IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
Vol. 5, Issue 5, May 2016

Feature Extraction Techniques for 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: Image mining is advancement in the field of data mining, in the domain of Image processing. Image mining is the additional pattern which is quite not clearly visible in the image, association of image data and extraction of hidden data. This field is interrelated and involves Database, Artificial Intelligence, Data Mining, Machine Learning, and Image processing. Image Mining has a lucrative point that without any information of the patterns it can generate all the significant patterns. This writing is done for a research on the data mining techniques and assorted image mining. The term data mining is the extraction of information/ knowledge from a wide database which is further stored in multiple heterogeneous databases. Information/ Knowledge are conveying of message through direct or indirect methods. These methods include clustering, correlation, association and neural network. This thesis provides with a basic informatory review on the applied fields of data mining which is varied into manufacturing, telecommunication, education, fraud detecting and marketing sector. In this method we use texture, dominant colour factors and size of an image. The feature which is used to determine the image texture is called as Gray Level Co-occurrence Matrix (GLCM). The texture, color and such relative features are normalized. Due to the use of the texture and color feature of the attached image due to the shape feature the image retrieval feature will be very sharp. Weighted Euclidean distance of color feature is utilized for the retrieving of features, of similar types of image shape and texture feature.

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

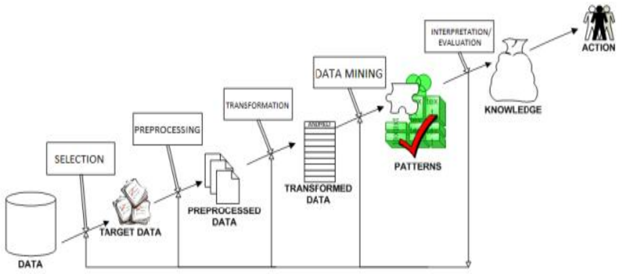


Fig.1. Knowledge Data Mining

- 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].

DOI 10.17148/IJARCCE.2016.55133

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 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.

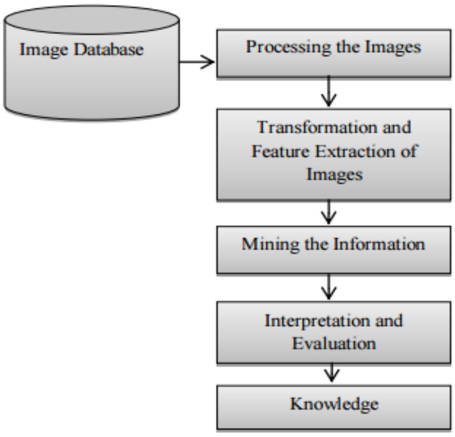


Image Database
Processing the Images
Transformation and Feature Extraction of Images
Mining the Information
Interpretation and Evaluation
Knowledge

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

DOI 10.17148/IJARCCE.2016.55133

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 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.

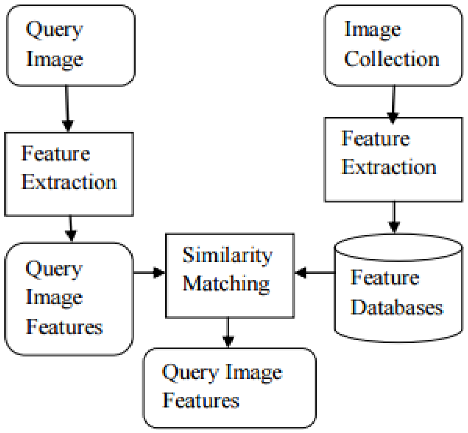
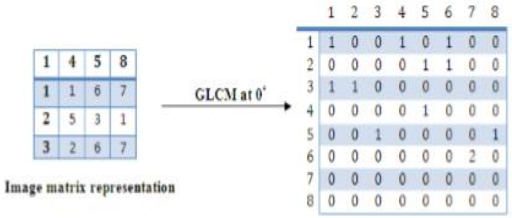


Fig.3. Content Based Image Retrieval System Architecture

III. METHODOLOGY



1	4	5	8
1	1	6	7
2	5	3	1
3	2	6	7

	1	2	3	4	5	6	7	8
1	1	0	0	1	0	1	0	0
2	0	0	0	0	1	1	0	0
3	1	1	0	0	0	0	0	0
4	0	0	0	0	1	0	0	0
5	0	0	1	0	0	0	0	1
6	0	0	0	0	0	0	2	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0

Fig. 4: Formation of Co-Occurrence Matrix for 0º Direction and a Single Gap Between the Pixels

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 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.

Texture Features from GLCM

After calculating GLCM in all four directions 0, 45, 90, 135 degrees features energy, Contrast, Homogeneity and Correlation are calculated as shown in equations:

$$\text{Energy}= \sum_{i,j=0}^{N-1}(P_{i,j})^2 \qquad (4)$$

$$\text{Contrast}= \sum_{i,j=0}^{N-1}P_{i,j}(i-j)^2 \qquad (5)$$

$$\text{Homogeneity}= \sum_{i,j=0}^{N-1}\frac{P_{i,j}}{1+(i-j)^2} \qquad (6)$$

$$\text{Correlation}= \sum_{i=1}^{N}\sum_{j=1}^{N}\frac{(i-m_r)(j-m_c)P_{ij}}{\sigma_r\,\sigma_c} \qquad (7)$$

$$m_r = \sum_{i=1}^{N} i \sum_{j=1}^{N} P_{ij} \qquad m_c = \sum_{j=1}^{N} j \sum_{i=1}^{N} P_{ij}$$

$$\sigma_r = \left(\sum_{i=1}^{N}(i-m_r)^2\sum_{j=1}^{N}P_{ij}\right)^{1/2} \quad \sigma_c = \left(\sum_{j=1}^{N}(j-m_c)^2\sum_{i=1}^{N}P_{ij}\right)^{1/2}$$

Here, N is the number of rows/columns of image matrix Q, Pij is the probability that a pair of points in Q will have values (Ni, Nj), mr and mc are the mean of rows and columns respectively, σr and σc are the standard deviation of rows and columns respectively.

Weighted Euclidean Distance

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

$$d_{x,y} = \sqrt{\sum_{j=1}^{J}\left(\frac{x_j}{s_j}-\frac{y_j}{s_j}\right)^2} \qquad ….(1.1)$$

DOI 10.17148/IJARCCE.2016.55133

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_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 $w_j = 1/s_j^2$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 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 2[8].

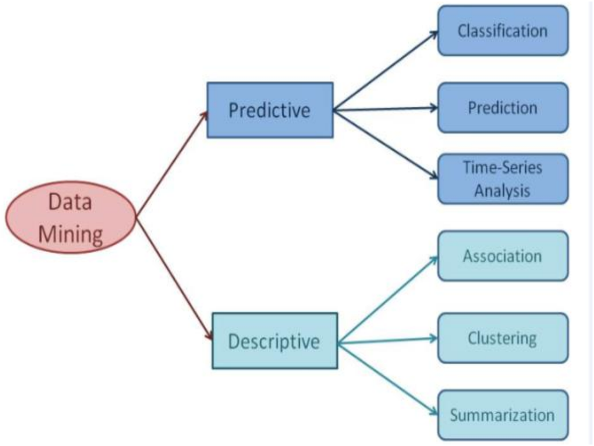


Fig.5: Data Mining Models

1.1 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.

1.2 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.

1.3 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.

1.4 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.

1.5 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.

1.6 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].

1.7 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.

RESULT

Table 5.1 We find the Result from the distance between data base images and query image by distance formula. Then we find the score of images and indexing the retrieve images according to their similarity.

S.No.	Data Base Images	Eculidean Distance= square root(feature test-feature vector)^2	Sort Distance
1	IMAGE(1)	2.2786	0
2	IMAGE(1)	2.9050	1.26
3	IMAGE(1)	3.2027	1.51
4	IMAGE(1)	2.1039	2.01
5	IMAGE(1)	2.017	2.10
6	IMAGE(1)	1.26	2.27
7	IMAGE(1)	1.51	2.41
8	IMAGE(1)	2.4109	2.45
9	IMAGE(1)	2.4570	2.90
10	IMAGE(1)	0	3.20

Table 5.2: Note-; Sorting the distance in ascending order and find the score

S.No.	Sort distance	Score=sort dist/max(sort dist)	percentage=score*100
1	0	0	0
2	1.26	0.0780	7.8
3	1.51	0.0936	9.3
4	2.017	0.1247	12.4
5	2.1039	0.1301	13.01
6	2.2786	0.1409	14.09
7	2.4109	0.1491	14.9
8	2.4570	0.1520	15.2
9	2.9050	0.1797	17.97
10	3.2027	0.1981	19.81

Indexing the images according their similarity result.

Image (1)

Plot of that Image

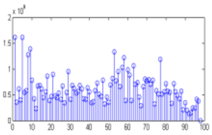


Image (2)

Plot of that Image

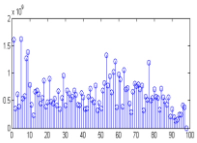
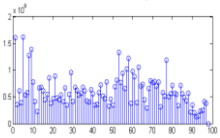


Image (3)

Plot of that Image



Image(4)

Plot of that image

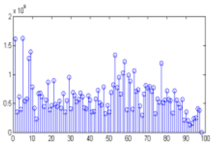


Image (5)

Plot of that image

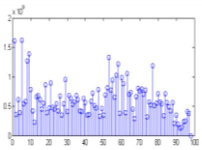


Image (6)

Plot of that image

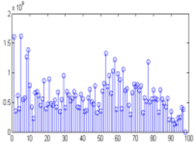


Image (7)

Plot of that image

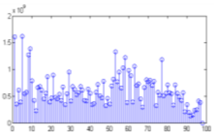


Image (8)

Plot of that Image

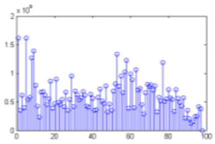


Image (9)

Plot of that Image

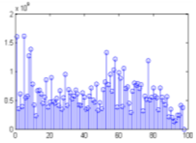
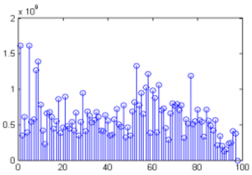


Image (10)

Plot of that Image



V. APPLICATIONS

1. In the military to find tanks or airstrips.
2. In urban development to determine the extent of housing sprawl.
3. In local government to track highway assets.

VI. CONCLUSION

We presented a novel approach for Content Based Image Retrieval by combining the colour shape, and texture features. Similarity between the images is ascertained by means of a distance function. The experimental result shows that the proposed method outperforms the other retrieval methods in terms of score. Moreover, the computational steps are effectively reduced the searching time of query image from database. As a result, there is a substation ally increase in the retrieval speed.

This paper proposed Texture Based Image Retrieval Using GLCM and image sub lock. The database used in this experiment is Texture library images database. The texture features are extracted based on GLCM (Gray Level Co occurrence Matrix) using four statistic features that are contrast, homogeneity, energy and correlation.

These four features are computed in four directions (00, 450, 900, and 1350). A total of 16 texture values are computed per an image sub block. The image is divided into nine sub-blocks in equal size. From the experiment of the study the retrieval result of image features using Euclidean distance. The effectiveness of retrieval result is evaluating by the distance and then find score of image. The retrieval result is based on similarity is computed after then retrieval results are ranked according to similarity

DOI 10.17148/IJARCCE.2016.55133

index. In order to improve the effectiveness of the study, method proposed should combine with other featuresultre and extent the number of image database to be tested.

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]. R. Datta, D. Joshi, J. Li and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age", ACM computing Survey, vol.40, no.2, pp.1-60, 2008.

[4]. 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.

[5]. A. M. Smeulders, M. Worring and S. Santini, A. Gupta and R. Jain, "Content Based Image Retrieval at the End of the Early Years", IEEE Transactions on Pattern Analysis and Machine Intelligence,22(12): pp. 1349-1380, 2000.

[6]. Y. Liu, D. Zang, G. Lu and W. Y. Ma, "A survey of content-based image retrieval with high level semantics", Pattern Recognition, Vol-40, pp-262-282, 2007.

[7]. T. Kato, "Database architecture for content-based image retrieval", In Proceedings of the SPIE – The International Society for Optical Engineering, vol.1662, pp.112-113, 1992.

[8]. M. Flickner, H Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafne, D. Lee, D. Petkovic, D. Steele and P. Yanker, "Query by Image and Video Content The QBIC System" IEEE Computer, pp-23-32, 1995

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

[10]. 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.

[11]. 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).

[12]. 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.

[13]. 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.

[14]. 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.

[15]. 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.

[16]. 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.

[17]. 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.

[18]. 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.

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

[20]. 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.

[21]. 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.

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

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

[24]. 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.

[25]. 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.

DOI 10.17148/IJARCCE.2016.55133