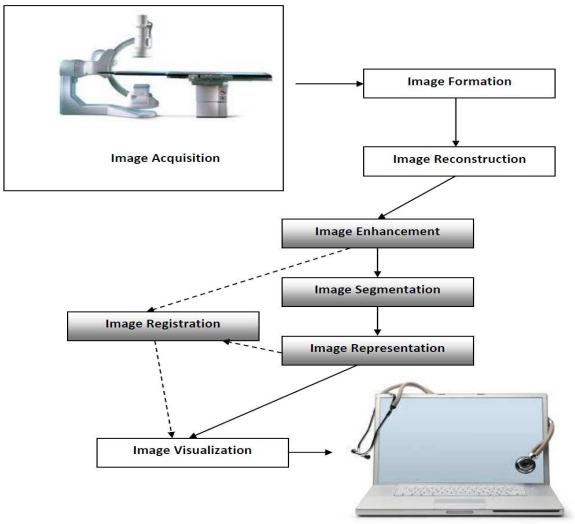

Fig 2.1: Medical Image Analysis

After acquisition of images by using one of the above technologies for putting automation in place, various image processing and image analysis steps should be followed as shown in figure 2.1. Image formation and image reconstruction relates to the formation of image whereas Image visualization takes care of visualizing the various aspects of the image. Other steps are discussed in following paragraphs.

Image enhancement is among the simplest and most appealing areas of digital image processing. Basically, the idea behind enhancement techniques is to bring out detail that is obscured, or simply to highlight certain features of interest in an image. It is a very subjective area of image processing. Even in case of medical image processing, image enhancement technique suitable for one image modality may not be suitable for other modality.

Image segmentation is a broad and active field, not only in medical imaging, but also in computer vision and satellite imagery. Its purpose is to divide an image into regions which are meaningful for a particular task. Autonomous segmentation is one of the most difficult tasks in digital image processing. A rugged segmentation procedure brings the process a long way toward successful solution of imaging problems that require objects to be identified individually. On the other hand, weak or erratic segmentation algorithms almost always guarantee eventual failure. In general, the more accurate the segmentation, the more likely recognition is to succeed. Thus segmentation of medical image is the most important step as it may result in success or failure of overall automation.

Representation almost always follow the output of a segmentation stage, which usually is raw pixel data, constituting either the boundary of a region (i.e., the set of pixels separating one image region from another) or all the points in the region itself. In either case, converting the data to a form suitable for computer processing is necessary. The first decision that must be made is whether the data should be represented as a boundary or as a complete region. Boundary representation is appropriate when the focus is on external shape characteristics, such as corners and inflections. If we consider example of one of the musculoskeletal image, it is suitable for focusing delineation between bone and cartilage. It is also suitable for identifying meniscus tears on the knee image. Regional representation is appropriate when the focus is on internal properties, such as texture or skeletal shape. This representation is suitable in case of human knee cartilage volume quantification.

Image registration is the process of matching two or more similar images. The images may be multiple snapshots of the same object with same or different viewing angle/direction. These images may be acquired by same device or different devices at same or different time. Basically, registration is a process of determining a coordinate transformation between two images that are misaligned.

Among these steps the most challenging steps for image analysis of medical images are image *enhancement*, image *segmentation*, image *representation* and image *registration*.

ISSN: 2277-9655
Impact Factor: 1.852

Main Technology Analyze

A. Preprocessing

Image mining deals with large collection of image datasets that are high-dimensional and have multiple features, so time and space cost are relative high when analysis them. In order to improve quality and efficiency of the following mining steps, it is vital to discover suitable preprocessing technologiesto clean up the un-related data and make useful hidden information more obvious. Traditional image processing technologies are applied to the image data ready to be mined. Some image preprocess has been introduced, for instance, Han propose a palm-print-based identification system in [1], thepre-processing steps including image-thresholding, bordertracing and wavelet-based segmentation, the preprocessing method is proved to be effective and can be simulated in other scenarios as well.

B. Feature extraction

One of the key problems is how to express image data, which can usually be represented by features such as texture, color, edge, shape. According to the mining object, extract the basic elements that can present the images, omit features inessential to mining result. In some cases, to get better mining result, it is necessary to converge many features to form multidimensional feature vectors. Color, edge, texture are very important features in image mining and are widely used. [2] presents a feature extraction method that uses a combination of features: color, edge, texture, the method achieves high recall and precision.

C. Image mining technology

Most commonly used image mining technologies are image classification, image clustering, mining association rules and neural network.

1) Image classification

Image classification is to do quantity analysis on image, and , it is a supervised learning method, a set of pre-labeled images are provided, then based on prior knowledge, tag the new images with suitable labels, often there are three processes involved in image classification:(1): Feature extraction, first build up a image representation model, extract features from sample images that are already labeled and establish feature description for each image; (2): Train the samples of each class and establish model description for each class; (3): Use the model to classify and index images that are not labeled. The most commonly used image classification technologies are as follows: Bayes, neural network, decision tree, support vector machine, K-nearest -neighbor-classifier, genetic algorithm etc. Performance of a classifier is normally measured by accuracy of prediction, speed, robustness, extensibility.

2) Image clustering

The process of grouping a set of images into classes of similar images without prior knowledge is called image clustering. It is a unsupervised learning method, images within a cluster have high similarity in comparison to one another but are very dissimilar to images in other clusters. The process normally comprises of 4 steps: (1) Image representation, feature extraction and selection; (2) Set up similarity metrics suitable for special application; (3) Image clustering; (4) Form cluster. After clustering, field experts are required to examine each cluster and label it with abstract concepts. Nowadays there are many clustering algorithms such as: partitioning methods, Association rule mining finds interesting associations and correlation relationships among large set of image data. Association rules show attribute value conditions that occur frequently together in a given dataset. Association rule mining consists of first finding frequent item sets, from which strong association rules are generated. These rules also satisfy a minimum confidence threshold. In recent years, association rule mining has become a heated problem and draws upon wide attention. Apriori is a typical algorithm for association rule mining, based on which there are a lot of improvement, for example, in [3], Anthony proposes a novel spatial mining algorithm to mine the spatial association rules from an image database, which prunes most of impossible candidates in the mining process, thus the proposed algorithm is more efficient than the Apriori algorithm. The experiment results show that it runs 2–5 times faster than the Apriori algorithm. hierarchical methods, grid-based methods, model-based methods, etc.

3) Association rules mining

Association rule mining finds interesting associations and correlation relationships among large set of image data. Association rules show attribute value conditions that occur frequently together in a given dataset. Association rule mining consists of first finding frequent item sets, from which strong association rules are generated. These rules also satisfy a minimum confidence threshold. In recent years, association rule mining has become a heated problem and draws upon wide attention. Apriori is a typical algorithm for association rule mining, based on which there are a lot of improvement, for example, in [3], Anthony proposes a novel spatial mining algorithm to mine the spatial association rules from an image database, which prunes most of impossible candidates in the mining process, thus the proposed algorithm is more efficient than the Apriori algorithm. The experiment results show that it runs 2–5 times faster than the Apriori algorithm.

ISSN: 2277-9655
Impact Factor: 1.852

Conclusions

Image mining is a promising field for research, however, it is still at the beginning and for future development the following problems need to be considered: Improvement on image preprocessing technologies, including feature extraction, image segmentation, object identification. Propose an united image representation model and representation techniques. Devise highly efficient and extensible image mining algorithms, for traditional techniques are hard to directly apply on image database. Introduce domain knowledge into image mining, which are essential for understanding mining results.

References

[1] Chin-Chuan Han, Hsu-Liang Cheng, et al. Personal authentication using palm-print features. Pattern Recognition 36 (2003) 371 – 381.
[2] Deepak Kolippakkam*, Huan Liu, et al. Feature Extraction for Image Mining.
[3] Anthony J.T. Lee, Ruey-Wen Hong, et al. Mining spatial association rules in image databases. Information Sciences 177 (2007) 1593–1608.
[4] P.Rajendran,M.Madheswaran.An Improved Image Mining Technique For Brain Tumour Classification Using Efficient classifier. (IJCSIS) International Journal of Computer Science and Information Security,Vol. 6, No. 3, 2009.
[5] Asanobu K . Data Mining for Typhoon Image[C] Proc on MDM/ KDD2001. San Francisco ,USA ,2001 :68277.
[6] Chang E , Chen Li , et al . Searching Near-Replicas of Images via Clustering [C]//Proc of SPIE. Boston ,MA ,1999 :2812292. 1870
[7] Klas W et aI, Using an Object-Oriented Approach to Model Multimedia Data, Computer Communications, I3(4) , I 990
[8] Ling CX, Li C. "Data mining for direct marketing: problems and solutions. " In: Proceedings of KDD'98, pp73-79
[9] Peng Hua, Xu Luping. Improved genetic FCM algorithm for color image segmentaion. International Conference on Signal Processing Proceedings, ICSP, v 2, 2007, 8th International Conference on Signal Processing, ICSP 2006.
[10] Yu Jian, Yang Miin-shea. A study on a generalized FCM. Rough Sets, Fuzzy Sets, Data Mining and Granular Computing,p390-393,2003.
[11] Junwei Tian, Yongxuan Huang. Histogram constraint based fast FCM cluster image segmentation. ISIE2007 ,Proceedings,p I 623-I 627.
[12] Srinivasan Dipti, Choy Min Chee. Neural networks for real-time traffic signal control. IEEE Transactions on Intelligent Transportation Systems, v7,n3,p261 -272,September 2006.
[13] Daggett T, Greenshields IR. A cluster computer system for the analysis and classification of massively large biomedical image data. Comput Bioi Med. 1998 Jan;28(l)47-60.
[14] Gunsel B, Umut U, Tekalp AM. "Robust watermarking of fingerprint images." Pattern Recognition 2002;35:2739-47.
[15] Almejalli, Khaled. Intelligent traffic control decision support system. Applications of Evolutionary Computing - EvoWorkshops p 688-701, 2007.
[16] M. Molina, 1. Hernbndez, and 1. Cuena, "A structure of problemsolving methods for real-time decision support in traffic control", International Journal of Human-Computer Studies, vol. 49, no. 4, pp. 577-600, Oct. 1998.
[17] AW. Sadek, MJ. Demetsky, and RL. Smith, "Case-based reasoning for real-time traffic flow management", Computer-Aided Civil and Infrastructure Engineering, vol. 14, no. 5, pp. 347-356,Sept. 1999
[18] Muhammad KK, Zhang J, Tian L. "Protecting biometric data for personal identification." In: Sinobiometrics'04. Lecture
[19] Notes in Computer Science, vol. 3383.springer-Verlag;pp. 629-38, 2004.