Image Mining Techniques – A Review

D. Angayarkanni[1*], L.Jayasimman[2]

[1] Dept. of Computer Science, Srimad Andavan Arts & Science College (Auto),Tiruchirappalli, Tamil Nadu.
[2] Dept. of Computer Science, Srimad Andavan Arts & Science College (Auto),Tiruchirappalli ,Tamil Nadu.

Abstract—In recent years, images play an important role in day to day life and with the advent of increasing multimedia content in the internet, manipulating information from the images in image database is necessary and useful task. Valuable informations are hidden in images and it reveal useful information to the user. Now a days, image mining is one of the research area for many researchers. It is a new focus to data mining and also utilizes the algorithms in data mining and image processing. Image mining is the process of searching and identifying the pattern present in the images, extracting useful information that is not explicitly present in the images. The goal of image mining is to mine knowledge from large image database. Various techniques has been used to discover knowledge from image database. This paper provides a review of various techniques used for discovery of knowledge from images using image mining techniques.

Keywords— Image Mining,Retrieval,Classification,Clustering,Indexing

I.   INTRODUCTION

Image mining extends data mining concepts from structured data (relational) to unstructured data (image) [1]. But there is a difference between structured data and unstructured data. In relational databases, the data value are semantically meaningful. But in image databases, the data value themselves may not be significant unless the context supports them and there are multiple interpretations for the same visual patterns and is critical because of implicit spatial information. It is an interdisciplinary field which utilizes methods from data mining, image processing, pattern recognition, image analysis, and artificial intelligence. . Image mining deals with the extraction of knowledge that is implicitly defined, relationship between image data, or other patterns that are implicitly stored in the images, between images and other alphanumeric data [2]. The focus of Computer vision and image processing techniques are understanding and/or extracting specific features from a single image. But Image mining is a different concept and it deals with extraction of patterns from large collection of images. Research in image mining can be classified into two directions namely Domain specific application which is used to extract relevant images and its features .The next one is General application which is used to extract relevant image patterns from a large collection of images without prior knowledge of image content. [3] Generally, Image mining uses two different approaches. First, extract images from the image database or collection of images. Second, mines information from the combination of alphanumeric data and collection of images [4]. The key objective of image mining is to expose how low level pixel representation surrounded in a raw image or image sequence can be processed to recognize high level image objects and the relationship that is to discover knowledge form large collection of images.

II.   PROCESS OF IMAGE MINING

Generally image database contains raw image data that can be directly acquired from the various devices. It is difficult to retrieve information directly from the raw image data. So it should be processed to get high level information. In the image mining process, the images in the database are pre-processed to improve its quality. Next, important features of images are produced using transformation and feature extraction methods. Feature extraction is applied to the images in the database for extracting essential features that describe a large set of data into smaller. So large and redundant data are transformed into reduced set features. Then retrieved information is used with mining techniques to discover significant patterns. And is followed by interpretation and evaluation of information. Hence knowledge is obtained which is an understandable form of information [5].

A.   Feature Extraction

To mine useful knowledge from images effectively and efficiently, first to solve the dimensionality problem i.e, reduce the information present in the image. This dimension reduction can be viewed as the transformation of high order dimension to low order dimension. This can be achieved by feature extraction methods.Various visual informations are presented in the images.The commonly used visual features are Color, Texture and Shape. These features can be extracted by various methodslike histogram, moments, GLCM, Wavelet, Edge detection methods, etc. This

extracted information constitute an input for image mining techniques.This can be stored in a vector form called feature vector.

B. Image Mining Techniques

In image mining, various techniques are used to mine information from the image database.

Object Recognition

Object recognition is the process of identifying a specific object from an image. It is basically imitator of the human expertise to distinguish different objects from an image. It is based on the model of known objects. This problem is mentioned as supervised learning problem. Object recognition algorithms depend on matching, learning, or pattern recognition algorithms using appearance-based or feature-based techniques [6,7,8,9].

Image Retrieval

Image retrieval is the process of searching and retrieving an image from the large image database. In image mining, image retrieval is the process of retrieving the images that have similar information or image patterns in a given collection of images based on some predefined features or requirements. The features or requirement can be categorized as

- Primary features (shape, color, texture)
- High level features (attributes retrieved based on purpose or scenes depicted).

On the other hand, image can be retrieved based the following criteria
1) Retrieval by description: The process is based on text descriptions. Description can be like size of images, type of images, date and time of image when it was captured, identity of the owner, keywords mapping of the image. Based on the contents,required image are retrieved from the database.
2) Retrieval by visual content: This method uses the visual features of both the query image and database images. Similar images in the database are retrieved based on the similarity distance measures between query and database images [6, 7].

Image Indexing

If size of the image database to be searched is very large and the feature vectors of images are of high dimension then the search complexity is high. Organizing the images in database is called as image indexing.Searching and retrieving the images efficiently from the that database is the main aim of indexing techniques . Reducing dimensionality and indexing high dimensional data are the

two main approaches used in indexing. Reducing the dimensions can be accomplished with the Singular Value Decomposition (SVD) update algorithm and clustering algorithms [2]. Each image has its unique feature. Image indexing can be implemented by the features, which are derived from the images. Now, the most image retrieval system is performed, by text, 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 multidimensional indexing techniques. Indexing techniques used range from standard methods such as signature file access method and inverted file access method, to multi-dimensional methods such as KD- B tree, R-tree, R* -tree and R+-tree. [6, 7, 8, 9].

Image clustering

Image clustering is the process of arranging a set of images into classes of images with same property without previous knowledge is called image clustering and is also denoted as unsupervised learning problem which groups a given set of unlabelled images into meaningful clusters according to the image content. Images in one cluster have high resemblance in comparison to one another but are very dissimilar to images in other clusters. Hierarchal, partitioning, density-based, grid based and fuzzy clustering methods are some techniques used in clustering of images.

Image Classification

The objective of classification is to label a newly encountered (unlabelled) image into a given collection of pre classified labelled images and it is known as supervised classification. The objective is to divide the detected objects into pre-defined categories. Methods of machine learning (decision tree, rule-based classification, support vector machines, neural networks) are applied to train the system based on the labelled (pre-classified) samples. The most commonly used classification techniques are Bayes, neural network, decision tree, support vector machine, K-nearest – neighbour-classifier, genetic algorithm, Artificial Neural Networks (ANN), Fuzzy measures, Fuzzy support Vector Machines (FSVM) and Genetic Algorithms.[2,8,9]

Association Rule Mining

Association Rule Mining (ARM) is a typical data mining technique that is used to discover frequent pattern appear in the large data set[10]. Assciation rule has LHS → RHS form. Support and confident measures are used to determine the item pattern present in the mining process.Support is defined as how frequently transactions(pattern) occur in the data set and confidence defined as how often items in RHS appear in transactions that contain LHS. This technique works in two step.First step to discover frequent itemset present in dataset.The second step generates rules from all the large item sets.In

images, association rules mining is used to capture frequently occurring pattern in the images.These patterns are characterised as a texture.Three main algorithms used to extract frequent pattern are Apriori,FP-Growth and Eclat.

III. LITERATURE STUDY

Numerous researches have been carried on image mining. This section of the paper presents a survey on various image mining techniques that were proposed earlier.

Kannan et al.[11] proposed a new clustering technique that combines Content based Image Retrieval and image mining technique. In this technique, Image features like color, texture can be extracted from the images. Initially pre-processing done with filtering and images are clustered based on RGB Components. Next retrieved images are clustered based on the texture, and are classified as high-texture detailed Image, average-texture detailed Image and low-texture detailed Image. Then Fuzzy C-means clustering is applied and similarity measurement entropy calculation is used to get the target image .This technique is compared with existing methods such as RGB Components, Entropy, Clustering methods and performance is measured using precision and recall.

Rajshree S. Dubey et al. [12] proposed a method which is based on the Color histogram, texture and shape of that Image. Color Histogram and Texture of the query image is created and compared with the image in database and the resultant Image is found. They used histogram-based search techniques for two different color spaces, RGB and HSV with color, shape and texture features. Histogram Euclidean distance and histogram intersection distance is used to find out similarity measurement. It is concluded that the Histogram Intersection-based image retrieval in HSV color space is most desirable among the other retrieval methods as compared with both computation time and retrieval effectiveness.

Hemalatha&Devasena et al. [13] developed an innovative technique for mining images by means of Lorenz Information Measure (LIM) dependent image matching method with neural networks. This approach is independent of several parameters to produce a robust solution. It is developed and implemented on MATLAB and is examined with several image databases. Suitable measures were developed to estimate the performance of the system. The performance results were significant and comparable with other methods.

ParagDhonde et al. [14] proposed a hierarchical K-means based image retrieval method system. Images are clustered in the database based on the features like color, texture and shape using hierarchical clustering. To reduce large quantities of data again, images are clustered using K-means algorithm. Features of query image is retrieved and is compared against the resulting clustered database. This method gives the better accuracy performance with the combined approach of the features such as color, texture, shape as compared to individual features.

Archana J. Waghchawre et al.[15] proposed a method for automatic indexing and annotation based retrieval of images from the image database using document retrieval technique. This method initially analyses the user query and tries to recognise the relevance between the index structure in the database and the user query for retrieving images. Markovian semantic Indexing (MSI) is used for indexing the images in the database and is dynamically trained by the queries of the users that will be served by the system This gives the better performance on recall and precision parameters as compared to other indexing techniques Latent Semantic Indexing (LSI) and Probalistic Latent Semantic Indexing(PLSI).This technique is also provide a good survey on the various aspects of information retrieval model.

Rahul Mehta et al. [16] presented a content based image retrieval system using the combination of color and texture property. To retrieve a image based on color feature, conventional color histogram(CCH) is used, and with Quadratic Distance Metric (QDM) measure, the similarity between the images are measured and the retrieved images are stored in a featured image database. To enhance the search result effectively, texture feature is retrieved from the above retrieved images using pyramid structure wavelet transform model and to measure the similarity between images Euclidean distance is used. Their experimental result gives better performance than the traditional one.

ManimalaSingham et al. [17] proposed a image retrieval using wavelet based colorhistogram(WBCH).In this system extract texture and color features through HAAR wavelet transformation and color histogram respectively. The extracted features both color and texture are combined to produce a new feature vector. Similarity between query and database images are measured with Histogram intersection distance method and this system used WANG database and implemented in MATLAB.

JyotiDeshmukh et al [18] presented an image mining system using association rule in mammogram images.Texture features from the images are extracted using GLCM method and optimised feature vectors are selected from all retrieved features which is given for transactional database for finding association rules using Extraction of strong and effective association rules(ESAR algorithm).The proposed algorithm gives better results and overcome the problem associated with association rule generated from Apriori algorithm.

Avinash N Bhute et al. [19] presented a survey of various methods and algorithms for the evaluation of image retrival system. Color, texture, edge and shape features are extracted from the query image and images in the database. Based on the features, Antipole tree method is used to create index structure of the images in the database. Similarity measurement is used to retrieve the relavant image from the image database.

P Thamilselvan et al.[20] made a comparative study on image classification using data mining algorithm such as

SVM, AdaBoost, CART, KNN, Artificial Neural Network, K-Means, Chaos Genetic Algorithm, EM Algorithm, C4.5.Based on the analysis of classification accuracy and kappa coefficient, they suggest SVM and KNN gives better performance result and in future, using of any two data mining algorithms to improve the classification accuracy and reduce the error rate.

NhatQuang Doan et al. [21] suggesteda Image indexing technique to retrieval of images from large data set. Dynamic Growing Self-organizing Tree (DyGSoT) is used for creating index for images.Color and texture features are extracted and is used for processing. Searching in term of topological and hierarchical relationships between images produce improved image retrieval result and searching time.

Vaibhavi S. Shukla et al.[22] presented a survey for the various techniques used in image mining applications and made a study of how to predict and analyse the occurrence of forest fire with different parameters.

Prof.RupaliSawant[23] illustrated a new image mining technique which uses concept lattice and cloud model to search the image in the image database.Image features such as color and texture are extracted and clustered using cloud model with concept lattice.Simillarity between query image and image in the database is performed with entropy to find relavent image.

Rajwinder Kaur Sandhar et al[24] presented a multiple object extraction from a single image using various threshold value.Color and edge features are extracted using various threshold values based on frequency images.

Joseph.CNeethu et al.[25] suggested technique for retrieval of images with the image mining technique clustering and association rule mining.This system extracts both textual and visual features and proposed the multimodal fusion of features give better results in achieving image retrieval process effecively.

IV. CONCLUSION

This paper focusesthe various techniques and algorithms used in knowledge discovery from images in different literatures. The main function in image mining system is to extract implicit knowledge, image data relationship, or other patterns that are not explicitly stored in the images and between images and other alphanumeric data. An Image mining process is very complicated because it uses various techniques such as image retrieval, indexing, clustering and utilizes various algorithms and approaches in data mining, pattern recognition, machine learning. A good image mining system should provide the users with an effective access to image database, produce useful knowledge and pattern from the images.

REFERENCES

[1] PrabhjeetKaur,Kamaljit Kaur *"Review of Different Existing Image Mining Techniques"* International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, Issue 6, June 2014,

[2] AboliW.Hole,Prabhakar L.Ramteke,*"Design and Implementation of Content Based Image Retrieval Using Data Mining and Image Processing Techniques"* , International Journal of Advance Research in Computer Science and Management Studies,Volume 3, Issue 3, March 2015.

[3] Shaikh NikhatFatma*"Image Mining Method and Frameworks"*, International Journal Of Computational Engineering Research (ijceronline.com) Vol. 2 Issue. 8 Dec 2012.

[4] D.N.D.Harini,Dr.D.LalithaBhaskari, *"Image Mining Issues and Methods Related to Image Retrieval System"*,International Journal of Advanced Research in Computer Science, Volume 2, No. 4, July-August 2011

[5] K.Saraswathi, Dr.V.GaneshBabu, *"A Review on Image Mining Techniques"* International Journal of Scientific & Engineering Research Volume 8, Issue 5, May-2017, 121 ISSN 2229-5518.

[6] RamadassSudhir, *"A Survey on Image Mining Techniques: Theory and Applications"*, Computer Engineering and Intelligent Systems, ISSN 2222-1719 (Paper) ISSN 2222-2863 (Online) Vol 2, No.6, 2011 .

[7] Sreelekshmi.U, Anil A.R *"A Survey on Feature Extraction Techniques for Image Retrieval using Data Mining & Image Processing Techniques"*, International Journal Of Engineering And Computer Science ISSN: 2319-7242 Volume 5 Issue 11 Nov. 2016, Page No. 18979-18982

[8] AnkitaTripathi, Hitesh Jangir*"A Study on Image Mining Methods and Techniques"*, International Journal of Innovative Research in Computer and Communication Engineering Vol. 4, Issue 4, April 2016 ISSN(Online) : 2320-9801 ISSN (Print) : 2320-9798

[9] Dr.Rina, ShailjaAgnihotri, *"A Review of Image Mining Concepts, Techniques and Challenges"* , International Journal of Innovative Research in Computer and Communication Engineering , Vol. 4, Issue 9, September 2016.

[10] V.K. Gujare, P. Malviya, "*Big Data Clustering Using Data Mining Technique*", International Journal of Scientific Research in Computer Science and Engineering, Vol.5, Issue.2, pp.9-13, 2017

[11] A.KannanDr.V.MohanDr.N.Anbazhagan*"Image Clustering and Retrieval using Image Mining Techniques"* IEEE International Conference on Computational Intelligence and Computing Research 2010

[12] Rajshree S. Dubey,Niket Bhargava ,RajnishChoubey*"Image Mining using Content Based Image Retrieval System"*International Journal on Computer Science and Engineering Vol. 02, No. 07, 2010, 2353-2356

[13] Hemalatha and C. Lakshmi Devasena (2011), *"A Hybrid Image Mining Technique using LIM based Data Mining Algorithm"*, International Journal of Computer Applications, vol. 25, no.2, pp. 1-5

[14] ParagDhonde,Prof. C.M. Raut*"Precise & Proficient Image Mining Using Hierarchical KMeansAlgorithm"*,International Journal of Scientific and Research Publications, Volume 5, Issue 1, January 2015

[15] Archana J. Waghchawre ,Prof. J.V. Shinde*"Query Mining for Image Retrieval System Using Markov Chain Model"*,International Conference on Emerging Trends in Computer Engineering, Science and Information Technology "-2015 ,Special Issue of International Journal of Electronics, Communication & Soft Computing Science and Engineering, ISSN: 2277-9477 .

[16] Rahul Mehta ,Nishchol Mishra ,Sanjeev Sharma ,*"Color - Texture based Image Retrieval System"*, International Journal of Computer Applications (0975 – 8887) Volume 24– No.5, June 2011.

[17] ManimalaSingham and K.Hemachandran*"Content Based Image Retrieval using Color and Texture"*, An International Journal of Signal & Image Processing (SIPIJ) Vol.3, No.1, Feb 2012.

[18] JyotiDeshmukh,, UdhavBhosle, *"Image Mining using Association Rule for Medical Image dataset"*, International Conference on Computational Modeling and Security

[19] Avinash N Bhute, B. B. Meshram*"Content Based Image Indexing and Retrieval",* International Journal of Graphics & Image Processing ,Vol 3,issue 4,Nov. 2013 .

[20] P Thamilselvan, Dr. J. G. R. Sathiaseelan*"A Comparative Study of Data Mining Algorithms for Image Classification"* International Journal of Education and Management Engineering, 2015.

[21] NhatQuang Doan, Thi Phuong Nghiem, Giang Son Tran *"Dynamic Indexing for Content-Based Image Retrieval Systems using Hierarchical andTopological Network",*IEEE,Eighth International Conference on Knowledge and Systems Engineering (KSE),2016

[22] Vaibhavi S. Shukla, Jay Vala*, "A Survey on Image Mining, its Techniques and Application"*, International Journal of Computer Applications (0975 – 8887) Volume 133 – No.9, January 2016

[23] Prof.RupaliSawant,"Image Mining Using Texture and Shape Feature", International Journal on Computer Science and Engineering Vol. 02, No. 09, 2010, 2808-2815

[24] Rajwinder Kaur Sandhar,GurbakashPhonsa , *"Distinctive Feature Mining based on Varying Threshold Based Image Extraction for Single and Multiple Objects"*,2014 IEEE International Conference on Advanced Communication Control and Computing Technologies.

[25] Joseph.cNeethu, Aswathy Wilson*"Retrieval of images using data mining techniques"*,IEEE,International Conference on Contemporary Computing and Informatics (IC3I), 2014.

Authors Profile

Ms.D.Angayarkanni is working as a Assistant Professor in Computer Science,Bharathidasan University Constituent College,Lalgudi.She is pursuing Ph.D in SrimadAndavan Arts and Science College,BharathidasanUniversity,Trichy,India.

Dr. L. Jayasimman is working as a Driector,MCA Department, SrimadAndavan Arts & Science College, Trichy, India. He received his M.Tech Degree in Bharathidasan University, Trichy, India in 2008 and has completed his PhD (Computer Science) in Bharathidasan University in 2014.He has more than 15 years of teaching experience and published more than 50 research papers in national and international journals. He has received an award Doctor of Letters from International Economics University,SAARCcountries. He is also a member in International Association of Engineers.