International Association of Scientific Innovation and Research (IASIR)
(An Association Unifying the Sciences, Engineering, and Applied Research)

ISSN (Print): 2279-0020
ISSN (Online): 2279-0039

International Journal of Engineering, Business and Enterprise Applications (IJEBEA)

www.iasir.net

Survey Paper on Image Retrieval Algorithms

Ms.Khushdeep Kaur, Mr.Hardeep Singh
Department of Computer Science
Lovely Professional University, Jalandhar, Punjab, INDIA

Abstract*: As we all know that popularity of images is increasing briskly from day to day because of improving technologies and suitable availability promoted by the internet. Therefore how to retrieve an image is a difficult task. In this paper I define various techniques for retrieving an image from database. Data mining is the technology that combines historic data analysis practices with practical algorithms for transforming enormous amount of data [3]. Image mining in simpler words is image processing in data mining.*

I. Introduction

Image mining is an evolution for data mining which comprehend various disciplines. The rapid evolution of multimedia and imaging technology tends to the growth of image data. Image mining is defined as the technology which helps in searching profitable information from huge image data [2]. The basic difference between image mining and image processing is that image mining focuses on huge data image set. The main aim of image data mining is to generate patterns [1].

The image database comprises of massive volume of information and it is becoming progressively complicated when its size continues to rise at a momentous rate. It becomes very difficult to retrieve an image from a large database. Huge number of images have acquired on internet and various other applications. Therefore how to mine different images from database is gaining more attention these days.

As we use advanced technology, the storing, acquiring and sharing images becomes easy and it helps in increasing the available images and their wide varieties. Therefore the significance of image retrieving algorithms is increasing. There are various techniques that are used for retrieving an image from database. There are different techniques and algorithm already defined for retrieving an image. These defined algorithms are being used in various fields like medical, diagnostics, space research, weather forecasting etc for retrieving the images. Various image retrieval algorithms and techniques used are text based image retrieval, content based image retrieval, region based image retrieval, hierarchical clustering, fuzzy shape clustering etc [3]. In this paper, I defined different techniques for retrieving an image from the database. This paper is basically a review on various techniques that already have been defined for retrieving an image from database. There are many techniques used these days for retrieving an image. In this paper I have explained five techniques out of many already defined techniques which are normally and widely used for retrieving images. In the next sections I am going to explain about the methods, algorithms being used for image retrieval.

II. Methods

Images should be retrieved according to the specifications of the requirement analysis. The requirement analysis can be categorized into various levels. These levels are based on the complexity [3].

Level-1: Image Retrieval by Primitive Feature:
Primitive features are defined as the color, shape, texture or the location of an image element. Example- retrieving an image with brown object in the left hand corner.
Level-2: Image Retrieval by Derived or Logical Feature:
These are defined as the person or object. Example- retrieving images of red car.
Level-3: Image Retrieval by Abstract Attribute:
These are defined as the scene or purpose of the object. Example- retrieving images of a badminton match.

III. Image Retrieval Algorithms

There are various algorithms that have already been defined. In this paper, I defined five usually used algorithms or techniques for retrieving images.

A. Text Based Image Retrieval

It was started with Boolean search of words with combination of AND, OR, NOT.
There are different techniques used for text based retrieval:

- Bag of words approach – This approach is used for representing a sentence, document or text in the form of bag of words.
  Example- Rohan likes to watch football matches. Sohan likes too.
  Rohan also likes to watch cricket matches.

Representation-

"Rohan" 1	"matches" 6
"likes" 2	"also" 7
"to" 3	"cricket" 8
"watch" 4	"sohan" 9
"Football" 5	"too" 10

- Stop words approach- This approach is used for the words which needs to be filtered out. There is not specific list of these words. In text based retrieval, stop words need to be eliminated.
- Stemming- It is the process of deriving the words from the root, stem or base form. Example- if we enter the word "stem" all the words associated with stem like "stemming", "stemmed", "stemmer" etc will be shown. Many research engines use this technique for searching the words.

Limitations of text based retrieval:

1. Difficult in case of large databases.
2. It is subjected to human perception.
3.

B. CONTENT BASED IMAGE RETRIEVAL

Recently used image retrieval technique is CBIR. CBIR techniques are more accurate that text based image retrieval. Content based image retrieval is defined as the process of retrieving the features and searching the database for the similar images [7]. Content based image retrieval extracts the images depending upon the visual features like shape, color, texture. The initial phase of CBIR is to evaluate the features and producing it in the form of numeric values. For better representation, more number of features should be evaluated [6]. The second phase is numeric values which are evaluated should be compared with images present in the database. The distance vector is computed. Different aspects of the image are:

- Color- Color is very essential level of any image. Color feature acquires more human attention. This feature is categorized into local and global descriptors. Local descriptors are defined as the descriptors which represent the color with reference to the spatial location. Local descriptors are more useful than global descriptors.
- Texture- Texture feature is basically defined as the merging of pixels that has occurred many times in an image. The importance of texture feature is that it helps in differentiating the objects with the backgrounds.
- Shape- Shape is an important aspect. It defines the relative frequency of the occurrence.

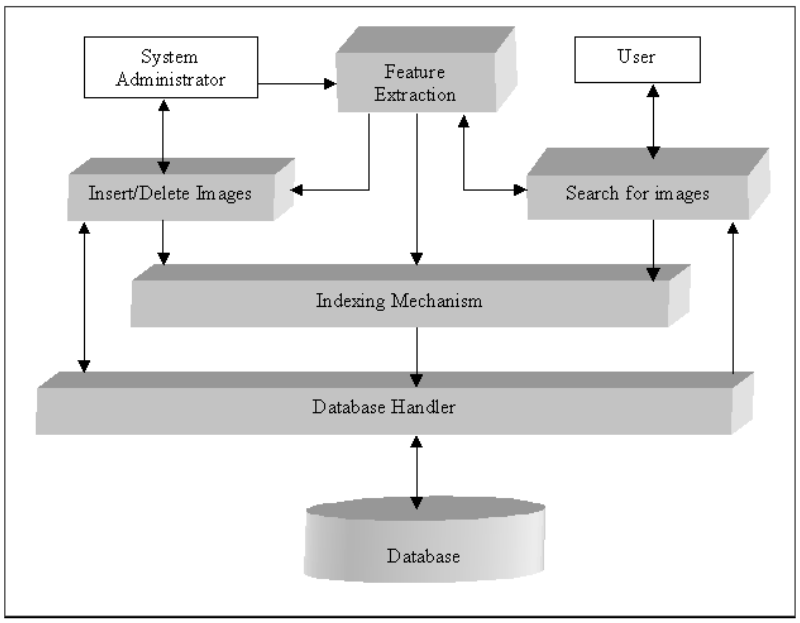

Figure 1: Architecture of Content based Image Retrieval

Advantages of Content based image retrieval:

The main advantage is the feasibility of an automatic retrieval process, instead of the text based approach, which usually requires very difficult and time-consuming.

C. REGION BASED IMAGE RETRIEVAL

For retrieving an image the most ordinary approach used is to work on low level aspects like color, texture and shape but still there is a scope of improvement as there is a difference between the low level and high level feature approaches. Researchers defined new methods which uses objects and regions to process the content of

images [8]. Then the region based methods come into consideration. The main aim is to embellish the ability to store and represent the image content. Region based image retrieval works as follows:

- Images are divided into assorted regions
- Features are drawn out from each region
- The combination of all the features represents the image content.

Features are retrieved first then compared with the sketch or region of an image provided by the user. RBIR provides high premium in segmentation quality which is difficult to achieve in practice.

D. HIERARCHICAL CLUSTERING

This approach is basically used for retrieving images from databases. The images in database are collectively associated into clusters of images having same features or properties. In case of searching the sample image is not compared with all the images but is compared with the subset [9]. Hierarchical clustering approach provides high response time and better accuracy in retrieval. Searching and retrieving for a particular image in large databases is a challenging task. Search engines usually compute the likenesses between sample image and the database images and rank them accordingly. The retrieval time for clustered images is

$T(cluster) = KT1(sim) + LT1(sim) + O(llogl)$

Where T(cluster) = time for clustered images

K= number of clusters

T1(sim) = time to calculate the similarity between two images

L = number of images in the clusters nearest to the query

O(nlogn) = time to sort n elements.

Let n= number of images present in the database. The hierarchical clustering is calculated as follows:

- The n images are located in n distinct clusters. These are grouped as (C1, C2, C3,….,Cn). For any Rth cluster,
- E(R) = all images contained in the cluster
- N(R) = number of images contained in the cluster
- C(p) and C(r) are the two clusters that are to be merged such that S(p,r) is the similarity measure which is largest and denoted as C(p+r). The set of all unmerged clusters are also calculated.
- The above steps are repeated till number of clusters has decreased to the defined number or the largest similarity measure between clusters has dropped to lesser threshold.

E. FUZZY SHAPE CLUSTERING

In order to retrieve an image from a database based on the shape, a sample shape is taken from the user and accordingly it is matched with the set of images present in the database [10]. Usually when user enters an input shape, the Fourier descriptors of that sampled shape (sp) is matched with the Fourier descriptors of all the images present in the database (sq). To perform matching process

$$M(pq)= 1-||sp – sq||$$

After the matching process, results obtained denoted as prototypes are sorted in the order of degree of matching. The prototypes obtained by the higher matching degree are provided as retrieval results.

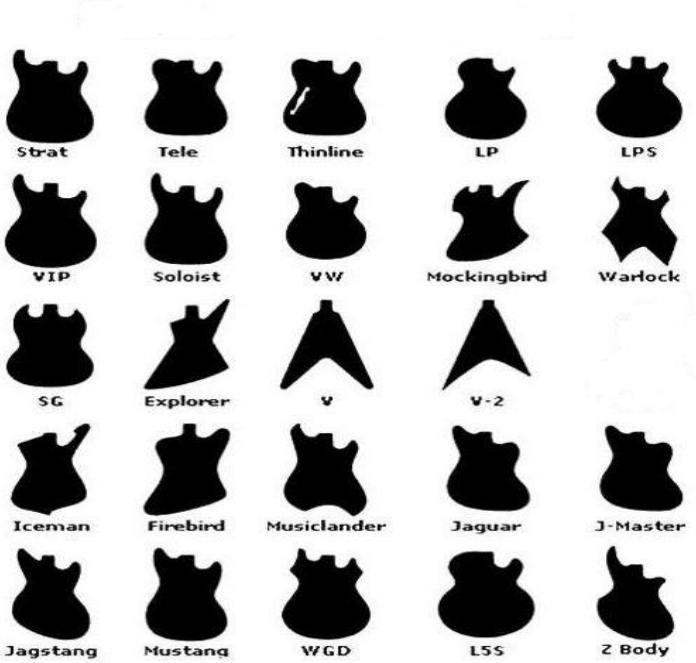

Figure 2: Different shapes of images of a Guitar

IV. Conclusion

In this paper, we have mentioned various algorithms or techniques which are being used these days in image mining for retrieving an image from a database. Image mining is the current topic in research. Image mining is a vital technique which is used to mine knowledge straightforwardly from image. Image mining is simply an expansion of data mining in the field of image processing. Image mining deals with extracting hidden knowledge, image data association and add on patterns which are not clearly gathered in the images. This paper is basically a review of some of the techniques being used to retrieve images from the database.

References

[1] Image mining: A new approach for data mining based on texture.
[2] A Survey on Image Mining Techniques: Theory and Applications.
[3] Image mining: Issues, frameworks and techniques.
[4] Image Mining by Data Compactness and Manifold Learning.
[5] Image Mining Using Association Rule.
[6] Content-Based Image Retrieval: Theory and Applications.
[7] Efficient Content Based Image Retrieval by Ruba A. A. Salamah.
[8] Region based image retrieval using probabilistic feature relevance learning by Byoungchul ko, jing peng, hyeran byun.
[9] Hierarchical clustering algorithm for fast image retrieval by Santhana Krishnamachari , Mohamed Abdel-Mottaleb Philips Research.
[10] Fuzzy Shape Clustering for Image Retrieval by G. Castellano, A.M. Fanelli, F. Paparella, M.A. Torsello.
[11] http://digital.cs.usu.edu/~xqi/Teaching/REU07/Notes/CBIR.pdf
[12] http://pages.csam.montclair.edu/~peng/publication/PAA.pdf

