IJIREEICE

ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 3, Issue 12, December 2015

Image Retrieval Using Data Mining and Image Processing Techniques

Preeti Chouhan[1], Mukesh Tiwari[2]

M. Tech Research Scholar, Digital Electronics, LNCT, Jabalpur, India[1]

Assistant Professor, Electronics & Communication Engineering, LNCT, Jabalpur, India[2]

Abstract: In the domain of Image processing, Image mining is advancement in the field of data mining. Image mining is the extraction of hidden data, association of image data and additional pattern which are quite not clearly visible in image. It's an interrelated field that involves, Image Processing, Data Mining, Machine Learning, Artificial Intelligence and Database. The lucrative point of Image Mining is that without any prior information of the patterns it can generate all the significant patterns. This is the writing for a research done on the assorted image mining and data mining techniques. Data mining refers to the extracting of knowledge /information from a huge database which is stored in further multiple heterogeneous databases. Knowledge/ information is communicating of message through direct or indirect technique. These techniques include neural network, clustering, correlation and association. This writing gives an introductory review on the application fields of data mining which is varied into telecommunication, manufacturing, fraud detection, and marketing and education sector. In this technique we use size, texture and dominant colour factors of an image. Gray Level Co-occurrence Matrix (GLCM) feature is used to determine the texture of an image. Features such as texture and color are normalized. The image retrieval feature will be very sharp using the texture and color feature of image attached with the shape feature. For similar types of image shape and texture feature, weighted Euclidean distance of color feature is utilized for retrieving features.

Keywords: Data Mining, Image Mining, Feature Extraction, Image Retrieval, Association, Clustering, knowledge discovery database,Gray Level Co-occurrence Matrix, centroid, Weighted Euclidean Distance.

I. INTRODUCTION

DATA MINING

In the real world, huge amount of data are available in education, medical, industry and many other areas. Such data may provide knowledge and information for decision making. For example, you can find out drop out student in any university, sales data in shopping database. Data can be analysed , summarized, understand and meet to challenges.[1] Data mining is a powerful concept for data analysis and process of discovery interesting pattern from the huge amount of data, data stored in various databases such as data warehouse , world wide web , external sources .Interesting pattern that is easy to understand, unknown, valid ,potential useful. Data mining is a type of sorting technique which is actually used to extract hidden patterns from large databases. The goals of data mining are fast retrieval of data or information, knowledge Discovery from the databases, to identify hidden patterns and those patterns which are previously not explored, to reduce the level of complexity, time saving, etc[2]. Sometimes data mining treated as knowledge discovery in database (KDD)[3] . KDD is an iterative process, consist a following step shown in

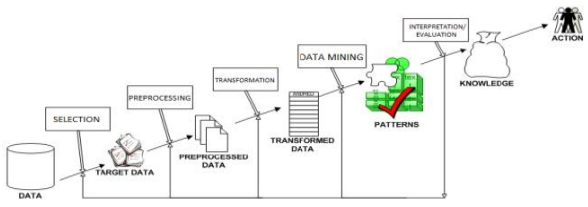


Fig.1. Knowledge Data Mining

- 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 report/ 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. And the 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

Image mining is the process of searching and discovering valuable information and knowledge in large volumes of data. Fig. 1 shows the Typical Image Mining Process. Some of the methods used to gather knowledge are, Image Retrieval, Data Mining, Image Processing and Artificial Intelligence. These methods allow Image Mining to have two different approaches. One is to extract from databases or collections of images and the other is to mine a combination of associated alphanumeric data and collections of images. In pattern recognition and in image processing, feature extraction is a special form of

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 3, Issue 12, December 2015

dimensionality reduction. When the input data is too large to be processed and it is suspected to be notoriously redundant, then the input data will be transformed into a reduced representation set of features. Feature extraction involves simplifying the amount of resources required to describe a large set of data accurately. Several features are used in the Image Retrieval system. The popular amongst them are Color features, Texture features and Shape features.

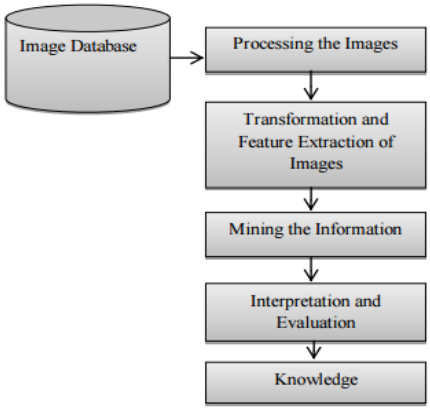


Fig.2. Image Mining Process

II. FEATURE EXTRACTION

Feature selection is an important problem in object detection, and demonstrates that Genetic Algorithm (GA) provides a simple, general and powerful framework for selecting good sets of features, leading to lower detection error rates. Zehang Sun et al., [13] discuss to perform Feature Extraction using popular method of Principle Component Analysis (PCA) and Classifications using Support Vector Machines (SVMs). GAs is capable of removing detection-irrelevant Features. The methods are on two difficult object detection problems, Vehicle detection and Face Detections. The methods boost the performance of both systems using SVMs for Classification. Patricia G. Foschi [10] discuss that Feature selection and extraction is the pre-processing step of Image Mining. Obviously this is a critical step in the entire scenario of Image Mining. The approach to mine from Images is to extract patterns and derive knowledge from large collections of images which mainly deals with identification and extraction of unique features for a particular domain. Though there are various features available, the aim is to identify the best features and thereby extract relevant information from the images. Increasing amount of illicit image data transmitted via the internet has triggered the need to develop effective image mining systems for digital forensics purposes. Brown, Ross A et al., [3] discuss the requirements of digital image forensics which underpin the design of our forensic image mining system. This system can be trained by a hierarchical SVM to detect objects and scenes which are made up of components under spatial or non-spatial constraints. Bayesian networks approach used to deal with information uncertainties which are inherent in forensic work. Image mining normally deals with the study and development of new technologies that allow accomplishing this subject. Image mining is not only the simple fact of recovering relevant images; but also the innovation of image patterns that are noteworthy in a given collection of images. Fernandez. J et al., [4] show how a natural source of parallelism provided by an image can be used to reduce the cost and overhead of the whole image mining process. The images from an image database are first pre-processed to improve their quality. These images then undergo various transformations and feature extraction to generate the important features from the images. With the generated features, mining can be carried out using data mining techniques to discover significant patterns.

A. Color Feature

Image mining presents special characteristics due to the richness of the data that an image can show. Effective evaluation of the results of image mining by content requires that the user point of view is used on the performance parameters. Aura Conci et.al, [2] proposed an evaluation framework for comparing the influence of the distance function on image mining by colour. Experiments with colour similarity mining by quantization on colour space and measures of likeness between a sample and the image results have been carried out to illustrate the proposed scheme. Lukasz Kobyli´nski and Krzysztof Walczak [9] proposed a simple but fast and effective method of indexing image meta databases. The index is created by describing the images according to their color characteristics, with compact feature vectors, that represent typical color distributions. Binary Thresholded Histogram (BTH), a color feature description method proposed, to the creation of a meta database index of multiple image databases. The BTH, despite being a very rough and compact representation of image colors, proved to be an adequate method of describing the characteristics of image databases and creating a meta database index for querying large amounts of data.
Ji Zhang, Wynne Hsu and Mong Li Lee [8] proposed an efficient information-driven framework for image mining. In that they made out four levels of information: Pixel Level, Object Level, Semantic Concept Level, and Pattern and Knowledge Level.

B. Texture Feature

The image depends on the Human perception and is also based on the Machine Vision System. The Image Retrieval is based on the color Histogram, texture. The perception of the Human System of Image is based on the Human Neurons which hold the 1012 of information; the Human brain continuously learns with the sensory organs like eye which transmits the Image to the brain which interprets the Image. Rajshree S. Dubey et.al, [12] examines the State-of-art technology Image mining techniques which are based on the Color Histogram, texture of Image. The query Image is taken then the Color Histogram and Texture is taken and based on this the resultant Image is output. Janani. M and Dr. Manicka Chezian. R [7] discusses Image mining is a vital technique which is used to mine knowledge from image. The development of the Image Mining technique is based on the Content Based Image Retrieval system. Color, texture, pattern, shape of

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 3, Issue 12, December 2015

objects and their layouts and locations within the image, etc are the basis of the Visual Content of the Image and they are indexed.

C. Shape Feature

Peter Stanchev [11] proposed a new method for image retrieval using high level semantic features is proposed. It is based on extraction of low level color, shape and texture characteristics and their conversion into high level semantic features using fuzzy production rules, derived with the help of an image mining technique. Dempster-Shafer theory of evidence is applied to obtain a list of structures containing information for the image high level semantic features. Johannes Itten theory is adopted for acquiring high level color features. Harini. D. N. D and Dr. Lalitha Bhaskari. D [5] discuss Image Retrieval, which is an important phase in image mining, is one technique which helps the users in retrieving the data from the available database. The fundamental challenge in image mining is to reveal out how low-level pixel representation enclosed in a raw image or image sequence can be processed to recognize high-level image objects and relationships.

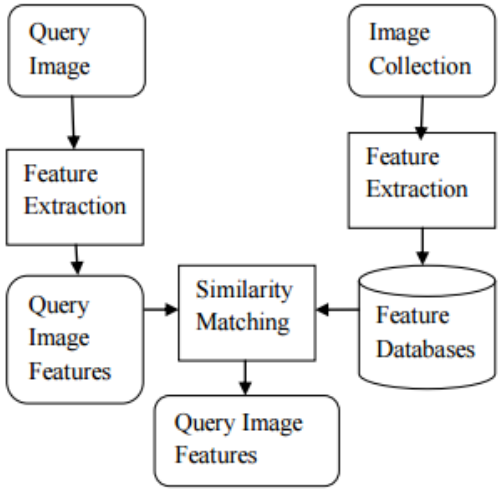


Fig.3. Content Based Image Retrieval System Architecture

III.METHODOLOGY

A statistical method of examining texture that considers the spatial relationship of pixels is the gray-level co-occurrence matrix (GLCM), also known as the gray-level spatial dependence matrix. The GLCM functions characterize the texture of an image by calculating how often pairs of pixel with specific values and in a specified spatial relationship occur in an image, creating a GLCM, and then extracting statistical measures from this matrix. (The texture filter functions, described in Texture Analysis cannot provide information about shape, i.e., the spatial relationships of pixels in an image.)

Understanding a Gray-Level Co-Occurrence Matrix

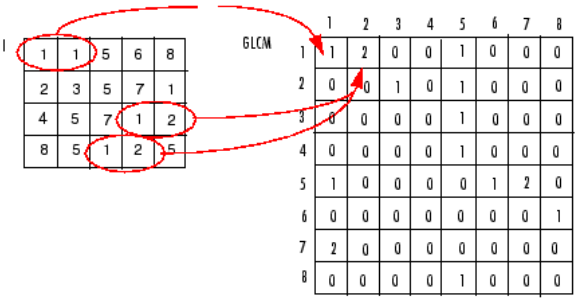
To create a GLCM, use the graycomatrix function. The graycomatrix function creates a gray-level co-occurrence matrix (GLCM) by calculating how often a pixel with the intensity (gray-level) value i occurs in a specific spatial relationship to a pixel with the value j. By default, the spatial relationship is defined as the pixel of interest and the pixel to its immediate right (horizontally adjacent), but you can specify other spatial relationships between the two pixels. Each element (i,j) in the resultant GLCM is simply the sum of the number of times that the pixel with value i occurred in the specified spatial relationship to a pixel with value j in the input image.

The number of gray levels in the image determines the size of the GLCM. By default, graycomatrix uses scaling to reduce the number of intensity values in an image to eight, but you can use the Num Levels and the Gray Limits parameters to control this scaling of gray levels. See the graycomatrix reference page for more information.

The gray-level co-occurrence matrix can reveal certain properties about the spatial distribution of the gray levels in the texture image. For example, if most of the entries in the GLCM are concentrated along the diagonal, the texture is coarse with respect to the specified offset. You can also derive several statistical measures from the GLCM. See Derive Statistics from GLCM and Plot Correlation for more information.

To illustrate, the following figure shows how 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 1 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



Specify Offset Used in GLCM Calculation

By default, the graycomatrix function creates a single GLCM, with the spatial relationship, or offset, defined as two horizontally adjacent pixels. However, a single GLCM might not be enough to describe the textural features of the input image. For example, a single horizontal offset might not be sensitive to texture with a vertical orientation. For this reason, graycomatrix can create multiple GLCMs for a single input image.

To create multiple GLCMs, specify an array of offsets to the graycomatrix function. These offsets define pixel relationships of varying direction and distance. For example, you can define an array of offsets that specify

DOI 10.17148/IJIREEICE.2015.31212

IJIREEICE
ISSN (Online) 2321 – 2004
ISSN (Print) 2321 – 5526
INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH IN ELECTRICAL, ELECTRONICS, INSTRUMENTATION AND CONTROL ENGINEERING
Vol. 3, Issue 12, December 2015

four directions (horizontal, vertical, and two diagonals) and four distances. In this case, the input image is represented by 16 GLCMs. When you calculate statistics from these GLCMs, you 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-th variable. Notice that we need not subtract the j-th mean from x_i and y_i 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 $w_i = 1/s_i^2$ is the inverse of the j-th variance. w_i as a weight attached to the j-th variable: in other words

IV. DATA MINING TECHNIQUES

Data mining means collecting relevant information from unstructured data. So it is able to help achieve specific objectives. The purpose of a data mining effort is normally either to create a descriptive model or a predictive model .A descriptive model presents, in concise form, the main characteristics of the data set. The purpose of a predictive model is to allow the data miner to predict an unknown (often future) value of a specific variable; the target variable [7]. The goal of predictive and descriptive model can be achieved using a variety of data mining techniques as shown in figure 5[8].

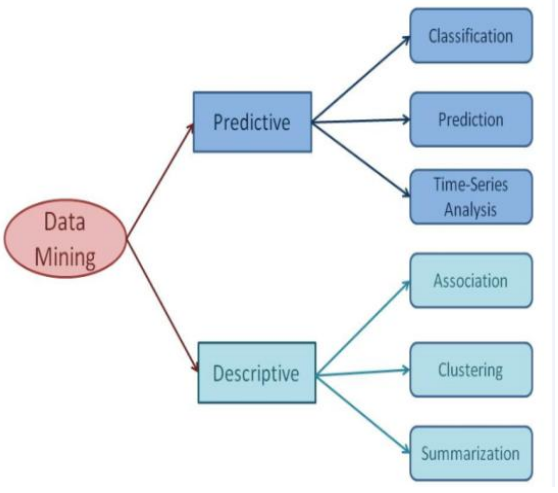


Fig.5. Data Mining Models

3.1 **Classification:** Classification based on categorical (i.e. discrete, unordered).This technique based on the supervised learning (i.e. desired output for a given input is known) .It can be classifying the data based on the training set and values (class label). These goals are achieve using a decision tree, neural network and classification rule (IF-Then).for example we can apply the classification rule on the past record of the student who left for university and evaluate them. Using these techniques we can easily identify the performance of the student.

3.2 **Regression:** Regression is used to map a data item to a real valued prediction variable [8]. In other words, regression can be adapted for prediction. In the regression techniques target value are known. For example, you can predict the child behaviour based on family history.

3.3 **Time Series Analysis:** Time series analysis is the process of using statistical techniques to model and explain a time-dependent series of data points. Time series forecasting is a method of using a model to generate predictions (forecasts) for future events based on known past events [9]. For example stock market.

3.4 **Prediction:** It is one of a data mining techniques that discover the relationship between independent variables and the relationship between dependent and independent variables [4].Prediction model based on continuous or ordered value.

3.5 **Clustering:** Clustering is a collection of similar data object. Dissimilar object is another cluster. It is way finding similarities between data according to their characteristic. This technique based on the unsupervised learning (i.e. desired output for a given input is not known). For example, image processing, pattern recognition, city planning.

3.6 **Summarization:** Summarization is abstraction of data. It is set of relevant task and gives an overview of data. For example, long distance race can be summarized total minutes, seconds and height. Association Rule: Association is the most popular data mining techniques and fined most frequent item set. Association strives to discover patterns in data which are based upon relationships between items in the same transaction. Because of its nature, association is sometimes referred to as "relation technique". This method of data mining is utilized within the market based analysis in order to identify a set, or sets of products that consumers often purchase at the same time [6].

3.7 **Sequence Discovery:** Uncovers relationships among data [8]. It is set of object each associated with its own timeline of events. For example, scientific experiment, natural disaster and analysis of DNA sequence.

V. DATA MINING APPLICATIONS

Various field adapted data mining technologies because of fast access of data and valuable information from a large amount of data. Data mining application area includes marketing, telecommunication, fraud detection, finance, and education sector, medical and so on. Some of the main applications listed below:

4.1 **Data Mining in Education Sector:** We are applying data mining in education sector then new emerging field called "Education Data Mining". Using these term enhances the performance of student, drop out student, student behaviour, which subject selected in the course. Data mining in higher education is a recent research Use of Data Mining in Various Field: A Survey Paper www.iosrjournals.org 20 | Page field and this area of research is gaining popularity because of its potentials to educational institutes. Use student's data to analyze their learning behaviour to predict the results [10].

4.2**Data Mining in Banking and Finance:** Data mining has been used extensively in the banking and financial markets [11]. In the banking field, data mining is used to predict credit card fraud, to estimate risk, to analyze the trend and profitability. In the financial markets, data mining technique such as neural networks used in stock forecasting, price prediction and so on.

4.3**Data Mining in Market Basket Analysis:** These methodologies based on shopping database. The ultimate goal of market basket analysis is finding the products that customers frequently purchase together. The stores can use this information by putting these products in close proximity of each other and making them more visible and accessible for customers at the time of shopping [12].

4.4**Data Mining in Earthquake Prediction:** Predict the earthquake from the satellite maps. Earthquake is the sudden movement of the Earth's crust caused by the abrupt release of stress accumulated along a geologic fault in the interior. There are two basic categories of earthquake predictions: forecasts (months to years in advance) and short-term predictions (hours or days in advance) [13].

4.5**Data Mining in Bioinformatics:** Bioinformatics generated a large amount of biological data. The importance of this new field of inquiry will grow as we continue to generate and integrate large quantities of genomic, proteomic, and other data [4].

4.6 **Data Mining in Telecommunication:** The telecommunications field implement data mining technology because of telecommunication industry have the large amounts of data and have a very large customer, and rapidly changing and highly competitive environment. Telecommunication companies uses data mining technique to improve their marketing efforts, detection of fraud, and better management of telecommunication networks [4].

4.7 **Data Mining in Agriculture:** Data mining than emerging in agriculture field for crop yield analysis a with respect to four parameters namely year, rainfall, production and area of sowing. Yield prediction is a very important agricultural problem that remains to be solved based on the available data. The yield prediction problem can be solved by employing Data Mining techniques such as K Means, K nearest neighbour (KNN), Artificial Neural Network and support vector machine (SVM) [14].

4.8**Data Mining in Cloud Computing:** Data Mining techniques are used in cloud computing. The implementation of data mining techniques through Cloud computing will allow the users to retrieve meaningful information from virtually integrated data warehouse that reduces the costs of infrastructure and storage [15].Cloud computing uses the Internet services that rely on clouds of servers to handle tasks. The data mining technique in Cloud Computing to perform efficient, reliable and secure services for their users.

VI. CONCLUSION

The expansion of image processing is presented as Image mining. This writing provides a research on the image techniques surveyed earlier. This review on image mining implies on challenges and accountability of various prospects.

This writing gives an idea on data techniques and mining in various projects. Its main task is to obtain information through current data. These programs utilize association, clustering, prediction and classification techniques and so on. In coming work efforts will be made on clustering algorithms and its classification importance.

REFERENCES

[1]. Janani M and Dr. Manicka Chezian. 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 Aditya Vailaya, "Image Retrieval using color and shape", In Second Asian Conference on Computer Vision, pp 5-8. 1995.

[4]. Harini. D. N. D and Dr. Lalitha Bhaskari. 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 Jagadeesh Pujari, "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, Niket Bhargava and Rajnish Choubey, "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. Rimmy Chuchra "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]. Lior Rokach and Oded Maimon, "Data Mining with Decision Trees: Theory and Applications (Series in Machine Perception and Artificial 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]. Xingquan Zhu, Ian Davidson, "Knowledge Discovery and Data Mining: Challenges and Realities", ISBN 978- 1-59904-252, Hershey, New York, 2007.
[14]. Zhao, Kaidi and Liu, 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.

BIOGRAPHIES

Preeti Chouhan obtained his B. E. (Electronics & Communication) from Gyan Ganga Institute of Technology and Science, Jabalpur, & pursuing M. Tech. in Digital Electronics from Lakshmi Narain College of Technology, Jabalpur, M.P.

Mukesh Tiwari is currently working as Assistant Professor in Department of Electronics and Communication Engineering in Lakshmi Narain College of Technology, Jabalpur, M.P. He obtained his M. Tech. in Instrumentation & Control from Jabalpur Engineering College, Jabalpur, M.P.