Image Mining Techniques: A Review

Hemavathy R , Dr. Shobha G
Computer science dept. R V College of Engineering.
Bengaluru, Karnataka, India.

***Abstract*—Image mining deals with extraction of implicit and useful data from images stored in the large data bases. Image mining finds application in almost all areas like in medical field, remote sensing, agriculture, industries and educational field. In this paper a study has been made to understand different mining methods used on image database in understanding and revealing useful information. Data mining methods can be used to mine Images after transforming Images in to some form of dataset. An overview of different methods used to extract the Image features and different mining techniques applied on it effectively and efficiently are briefed.**

Keywords- Image mining; Data mining,Dataset.

I. INTRODUCTION

Image mining involves basic two processes first is transformation of Image in to some form of quantitative data; second is efficient method of data mining techniques applied on the transformed data . The result of this process reveals useful information which we term it as knowledge. Decision makers in every field like business, public sector, Medical etc. are trying to get useful and implicit information from the already existing digital data bases[1,2,3,4]. Image mining is the concept used to extract implicit and useful data from images stored in the large data bases. This paper elaborates the research works already done in image mining summarizes different tool developed and its applications. 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 is not explicitly found in the images

In the following section a description of different methods applied on various applications are explained. Efficient technique can be applied according to the requirement.Surey done may be helpful in developing new image mining algorithms, refineing existing ones, or act as an effective filter.

II. RELATED WORK

A new technique called Image retrieval based on optimum clusters[1] was proposed for improving user interaction with image retrieval systems by fully exploiting the similarity information. Given a query Image it searches image database for similar to query images. The index is created by describing the images according to their color characteristics, with compact feature vectors, that represent typical color distributions [8].

In the Pre-procesing stage, texture features are extracted on the basis of statistical approach, the Grey Level Co occurrence Matrix (GLCM) is used to extract second order statistics from an image. Different texture parameters like entropy, contrast, dissimilarity, homogeneity, standard deviation, mean, and variance of both query image and target images are calculated. Then, the pre-processed images in the database are classified as low-texture, average-texture and high-texture detailed images respectively on the basis of some factor like MLE (Maximum Likelihood Estimation) estimation. The classified images are then subjected to color feature extraction where average value of R, G, and B values for both query image and target images are calculated stored and considered as features. The retrieved result is pre-clustered by Fuzzy-C means technique. The resulted values of both the query image and target images are compared by Euclidean distance method. Implementation reduces the searching time space since it uses both texture feature and RGB color space. This method is very effectively reduces the searching time for colored image and used in Medical transcription .

Decision tree induction[9,10,11] is a well-known methodology used widely on various kinds of domain, such as artificial intelligence, machine learning, data mining, and pattern recognition.

A general mining approach based on decision trees for segmenting image data is proposed [2] pixel wise image feature is extracted and decision trees is applied on it.

Pixel-wise image features are extracted and transformed into a database-like table. Each tuple in the table has a feature descriptor consisting of a set of feature values for a given pixel along with its label. The decision tree methodology is used to mine the data to discover relationship between the attributes of pixels and their target labels, and to build a model for image processing by using the training data set.

The results in this model is very efficient and effective for image mining and image segmentation. It can also be used to develop new image processing algorithms, refine existing algorithms, or act as an effective filter.The misclassification cost could be artificially tuned in the decision tree training process to minimize the penalty when the application uses different weights with different output classes. The controllability is a big advantage of decision tree over other algorithms such as a neural network.

It can easily extend from 2D to 3D image processing without making a revolution and the created model can generate very efficient and compact code. Disadvantage in this is adjustment is required in the proposed model to specialize for a particular case.

Application in medical field is considered in [3] image of CT brain scan is used to determine the tumor. Novel Fuzzy Association Rule Mining (NFARM) applied on the image database which contains the features that are extracted from

the CT scan brain images. This method selects only the most relevant features to be used during the mining process [12,13]. Moreover this approach is optimized because it performs two operations in a single step that is feature selection and discretization [14]. This makes the mining algorithm faster, and also improves the Classification accuracy.

In the pre-processing technique the Regions of Interest are extracted and stored as attribute. Objects attribute value in each ROI differs in its pixel value. Novel Fuzzy Association Rule Mining algorithm is applied on this data set to identify the cancerous cells. Advantage is it provides better efficiency of the proposed Novel Fuzzy Association Rule Mining algorithm compared to the traditional Fuzzy Apriori algorithm. Fuzzy apriori algorithm has been well supported for the numerical analysis [15], but where as in medical image classifications [16] it is less efficient. it is best suited for CT scan images which contains only gray level pixels.

Fingerprint, iris, hand impression patterns used in Biometrics systems images are represented by feature vectors[4]. The candidates patterns are then retrieved from Database by comparing the distance of their feature vectors. Visual features such as shape, color and texture are extracted to characterize images.

Feature extraction is the process of generating features to be used in the selection and classification tasks. Feature selection reduces the number of features provided to the classification task. Those features which are likely to assist in discrimination are selected and used in the classification task.

In the preprocessing stage: - Image is transformed to data set in terms of color, texture and shape feature .**Colo**r feature is extracted using color histogram, **texture** feature can be used to classify textured images from non-textured ones and then be combined with another visual attribute like color to make the retrieval more effective. Texture representation methods can be classified into two categories: structural and statistical. **Shape** is an important visual feature and it is one of the primitive features for image content description. Shape descriptors can be divided into two main categories: region based and contour-based methods. Region-based methods use the whole area of an object for shape description, while contour-based methods use only the information present in the contour of an object.

The classifier used commonly is Nearest Neighbor classifier. It compares the feature vector of the prototype with image feature vectors stored in the database. It is obtained by finding the distance between the prototype image and the database. This system identifies Iris pattern in Biometrics systems.

A framework of image mining based on concept lattice is presented [5]. Main use is 1) The cloud concept provides a means of both qualitative and quantitative characterization of linguistic terms. This method can reflect the distribution of data in that domain, while keeping the soft boundaries. 2) We can find the association rule from the hasse diagram, and can also choose the concept hierarchy to find the association rules.

Concept lattice is proposed by Wille R. [19]. It reflects the process of human's concept formation with mathematical formal language. Based on binary relationship, concept lattice embodies the unification of intension and extension of concepts, reflects the relationships between objects and characteristics and the relationships between generalization and characterization among concepts. With corresponding Hasse graph, concept lattice can implement the visualization of the hierarchies of data concepts, and it is suit to find the latent concepts from image data [20].The cloud model is an effective tool in transforming between qualitative concepts and their quantitative expressions. Cloud model is used to calculate support, confidence and relationship in the field of association rules mining. The hasse diagram of concept lattice reveals the concept hierarchy of the context.

The algorithm of image texture feature association rule mining is outlined as follows.

1. Read Image and pre-process it.
2. Use cloud model to extract a series of concepts from the pre-processed image.
3. Create transaction database D.
4. Take D as source dataset, build concept lattice C, and draw Hasse graph.
5. According to concept lattice C, generate texture feature association rules.
6. Choose some rules as texture feature knowledge.

The method of texture feature data mining is similar to the following methods of color feature data mining, shape feature data mining and spatial relationship feature mining

conceptlattice and cloud model provides the tool for formal conceptanalysis. The advantages of the algorithm is that cloud concept provides a means of both qualitativeand quantitative characterization of linguistic terms . The hasse diagram finds the association rule

A new technique called image retrieval from clusters is introduced in order to reduce the searching time space Images are to be clustered based on the RGB components[6].

The given query image to be retrieved from the concerned cluster based on the texture feature In preprocessing stage noise is removed in the image using Average and median filters.

Images are to be clustered based on the RGB components. Calculate mean values of Red, Blue and Green components to group the image in to its respective groups to form clusters. Whenever the query image is given calculate the RGB components average values. Then compare this with the stored values.The given query image to be retrieved from the concerned cluster based on the texture features. Texture represents the energy content of the image

Experiments have been conducted by using shape outline as the features[7]. Shape outline readings are put through normalization and dimensionality reduction process using an eigenvector based method to produce a new set of readings. After this pre-processing step data will be grouped through their shapes. Using statistical analysis, these readings together with *peak measures* a robust classification and recognition process is achieved. After preprocessing stage the image shape outline technique is used it is automatic method of finding the

initial point to start the shape outline. Outline readings went through a process of transformation which involved normalization and dimensionality reduction. Transformation uses eigenvector , which can reduce the computational burden of pattern recognition algorithms and the image mining process. The data will then go through the next stage that is the shape categorization process. Advantages are it reduces the searching time space and simple to implement. Disadvantage is, it is applicable to only colored images as, in first stage is clustering using RGB.

Object with no back ground in an image can be identified [7] using shape outliner algorithm. It identifies simple, complex shaped objects in images and can be categorized. Disadvantage of the algorithm is it works only for no background Image. It cannot identify the overlapped images.

Fuzzy clustering algorithm on spatial contextual information in image data is introduced in [18]. Unsupervised Clustering, classification of image data is done using both the feature space information and the spatial contextual information. Algorithm uses dissimilarity index that influence the neighbouring pixels on the centre pixel in a 3 x 3 window. If the window is in non homogeneous region the influence of the neighbouring pixels on the centre pixel is suppressed; else, the centre pixel is smoothed by its neighbouring pixels during the computation of the membership values and the cluster centroids. It resolves classification ambiguity for data in the overlapping region of two clusters. Clusters are merged if they close together with significant overlap.

Experimental results with synthetic and real images indicate that the proposed algorithm is more tolerant to noise, better at resolving classification ambiguity and coping with different cluster shape and size than the conventional fuzzy c-means algorithm.

As there is an increase in the amount and resolution of remotely sensed imagery it necessitates in the development of automatic processing and classification system [17].

Bayesian framework [17] uses spatial information for classification of high-resolution images. First, spectral and textural features are extracted for each pixel and quantized, then train Bayesian classifiers on this with discrete non-parametric density models. An iterative split-and-merge algorithm is used to map the pixel level classification into contiguous regions. Then, the resulting regions are modeled using the statistical summaries of their spectral, textural and shape properties, and used with Bayesian classifiers to compute the final classification maps. Experiments with three ground truth data sets show the effectiveness of the proposed approach over traditional techniques that do not make strong use of region-based spatial information. Proposed region level features and Bayesian classifiers performed better than the traditional pixel level classification techniques. Even though the numerical results already look quite impressive, selection of discriminative subset of features and better segmentation of regions is obtained.

Shared structures embody the consistence and coherence of features that repeatedly co occurs at an object class[21].They can be used as discriminative information to separate the various objects contained in unlabeled images.The method[21] is able to separate the original object from the distracting objects. Steps to maximize the likelihood for shared structure learning: feature extraction; clustering; exploring consistent pair wise relationships; combining pairwise relationships into high order structures Histogram of Oriented Gradient (HOG) descriptor is used to extract image features at multiple scales. HOG descriptors describe local object appearance and shape within an image using the distribution of intensity gradients or edge directions. Finally, provide quantitative results on Caltech dataset.If the shared structures that correspond to the same classes are combined together, object location or object categorization will be more accurate. And more primitive features such as edges might improve robustness to background clutter and shape ambiguity.

III. CONCLUSION

The mining techniques explored in these papers are useful for particular application and on particular images, mining techniques aims at reducing the time of searching and mining. These techniques can be effectively used by combining different algorithm and concepts. The image considered is usually colored and gray scale. Technique can be improved and could be applied on camera captured images like images of ship, submarines, under water object images main goal of any mining techniques is not only just mining it should also be efficient. Mining technique efficiency is measured in terms of – minimizing time in transforming the image to data set .Reduction in the size of the transformed image data and less time to mine the bulk data with good performance

REFERENCES

[1] 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.

[2] Kun-Che Lu, Don-Lin Yang, and Ming-Chuan Hung , “Decision Trees Based Image Data Mining and Its Application on Image Segmentation”.

[3] 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.

[4] Ryszard S. Chora´s, “Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems “,INTERNATIONAL JOURNAL OF BIOLOGY AND BIOMEDICAL ENGINEERINGIssue 1, Vol. 1, 2007.

[5] Ruchi Bhatt , S.M.Patil ,”Image Mining Based on Concept Lattice Theory Using Texture Feature”.

[6] A.Kannan, Dr.V.Mohan, Dr.N.Anbazhagan, “Image Retrieval Based on Clustering and Non-Clustering Techniques using Image Mining”.M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[7] R. A. Salam, M.A. Rodrigues , “Mining Image Features in an Automatic Two-Dimensional Shape Recognition System”.

[8] Lukasz Kobyli´nski and Krzysztof Walczak, “Color Mining of Images Based on Clustering”, Proceedings of the International Multiconference on Computer Science and Information Technology pp. 203–212.

[9] John C. Shafer, Rakesh Agrawal, Manish Mehta.: SPRINT: A Scalable Parallel Classifierfor Data Mining. VLDB 1996, p. 544-555. M. Mehta, R. Agrawal, J. Rissanen.: SLIQ: A fast scalable classifier for data mining. In.

[10] Proc. 1996 Intl. Conf. on Extending Database Technology (EDBT'96), Avignon,France, March 1996

[11] S. M. Weiss, C. A. Kulikowski.: Computer System that Learn: Classification and Prediction Methods from Statistics, Neural Nets, Machine Learning, and Expert System. Morgan Kaufman, 1991.

[12] Agrwal. R., Imielinsk., T, Swam., A.N. : Mining Association Rules between sets of Items in Large Databases.:. S/G Mod Record 22, 207-216(1993)

[13] Laila Elfangary and Walid Adly Atteya: Mining MedicalDatabases using Proposed Incremental Association Rules Algorithm (PIA).: Second International Conference on the Digital Society ,IEEE Computer Society(2008)

[14] Pudi. V., Harilsa., j. : On the optimality of association rule mining algorithms. Technical Report TR-2001-01, DSL, Indian Institute of Science (2001).

[15] Hoppner. F., Klawonn. F.; Kruse, R, Rurkler, T.: Fuzzy cluster Analysis, methods for classification Data Analysis and Image recognition.: Wiley, New York (1999)

[16] Maria-Luiza Antonie, Osmar R. Zaiane, Alexandra Coman,: Application of data mining techniques for medical image classification. In : Proc. Second Int. WorkShop on Multimedia Data Mining (MDM/KDD' 2001)

[17] Selim Akshoy ;Spatial Techniques for Image Classification.

[18] A.W.C.Liew, S.H.Leung and W.H.Lau ; Fuzzy image clustering incorporating spatial continuity. In IEE Proc.-Vis. Image Signal Process.. Vol. 147. hb. 2, April 2000.

[19] Bernhard Ganter and Rudolf Wille. *Formal Concept Analysis* Mathematical Foundations. Springer-Verlag Berlin Heidelberg. 1999.

[20] Hu Keyun. ;Research on Concept Lattice and Rough Set based on Data Mining Methods . International Journal of Computer Science and its Applications(2001).

[21] Fenglei Yang · Baomin Li; Unsupervised learning of spatial structures shared among images Published online: 9 August 2011.© Springer-Verlag 201Vis Comput (2012) 28:175–180.