Image Mining: Issues, Frameworks and Techniques

Ji Zhang   Wynne Hsu   Mong Li Lee
Department of Computer Science, School of Computing
National University of Singapore
Singapore, 117543

{zhangji, whsu, leeml}@comp.nus.edu.sg

Abstract

Advances in image acquisition and storage technology have led to tremendous growth in significantly large and detailed image databases. These images, if analyzed, can reveal useful information to the human users. Image mining deals with the extraction of implicit knowledge, image data relationship, or other patterns not explicitly stored in the images. 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. In this paper, we will examine the research issues in image mining, current developments in image mining, particularly, image mining frameworks, state-of-the-art techniques and systems. We will also identify some future research directions for image mining at the end of this paper.

Keywords: Image mining, image indexing and retrieval, object recognition, image classification, image clustering, association rule mining.

1. Introduction

Advances in image acquisition and storage technology have led to tremendous growth in significantly large and detailed image databases [36]. The World Wide Web is regarded as the largest global image 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 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 [1]. While some of the individual fields in themselves may be quite matured, image mining, to date, is just a growing research focus and is still at an experimental stage. The main obstacle to rapid progress in image mining research is the lack of understanding of the research issues involved in image mining. Many researchers have the wrong impression that image mining is just a simple extension of data mining applications; while others view image mining as another name for pattern recognition. In this paper, we attempt to identify the unique research issues in image mining. This will be followed by a review of what are currently happening in the field of image mining, particularly, image mining frameworks, state-of-the-art techniques and systems. We will also identify possible research directions to bring image mining research to a new height.

The rest of the paper is organized as follows. Section 2 will discuss research issues that are unique to image mining. Section 3 discusses two possible frameworks for image mining: the functionality framework versus the information-driven framework. Section 4 gives an overview of the major image mining approaches and techniques used in image mining including object recognition, image indexing and retrieval, image classification and clustering, association rules mining, and neural networks. Finally, section 5 concludes with some future research directions for image mining.

2. Research Issues In Image Mining

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, 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.

Perhaps, 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.

(a) 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.

(b) 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.

(c) 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.

In addition to the need for new discovery algorithms for mining patterns from image data, a number of other related research issues also need to be resolved. For instance, for the discovered image pattern to be meaningful, they must be presented visually to the users. This translates to following issues:

(a) Image pattern representation.

How can we represent the image pattern such that the contextual information, spatial information, and important image characteristics are retained in the representation scheme?

(b) Image features selection.

Which are the important image features to be used in the mining process so that the discovered patterns are meaningful visually?

(c) Image pattern visualization.

How to present the mined patterns to the user in a visually-rich environment?

3. Image Mining Frameworks

Early work in image mining has focused on developing a suitable framework to perform the task of image mining. An image database containing raw image data cannot be directly used for mining purposes. Raw image data has to be first processed to generate the information usable for high-level mining modules. 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. 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.

At present, we can distinguish two kinds of frameworks used to characterize image mining systems: function-driven versus information-driven image mining frameworks. The former focuses on the functionalities of different component modules to organize image mining systems while the latter is designed as a hierarchical structure with special emphasis on the information needs at various levels in the hierarchy.

3.1 Function-Driven Frameworks

The majority of existing image mining system architectures [8, 36] fall under the function-driven image mining framework. These descriptions are exclusively application-oriented and the framework was organized according to the module functionality. For example, Mihai Datcu and Klaus Seidel [8] propose an intelligent satellite mining system that comprises two modules:

(a) A data acquisition, preprocessing and archiving system which is responsible for the extraction of image information, storage of raw images, and retrieval of image.

(b) An image mining system, which enables the users to explore image meaning and detect relevant events.

Figure 1 shows this satellite mining system architecture.

Similarly, the MultiMediaMiner [36] comprises four major components:

(a) Image excavator for the extraction of images and videos from multimedia repository.

(b) A preprocessor for the extraction of image features and storing precomputed data in a database.

(c) A search kernel for matching queries with image and video features in the database.

(d) The discovery modules (characterizer, classifier and associator) exclusively perform image information mining routines to intelligently explore underlying knowledge and patterns within images.

3.2 Information-Driven Frameworks

While the function-driven framework serves the purpose of organizing and clarifying the different roles and tasks to be performed in image mining, it fails to emphasize the different levels of information representation necessary for image data before meaningful mining can take place. Zhang et.al. [18] proposes an information-driven framework that aims to highlight the role of information at various levels of representation. The framework, as shown in Figure 2, distinguishes four levels of information as follows.
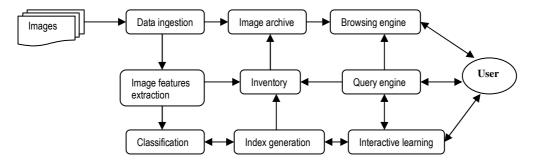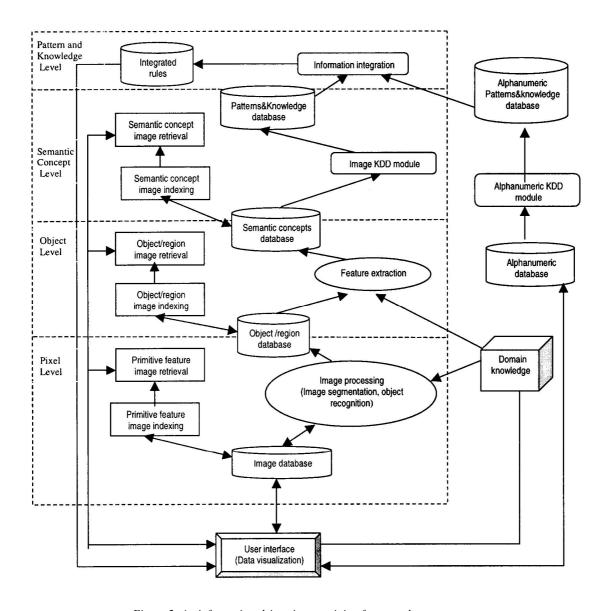
Figure 1. Functionality architecture of an intelligent satellite information mining system

Figure 2: An information-driven image mining framework

(a) Pixel Level, also the lowest level, consists of the raw image information such as image pixels and the primitive image features such as color, texture, and shape;

(b) Object Level deals with object or region information based on the primitive features in the Pixel Level;

(c) Semantic Concept Level takes into consideration domain knowledge to generate high-level semantic concepts from the identified objects and regions;

3

(d) Pattern and Knowledge Level incorporates domain related alphanumeric data and the semantic concepts obtained from the image data to discover underlying domain patterns and knowledge.

The four information levels can be further generalized to two layers: the Pixel Level and the Object Level form the lower layer, while the Semantic Concept Level and the Pattern and Knowledge Level form the higher layer. The lower layer contains raw and extracted image information and mainly deals with images analysis, processing, and recognition. The higher layer deals with high-level image operations such as semantic concept generation and knowledge discovery from image collection. The information in the higher layer is normally more semantically meaningful in contrast to that in the lower layer.

4. Image Mining Techniques

Besides investigating suitable frameworks for image mining, early image miners have attempted to use existing techniques to mine for image information. The techniques frequently used include object recognition, image indexing and retrieval, image classification and clustering, association rules mining, and neural network.

4.1 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 the domain of 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. Specifically, 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. Models of known objects are usually provided by human input a priori.

In general, an object recognition module consists of four components, namely, model database, feature detector, hypothesizer and hypothesis verifier. The model database contains all the models known to the system. 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.

Recently, Jeremy S. De Bonet [17], aiming to locate a particular known object in an image or set of images, design a system that processes an image into a set of "characteristic maps". Michael C. Burl et al. [1] pursue an approach to generate recognizers automatically through learning techniques. The domain expert knowledge is captured implicitly through a set of labeled examples. Stephen Gibson et al. [13] explore the possibility of finding common pattern in several images, which is an important part of image mining. Stephen Gibson develops and tests an optimal FFT-based mosaicing algorithm that has been shown to work well on all kinds of images.

4.2 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 [1]:

(a) Level 1 comprises image retrieval by primitive features such as color, texture, shape or the spatial location of image elements. Examples of such queries are "Retrieve the images with long thin red objects in the top right-hand corner" and "Retrieve the images containing blue stars arranged in a ring"

(b) Level 2 comprises image retrieval by derived or logical features like objects of a given type or individual objects or persons. Examples include "Retrieve images of round table" and "Retrieve images of Jimmy"

(c) 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. For example, we can have queries such as "Retrieve the images of football match" and "Retrieve the images depicting happiness".

Rick Kazman and John Kominek [20] describe three query schemas for image retrieval: Query by Associate Attributes, Query by Description, and Query by Image Content. In Query by Associate Attributes, only a slight adaptation of conventional table structure is needed to tailor it to fit the image needs. The images are appended as extra field. Image retrieval is performed based on other associated attributes within the same table. In Query by Description, the basic idea is to store image descriptions, also known as labels or keywords, along with each image so that users can locate the images of interest using the descriptions. The image descriptions are normally generated manually and assigned to each image in the image preprocessing stage. It suffers from the drawbacks of the "vocabulary problem" [20] and non-scalability. In the early 1990's, because of the emergence of large-scale image repository, the two difficulties of vocabulary problem and non-scalability faced by the manual annotation approach became more and more acute. 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 [28]. Many techniques have been developed in this direction, and many image retrieval systems, both research and commercial, have been built.

In the area of commercial systems, IBM's QBIC system is probably the best known of all image content retrieval systems. It offers retrieval by any combination of color, texture or shape, as well as text keyword. It uses R*-tree indexes to improve search efficiency. More efficient indexing techniques, an improved user interface, and the ability to search grey-level images are incorporates in the latest version. Virage is another well-known commercial system. This 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. As far as the experimental systems, there have been a large number of such systems available. The representatives are Photobook, Chabot, VisualSEEk, MARS, Informedia, Surfimage and Synapse.

4.3 Image Indexing

Image mining systems require a fast and efficient mechanism for the retrieval of image data. Conventional database systems such as relational databases facilitate indexing on primary or secondary key(s). Currently, the retrieval of most image retrieval system is, by nature, similarity-based retrieval. In this case, indexing has to be carried out in the similarity space. One promising approach is to first perform dimension reduction and then use appropriate multi-dimensional indexing techniques that support Non-Euclidean similarity measures [28]. Indexing techniques used range from standard methods such as signature file access method and inverted file access method, to multi-dimensional methods such as K-D-B tree [26], R-tree [10], R*-tree [2] and R+-tree [29], to high-dimensional indexes such as SR-tree [19], TV-tree [21], X-tree [3] and iMinMax [24].

Other proposed indexing schemes focus on specific image features. [24] presents an efficient color indexing scheme for similarity-based retrieval which has a search time that increases logarithmically with the database size. [31] proposes a multi-level R-tree index, called the nested R-trees for retrieving shapes efficiently and effectively. With the proliferation of image retrieval mechanisms, [32] give a performance evaluation of color-spatial retrieval techniques which serves as guidelines to select a suitable technique and design a new technique.

4.4 Image Classification and Image Clustering

Image classification and image clustering are the supervised and unsupervised classification of images into groups respectively. In supervised classification, one is provided with a collection of labeled (pre-classified) images, and the problem is to label newly encountered, unlabeled images. Typically, the given labeled (training) images are used to do the machine learning of the class description which in turn are used to label a new image. In image clustering, the problem is to group a given collection of unlabeled images into meaningful clusters according to the image content without a priori knowledge [15]. The fundamental objective for carrying out image classification or clustering in image mining is

to acquire content information the users are interested in from the image group label associated with the image.

Intelligently classifying image by content is an important way to mine valuable information from large image collection. The classification module in the mining system is usually called classifier. [33] recognizes 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. [7] develops a variety of classifiers to label the pixels in a Landset multispectral scanner image. MM-Classifier, the classification module embedded in the MultiMedia Miner developed by Osmar R.Zaiane et al. [36], classifies multimedia data, including images, based on some provided class labels. James Ze Wang et al. [35] propose IBCOW (Image-based Classification of Objectionable Websites) to classify whether a website is objectionable or benign based on image content. [33] uses binary Bayesian classifier to attempt to perform hierarchical classification of vacation images into indoor and outdoor categories. An unsupervised retraining technique for a maximum likelihood (ML) classifier is presented to allow the existing statistical parameter to be updated whenever a new image lacking the corresponding training set has to be analyzed [4].

Image clustering is usually performed in the early stages of the mining process. Feature attributes that have received most attention for clustering are color, texture and shape. Generally, any of the three, individually or in combination, could be used. There is a wealth of clustering techniques available: hierarchical clustering algorithms, partition-based algorithms, mixture-resolving and mode-seeking algorithms, nearest neighbor clustering, fuzzy clustering and evolutionary clustering approaches. 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. Edward Chang et al. [4] use clustering technique in an attempt to detect unauthorized image copying on the World Wide Web. [15] uses clustering in a preprocessing stage to identify pattern classes for subsequent supervised classification. Lundervold et al. [15] describe a partition-based clustering algorithm and manual labeling technique to identify material classes of a human head obtained at five different image channels (a five-dimensional feature vector).

4.5 Association Rule Mining

An association rule is an implication of the form $X \rightarrow Y$, where X, $Y \subset I$ and $X \wedge Y = \phi$. 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 call an itemset. The rule $X \rightarrow Y$ has support s 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 support X also Support Y. Association rule mining generate rules that have support 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 itemsets that meet the minimum support constraint. The second step generates rules from all the large itemsets that satisfy the minimum confidence constraint.

Association rule mining is a typical approach used in data mining domain for uncovering interesting trends, patterns and rules in large datasets. Recently, association rule mining has been applied to large image databases [25,22,36]. 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 the combined collections of images and associated alphanumeric data [25]. C. Ordonez et al. [25] present an image mining algorithm using blob needed to perform the mining of associations within the context of images. A prototype has been developed in Simon Fraser University called Multimedia Miner [36] where one of its major modules is called MM-Associator. It uses 3-dimensional visualization to explicitly display the associations. In another application, Vasileios M. et al. [22] use association rule mining to discover associations between structures and functions of human brain. An image system called BRAin-Image Database has also been developed. Though the current image association rule mining approaches are far from mature and perfection compared its application in data mining field, this opens up a very promising research direction and vast room for improvement in image association rule mining.

4.6 Neural network

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 [14]. 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.

Even though there has been a lot of research work with regard to neural network and its applications, it is relatively new in the image mining domain. A noteworthy research work that applied neural network to image mining is the Artificial Neural Network (ANN) developed by G.G. Gardner et al [12] which provides a wholly automated approach to fundus image analysis. A Site Mining Tools, based upon the Fuzzy ARTMAP neural network [6], provides an intuitive means by which an image analyst can efficiently and successfully mine large amounts of multi-sensor imagery for Feature Foundation Data (e.g. roads, rivers, orchards, forests) [30].

5. Conclusions

In this paper, we have highlighted the need for image mining in view of the rapidly growing amounts of image data. We have pointed out the unique characteristics of image databases that brought a whole new set of challenging and interesting research issues to be resolved.

In addition, we have also examined two frameworks for image mining: function-driven and information-driven image mining frameworks. We have also discussed techniques that are frequently used in the early works in image mining, namely, object recognition, image retrieval, image indexing, image classification and clustering, association rule mining and neural network.

In summary, image mining is a promising field for research. Image mining research is still in its infancy and many issues remain solved. Specifically, we believe that for image mining research to progress to a new height, the following issues need to be looked at:

(a) Propose new representation schemes for visual patterns that are able to encode sufficient contextual information to allow for meaningful extraction of useful visual characteristics;

(b) Devise efficient content-based image indexing and retrieval techniques to facilitate fast and effective access in large image repository;

(c) Design semantically powerful query languages for image databases;

(d) Explore new discovery techniques that take into account the unique characteristics of image data;

(e) Incorporate new visualization techniques for the visualization of image patterns.

References

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

[2] N. Beckmann, H. P. Kriegel, R. Schneider, and B. Seeger. The R*-tree: an efficient and robust access method for points and rectangles. *In Proc ACM SIGMOD*, 1990.

[3] S. Berchtold, D. A. Keim and H. P. Kriegel. The X-tree: An index structure for high-dimensional data. *In Proceedings of the 22nd VLDB Conference*, pages 28-39, Mumbai, India, September 1996.

[4] L. Bruzzone and D. F. Prieto. Unsupervised retraining of a maximum likelihood classifier for the analysis of multitemporal remote sensing images, *IEEE Transactions on Geoscience and Remote Sensing*, Volume: 39 Issue: 2, pp 456 –460, Feb. 2001.

[5] E. Chang, C. Li and J Wang. Searching Near-Replicas of Image via Clustering. *SPIE Multimedia Storage and Archiving Systems VI*, Boston, MA, USA, 1999.

[6] G. A. Carpenter, S. Grossberg and J. H. Mrkuzon. Fuzzy ARTMAP: a neural architecture for incremental supervised learning of analog multimensional maps, *IEEE Transactions on Neural Networks*, 3(5), 698-713, 1688-1692,1998.

[7] R. F. Cromp and W. J. Campbell. Data mining of multi-dimensional remotely sensed images. *International Conference on Information and Knowledge Management (CIKM),* 1993.

[8] M. Datcu and K. Seidel. Image information mining: exploration of image content in large

archives. *IEEE Conference on Aerospace*, Vol.3, 2000.

[9] J. P. Eakins and M. E. Graham. Content-based image retrieval: a repot to the JISC technology applications program. *Northumbria Image Data Research Institute*, 1999.

[10] A. Guttman. R-tree: a dynamic index structure for spatial searching. *In Proc ACM SIGMOD*, 1984.

[11] D. Greene. An implementation and performance analysis of spatial data access. *In Proc ACM SIGMOD*, 1989.

[12] G. G. Gardner and D. Keating. Automatic detection of diabetic retinopathy using an artificial neural network: a screening tool. *British Journal of Ophthalmology*, 1996.

[13] Gibson, S et al. Intelligent mining in image databases, with applications to satellite imaging and to web search, *Data Mining and Computational Intelligence*, Springer-Verlag, Berlin, 2001.

[14] S. Haykin. Neural Networks: a comprehensive foundation. *Prentice Hall International, Inc.* 1999.

[15] A. K. Jain, M. N. Murty and P. J. Flynn. Data clustering: a review. *ACM computing survey, Vol.31, No.3*, September 1999.

[16] R. Jain, R. Kasturi and B. G. Schunck. Machine Version. *MIT Press and McGraw-Hill Press*, 1995.

[17] J. S. D. Bonet. Image preprocessing for rapid selection in "Pay attention mode". *MIT*, 2000.

[18] J. Zhang, W. Hsu and M. L. Lee. An Information-driven Framework for Image Mining, *in Proceedings of 12th International Conference on Database and Expert Systems Applications (DEXA)*, Munich, Germany, September 2001.

[19] N.Katayama and S. Satoh. The SR-tree: An index structure for high-dimensional nearest neighbor queries. *In proceedings of the 1997 ACM SIGMOD Conference*, pages 369-380, Tucson, Arizona, May 1997.

[20] R. Kazman and J. Kominek. Information organization in multimedia resources. *Proceedings of the 11th annual international conference on Systems documentation*, Pages 149 – 162,1993.

[21] K. Lin, H. V. Jagadish and C. Faloutsos. The TV-tree: An index structure for high-dimensional data. *The VLDB Journal*, 3 (4): 517-542,1994.

[22] V. Megalooikonomou, C. Davataikos and E. H. Herskovits. Mining lesion-deficit associations in a brain image database. *KDD*, San Diego, CA USA, 1999.

[23] W. Y. Ma and B. S. Manjunath. A texture thesaurus for browsing large aerial photographs, *Journal of the American Society for Information Science* 49(7), 633-648,1998.

[24] B. C. Ooi, K. L. Tan, C. Yu and S. Bressan. Indexing the Edges - A Simple and Yet Efficient Approach to High-Dimensional Indexing, *19th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems,* pp. 166-174, Dallas, Texas, May 2000.

[25] C. Ordonez and E. Omiecinski. Discovering association rules based on image content.

Proceedings of the IEEE Advances in Digital Libraries Conference (ADL'99), 1999.

[26] J. T. Robinson. The K-D-B tree: A search structure for large multidimensional dynamic indexes. *In Proceeding of the 1981 ACM SIGMOD Conference*, pages 10-18, June 1981.

[27] Y. Rui, K. Chakrabarti, S. Mehrotra, Y. X. Zhao, and T. S. Huang. Dynamic clustering for optimal retrieval in high dimensional multimedia databases. *In TR-MARS-10-97*, 1997.

[28] Y. Rui, T. S. Huang et al. Image retrieval: Past, present and future. Invited paper in *Int Symposium on Multimedia Information Processing*, Taipei, Taiwan, Dec 11-13, 1997.

[29] T. Sellis, N. Roussopoulos and C. Faloutsos. The R+ tree: A dynamic index for multi-dimensional objects. *In Proc 12th VLDB*, 1987.

[30] W. Strelilein and A. Waxman. Fused multi-sensor image mining for feature foundation data. *Proceedings of the Third International Conference on Information Fusion (FUSION 2000),* Volume: 1, 2000.

[31] K. L. Tan, B.C. Ooi and L. F. Thiang. Retrieving similar shapes effectively and efficiently, *Multimedia Tools and Applications*, Kluwer Academic Publishers, The Netherlands, 2001.

[32] K. L Tan, B. C. Ooi and C. Y. Yee. An Evaluation of Color-Spatial Retrieval Techniques for Large Image Databases, *Multimedia Tools and Applications*, Vol. 14, No. 1, pp. 55-78, Kluwer Academic Publishers, The Netherlands, 2001.

[33] A.Vailaya, A. T. Figueiredo, A. K. Jain, and H. J. Zhang. Image classification for content-based indexing. *IEEE Transactions on Image Processing*, Volume: 10 Issue: 1, pp 117 –130, Jan. 2001.

[34] D. White and R. Jain. Similarity indexing: Algorithms and performance. *In Proc. SPIE Storage and Retrieval for image and Video Databases*, 1996.

[35] J. Z. Wang, J. Li et al. System for Screening Objectionable Images Using Daubechies' Wavelets and Color Histograms. *Interactive Distributed Multimedia Systems and Telecommunication Services, Proceedings of the Fourth European Workshop (IDMS'97)*, 1997.

[36] O. R. Zaiane, J. W. Han et al. Mining MultiMedia Data. *CASCON'98: Meeting of Minds*, pp 83-96, Toronto, Canada, November 1998.