www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 11 Nov. 2016, Page No. 18979-18982

A Survey on Feature Extraction Techniques for Image Retrieval using Data Mining & Image Processing Techniques

Sreelekshmi.U[1], Anil A.R [2]
[1]lekshmishibin@gmail.com [2]anilar123@gmail.com

Computer Science & Engineering from Sreebuddha College of Engineering, Pattoor, Alappuzha, kerala

Abstract

There is an interesting field in data mining called image mining for image processing. Image mining is the association of image data and extraction of hidden data. Data mining is the process of extracting information or knowledge from a wide database. Image mining makes use of texture, color factors and size of an image. Image texture is determined by a feature called Gray Level Co-occurrence Matrix (GLCM).In this case image retrieval feature will be sharp. In order to retrieve features of similar types of image shapes and texture, a feature called Weighted Euclidean distance is used.

***Keywords**: Data Mining, Image Mining, Image retrieval, Gray Level Co-occurrence Matrix, Weighted Euclidean distance*

I. Introduction

Data mining is a process of discovering patterns from huge data stored in various databases such as data warehouse, worldwide web, and external sources. Image mining is the process of finding relevant information and knowledge in large volumes of data.

In the image mining process, the images are retrieved from the database and they are preprocessed to improve its quality. The transformation and feature extraction of images generates important features from images. If the data are large or redundant, then the data are transformed into a reduced representation set of features. Feature extraction is the process of extracting essential features that describe a large set of data. Then information is mined using data mining techniques. This is followed by interpretation and evaluation of information. Hence knowledge is obtained which is an understandable form of information.

The important features extracted during image mining are color, texture and shape features. Based on color feature, the image is extracted by color similarity mining by quantization on color space. The texture feature is extracted based on the color histogram texture. In content based image retrieval system. The features are extracted from query image and image collection. Then its corresponding image features are compared to find similarity and relevant images features are retrieved. GLCM matrix is used to store the features of an image. For a single image there are multiple GLCMs to specify four directions (horizontal, vertical, and two diagonals) and four distances.

After calculating GLCM in all four directions, the Weighted Euclidean distance for a point(x,y) is calculated by subtracting the jth mean from x(j) and y(j) and taking square, then multiplying a weight(j) attached to j. The weight is the inverse of jth variance. Here s is the standard deviation of jth variable.

$$d_{x,y} = \sqrt{\sum_{j=1}^{J} \frac{1}{s_j^2}(x_j - y_j)^2}$$

$$= \sqrt{\sum_{j=1}^{J} w_j (x_j - y_j)^2}$$

Thus, with the help of GLCM, image textures are determined. Moreover the Weighted Euclidean distance seeks to get similar image features easily. Hence, the

process of image retrieval is found more relevant. The image and its corresponding histogram help in extracting the various features like color, shape and texture.

This paper is organized as follows: In Section II the survey of different methods is described and section III includes the conclusion.

II. Literature Survey

Y.Liu *et.al* proposed [1] Content based image retrieval with high level semantics which involves the extraction of low level image features, similarity measurements and deriving high level semantics features. The low level image features like color, texture, shape or spatial location can be extracted from segmented image. Image similarity is measured by two levels. The first is a region level that measures distance between two regions based on low level features. The second is image level that measures overall similarity of two images which consists of different number of regions. High level semantic features are defined as the representative feature of a concept calculated from a collection of sample images.

In the paper image retrieval using color and shape[2], a combination of clustering and a branch and bound matching scheme helps in improving the speed of image retrieval.For an image color information is represented by histograms.Euclidean distance is used as a means to compute the distance between the image features.A histogram with edge directions is used to represent shape attribute of an image.For shape based image retrieval,a histogram intersection technique is used.To reduce the wastage of space,the database images are clustered.For optimal search,branch and bound method is used in which the database images are divided into a hierarchy of disjoint subsets of images.

Janani M *et.al* introduced [3] content based image retrieval system, in which an image is retrieved with the help of image features like color, texture, pattern and shape of objects. The color similarity of images is done by quantization on color space and measures the similarity between sample and image results. In order to make the process more fast, sometimes a specific image like logo is selected in which case the target image can be retrieved with fewer iterations. Some clustering algorithms like hierarchical and K-means are used. These clustering algorithms group the similar images into a dataset which forms a cluster and thus forming various clusters.

In the paper, image mining using content based image retrieval system [4], the image retrieval is based on the color histogram texture. Initially the images from image database are retrieved. Then the desirable features of image are extracted. The features like color, texture, patterns, image topology, shape of objects and their layouts and locations within the image are indexed. Then they are stored in image meta-data database. Whenever a query image arrives the features are extracted. Then it measures the similarity of features with those in meta-data database. If found any match then those images are retrieved.

Hiremath P.S *et.al* introduced [5] Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement which uses an integrated image matching procedure. For this an image is taken and it is represented at different resolutions. Even if resolutions differ all the images have same significance. The matching is done by comparing tiles of target image with the tiles of query image. This is represented in the form of a bipartite graph which shows the matching between query image tile and target image tile. A bipartite graph has source image tile on the left side and target image tile on the right side. There occurs a maximum bipartite matching between the source and target images to find the relevant image.

In the paper, Query by Image and video content the QBIC [6] system is explained as a content based retrieval method that allows querying on large image and video database. The properties of QBIC are:(1) Uses image and video content and properties like color, texture, shape and motion of images, videos and their objects in queries(2)It has a graphical query language in which queries are induced in the form of drawing, selecting and other graphical means. The main components of QBIC are: database population and database query. In database population, the images and

videos are processed to extract features like color, texture, shapes, camera and object motion. These features are stored in a database. In database query, the user creates a graphical query. The features are extracted for the graphical query and are fed to a matching engine which retrieves images or videos from database having similar features.

Aboli W.Hole *et.al* proposed [7] n the design and implementation of content based image retrieval using data mining and image processing techniques , the image is divided into coarse partitions. Here the image is identified by retrieving the dominant color from the centroid of each of its partition. Thus the query image and images retrieved from database are checked for similarity with respect to its dominant color in their coarse partitions. The image mining techniques employed are object recognition, image retrieval and image indexing. The images are retrieved by identifying specific objects. The images are indexed inside database for fast image retrieval. Thus by dividing images into specific components like object recognition, dominant color has improved the image retrieval to a great extent.

R. Datta, *et.al* introduced [8] image retrieval: Ideas, influences, and trends of the new age" that discuss several image search domains like narrow and broad image search domains. The narrow image search domains have limited variability but better visual features while the broad image search domains has high variability but unpredictable visual features. Some of the image search categories are: search by association, aimed search and category search. There happens repeated browsing to retrieve an image in search by association. A specific image search is done for aimed search. For category search, a class of images of a group is searched for retrieve the relevant image features.

Another paper, content based image retrieval at the end of the early years [9] highlights the importance of image processing with the extraction of image features like color, texture and shape. An image has features like global, salient sign and object which can be extracted for image retrieval process. Images are partitioned to acquire global features that find a match between query image and images in database. The salient features of an image is retrieved by a process called weak segmentation. For acquiring sign probabilities of an image, the various sign locations are identified.

T. Kato proposed [10] the database architecture for content based image retrieval that focus on color and texture extraction based on Discrete Wavelet Transform (DWT) and Self Organizing Map (SOM).There are several visual interaction mechanisms like query by example, subjective descriptions. Several functions like sketch and similarity are made use in query by example. A function like sense retrieval is used to highlight subjective descriptions of an image. An image model is generated from the graphical features like texture, color, shape. On the other hand, a model called user model exhibits the visual ideas of a user. Some of the experimental databases used are Trademark and Art Museum to perform image processing.

III. Conclusion

This survey has been performed for identifying the various image retrieval methods which are useful for image mining. It was found that the content based image retrieval is based on the various features of images like color, shape and texture. In image processing, the similarity between query image and images in database are measured with the help of Weighted Euclidean distance. It reduces the query image searching time which leads to an increase in the image retrieval speed. Moreover it highlights the benefits of the proposed approach namely GLCM (Gray Level Co occurrence Matrix). The texture extraction of an image is done with the help of GLCM which makes use of four statistic features of an image like contrast, homogeneity, energy and correlation. Some advantages of this approach are: there is a large coverage of domains e.g. Entertainment, sports etc.,. It was found more scalable i.e. it covers large number of topics .It is not at all biased by any editor's interest. Some of the applications of this approach are: in military to find tanks or airstrips, in Government to track highway assets and in urban development to search housing sprawl.

References

[1] .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 .

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

[3] Janani M and Dr. Manicka Chezian. R, "A Survey On Content Based Image Retrieval System", International Journal of Advanced Research in Computer Engineering & ICWSM,11:401-408,2011 Technology, Volume 1, Issue 5, pp 266, July 2012.

[4]. 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.

[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]. 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

[7]. 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.

[8]. 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.

[9]. 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.

[10]. 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.

BIOGRAPHIES

Sreelekshmi.U obtained B. tech. (Computer Science & Engineering) from Sreebuddha College of Engineering, Pattoor, Alappuzha, kerala & pursuing M. Tech. in Computer Science and Engineering from Sreebuddha College of Engineering, Pattoor, Alappuzha,kerala .

Anil A.R received his Master (MTech) degree in Computer Science from University of Kerala. Currently he is working as Associate Professor,Department of Computer Science & Engineering at SreeBuddha College of Engineering,Alappuzha,kerala. He is doing PhD in Computer Science at Bharathiar University,Coimbatore,under the supervision of Dr.R.Rajesh (co-author).His research interests include Medical Image Processing.