RESEARCH ARTICLE OPEN ACCESS

An Overview on Image Retrieval Using Image Processing Techniques

Mylavarapu Kalyan Ram [1], Dr.M.Venkateswara Rao [2], Challapalli Sujana [3]
PhD [1] Research Scholar, Gitam University,Vishakapatnam,
Asst Professor in Aditya Engineering College, Surampalem,Kakinada,
Associate Professor [2], Gitam University,Vishakapatnam
Asst Professor [3], Aditya Engineering College,Surampalem, Kakinada
Andhra Pradesh -India

ABSTRACT
Data mining refers to the extracting of knowledge /information from a huge database. There are number of topics in data mining such as Clustering, Classification, Association, Decision Tree, Graph mining, Multimedia mining and Image Mining. In above topics Image mining plays a vital role in every aspect. Image mining is the process of searching and discovering valuable information and knowledge in large volumes of data. Image mining draws basic principles from concepts in databases, machine learning, statistics, pattern The demand of image mining increases as the need of image data is growing day by day. There are many techniques developed in the earlier researches and eventually these techniques can reveal useful information according to the human requirements, but Image Mining still require more development especially in the area of web images. Image mining contains different research areas like Space, remote sensing, medical diagnosis etc. These techniques include neural network, clustering, correlation and association. This writing gives a 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.
Keywords :- Data Mining, Image Mining, Feature Extraction, Image Retrieval, Gray Level Co-occurrence Matrix (GLCM).

I. INTRODUCTION

1.1Preprocessing*:* In the image database, the spatial segmentation can be done at the region or edge level based on the requirements of the application. It can be done automatically or manually and it should be resemblance enough to retrieve the features that can reasonably capture the image content

Image Cleaning:

Image cleaning is the process of detecting and correcting (or removing) corrupt or inaccurate images from the set of images and refers to identify the unclear, incorrect, or irrelevant parts of the images and then replacing, modifying, or deleting the dusty or fouled image data.

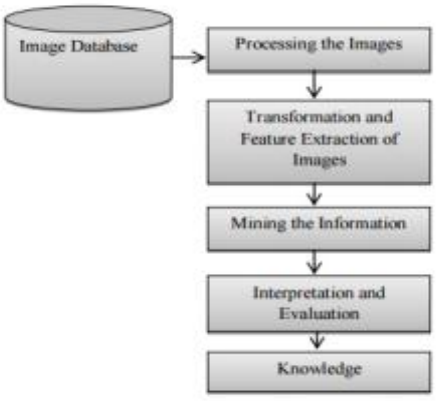

Fig 1: Image Mining Process

1.2Feature Selection and Extraction:

Feature selection and extraction a type of dimensionality reduction that efficiently represents the interesting parts of an image as a feature vector. This approach is useful when image sizes are large and a reduced feature representation is required to quickly complete tasks such as image matching and retrieval.

A.Colors Characteristics

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, [8] proposed an evaluation framework for comparing the influence of the distance function on image mining by color. 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 Krzyszt of Walczak proposed a simple but fast and effective method of indexing image metadata bases. The index is created by describing the images according to their color characteristics, with compact feature vectors, that represent typical color distributions. Binary Threshold Histogram (BTH), a color feature description method proposed, to the creation of a metadata base 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 metadatabase index for querying large amounts of data.
Fig 2: An example of color characteristics

Fig 3: An example of color characteristics

A. Texture Characteristics

Texture measures look for visual patterns in images and how they are spatially defined. Textures are represented by texels which are then placed into a number of sets, depending on how many textures are detected in the image. These sets not only define the texture, but also where in the image the texture is located.[5]
Texture is a difficult concept to represent. The identification of specific textures in an image is achieved primarily by modeling texture as a two-dimensional gray level variation. The relative brightness of pairs of pixels is computed such that degree of contrast, regularity, coarseness and directionality may be estimated.[3][11] The problem is in identifying patterns of co-pixel variation and associating them with particular classes of textures such as *silky*, or *rough*.
Other methods of classifying textures include:

- Co-occurrence matrix
- Laws texture energy
- Wavelet transform

B. Shape Characteristics

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. DempsterShafer 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.

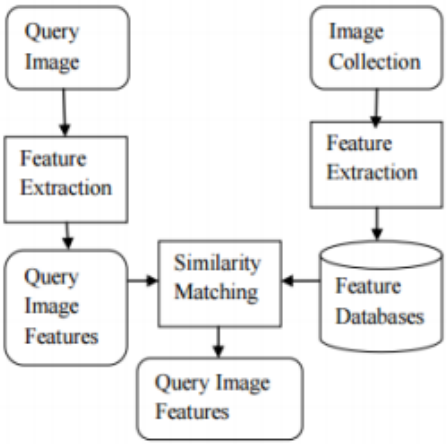

Fig 4: Content Based Image Retrieval (CBIR) System architecture

We discuss Image Retrieval, which is an important phase in image mining, it is one of the technique which helps the users in retrieving the data from the available database. The significant challenge in image mining is to reveal out how low-level pixel illustration enclosed in a raw image or image sequence can be processed to recognize high-level image objects and relationships.

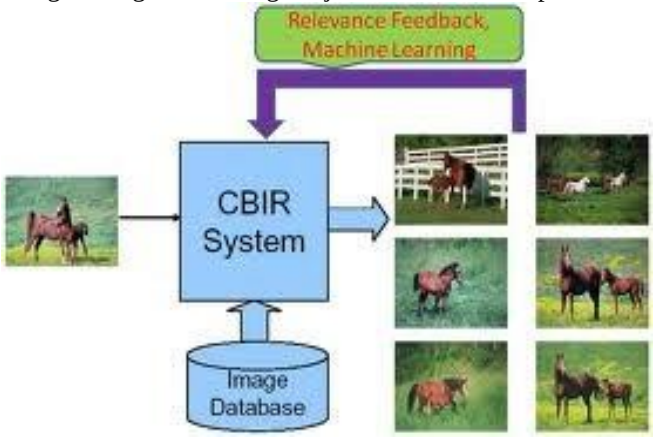

Fig 4. An Example for Content Based Image Retrieval (CBIR) System

Anil K. Jain and Aditya Vailaya [3] propose that the Color and Shape based queries provide better performance than either of the individual feature based queries. Combination of simple features which can be easily extracted is more efficient. The speed of retrievals can be also be increased by using Branch and Bound to compute the nearest neighbour for the query image without affecting the robustness of the system.

Applications:

- Potential uses for CBIR include:[2]
- Architectural and engineering design
- Art collections
- Crime prevention
- Geographical information and remote sensing systems
- Intellectual property
- Medical diagnosis
  - Military

II.METHODOLOGY

A analytical method of exploring 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 symbolize 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.

Texture Analysis Using the Gray-Level Co-Occurrence Matrix:

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.)

After you create the GLCMs, using graycomatrix, you can derive several statistics from them using graycoprops. These statistics provide information about the texture of an image. The following table lists the statistics.

Create a Gray-Level Co-Occurrence Matrix

To create a GLCM, use the `graycomatrix` function. The 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 lrelationship to a pixel with the value *j*. 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 `NumLevels` and the `GrayLimits`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. You can also derive several statistical measures from the GLCM. 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

Create the GLCM

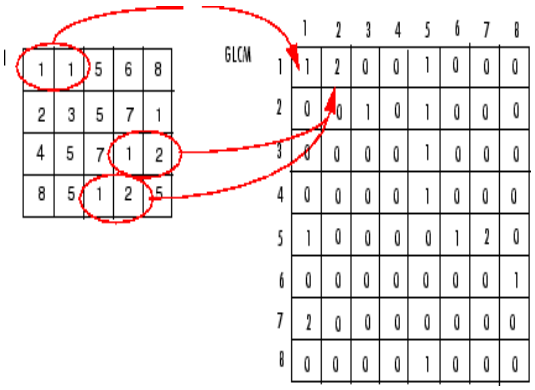

Fig 5: Process used to create the GLCM

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 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 Jdimensional 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 sj is the sample standard deviation of the j-th variable. Notice that we need not subtract the j-th mean from xj and yj 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 wj = 1/sj 2 is the inverse of the j-th variance. wj as a weight attached to the j-th variable: in other words

III PROPERTIES OF THE GLCM

1. It is square:

The reference pixels have the same range of values as the neighbour pixels, so the values along the top are identical to the values along the side.

2. Has the same number of rows and columns as the quantization level of the image:

The test image has four grey level values (0,1,2 and 3). Eight bit data has 256 possible values, so would yield a 256 x 256 square matrix, with 65,536 cells. 16 bit data would give a matrix of size 65536 x 65536 = 429,496,720 cells!

3. It is symmetrical around the diagonal:

The diagonal elements all represent pixel pairs with no grey level difference (0-0, 1-1, 2-2, 3-3 etc.). If there are high probabilities in these elements, then the image does not show much contrast: most pixels are identical to their neighbours.

When values in the diagonal are summed, the result is the probability of *any* pixel's being the same grey level as its neighbour.

Look at lines parallel to the diagonal. Cells one cell away from the diagonal represent pixel pairs with a difference of only one grey level (0-1, 1-2, 2-3 etc.). Similarly, values in cells two away from the diagonal show how many pixels have 2 grey level differences, and so forth. **The farther away from the diagonal, the greater the difference between pixel grey levels.**

IV DATA MINING TECHNIQUES

There are several major data mining techniques have been developing and using in data mining projects recently including
association, classification, clustering, prediction,sequential patterns and decision tree.

Association

Association is one of the best-known data mining technique. In association, a pattern is discovered based on a relationship between items in the same transaction. That's the reason why association technique is also known as *relation technique*. The association technique is used in *market basket analysis* to identify a set of products that customers frequently purchase together.

Classification

Classification is a classic data mining technique based on machine learning. Basically, classification is used to classify each item in a set of data into one of a predefined set of classes or groups. Classification method makes use of mathematical techniques such as decision trees, linear programming, neural network and statistics. In classification, we develop the software that can learn how to classify the data items into groups.

Clustering

Clustering is a data mining technique that makes a meaningful or useful cluster of objects which have similar characteristics using the automatic technique. The clustering technique defines the classes and puts objects in each class, while in the classification techniques, objects are assigned into predefined classes.

Sequential Patterns

Sequential patterns analysis is one of data mining technique that seeks to discover or identify similar

patterns, regular events or trends in transaction data over a business period

Decision trees

The A decision tree is one of the most common used data mining techniques because its model is easy to understand for users. In decision tree technique, the root of the decision tree is a simple question or condition that has multiple answers. Each answer then leads to a set of questions or conditions that help us determine the data so that we can make the final decision based on it.

IV.DATA MINING APPLICATIONS

Banking / Finance

Several data mining techniques e.g., distributed data mining have been researched, modelled and developed to help credit card fraud detection.

Data mining is used to identify customers' loyalty by analyzing the data of customer's purchasing activities such as the data of frequency of purchase in a period of time, a total monetary value of all purchases and when was the last purchase. After analyzing those dimensions, the relative measure is generated for each customer.

Health Care and Insurance

The growth of the insurance industry entirely depends on the ability to convert data into the knowledge, information or intelligence about customers, competitors, and its markets. Data mining is applied in insurance industry lately but brought tremendous competitive advantages to the companies who have implemented it successfully. The data mining applications in insurance industry are listed below:

- Data mining is applied in claims analysis such as identifying which medical procedures are claimed together. Data mining enables to forecasts which customers will potentially purchase new policies.
- Data mining allows insurance companies to detect risky customers' behaviour patterns.
- Data mining helps detect fraudulent behaviour.

Transportation

- Data mining helps determine the distribution schedules among warehouses and outlets and analyze loading patterns.

Medicine

- Data mining enables to characterize patient activities to see incoming office visits.
- Data mining helps identify the patterns of successful medical therapies for different illnesses.

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.

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

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)

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

V. CONCLUSION

The expansion of image processing is presented as Image mining. This writing provides research on image techniques surveyed earlier. Image mining implies on challenges and accountability of various prospects. Its main task is to obtain information through current data. These programs utilize association, clustering and classification techniques and so on

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.