RESEARCH ARTICLE OPEN ACCESS

Amalgamation of Data Mining and Image Processing Techniques in Image Retrieval

K. Kamala Kannan[1], Dr. G. Anandharaj[2]

Assistant Professor, PG & Research Department of Computer Science, Sun Arts and Science College, Vettavalam Road, Keeranoor Village, Rajapalayam Post, Tiruvannamalai - 606 755,
Assistant Professor & Head, PG & Research Department of Computer Science, Adhiparasakthi College of Arts and Science (Autonomous), G.B.Nagar, Kalavai, Vellore Dt – 632 506,

Abstract: In the field of Image Processing, Mining is a new technique. Image mining is the extraction of latent data, amalgamation of image data and surplus patterns which are invisible in images. This involves Image Processing, Data Mining etc., without any prior notice, it can create all the important patterns. This research paper focuses on the extracting of knowledge from a huge database. Information is transferring message through direct or indirect technique. These techniques are neural network, clustering, correlation and association. This writing explains how data mining is used in the fields of telecommunication, fraud detection, manufacturing, marketing and education sector. Through this technique, we can use size, texture and dominant colour factors of an image. Gray Level Co-Occurance Matrix(GLCM) feature determines the texture of an image. Texture and colour features are normalized. Data Mining predicts the earth quake and Bioinformatics. It also analyses agricultural fields. This is used in cloud computing to retrieve meaningful information from data storage.

Keywords: Data Mining, Feature Extraction, Image Retrieval, Clustering, knowledge discovery database, Gray Level Co-Occurance Matrix and Cloud Computing,.

I. INTRODUCTION

In the actual world, massive amount of data is found in the education, industry, medical and many other branches. These data may give knowledge and information for decision making. For instance, we can detect the drop out students in any college or university and discover the sales details in the shopping databases. These data can be analysed, shortened or understood to meet the challenges.[1] Data Mining is the significant idea for data analysis, discovering amazing patterns from the large data, knowing the data stored in various databases, like warehouse, World Wide Web and external sources. The pattern is to understand the unknown valid and the potential data. Data Mining is a kind of sorting techniques used to extract the hidden patterns. Their goals are past recovery of data or information. They help us identify the hidden patterns and reduce the level of complexity. They also save time.[2] Data Mining is sometimes treated as Knowledge Discovery in Database(KDD).[3] KDD process consists of following steps shown below.

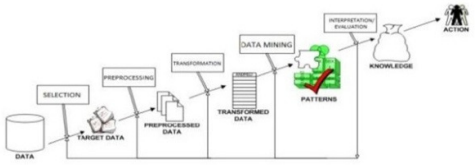

Fig.1. Knowledge Data Mining

- **Selection:** Select data from various resources where operation to be performed.
- **Preprocessing**: Also known as data cleaning in which remove the unwanted data.
- **Transformation:** Transform/consolidate into a new format for processing.
- **Data mining:** Identify the desire result.
- **Interpretation/Evaluation:** Interpret the result/query to give meaningful information.

Various algorithms and techniques like Classification, Clustering, Regression, Artificial Intelligence, Neural Networks, Association Rules, Decision Trees, Genetic Algorithm, Nearest Neighbor method etc., are meant for knowledge discovery from databases[5].The main objective of this paper learns about the data mining andthe rest of this Section 2 discusses data mining models and techniques. Section 3 explores the application of data mining. Finally, we conclude the paper in Section 4.

IMAGE MINING

This is to search and discover the valid and hidden data largely. The above figure (Fig.1) exhibits the different process of Image mining system. Some other methods too used to gather knowledge. They are, Image Retrieval, Data Mining, Image Processing and Artificial Intelligence. The methods permit Image mining to follow two different approaches. The first is to extract from databases or images. The second is to mine the alphanumeric data or images. Here the feature extraction reduces dimensionally. If the input data is more to be accessed, it is doubted as notoriously repeated, then the input data will be changed into a reduced set of features. It simplifies the quantity of resources needed to locate a lot of data clearly. Many other features are used in Image Retrieval system. The most famous are color features, shape features and texture ones.

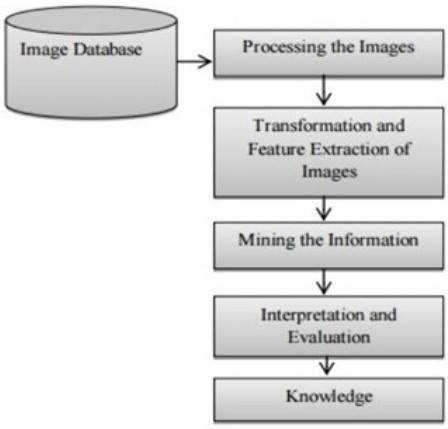

Fig.2. Image Mining Process

II. FEATURE EXTRACTION

Generally the feature extraction has got a major problem in detecting the objects but the Genetic Algorithm (GA) gives an easy common and powerful framework for detecting the better sets of features. Therefore it leads to lower detection error rates. Zehange Sun et al., [13] debate to carry out the method using Principle Component Analysis (PCA) and Classifications using Support Vector Machines (SVMs). Hence GA can remove the detection and unwanted features. The methods have two difficult objects detection problems – vehicle detection and face detection. They boost the performance of both systems using SVMs for classification. According to Patricia G.Foschi [10], feature selection and extraction is the pre-processing step of Image Mining. It is a critical step. Mining from images is to extract patterns and derive the images. Its aim is to identify the best ones. In the views of Broun, Ross A et al., [3] discuss the need of digital images forensics which underpin the design of mining system. It can be trained by a hierarchical SVMs to detect objects. Image mining generally deals with the study and development of new technologies. It is not only to rediscover relevant images; but also to innovate the image patterns. Fernandez.J et al.,[4] exhibit how a natural source of parallelism can be utilized to reduce the cost of mining. The images from the database are first pre-processed to improve their quality. They undergo several transformations to generate the important features from the images with the help of generated features.Mining can be done using data mining techniques to discover important patterns.

A. Color Features

Image Mining gives unique characteristics due to the richness of the data shown. Its evaluation result needs the performance parameters. Aura Conci et al.,[2] refer an evaluation for comparing the function by colour. Experiments with colour affinity mining by quantization on colour space and measures of similarity illustrate the scheme. Lukaz Kobylinski and Krzysztof Walczak[9] proposed a fast and effective method of indexing image metadatabases. The index is made to their colour characteristics Binary Thresholded Histogram (BTH), a color feature description method to create a metadatabase index. The BTH is proved to be a sufficient method to show the characteristics of image databases.

Ji Zhang, Wynne Hsn and Mong Li Lee [8] recommended an effective information driven framework for image mining. They divided four levels of information: Pixel Level, Object Level, Semantic Concept Level and Pattern and Knowledge Level.

B. Texture Feature

The Human percepts the images which is based on the color histogram texture. The Human Neurons hold the 1012 of information; the Human brain knows everything with the sensory organs like eye which transfers the image to the brain which interprets the image. According to Rajshree S. Dubey et al., [12] the mining images are based on the color Histogram, texture of Image. Janani. M and Dr. Manicka Chezian. R [7] refer that Image Mining is a pivotal method used to mine

knowledge from Images. This is based on the content based Image Retrieval system. Color, texture, pattern and shape of objects are the basis of visual content.

C. Shape Feature

According to Peter Stanchev [11] a new method is proposed on extraction of low level color, shape and texture into high level semantic features with the help of an image mining method. Johannes Itten's theory is offered for getting high level shape features. Harini D.N.D and Dr. Lalitha Bhaskari. D [5] argue that Image Retrieval is simply to reveal out low level pixel representation to recognize high level image objects and their relationships [22,23].

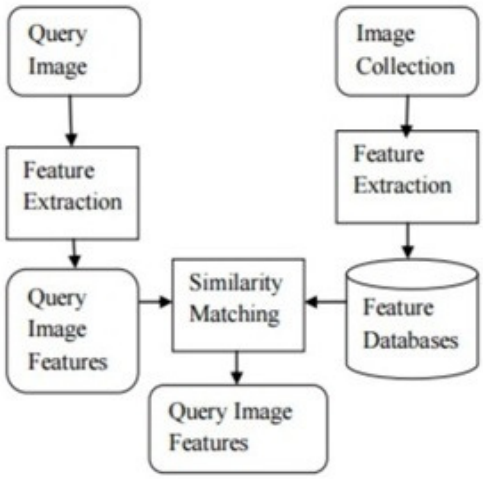

Fig.3. Content Based Image Retrieval System Architecture

III.METHODOLOGY

The gray-level co-occurrence matrix (GLCM) considers the relationship of pixels. This calculates how often the pairs of pairs of pixel with specific values and in a specified spatial relationship in an image.

Understanding a Gray-Level Co-Occurrence Matrix

We use the graycomatrix functionto make a GLCM. It creates GLCM by calculating how often a pixel with the intensity (Gray Level) value i occurs in a default. Each element (i,j) is the sum of the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image. Graycomatix uses scaling to reduce the number of intensity values. The Num levels and the Gray Limits control this scaling of gray level. Let us understand the process through the following diagram. The following figures explains how graycomatrix calculates the first three values in a GLCM.

To illustrate, the following figure show graycomatrix calculates the first three values in a GLCM. In the output GLCM, element (1,1) contains the value 1 because there is only one instance in the input image where two horizontally adjacent pixels have the values 1 and 1, respectively. glcm(1,2) contains the value 2 because there are two instances where two horizontally adjacent pixels have the values 1 and 2. Element (1,3) in the GLCM has the value 0 because there are no instances of two horizontally adjacent pixels with the values and 3.graycomatrix continues processing the input image, scanning the image for other pixel pairs (i,j) and recording the sums in the corresponding elements of the GLCM.

Process Used to Create the GLCM

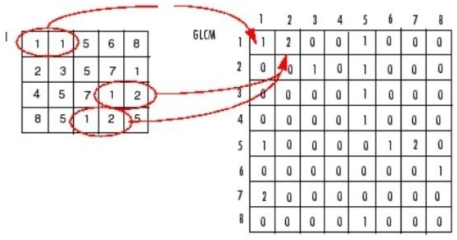

Fig.4. Process used to create the GLCM

Specify Offset Used in GLCM Calculation

By default, a single GLCM is created by the graycomatrix with offset as two horizontally adjacent pixels. A single GLCM might not adequate to describe the texture features of the input image. A single offset might not be sensitive to texture. So graycomatrix can make multiple GLCM for a one input image. The offsets produce multiple GLCM to graycomatrix function. They define mainly pixel relationships of different directions (Horizontal, Vertical and Two diagonals) and four distances. In this way, the input image is shown by 16 GLCMs. When we calculate statistics from these GLCMs, we can take the average.

Weighted Euclidean Distance

The standardized Euclidean distance between two J-dimensional vectors can be written as:

$$d_{x,y} = \sqrt{\sum_{j=1}^{J} \left(\frac{x_j}{S_j} - \frac{y_j}{S_j}\right)^2}$$

Where s_j is the sample standard deviation of the j-thvariable. Notice that we need not subtract the j-th mean from x_j and y_j because they will just cancel out in the differencing. Now (1.1) can be rewritten in the following equivalent way:

$$d_{x,y} = \sqrt{\sum_{j=1}^{J} \frac{1}{S_j^2}(x_j - y_j)^2}$$

$$= \sqrt{\sum_{j=1}^{J} w_j(x_j - y_j)^2}$$

Where $w_j = 1/s_{j2}$is the inverse of the j-th variance. w_j as a weight attached to the j-th variable: in other words

IV. DATA MINING TECHNIQUES

Data mining is gathering relevant information from disordered data. So it helps achieve specific objectives. Its aim is simply either to create a descriptive model or a predictive model. A description refers the main characteristics and a prediction allows the data miner to predict an unknown (often future) values of specific or the target variable.[7] Simply their goals are to use a variety of data mining techniques as shown in the figure 5[8].

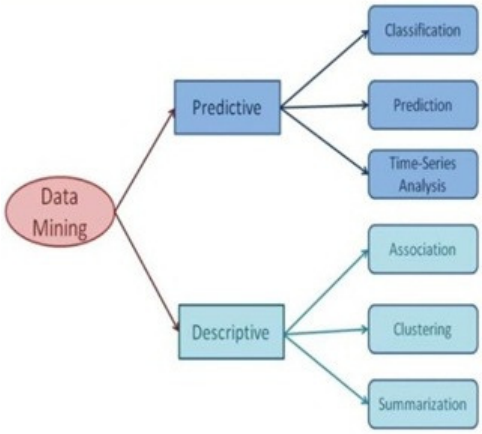

Fig.5. Data Mining Models

3.1 Classification: It is based on discrete and unordered. This is based on the desired output. It classifies the data based on the training set and values. These goals achieve using a decision tree, neural network and classification rule(If-Then). For instance we can apply this rule on the past record of the students who left for university and evaluate them. This helps us identify the performance of the students.

3.2 Regression: It is utilized to map a data part to a real valued prediction variable. [8] It can be used for prediction too. Here, the target values are known, for example, we can predict the child behavior based on family history.

3.3 Time Series Analysis: This process used the statistical techniques to model. It explains a time dependent series of data points. It is a method of using a model to create prediction (forecasts) for future happenings based on known past events.[9] Stock market is a good example.

3.4 Prediction: This techniques discover the relationship between independent variables and dependent and independent variable.[4] This model is based on continuous or organized value.

3.5 Clustering: It is a gathering of some data objects. Another cluster is dissimilar object. It generally finds out the similarities between the data of the same qualities. This is based on the unsupervised learning. For instance, city planning, image processing, pattern recognition, etc.,

3.6 Summarization: This is abstraction of data. It is formed as a set of related task. It provides an overview of data. For example, running race for long distance can be shortened total minutes or seconds. Association rule is another famous technique to mine the data. It find the most frequent item set. It discovers the patterns in data of relationships between items in the same transaction. It is also referred as "relation technique" as it relates the sets/items.[6,26]

3.7 Sequence Discovery: This sequence discloses the relationships among data.[8] It is a set of object associated with its own timeline of events. Natural disaster and analysis of DNA sequence and scientific experiments are best examples.

V. DATA MINING APPLICATIONS

Data mining is applied for fast access of data and valid information from a huge amount of data. Its main area includes marketing, fraud detection, finance, telecommunication, education sector, medical field, etc, some of the main applications are categorized below.

4.1 Data Mining in Education Sector: This is used in new emerging field called "Education Data Mining". This helps know performance of students, dropouts, students' behaviour and their choice of different courses. It is highly used in higher education sector.[10, 22]

4.2 Data Mining in Banking and Finance: It is

used largely in the Banking and Financial market.[11,27] It mines the credit card fraud, estimate risk and trend and profitability. In financial markets, it plays as a neural networks in stock forecasting price prediction etc.,

4.3 Data Mining in Market Basket Analysis: These methodologies are based on the shopping database. Their goal is to find out the products and the customers purchase. The shopping can utilize this information by putting these products more visible and accessible for customers.[12, 28]

4.4 Data Mining in Earthquake Prediction: This predicts the earthquake from the satellite maps. The quake is the sudden movement of the Earth's crust caused by the abrupt release of stress of a geologic fault in the interior. This is done in two types of prediction: forecasts (months to years in advance) and short-term predictions (hours or days in advance) [13, 29].

4.5 Data Mining in Bioinformatics: Bioinformatics created a huge amount of biological data. This is a new field of inquiry to generate and integrate large quantities of proteomic, genomic and other data.[4]

4.6 Data Mining in Telecommunication: This field has large amount of data consisting of huge customers. So it is need to mine the data to limit the fraudulent, improve their marketing efforts and better management of networks[4, 24].

4.7 Data Mining in Agriculture: This is mainly used to produce more crop yields. This is done in four parameters namely year, rainfall, production and area of sowing. It is used to improve yield from the prediction data. It can be promoted by using data mining techniques such as K Means, K nearest neighbor (KNN), Artificial Neural Network and support vector machine (SVM) [14,20,25].

4.8 Data Mining in Cloud Computing: This technique is used in cloud computing. Through cloud computing, the mining technique will permit the users to retrieve correct information from the data warehouse. This lessens the cost of infrastructure and storage[15, 21]. It utilizes the internet services to relay on clouds of servers to manage tasks. This techniques in cloud computing performs efficient, reliable and secure services for their users [21, 30].

VI. CONCLUSION

This article presents the expansion of Image mining. It gives a research on the image techniques measured earlier. This review finds on challenges and accountability of different prospects. This mainly focuses on data mining techniques in various projects. Its aim is to get information by current data. People from different fields can utilize association, clustering, prediction and classification techniques.

REFERENCES

[1]. Janani M and Dr. ManickaChezian. R, "A Survey On Content Based Image Retrieval System", International Journal of Advanced Research in Computer Engineering & Technology, Volume 1, Issue 5, pp 266, July 2012.

[2]. Aboli W. 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 pg. 219-224

[3]. Anil K. Jain and AdityaVailaya, "Image Retrieval using color and shape", In Second Asian Conference on Computer Vision, pp 5-8. 1995.

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

[5]. Hiremath. P. S and JagadeeshPujari, "Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement", International Journal of Computer Science and Security, Volume (1) : Issue (4).

[6]. Brown, Ross A., Pham, Binh L., and De Vel, Olivier Y, "Design of a Digital Forensics Image Mining System", in Knowledge Based Intelligent Information and Engineering Systems, pp 395-404, Springer Berlin Heidelberg, 2005.

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

[8]. Aura Conci, Everest Mathias M. M. Castro, "Image mining by Color Content", In Proceedings of 2001 ACM International Conference on Software Engineering and Knowledge Engineering (SEKE), Buenos Aires, Argentina Jun 13-15, 2001.

[9]. Er. RimmyChuchra "Use of Data Mining Techniques for the Evaluation of Student Performance: A Case Study" International Journal of Computer Science and Management Research Vol. 01, Issue 03 October 2012.

[10]. Ji Zhang, Wynne Hsu and Mong Li Lee, "An Information-Driven Framework for Image Mining" Database and Expert Systems Applications in Computer Science, pp 232 - 242, Springer Berlin Heidelberg, 2001.

[11]. LiorRokach and OdedMaimon, "Data Mining with Decision Trees: Theory and Applications (Series in Machine Perception andArtificial Intelligence)", ISBN: 981-2771-719, World Scientific Publishing Company, 2008.

[12]. Venkatadri.M and Lokanatha C. Reddy ,"A comparative study on decision tree classification algorithm in data mining" , International Journal Of Computer Applications In Engineering ,Technology And Sciences (IJCAETS), Vol.- 2 ,no.- 2 , pp. 24- 29 , Sept 2010.

[13]. XingquanZhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978-1-59904-252, Hershey, New York, 2007.

[14]. Zhao, KaidiandLiu, Bing, Tirpark, Thomas M. and Weimin, Xiao, "A Visual Data Mining Framework for Convenient Identification of Useful Knowledge", ICDM '05 Proceedings of the Fifth IEEE International Conference on Data Mining, vol.-1, no.-1,pp.- 530-537,Dec 2005.

[15]. Li Lin, Longbing Cao, Jiaqi Wang, Chengqi Zhang, "The Applications of Genetic Algorithms in Stock Market Data Mining Optimisation", Proceedings of Fifth International Conference on Data Mining, Text Mining and their Business Applications, pp-593-604,sept 2005.

[16]. V. Gudivada and V. Raghavan. Content-based image retrieval systems. IEEE Computer, 28(9):18-22, September 1995.

[17]. J. Han and M. Kamber. "Data Mining, Concepts and Techniques", Morgan Kaufmann, 2000.

[18]. Nikita Jain, Vishal Srivastava "DATA MINING TECHNIQUES: A SURVEY PAPER" IJRET: International Journal of Research in Engineering and Technology, Volume: 02 Issue: 11 | Nov-2013.

[19]. Peter Stanchev, "Image Mining for Image Retrieval", In Proceedings of the IASTED Conference on Computer Science and Technology, pp 214-218, 2003.

[20]. Jeyakumar, Balajee, MA SaleemDurai, and Daphne Lopez. "Case Studies in Amalgamation of Deep Learning and Big Data." HCI Challenges and Privacy Preservation in Big Data Security, pp 159, 2017.

[21]. Ranjith, D., J. Balajee, and C. Kumar. "In premises of cloud computing and models." International Journal of Pharmacy and Technology 8, no. 3, pp. 4685-4695, 2016.

[22]. Kamalakannan, S. "G., Balajee, J., SrinivasaRaghavan., "Superior content-based video retrieval system according to query image"." International Journal of Applied Engineering Research 10, no. 3, pp7951-7957, 2015.

[23]. Rajeshwari, A., T. C. Prathna, J. Balajee, N. Chandrasekaran, A. B. Mandal, and A. Mukherjee. "Computational approach for particle size measurement of silver nanoparticle from electron microscopic image." Int. J. Pharm. Pharm. Sci. 5, no. 2 pp619, 2013.

[24] Janarthanan Y, Balajee J.M, and Srinivasa Raghava S. "Content based video retrieval and analysis using image processing: A review."International Journal of Pharmacy and Technology 8, no.4 (2016): 5042-5048.

[25] Ushapreethi P, Balajee Jeyakumar and BalaKrishnan P, Action Recongnition in Video Survillance Using Hipi and Map Reducing Model, International Journal of Mechanical Engineering and Technology 8(11), 2017,pp. 368–375.

[26] D. Ranjith, J. M. Balajee and C. Kumar," Trust computation methods in mobile ADHOC network using glomosim: A Review" International Journal of Scientific Research and Modern Education, Vol. I, Issue I, pp. 777-780, Nov.2016.

[27] Rangith. D Lakshmi narayanan. J and Balajee. J," A study of behavior on information system in a university campus by analysis of people mobility" International journal of research in computer application & management, Vol. 6, Issue 7, pp. 29-31, Jul.2016.

[28] Sethumadahavi R Balajee J "Big Data Deep Learning in Healthcare for Electronic Health Records," International Scientific Research Organization Journal, vol. 2, Issue 2, pp. 31–35, Jul. 2017.

[29] Witten, Ian H., Eibe Frank, Mark A. Hall, and Christopher J. Pal. *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[30] Conte, Samuel Daniel, Carl De Boor, S. D. Conte, Carl De Boor, Etats-Unis Mathématicien, and Carl De Boor. *Elementary numerical analysis: an algorithmic approach*. Vol. 3. New York: McGraw-Hill, 1980.