



## A SURVEY OF VARIOUS FEATURE EXTRACTIONS USING IMAGE MINING TECHNIQUES

**Dr(Mrs).A.Hema**

*Associate Prof & Head, Department of Computer Application(UG),Kongunadu Arts and Science College, Coimbatore .*

### **Abstract**

*Image mining is the concept used to detect unusual patterns and extract implicit and useful Information. Image mining is an extension of Data mining. This technology has been considered as an advanced field for discovering information related to the images. Image Mining is a process of extracting knowledge concerning images. Mining Image data is the one of the essential features in this present scenario. Image data plays vital role in every aspect of the system such as business for marketing, hospital for surgery, engineering for construction, Web for publication and so on. It is the concept which can be used to extract potential information from the general collection of images. Image mining is a development for data mining which involves in multiple disciplines. It is a challenging field which extends traditional data mining from structured data to unstructured data such as image data. This paper deals a survey of extracting various features like color, text, shape, edge, from the images.*

**Keywords:** *Feature extraction, Image mining, Edge, Color, Shape, Text, Size, Data Mining.*

### **I. Introduction**

Image Mining is an extended branch of data mining that is concerned with the process of knowledge discovery concerning images. It deals with the extraction of image patterns from a large collection of images. Also deals with extraction of knowledge, image data relationship and other required patterns and uses ideas from image processing, image retrieval and machine learning, databases. The focus of image mining is on the extraction of knowledge patterns from a large collection of images. The goal is to discover image patterns that are significant in a given collection of images and the related alphanumeric data .The fundamental challenge in image mining is to reveal out the knowledge relating to the images from the web pages. It is the process of searching and discovering valuable information and knowledge in large volumes of data. Also, it is an interdisciplinary endeavor that draws upon expertise in various fields like computer vision, image retrieval, and matching and pattern recognition. Some methods allow image mining to have two different approaches. First method extracts images from image databases or collection of images. Second method mines a combination of associated alphanumeric data and collection of images. Image mining normally deals with the study and development of new technologies that allow easy analysis and interpretation of the images. Mining Image is not only the simple fact of recovering relevant images but is the innovation of image patterns that are noteworthy in a given collection of images. The number of features required to represent an image can be very huge. Using all available features to recognize objects can suffer from curse dimensionality. Feature selection and extraction is the pre-processing step of image mining. The main issues in analyzing images are the effective identification of features and another one is extracting them. Image mining is one in which, it involves general application where the focus is on the process of generating image patterns that may be helpful in the understanding of the interaction between high-level human perception of images and low-level features .So, the latter may be the best one to lead the improvement in the accuracy of images retrieved from image databases. The focus of image mining is the extraction of patterns from a large collection of images.

### **II.Literature Review**

#### **Feature Extraction**

Feature plays a very important role in images. Feature extraction techniques are applied to get features that will be useful in classifying and recognizing of images. As features define the behavior of an image, they show its place in terms of storage taken, efficiency in classification and obviously in time consumption. Feature extraction is the process of generating features to be used in the selection and classification tasks. Extracting feature is most critical because the particular features made available for discrimination directly influence the efficacy of the classification task. The end result of the extraction task is a set of features, commonly called a feature vector, which constitutes a representation of the image. Feature is a distinctive attribute or aspect of something.

In the field of images, features might be raw pixels for simple problems like digit recognition of well-known mining dataset. Feature extraction based on these attributes may be performed at the global or the local level. There are obvious trade-offs between global and local descriptors. Global descriptors are generally easy to compute, provide a good representation, but they tend to integrate and therefore are often unable to discover subtle patterns or changes in shape. Local descriptors, on the other hand, tend to generate more elaborate representations and can yield useful results even when part of the underlying attribute. To detect and describe the local features in the images SIFT algorithm is applied in computer vision. Interesting points also called as edge points of any object in an image can be extracted using the algorithm that helps to offer a feature



description of the object which is extracted from a training image. When it comes to locating a particular object in a test image that has many other objects already, the algorithm is useful to identify the particular object easily. SIFT helps to detect and use a good amount of features from the images which decrease errors that can be caused because of local variations. In general, images have the following features like color, text, shape, edge, etc. and these features are the common image attributes that are used to extract features of mining.

### **Edge Feature Extraction**

An edge is a location of strong intensity expressed as a fast change in intensity gradient- thin, long shape. In image data, the spatial segmentation can be done at region and or edge level based on the requirements of the application. It can be automatic or manual and should be approximate enough to yield features that can reasonably capture the image content. Also an edge in an image may point to a variety of directions, so the canny algorithm uses four filters to detect horizontal, vertical and diagonal edges in the blurred image. The edge detection operator returns a value for the first derivative in the horizontal direction and the vertical direction. One of the relevant hand engineered method is SIFT ( Scale-invariant feature transform). SIFT starts by detecting edges and corners in the image. On the resulted image, SIFT tries to find interesting points which is refers to high frequency pixel values which differentiating that image from the others.

### **Text Feature Extraction**

Text feature extraction is a very efficient and easy technique. Basically, the approach emphasizes on keywords given for a specific image. Text data present in images contain useful information for, keyword additions, Standard titles of content, caption indexing, etc. There are differences in text style, text size, text orientation and alignment and all these can be used for image indexing. In text feature extraction, initially, only foreground of the image is considered by removing background pixels. Gradient vector is applied to determine the center position of the image. These values give center position of image. Then, canny edge detection algorithm is used to identify high frequency pixels in the image. Finally, Euclidean formula is applied to compute the distance between center position and high frequency position in text features. Finally the resultant value is stored in the form of vectors.

### **Shape Feature Extraction**

In order to extract the shape features of the image, the methodology Geometrical gradient is used and applied to the background of the image by removing the foreground pixels. Shape based image retrieval is the measuring of similarity between shapes represented by their features. Shape is an important visual feature and it is one of the primitive features for image content description. Shape content description is difficult to define because measuring the similarity between shapes is difficult. Therefore, two steps are essential in shape based image retrieval, they are: feature extraction and similarity measurement between the extracted features. Shape descriptors can be divided into two main categories: region - based and contour-based methods. Region-based methods use the whole area of an object for shape description, while shape descriptors - features calculated from objects con-tour: circularity, aspect ratio, discontinuity angle irregularity, length irregularity, complexity, right angularness, sharpness, directedness. Those are translation, rotation, and scale invariant shape descriptors. It is possible to extract image contours from the detected edges. Here also, using Gradient vector and canny method used to find the high frequency pixel values in the images. And by applying Euclidean formula the distance between center position and high frequency position in shape features be computed. Finally the resultant value is stored in the form of vectors.

### **Color Feature Extraction**

The color feature is one of the most widely used visual features in image retrieval. Images characterized by color features have many advantages like Robustness in which the color histogram is invariant to rotation of the image on the view axis, and changes in small steps when rotated otherwise or scaled. It is also insensitive to changes in image and histogram resolution and occlusion. Effectiveness is high percentage of relevance between the query image and the extracted matching images. Implementation simplicity is a straightforward process, including scanning the image, assigning color values to the resolution of the histogram, and building the histogram using color components as indices.

A more common approach to comparing the color content of a query image to that of database images is that of comparing color histograms. The methodology relies on the fact that images are generally represented as a series of pixel values, each corresponding to a visible color. Color histograms are computed for each image so as to identify relative proportions of pixels within certain values. A most basic form of color retrieval involves specifying color values that can be searched for an images from a database. Computers represent all visible colors with a combination of some set of base color components, generally Red, Green and Blue (**RGB**).The image retrieval utilized during their experimentation computes similarity based on three different histograms, one for each component of a RGB pixel. Histogram is a demonstration of sharing of color in an image. With the help of color histogram method, the images that have been matched with text and shape feature can be compared. Histogram method enables for summarization of data distribution by identifying color matching of image. The



repetition of values in the range of those pixels is calculated by considering some range of pixel values. The brightness distribution for these colors can be finest which is described by the color histogram which will be more helpful to assess the clipping of the individual colors. The histogram for Red, Green, Blue color values of pill images can be individually calculated separately by taking grayscale value in X-axis and number of recurrence of that value at image in Y-axis. Together of those values get varied for every image. The histogram is a very common approach to compare the color component of a query image to that of database images. Hence, this methodology relies on such pixel values. Color histograms are measured for each image so that to identify relative proportions of pixels within certain values. The specific color values helps for searching images from a database and it is the most basic form of color retrieval. All visible colors along with the basic combination of some set of basic color namely Red, Green and Blue (RGB) are represented by the Computers. Based on the similarity of three unlike histograms in which one for RGB pixel, the image retrieval measures likeness for being utilized during their experimentation. The validation images have the same size and hence can be compared with the extracted images in terms of blocks.

### Performance Evaluation

The performance can be evaluated by the four possible outcomes of a prediction. The outcomes are True Positives (TP), True Negatives (TN), False Positives (FP) and False Negatives (FN), where TP are those extracted regions that are correct, TN are the regions that are incorrect and are not retrieved, FP are regions that are actually incorrect, but have been extracted and FN are regions which were supposed to be extracted but were missed. Only based on this performance ratio the accuracy been calculated.

### Conclusion

Image mining is a progressive field that retrieves images. The main objective of the image mining is to remove the data loss and extracting the meaningful potential information to the human expected needs. Image Mining is the advanced field of Data Mining technique. It is a promising field for research. Image mining research is still in its infancy and many issues remain solved. Specifically, it believes that for Image Mining research to progress to a new height, the following issues need to be investigated. Design semantically powerful query languages for image databases. Exploring new discovery techniques that take into account the unique characteristics of image data. Incorporating new visualization techniques for the visualization of image patterns. The process model represents about storing the various type or level of information already present in image dataset and pointing the issues of analyzing and retrieving useful patterns or knowledge from each level. In this paper, the need for image mining due to boom in growing image datasets been specified. And also pointed the unique features in the area of image mining that brings a whole new set of challenging research issues to be resolved. Finally, it ensures that the applications of image mining have a very dynamic and deep future scope.

### References

1. "Feature Extraction Techniques for Image Retrieval Using Data Mining and Image Processing Techniques" by Preeti Chouhan<sup>1</sup>, Mukesh Tiwari, in International Journal of Advanced Research in Computer and Communication Engineering, Vol. 5, Issue 5, May 2016.
2. "Review of Different Existing Image Mining Techniques" by Prabhjeet Kaur Kamaljit Kaur International Journal of Advanced Research in Computer Science and Software Engineering Volume 4, Issue 6, June 2014.
3. "Automatic Text Extraction in Video Based on the Combined Corner Metric and Laplacian Filtering Technique" by Kaushik K., Suresha, in International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 2, Issue 6, June 2013.
4. A survey in need of image mining techniques" by A.Hema, E.Annasaro, in International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 2, February 2013.
5. "Pill-ID: Matching and retrieval of drug pill images" by Young-Beom, Lee, Unsang Park, Anil K. Jain, Seong-Whan Lee, journal, [www.elsevier.com/locate/patrec](http://www.elsevier.com/locate/patrec), pattern recognition Letters (2011).
6. "Image Mining Issues and Methods Related to Image Retrieval System", by Harini. D. N. D and Dr. Lalitha Bhaskari. D in International Journal of Advanced Research in Computer Science, Volume 2, No. 4, 2011.
7. A Study of Color Histogram Based Image Retrieval" by Rishav Chakravarti, Xiannong Meng, in 2009 Sixth International Conference on Information Technology: New Generations.
8. "Design of a Digital Forensics Image Mining System", by Brown, Ross A., Pham, Binh L., and De Vel, Olivier Y in Knowledge Based Intelligent Information and Engineering Systems, Springer Berlin Heidelberg, 2005.
9. "Content based information retrieval in forensic image databases" by, Geradts, Z. And Bijhold, J. in J. Forensic Sci., Vol. 47, No.2, (2002).