*Available online a*t www.scientiaresearchlibrary.com

Scientia Research Library **ISSN 2348-0424**
USA CODEN: JETRB4

Journal of Engineering And Technology Research,
2014, 2 (5):8-21
(http://www.scientiaresearchlibrary.com/arhcive.php)

Image Mining Architecture, Techniques and Algorithms

Balu Ramu
Faculty, Department of Computer Applications
School of Computer Science and Engineering,
Bharathiar University, Coimbatore

ABSTRACT
In this paper highlights a four-level information-driven framework for image mining systems, various methods available from literature, historical background, related work, algorithms and techniques that are frequently used in image mining, namely object recognition, image retrieval, image indexing, image classification and clustering, association rule mining and neural network.

Keywords: Image Mining, Framework, Techniques, Algorithms

INTRODUCTION

Progresses in image acquirement and storage technology have led to an incredible growth in very large and detailed image databases. These images, if analyzed, can expose useful information to the human users. Image mining deals with extracting inherent and embedded knowledge, image data relationship, or other patterns which are not explicitly found in the images [Zhang et al., 2001]. Zhang et al., (2002) have examined the research issues in image mining, developments in image mining, predominant image mining frameworks and suggested some future research guidelines for image mining.

The fundamental challenge in image mining is to reveal out how low-level pixel representation enclosed in a raw image or image sequence can be processed to recognize high-level image objects and relationships. Zhang et al., (2001) have proposed an efficient information-driven framework for image mining that includes four levels of information: Pixel Level, Object Level, Semantic Concept Level, and Pattern and Knowledge Level. To achieve that High-dimensional indexing schemes, retrieval techniques are incorporated in the framework to maintain the flow of information. Zhang et al., (2001) have highlighted the need for image mining in the era of rapidly growing amounts of image data and pointed out the unique characteristics of image databases. In addition, it also examines function-driven and information-driven frameworks for image mining.

Image mining normally deals with the study and development of new technologies that allow to accomplish this subject. Image mining is not only the simple fact of recovering relevant images; the aim is the innovation of image patterns that are noteworthy in a given collection of images. Fernandez et al., (2001) show how a natural source of parallelism provided by an image can be used to reduce the cost and overhead of the whole image mining process.

The remainder of this paper is organized as follows: Section II describes an overview of the Image Mining Framework, Section III represents various Existing Methods for Image Mining, Section IV demonstrates a Image Mining and Relevant Frameworks, Section V shows a Historical Background of Image Mining; Section VI describes various Image Mining Techniques. Section VII details about the Image Mining Algorithms, Section VIII concludes the paper.

MATERIALS AND METHODS

Image Mining Framework

An image mining system is often complicated as it employs various approaches and techniques ranging from image retrieval and indexing schemes to data mining and pattern recognition. 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 perspectives [Zaiane et al., 1998, Burl 1999, Carlos Ordonez et al., 1998]. While such functional-based framework is easy to understand, it fails to emphasize the different levels of information representation necessary for image data before meaningful mining takes place.

Figure 1 shows information-driven framework for image mining. Inputs from domain scientists are needed to identify domain specific objects and semantic concepts. At the Pixel Level, authors are dealing with information relating to the primitive features such as color, texture and shape. At the Object Level, simple clustering algorithms and domain experts help to segment the images into some meaningful regions/objects. At the Semantic Concept Level, the objects/regions identified earlier are placed in the context of the scenes depicted. High-level reasoning and knowledge discovery techniques are used to discover the interesting patterns. Finally, at the Pattern and Knowledge Level, the domain-specific alphanumeric data are integrated with the semantic relationships discovered from the images and further mining are performed to discover useful correlations between the alphanumeric data and those found in the images. Such correlations discovered are particularly useful in the medical domain [Zhang et al., 2001].

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 is, perhaps, the most widely used visual features in most image management database system. Color is widely represented by its RGB values. Subsequent improvements include the use of cumulative color histogram [Stricker et al., 1995], and spatial histogram intersection [Stricker et al., 1996]. Texture is the visual pattern formed by a sizable layout of color or intensity and homogeneity. It contains important information about the structural arrangement of surfaces and their relationship to the surrounding environment [Rui et al., 1997]. Common representations of texture information include: the co-occurrence matrix representation proposed by Haralick et al. (1973), the coarseness, contrast, directionality, line likeness, regularity, and roughness measures proposed By Tamura et al., (1978), the use of Gabor filter [Manjunath et al., 1996], and fractals [Kaplan 1998].

Ma et al., (1998) have developed a texture thesaurus that was able to automatically derive code words representing important classes of texture within the collection. Edge information is an important visual issue to the detection and recognition of objects in an image. Researchers try to identify a small subset of primitive features that can uniquely distinguish images of one class from another class. While there has been some success in improving the retrieval precision and recall of images, researchers realize that primitive image features have their limitations. In particular, the primitive image features are typically global and they do not have the concept of objects/regions as perceived by a human user. This lack of objects/regions concept means that the pixel level is unable to answer simple queries such as “retrieve the images with a girl and her dog” and “retrieve the images containing blue stars arranged in a ring” [Zhang et al., 2001].

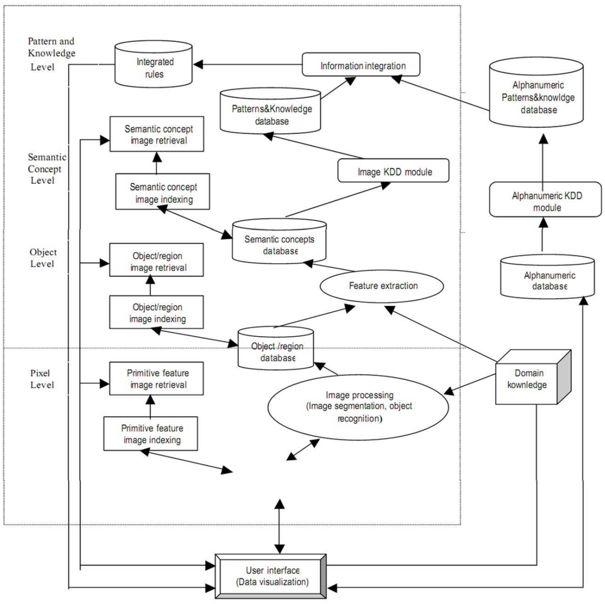

Figure 1 An information-driven image mining framework [Zhang et al., 2001]

In general, an object recognition module consists of four components: model database, feature detector, hypothesizer and hypothesis verifier [Jain et al., 1995]. The model database contains all the models known to the system and the 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. 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. Object recognition is closely tied to image segmentation. To improve the accuracy of object recognition, image segmentation is performed on partially recognized image objects rather than randomly segmenting the image. In addition, several techniques such as characteristic maps, machine learning and common objects have been proposed to improve object recognition rate. Characteristic maps locates a particular known object in images [Bonet 2000], machine learning techniques generate recognizers automatically [Burl 1999] and common objects find images by using a set of examples already labeled by the domain experts [Gibson 2001].

While objects are the fundamental building blocks in an image, there is “semantic gap between the Object level and Semantic Concept level. Abstract concepts such as happy, sad and the scene information are not captured at the Object level. Such information requires domain knowledge as well as state-of-the-art pattern discovery techniques to uncover useful patterns that are able to describe the scenes or the abstract concepts [Zhang et al., 2001]. To support all the information needs within the image mining framework, author need the fourth and final level pattern and knowledge level. At this level, authors are concerned with not just the information derivable from images, but also all the domain-related alphanumeric data. The key issue here is the integration of knowledge discovered from the image databases and the alphanumeric databases.

IRIS, an integrated retinal image information system developed in 2000 at the School of Computing, National University of Singapore tried to integrate both patient data and the corresponding retinal images to discover interesting patterns and trends on diabetic retinopathy in the local population, and the risk factors for disease occurrence and disease progression [Hsu et al., 2000]. Brain-Image Database is another image mining system developed to discover associations between structures and functions of human brain [Megalooikonomou et al., 1999].

Existing Methods for Image Mining

Magnetic resonance imaging of the brain, followed by automated segmentation of the corpus callosum in mid sagittal sections has momentous applications in neurology and neuro cognitive study since the size and shape of the Corpus Callosum are shown to be interrelated to sex, age, neuro degenerative diseases and various lateralized behaviour in people. The segmentation of the Corpus Callosum is considered as a crucial step in image mining frameworks to classify the brain MR images. Ashraf Elsayed et al., (2009) have proposed a segmentation algorithm that operates by first extracting regions satisfying the statistical characteristics of the Corpus Callosum that have relatively high intensity values.

Hsu et al., (2000) have proposed a system called IRIS, an Integrated Retinal Information system, which has been built up to afford medical professionals trouble-free and unified access to the screening, trend and development of diabetic-related eye diseases in a diabetic patient database. Rajendran et al., (2010) have proposed a method that deals with the detection of brain tumor in the computed tomography scan brain images. First preprocessing technique applied on the images eliminates the inconsistent data from the computed tomography scan brain images. Then feature extraction process is applied to extort the features from the brain images. A novel fuzzy association rule mining applied on the image transaction database contains the features extracted from the computed tomography scan brain images. The novel fuzzy association rule mining gives the diagnosis keywords to physicians for making a better diagnosis system. Automated detection of tumors in different medical images is motivated by the inevitability of high accuracy when author is dealing with human life.

Jeba Sheela et al., (2005) have proposed a system using image mining techniques to categorize the images either as normal or abnormal and then classify the tissues of the abnormal brain magnetic resonance imaging to identify brain related diseases.

Rajendran et al., (2010a) have proposed a method, which is concerned with the classification of brain tumor in the computed tomography scan brain images. The main steps involved in the system are: pre-processing, feature extraction, association rule mining and hybrid classifier. The pre-processing is done using the median filtering and edge features that are extracted using canny edge detection technique. The combination of two image mining approach has been proposed. The frequent patterns from the computed tomography scan images are produced by frequent pattern tree algorithm that mines the association rules. The decision tree method is used to categorize the medical images for diagnosis. This system improves the classification process to be more accurate and the hybrid method enhances the efficiency of the proposed method than the traditional image mining methods.

Aswini Kumar Mohanty et al., (2010) have applied image mining in the domain such as breast mammograms to classify and detect the cancerous tissue. A hybrid approach of feature selection using fast branch and bound algorithm and a hybrid genetic algorithms are used which approximately reduces 75% of the features and new decision tree is used for classification and provide promising results. Balu et al., (2010) are exploring the image mining in depth in order to propose algorithms for improving the efficiency and effectiveness of image mining.

Image Mining and Relevant Frameworks

Several image mining systems have been developed for different applications: The MultiMediaMiner mines high-level multimedia information and knowledge from large multimedia database [Zaiane et al., 1998]. Datcu et al., (2000) describe an intelligent satellite mining system that comprises of two modules: a data acquisition, preprocessing and archiving system which are responsible for the extraction of image information, storage of raw images, and retrieval of image, and an image mining system. The Diamond Eye is an image mining system that enables scientists to locate and catalog objects of interest in large image collections. These systems incorporate novel image mining algorithms, as well as computational and database resources that allow users to browse, annotate, and search through images and analyze the resulting object catalogs. The architectures in these existing image mining systems are mainly based on module functionality [Burl 1999].

Carlos Ordonez et al., (1999), provide a different perspective to image mining with four level information image mining framework. As per Carlos Ordonez et al., (1999), the primary focus of applications are the Pixel and Object level. Zaiane et al., (1999) have focused on MutiMediaMiner at the Semantic Concepts level with some brief mentions of the supports from the Pixel and Object levels. Burl (1999) Diamond Eye system primarily stresses on the Pixel level information. It is mentioned that by proposing a framework based on the information flow, authors are able to focus on the critical areas to ensure that all the levels can work together seamlessly. More research is needed at the Semantic Concept level and the Knowledge and Pattern level.

Aura Conci et al., (2001) have proposed a framework for mining images by colour content. Androutsos et al., (1999) have presented a scheme which implements a recursive HSV-space segmentation technique to identify perceptually prominent color areas. The average color vector of these extracted areas is then used to build the image indices, requiring very little storage. Sanjay Silakari et al., (2009) have proposed a framework that focuses on color as feature using color moment and block truncation coding to extract features for image dataset. The K-Means clustering algorithm is used to group the image dataset into various clusters. Sangkyum Kim et al., (2010) have proposed discriminative frequent pattern-based image classification.

Rajshree (2010) describes about an image mining techniques which is based on the color histogram, texture of that image. Carlos Ordonez et al., (1998) have presented a data mining algorithm to find association rules in 2-dimensional color images. The algorithm has four main steps: feature extraction, object identification, auxiliary image creation and object mining and it does not rely on any type of domain knowledge. Vitorino Ramos et al., (2000) have formulated the segmentation problem upon images as an optimization problem and espouses evolutionary strategy of Genetic Algorithms for the grouping of small regions in colour feature space. The approach uses k-Means unsupervised clustering methods into genetic algorithms, for guiding Evolutionary Algorithm in the search for finding the finest or best data partition. Mohan et al., (2010) have proposed color image classification and retrieval using an image for improving user interaction with image retrieval systems by fully exploiting the similarity information. In that technique, retrieving the images from the image collection involves the steps like preprocessing, color image classification, preclustering, texture feature extraction, similarity comparison and neighboring target image selection.

Historical Background of Image Mining

Hassan Malik (2005) provides a simple and easy to learn language that facilitates mining association rules from images called iARM. iARM is a scripting language that makes it easier to dig out association rules from images. It allows one to make a list of source image files and customize association rule parameters similar to number of terms, filters, support and confidence. iARM can be customized according to the end-user needs and can extract flexible rules utilizing both textual and signal features. Extracted rules symbolize implicit knowledge contained in images, and implicit relationship that exists in a set of images. This extracted data could further assist in classification and clustering of images.

Guadagnina et al., (2010) have proposed an integration of geoprocessing and image mining to support image based decisions in several domains such as healthcare. Ana Benitez et al., (1998) describe MetaSEEk, a meta-search engine used for extracting images based on their visual content on the Web. MetaSEEk has been designed to astutely select and interface with multiple on-line image search engines and this is carried out by ranking their performance for diverse classes of user queries. User response is also integrated in the ranking refinement. MetaSEEk has been built up to discover the issues involved in querying large, distributed, on-line visual information system sources.

Nick Morsillo e al., (2008), have presented a technique that allows a user to reduce noisy search results and characterize a more precise visual object class. This approach is based on semi-supervised machine learning in a novel probabilistic graphical model made of both generative and discriminative elements. Keiji Yanai (2008) proposes a new method to select relevant images to the given keywords from images acquired from the web based on the probabilistic latent semantic analysis model. It is proved that this method can select various images compared to the existing support vector machine based methods.

Bingbing Ni et al., (2009) have presented an automatic image mining system in web which builds a universal human age estimator based on facial information, which can be used to all racial groups and various image qualities.

Keiji Yanai (2005) has proposed a new method for automated large scale gathering of web images relevant to specified concepts. In that, good quality candidate sets of images for each keyword are collected as a function of analysis of the surrounding HTML text. The gathered images are then segmented into regions of images by using a Gaussian mixture. Ruhan He et al., (2009) investigated the multi-modal associations between two basic modalities of Web images, i.e. keyword and visual feature clusters, by multi-model association rule. Automatic image classification is a demanding research topic in web image mining.

Rong Zhu et al., (2009) have formulated image classification problem as the calculation of the distance measure between training manifold and test manifold. The author has proposed an improved nonlinear dimensionality reduction algorithm based on neighborhood optimization to decrease feature dimensionality and complexity. Since the web images are not annotated, it very difficult to get user intended image from web.

Zheng Chen et al., (2008) have presented an effective approach and a prototype system for image retrieval from the internet using web mining. This system can serve as a web image search engine. Key idea in this approach is to extract the text information on the web pages to semantically describe the images. The text description is then pooled with other low-level image features in the image similarity assessment.

Chunjie Zhang et al., (2009) have proposed a concept in which Google Image Searcher is used to find the relevant images and then the output image set is explored to learn sensitive markov stationary feature to represent images by the algorithm of random walk with restart, in which the spatial co-occurrence of the bag-of-words representation and the concept information are integrated and classified with a support vector machine based classifier for web image mining. Seong-Yong Hong (2009) has presented an intelligent web e-catalog image retrieval system using metadata and user log. The texture and color based image classification and indexing techniques like bit vector indexing are used to represent schemes of user usage patterns. Mohamed Eldib et al., (2011) have introduced a fully automated age estimation engine that is competent of collecting images using human age related text queries from flickr photo sharing website.

Wichian Premchaiswadi et al., (2010) have presented an online content based image retrieval system using joint query and relevance feedback concept based on high level and low level features. This approach has also used fast and efficient color feature extraction namely auto color correlogram for extracting and indexing low-level features of images and correlation based on color correlogram. To integrate an image analysis algorithm into the text-based image search engines without debasing their response time, the framework of multi-threaded processing has been proposed.

Image Mining Techniques

Brown et al., (2005) have described in detail a general hierarchical image classifier approach, and illustrated the ease with which it can be trained to find objects in a scene using support vector machine concept. Ji Zhang et al., (2001) have proposed new representation schemes for visual patterns and to facilitate fast and effective access needed to frame efficient content-based image indexing and retrieval techniques. Victor et al., (2010) have proposed an efficient retrieval technique for images using enhanced univariate transformation approach. In this method they have treated images as a compilation of the representative prototypes selected from the training image corpus, and then used the resulted distribution in the descriptor space as a characterization of the image.

Peter Stanchev (2003) has proposed a new method for image retrieval using high level semantic features. It is based on removal of low level color, shape and texture characteristics and their translation into high level semantic features using fuzzy production rules derived using an image mining technique. Jen-Hao Hsiao et al. (2008) have proposed a novel language model based approach with pseudo-relevance feedback for tackling the vocabulary problem in visual mining. Iwaszko et al., (2010) have suggested a novel general approach applicable to image mining and retrieval, using compact geometric structures which can be recomputed from a database. Rosalina Abdul Salam et al., (2004) have proposed a technique to test the capability of producing an automatic shape recognition system by mining relevant image features. The method has the capability to be extended to three-dimensional objects, which are currently under investigation. Color, depth and texture can be grouped together to form a set of new features.

Kannan et al., (2010) have formulated a new technique called image retrieval based on optimal clusters is proposed for improving user interaction with image retrieval systems by fully exploiting the similarity information. Dipesh Dugar et al., (2010) have proposed a novel discriminative learning framework based on canonical correlation for object recognition with image sets. Kun-Che Lu et al., (2009) have adopted the decision tree induction to recognize relationships between attributes and the target label from image pixels, and constructed a model for pixel-wised image processing according to a given training image dataset. And it is proved that the proposed model can be very competent and effectual for image processing and mining. Victor et al., (2010) have proposed minimum spanning tree based clustering algorithm using weighted Euclidean distance for edges to segment the image. John Peter (2010) proposes a novel algorithm, minimum spanning tree based Structural similarity clustering for image mining with local region outliers to segment the given image and to detect anomalous pattern.

Ashok Srivastava et al., (2004) have presented a methodology for automatic knowledge driven image mining using the theory of Mercer Kernels. Mercer Kernels are extremely nonlinear symmetric positive definite mappings from the original image space to a very high, probably infinite dimensional feature space. Rupali Sawant (2010) has presented a framework of image mining based on concept lattice and cloud model theory. The methods of image mining from image texture and shape features are introduced here, which include the basic steps: pre-processing the images, using cloud model to extract concepts, and then using concept lattice to extract a series of image knowledge. Early image miners have attempted to use existing techniques to mine for image information. The following techniques are applied to image mining: object recognition, image indexing and retrieval, image classification and clustering, association rules mining, and neural network.

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. The object recognition problem can be referred to as a supervised labeling problem based on models of known objects
[Zhang et al., 2002].

An object recognition system typically consists of four components: model database, feature detector, hypothesizer and hypothesis verifier. The model database contains all the models known to the system and 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. 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. In order to locate a particular known object in an image or set of images, Bonet (2000) has designed a system that processes an image into a set of "characteristic maps". Burl (1999) employed learning techniques to generate recognizers automatically and domain knowledge is captured implicitly through a set of labeled examples. Stephen Gibson et al., (2001) have developed an optimal Fast Fourier Transform (FFT) based mosaicing algorithm to find common patterns in images and show that it works well on various kinds of images.

Image Retrieval

Image mining requires that images be retrieved according to some requirement specifications. The requirement specifications can be classified into three levels of increasing complexity [Burl 1999]. Level 1 comprises image retrieval by primitive features such as color, texture, shape or the spatial location of image elements. Level 2 comprises image retrieval by derived or logical features like objects of a given type or individual objects or persons. Level 3 comprises image retrieval by abstract attributes, involving a significant amount of high level reasoning about the meaning or purpose of the objects or scenes depicted.

Rick Kazman et al., (1993) have proposed three query schema for image retrieval: query by associate attributes, query by description and query by image content. In practice, this approach suffers from the drawbacks of the "vocabulary problem" and non-scalability. With the emergence of large-scale image repositories, the problems of vocabulary and non-scalability faced by the manual annotation approach have become more pronounced. Content-based image retrieval is thus proposed to overcome these difficulties. There are three fundamental bases in content based image retrieval, namely, visual information extraction, image indexing and retrieval system application [Rui et al., 1997]. Many techniques have been developed in this direction, and many image retrieval systems, both research and commercial, have been built. Commercially, IBM's Query By Image and Video Content (QBIC) system is probably the best known of all image content retrieval systems [Flickner et al., 1995).

It offers retrieval by any combination of color, texture or shape, as well as text keyword. Virage is another well-known commercial system [Bach et al., 1996] and is available as a series of independent modules, which system developers can build into their own programs. Excalibur by virtue of its company's pattern recognition technology offers a variety of image indexing and matching techniques [Feder et al., 1996]. There are also a large number of University prototypes and experimental systems available, the representative ones being Photobook [Pentland et al., 1994], Chabot [Datcu et al., 2000], VisualSEEk [Smith et al., 1996], MARS [Mehrotra et al., 1997], Surfimage [Nastar et al., 1998] and Synapse [Manmatha et al., 1997].

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 search complexity. 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 update algorithm and clustering [Salton et al., 1983].

High-dimensional indexing schemes include SR-tree [Katayama et al., 1997], TV-tree [Lin et al., 1994] X-tree [Berchtold et al., 1996] and iMinMax [Ooi et al., 2000]. Current image systems retrieve images based on similarity. The filters operate on an approximation of the high dimension data that represents the images and reduces the search space so that the computationally expensive comparison is necessary for only a small subset of the data. Haykin (1998) has presented a new compressed image indexing technique by using compressed image features as multiple keys to retrieve images and other indexing schemes focus on specific image features.

Lin et al., (1994) give an efficient color indexing scheme for similarity-based retrieval which has a search time that increases logarithmically with the database size. Tan et al., (2001) have proposed a multi-level R-tree index, called the nested R-trees for retrieving shapes efficiently and effectively. With the proliferation of image retrieval mechanisms, a performance evaluation of color-spatial retrieval techniques that serves as a guideline to select a suitable technique is given by Tan et al., (2001).

Image Classification and Image Clustering

Image classification and clustering are the supervised and unsupervised classification of images into groups. In supervised classification, author is given a collection of labeled images, and the problem is to label a newly encountered, yet unlabeled images. Typically, the given labeled images are used to do the machine learning of the class description which in turn are used to label a new image. In unsupervised classification, the problem is to group a given collection of unlabeled images into meaningful clusters according to the image content without a priori knowledge [Jain et al., 999]. Uehara et al., (2001) recognize the challenge that lies in grouping images into semantically meaningful categories based on low-level visual features. Currently, there are two major types of classifiers, the parametric classifier and non-parametric classifier.

Bruzzone et al., (2001) develop a variety of classifiers to label the pixels in a landset multispectral scanner image. MM-Classifier, a classification module embedded in the MultiMedia Miner developed by Zaiane et al., (1998) classifies multimedia data, including images, based on some provided class labels. James Z Wang et al., (1997) have proposed Image-based Classification of Objectionable Websites (IBCOW) to classify whether a website is objectionable or benign based on image content. Carpenter et al., (1998) have applied clustering methods such as k-means and the self-organizing map for visualizing the distribution of typhoon cloud patterns on a two-dimensional space and the following generic steps are required in the image classification and clustering: (a) Pattern representation: this may involve image processing such as image segmentation, feature extraction and selection; (b) Definition of image proximity measure appropriate to the domain; (c) Classification or clustering; (d) Group abstraction or adaptation.

Association Rule Mining

An association rule is an implication of the form X□ Y, where X, Y I and X Y=f, I is the set of objects, also referred as items D is a set of data cases. X is called the antecedent and Y is called the consequent of the rule. A set of items, the antecedent plus the consequent, is called an item set. The rule X□ Y has supports in D if s% of the data case in D contains both X and Y, and the rule holds in D with confidence c if c% of the data base in D that supports X also Supports Y. Association rule mining generates rules that have supported and confidence greater than some user specified minimum support and minimum confidence thresholds. A typical association rule mining algorithm works in two steps. The first step finds all large item sets that meet the minimum support constraint. The second step generates rules from all the large item sets that satisfy the minimum confidence constraint
[Zaiane et al., 1998].

Association rule mining is frequently used in data mining to uncover interesting trends, patterns and rules in large datasets. Recently, association rule mining has been applied to large image databases [Megalooikonomou et al., 1999]. Although the current image association rule mining approach is far from mature and perfection, compared to its application in data mining field, there opens up a very promising research direction and vast room for image association rule mining. There are two main approaches: the first approach is to mine from large collections of images alone, and the second approach is to mine from a combined collection of images and associated alphanumeric data. Association mining from transaction database is a typical case of mining association rules from large database. Ordonez et al. present an image mining algorithm using blob needed to perform the mining of associations within the context of images [Ordonez et al., 1999].

A prototype has been developed in Simon Fraser University called Multimedia Miner where one of its major modules is called MM-Associator and it uses 3- dimensional visualization to explicitly display the associations [Zaiane et al., 1998]. Antonie et al., (2001) use the apriori algorithm to discover association rules among the features extracted from mammography database and category to which each mammography belongs.

Neural Networks

Artificial neural network models have been studied for many years with the hope of achieving human like performance in several fields such as speech and image understanding [Antonie et al., 2001]. A neural network, by definition, is a massively parallel distributed processor made up of simple processing units, each of which has a natural propensity for storing experiential knowledge and making the knowledge available for use [Haykin 1998]. Neural networks are fault tolerant and are good at pattern recognition and trend prediction. In the case of limited knowledge, artificial neutral network algorithms are frequently used to construct a model of the data.

A noteworthy research work that applied neural network to image mining is the artificial neural network developed by Gardner et al., which provides a wholly automated approach to fundus image analysis by computer that could improve the efficiency of the assessment work of the image by offering an immediate classification of the fundus of the patient at the time of acquisition of the image [Gardner et al., 1996]. A site mining tools, based upon the Fuzzy ARTMAP neural network [Carpenter et al., 1998] provides an intuitive means by which an image analyst can efficiently and successfully mine large amounts of multi-sensor imagery for feature foundation data [Strelilein et al., 2000].

Zhang et al., (1995) proposed to use self organization map neural nets as the tool for constructing the tree indexing structure; the advantages of using SOM were its unsupervised learning ability and dynamic clustering nature. Antonie et al (2001) exploited the use of neural networks in classification of breast cancer images using back-propagation which proved to be less sensitive the database imbalance at a cost of high training time. A three-step method is proposed for discovering the relationship between visual image features and feature of related data [Uehara et al., 2001]. User is able to make use of the computer-aided visual exportation system named MIRACLES to facilitate the tasks of feature selection and hypothesis formulation.

Image Mining Algorithms

Image mining, a broader view of data mining technique can help us find meaningful relationship among various images generated by the image mining algorithm [Carlos Ordonez et al., 1998]. The various image mining algorithms available from literature are: segmentation, searching images based on various region descriptors and refining the search operation [Nitin et al., 2010].

Image segmentation is an initial and vital step in a series of processes aimed at overall image understanding. The image is segmented into various regions and the purpose of segmentation is to partition an image into meaningful regions with respect to a particular application. The segmentation is based on measurements taken from the image and might be grey level, color, texture, depth or motion. According to the varied regions of the above mentioned image, to find the related object images which are similar to it. Compare objects in one image to objects in every other image. In this algorithm propose to find the images which are having the same objects as that of our segmented image by Nitin et al., (2010).

CONCLUSION

Image mining is currently a growing yet active research focus in computer science. In this paper, a four-level information-driven framework for image mining systems has been highlighted. High-dimensional indexing schemes and retrieval techniques are also included in the framework to support the flow of information among the levels. Various existing methods available from literature and historical background of image mining are pointed out. In addition, related work in image mining and techniques that are frequently used in the early works in image mining - object recognition, image retrieval, image indexing, image classification and clustering, association rule mining and neural network are discussed in detail. Various algorithms in image mining that are currently and are introduced. The author is exploring the image mining in depth in order to propose algorithms for improving the efficiency and effectiveness of image mining.

REFERENCES

[1]. Ji Zhang, Wynne Hsu and Mong Li Lee "An Information-Driven Framework for Image Mining" Database and Expert Systems Applications in Computer Science, 2001, Volume 2113/2001, 232-242.

[2]. J. Fernandez, N. Miranda, R. Guerrero, F. Piccoli "Appling Parallelism in Image Mining",http://www.ing.unp.edu.ar/wicc2007/trabajos/PDP/120.pdfMarch 26, 2007

[3]. Carlos Ordonez and Edward Omiecinski "Image Mining: A New Approach for Data Mining", Technical Report GITCC -98-12,1998

[4]. Swain, M J and Ballard, D H. "Color indexing" International Journal of Computer Vision, Kluwer Academic Publishers, 7(1), 11-32,1991.

[5]. Stricker, M and Orengo, M. "Similarity of color images", Proc SPIE 2420, pp 381-392,1995.

[6]. Stricker, M and Dimai, A. "Color indexing with weak spatial constraints", Proc SPIE 2670, 29-40,1996.

[7]. Y. Rui, T. S. Huang and Chang, "Image Retrieval: Past, and Future" , Int Symposium on Multimedia Information Processing, Taipei, Taiwan, Dec 11-13, 1997.

[8]. Haralick, R M and Shanmugam, K. "Texture features for image classification"IEEE Transaction on Sys, Man, and Cyb, SMC-3 (6), 1973.

[9]. Tamura, H "Textural features corresponding to visual perception",IEEE Transactions on Systems, Man and Cybernetics 8(6), 460-472,1978.

[10]. Manjunath, B S and Ma, W Y. "Texture features for browsing and retrieval of large image data" IEEE Transactions on Pattern Analysis and Machine Intelligence, 18, 837842,1996.

[11]. Kaplan, L M, "Fast texture database retrieval using extended fractal features" , Proc SPIE 3312, 162-173,1998.

[12]. P. Rajendran, M. Madheswaran " Novel Fuzzy Association Rule Image Mining Algorithm for Medical Decision Support System" 2010 International Journal of Computer Applications (0975 - 8887) Volume 1 – No. 20

[13]. L. Jeba Sheela, Dr. V.Shanthi, “Image Mining Techniques for Classification and Segmentation of Brain MRI Data” 2005 – 2007 JATIT.
[14]. Aura Conci, Everest Mathias M. M. Castro, “Image Mining by Color Content”, SEKE 2001: 405-411
[15]. D. Androutsos, K. N. Plataniotis and A. N. Venetsanopoulos “A Novel Vector-Based Approach to Color Image Retrieval Using a Vector Angular-Based Distance Measure” Computer Vision and Image Understanding Vol. 75, Nos. 1/2, July/August, pp. 46–58, 1999
[16]. Dr. Sanjay Silakari, Dr. Mahesh Motwani and Manish Maheshwari “Color Image Clustering using Block Truncation Algorithm” IJCSI International Journal of Computer Science Issues, Vol. 4, No. 2, 2009 ISSN (Online): 1694-0784 ISSN (Print): 1694-0814
[17]. Rajshree S. Dubey “Image Mining using Content Based Image Retrieval System” (IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, 2010, 2353-2356
[18]. Vitorino Ramos, Fernando Muge, “Image Colour Segmentation" by Genetic Algorithms”, RecPad´2000 - ISBN 972-96883-2-5, pp. 125-129.
[19]. Dr.V.Mohan and A.Kannan “Color Image Classification and Retrieval using Image mining Techniques” International Journal of Engineering Science and Technology Vol. 2(5), 2010, 1014-1020.
[20]. Pham, N.K. Morin, A. Gros, P. “CAViz, an interactive graphical tool for image mining“IEEE, June 2008,pages 371 - 376 ISSN: 1330-1012 ISBN: 978-953-7138-12-7
[21]. Bingbing Ni, Zheng Song, Shuicheng Yan “Web Image Mining Towards Universal Age Estimator” Proceeding MM '09 Proceedings of the 17th ACM international conference on Multimedia ACM New York, NY, USA ©2009 ISBN: 978-1-60558-608-3.
[22]. Sangkyum Kim, Xin Jin, Jiawei Han “DisIClass: Discriminative Frequent Pattern-Based Image Classification”, ACM , MDMKDD’10, 978-1-4503-0220-3, 2010
[23]. R. Guadagnina, L. Santanab, E. Fernedaa, and H. Pradoa “Improving Image Mining through Geoprocessing” Pattern Recognition and Image Analysis, 2010, Vol. 20, No. 1, pp. 81–85.
[24]. Ana B. Benitez, Mandis Beigi, and Shih-Fu Chang “A Content-Based Image Meta-Search Engine using Relevance Feedback” , IEEE , 59-69, 1998
[25]. Nick Morsillo, Chris Pal, Randal Nelson “Mining the Web for Visual Concepts”Technical Report TR-2008-931 March 18, 2008
[26]. Keiji Yanai “Automatic Web Image Selection with a Probabilistic Latent Topic Model”, ACM, April 21–25, 2008, 978-1-60558-085-2/08/04.
[27]. Keiji Yanai “Probabilistic Web Image Gathering”, MIR’05, November 10–11, 2005, Singapore, ACM 1-59593-244
[28]. Chunjie Zhang, Jing Liu, Hanqing Lu, Songde Ma “Web Image Mining Using Concept Sensitive Markov Stationary Features”,ICME 2009, 978-1- 4244-4291-1/09
[29]. Ruhan He Wei Zhan “Multi-Modal Mining in Web image retrieval”, PACIIA 2009. IEEE Issue Date: 28-29 Nov. 2009 Volume: 2 On page(s): 425 - 428 ISBN: 978-1-4244-4606-3
[30]. Rong Zhu, Min Yao and Yiming Liu, “Image Classification Approach Based on Manifold Learning in Web Image Mining” ADMA '095th International Conference on Advanced Data Mining and Applications, 2009, ISBN: 978-3-642-03347-6
[31]. Zheng Chen, Liu Wenyin, Feng Zhang, Mingjing Li, Hongjiang Zhang “Web Mining for Web Image Retrieval” , Journal of the American Society for Information Science and Technology, Volume 52 Issue 10, 2001
[32]. Seong-Yong Hong “Web Image Information Mining System Using Uup For E-Business Intelligence” Issues in Information Systems Volume X, No. 2, 2009
[33]. Chu-Hong Hoi, Michael R. Lyu “Web Image Learning for Searching Semantic Concepts in Image Databases”, WWW2004, ACM, May 17 22, 2004.
[34]. Mohamed Y. Eldib and Hoda M. Onsi “Web Image Mining Age Estimation Framework” ICGST-GVIP Journal, Volume 11, Issue 1, March 2011

[35]. Wichian Premchaiswadi, Anucha Tungkatsathan "On-line Content-Based Image Retrieval System using Joint Querying and Relevance Feedback Scheme" Wseas Transactions On Computers Issue 5, Volume 9, May 2010 ISSN: 1109-2750
[36]. T. Iwaszko, Mahmoud Melkemi, Lhassane Idoumghar "A Geometric Data Structure Applicable to Image Mining and Retrieval", Proceedings of ICIAR (1)'2010. pp.364-373
[37]. Rosalina Abdul Salam, Abdullah Zawawi Hj. Talib, "Feature Extraction In Automatic Shape Recognition System", International Conference on Information and Communication Technologies, 2004
[38]. A.Kannan, Dr.V.Mohan, Dr.N.Anbazhagan "An Effective Method of Image Retrieval using Image Mining Techniques" The International journal of Multimedia & Its Applications (IJMA) Vol.2, No.4, November 2010
[39]. Dipesh Dugar M , Dinesh Kumar Jain N , Raaja Sarabhoje Gopal , Haritha Ravi , Baskaran Alagappan "Object Recognition of Image Sets Using Discriminative Canonical Correlations" International MultiConference of Engineers and Computer ScientistsIMECS 2010..
[40]. Stephen Gibson, Vladik Kreinovich , Luc Longpre , Brian Penn, Scott A. Starks "Intelligent Mining in Image Databases, with Applications to Satellite Imaging and Web Search", Book on Data mining and computational intelligence, Germany, ©2001ISBN:3-7908-1371-0
[41]. R. Kazman and J. Kominek, "Information Organization in Multimedia Resources", Proceedings of the 11th annual international conference on Systems documentation, Pages 149 – 162,1993.
[42]. M. Flickner, H.S. Sawhney, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, P. Yanker, "Query by Image and Video Content: The QBIC system", IEEE Computer Volume 28 No. 9, pp 23-32, 1995.
[43]. R. Manmatha and S. Ravela, "A Syntactic Characterization of Appearance and its Application to Image Retrieval", Proceedings of the SPIE conference on Human Vision and Electronic Imaging II, Vol, 3016, San Jose, CA, Feb. '97, 1997.

Dr. R. Balu M.Sc., M.Phil., Ph.D.

He received his Doctoral Degree in Computer Science from Bharathiar University (2013) in the area of Image Mining. He obtained his M.Phil degree from Bharathiar University, Coimbatore in 2008, M.Sc Degree in Computer science from Bharathidasan University, Trichy in 2005. He is working as a Faculty in Department of Computer Applications, School of Computer Science and Engineering, Bharathiar University. He has totally 9 years of experience out of which three years as a system administrator and the remaining in teaching and research. He has contributed more than 15 papers in various National / International conference / journals. He is a member of IEEE, CSS, IET and IAENG.