Overview of Content Base Image Retrieval

Manisha Raut

Dept. Of Electronics And Telecommunication
Rajiv Gandhi College of Engineering & Research
Nagpur, India

Pallavi Dhok

Dept. Of Electronics And Telecommunication
Rajiv Gandhi College of Engineering & Research
Nagpur, India

Shital Bramhe

Dept. Of Electronics And Telecommunication
Rajiv Gandhi College of Engineering & Research
Nagpur, India

Dhanashree Tajne

Dept. Of Electronics And Telecommunication
Rajiv Gandhi College of Engineering & Research
Nagpur, India

***Abstract*— CBIR is used to search and retrieve the query image from wide range of databases. Many Features and algorithms can be used for efficient image retrieval. Content-Based Image Retrieval (CBIR) uses the visual contents of an image such as color, shape, texture, and spatial layout to represent and index the image. In this paper we survey a CBIR system. Discuss a various challenges in CBIR techniques. This paper surveys recent studies on content-based retrieval for multimedia databases from the point of view of fundamental issues. We review and discuss the advantages, disadvantages and application of these methods for images.**

Keywords- Content based image retrieval (CBIR), Pixels, Image

I. INTRODUCTION

Content-based image retrieval (CBIR), also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) is the application of computer vision techniques to the image retrieval problem, that is, the problem of searching for digital images in large databases

Content Based Image Retrieval is the process of retrieving the desired query image from a huge number of databases based on the contents of the image. Colour , texture, shape and local features are some of the general techniques used for retrieving a particular image from the images in the database. Content Based Image Retrieval systems works with all the images and the search is based on comparison of features with the query image [1]

Images have always been an inevitable part of human communication and its roots millennia ago. Images make the communication process more interesting, illustrative, elaborate, understandable and transparent In CBIR system, it is usual to group the image features in three main classes: color, texture and shape [2,3]. Ideally, these features should be integrated to provide better discrimination in the comparison process.

Color is by far the most common visual feature used in CBIR, primarily because of the simplicity of extracting color information from images [4,6]. To extract information about shape and texture[6]feature are much more complex and costly tasks, usually performed after the initial filtering provided by color features[7]

Different CBIR systems have adopted different techniques. These technique are generally divided into two groups 1) global color and texture features2) local color and texture features.

Definition: In CBIR method a Visual features as color, shape and texture are implemented for retrieval of images. Traditional methods of image indexing have been proven neither suitable nor efficient in terms of space and time. so it needed to Develop a new technique. It is a 2 step process .In first step image features are extracted to a Distinguishable extent .In second step matching of features which are visually similar is done

II. LITERATURE REVIEW

Aura Conci and Castro [8] has proposed a novel approach based on calculating the distance between the images. Chee Sun Won [9] proposed an efficient use of MPEG-7 Color Layout and Edge Histogram Descriptors in CBIR Systems.

H.B.Kekre et.al.[10]proposed a Performance Comparison of Image Retrieval Techniques using Wavelet Pyramids of Walsh, Haar and Kekre Transforms. Hemalatha and Devasena[11]proposed a research to find out the accurate images while mining an image (multimedia) database and developed an innovative technique for mining images by means of LIM dependent image matching method with neural networks. Hiremath P. S and JagadeeshPujar [12] proposed Content based Image Retrieval based on Color, Texture and Shape features using Image and its complement. Color, texture and shape information have been the primitive image descriptors in Content based Image Retrieval systems. Another methodology is proposed using hierarchical and K-Means clustering technique by Murthy et al. [13].N S T Sai and R C Patil [14] proposed an Image Retrieval using DWT with Row

And Column Pixel Distributions of BMP Image Neetu Sharma. S et al.[15] proposed an Efficient CBIR Using Color Histogram Processing. Ramamurthy, B. and K.R. Chandran [16] proposed a Content Based Medical Image Retrieval with Texture Content Using Gray Level Co-occurrence Matrix and K-Means Clustering Algorithms . Rajshree S and Dubey [17] illustrated about an Image mining methods which is dependent on the Color Histogram. They have examined a histogram-based search techniques and color texture techniques in two different color spaces, RGB and HSV. Histogram search distinguish an image through its color distribution. Another method [18]

is discussed using the edge histogram Rajendran and Madheswaran discussed an improved image mining technique. An enhanced image mining technique for brain tumor classification is proposed using pruned association rule with MARI algorithm. Rajendran P and Madheswaran M proposed a Hybrid Medical Image Classification Using Association Rule Mining with Decision Tree Algorithm

III. NEED & MOTIVATION

Advances in processor speed & digital sensor technology, expansion & availability of internet have tremendously increased volume & accessibility of digital images. This has caused a need to have a system that returns images based on the given query considering visual similarity among them to facilitate image searching /cataloging. The conventional image search techniques involves search on metadata comprises of textual annotation associated with images, having severe drawbacks and limitations .The CBIR system aims to overcome these limitations. The vast application areas and lack of availability of general purpose CBIR system imply huge research potential on the subject.

IV. APPLICATIONS OF CONTENT-BASED RETRIEVAL

There is three broad categories of user aims when using CBIR the system,

1) There is a broad class of methods and systems aimed at browsing through a large set of images from unspecified sources. Users of search by association at the start have no specific aim other than find interesting things.

2) Another class of users targets the search at a specific image. The search maybe for a precise copy of the image in mind, as in searching art catalogues

3) The third class of applications, category search, aims at retrieving an arbitrary image representative of a specific class. It may be the case that the user has an example and the search is for other elements of the same class.

Application Areas

Typical areas involving content based image retrieval are

1. Art galleries and museum management
2. Architectural and engineering design
3. Interior design
4. Remote sensing and management of earth resources
5. Geographic information systems
6. Scientific database management
7. Weather forecasting
8. Retailing
9. Fabric and fashion design
10. Trademark and copyright database management
11. Law enforcement and criminal investigation
12. Picture archiving and communication systems.
13. Other applications reported in the literature are
14. Web image searching
15. Education and training
16. Home applications like digital photograph cataloging and retrieval
17. Medical image database maintenance and medical diagnosis
18. Military applications
19. Pornography detection & elimination

V. CBIR SYSTEM

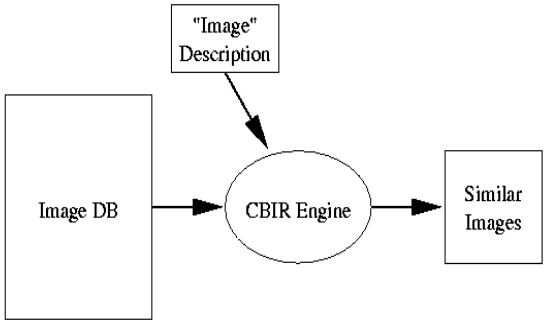

Fig 1. CBIR system

CBIR involves the subsequent four parts in system realization , the first one is collection of data than build up database of feature. Afterwards we will search in the database, arrange the order and deal with the results of the retrieval

1) Data gathering:-By using Internet spider program that can collect webs automatically to interview Internet and do the gathering of the images on the web site, then it will go over all the other webs through the URL, repeating this process and collecting all the images it has reviewed into the server.
2) Extract feature database:-Using index system program do analysis for the collected images and extract the feature information. At this time, the features that use widely involve low-level features such as color, texture and so on, the middle-level features such as shape.
3) Searching in the Database:-System extract the feature of image that waits for search when user input the image sample that need search, then the search engine will search the suitable feature from the database and calculate the similar distance, then find some related webs and images with the lowest similar distance.
4) Process and index the results:-Index the image obtained from searching due to the similarity of features, and then returns the retrieval images to the user and allows the user select. If the user is not pleased with the searching result, he can re-retrieval the image again, and searches database again

VI. CHALLENGES

Following issues make development of versatile techniques for image feature extraction and hence retrieval difficult and challenging.

1. Semantic gap: It is a most crucial factor affecting the relevant-image-retrieval performance of the image

retrieval. The semantic gap - as defined in , is the lack of coincidence between the information that one can extract from visual data and the interpretation that the same data have for a user in a given situation. It is also described in the literature as a gap between human perception for the image content description and its feature representations.

2. Subjectivity: The subjectivity of human being for the content analysis and description, which is characterized by human psychology, emotions and imaginations, is a second most crucial factor affecting the relevant-image-retrieval performance of the system.
3. Inter tuning of various phases: The feature extractor module generally consists of a series of operations /phases whose proper tuning is important for better over-all performance of the system. Inter tuning of phases plays very important role for example in CBIR systems incorporating relevance feedback or hierarchical frame work for feature extraction & representation.
4. Variety of image categories & characteristics: Different image categories & varieties of image characteristics add to the difficulty levels for development of versatile image feature extraction algorithms.

VII. DIFFERENT CBIR ALGORITHM

Different algorithms and models for the retrieval of images have been explored over the last twenty years. We will discuss some of algorithm and model.

1. Content –Based Retrieval Of Ottoman Archive

In our content-based retrieval system, the query specification is performed by identifying a region containing a word, a phrase, or a sentence in the document image. Any keyword or pattern search algorithm can be carried out over the code-books of the document images. The template image is analyzed, and distance and angular span vectors in spatial domain, and wavelet extreme information are determined. Once the symbols forming the template keyword image are identified according to the extracted features, the locations of the symbols within the documents are determined. If the symbols of a keyword image appear consecutively in the codebook of a document image, then that document is a match for the query. In our system, the documents in the archive are ranked according to the partial symbol wise matching scheme. As the name of this method implies, the similarity of a Document to the query image is calculated symbol wise. In other words, the system orders the resulting document images in the decreasing order of similarities with respect to the symbols, since the query image is also split into symbols[19]

2. Histogram Refinement for Content-Based Image Retrieval

In histogram refinement the pixels of a given bucket are subdivided into classes based on local features. There are many possible features, including texture, orientation, distance from the nearest edge, relative brightness, etc. Histogram refinement prevents pixels in the same bucket from matching each other if they do not fall into the same class. Pixels in the same class can be compared using any standard method for com-paring histogram buckets (such as the L1 distance).

This allows the distinctions that cannot be made with color histograms. As a simple example of histogram refinement, consider a positional refinement where each pixel in a given color bucket is classified as either \ in the center" of the image, or not. Specially, the centermost 75% of the pixels are defined as the \center". This produces a split histogram in which the pixels of color buckets are loosely constrained by their location in the image. The resulting split histograms can be com-pared using the L1 distance. [20]

3. Interactive Content Based Image Retrieval using Ripplet Transform and Fuzzy Relevance Feedback

Ripplet Transform Type-I (RT)-Conventional transforms like (Fourier Transform) FT and Wavelet transform (WT) suffer from discontinuities such as edges and contours in images. To address this problem, Jun Xu et al. proposed a new MGA-tool called RT. RT is a higher dimensional generalization of the Curvelet Transform (CVT), capable of representing images or 2 D signals at different scales and different directions. To achieve anisotropic directionality, CVT uses a parabolic scaling law. From the perspective of micro local analysis, the anisotropic property of CVT guarantee resolving 2D singularities along C2 curves. On the other hand, RT provides a new tight frame with sparse representation for images with discontinuities along Cd curves. There are two questions regarding the scaling law used in CVT: 1) Is the parabolic scaling law optimal for all types of boundaries? and if not, 2) What scaling law will be optimal? To address these questions, Jun Xu et al. intended to generalize the scaling law, which resulted in RT. RT decentralizes CVT by adding two parameters, i.e., support c and degree d. CVT is just a special case of RT with c=1and d= 2. The anisotropy capability of representing singularities along arbitrarily shaped curves of RT, is due to these new parameters c and d [21]

4. Data Clustering

Data Clustering is often took as a step for speeding -up image retrieval and improving accuracy especially in large database. In general, data clustering algorithms can be divided into two types: Hierarchical Clustering Algorithms and Non hierarchical Clustering Algorithms. However, Hierarchical Clustering is not suitable for clustering large quantities of data. Non-hierarchical is suggested to cluster large quantities of data. The most adapted method for non-hierarchical clustering is the K-Means clustering algorithm proposed by James MacQueen in 1967. K-means clustering algorithm first defined the size of K clusters. Based on the features extracted from the images themselves, K-means allocates those into the nearest cluster. The algorithm calculates and allocates until there is little variation in the movement of feature points in each cluster. This paper applies k-means clustering algorithm for color clustering module based on this concept [22]

VIII. CONCLUSION

The Ottoman documents compressed in this special textual form are effectively queried by specifying rectangular regions in document images without any limitation on the size of the query region. The querying takes place using only the com-pressed data. Our symbol matching scheme ranks the resulting document images in the decreasing order of similarities with respect to the number of symbols matched. The total similarity is computed by comparing the symbols in the query region with the pointers to the codebook. The resulting documents are presented by identifying the matched region of each document in a rectangle [19]

Histogram refinement idea can be extended by placing further constraints on the split histogram itself. Both histogram refinement and successive refinement are general methods for improving the performance of histogram based matching. If the initial histogram is a color histogram, and it is refined based on coherence, then the resulting split histogram is a CCV. But there is no requirement that this refinement be based on coherence, or even that the initial histogram be based on color [20]

Ripplet transform based image coding is suitable for representing low level features (color, texture, edge etc.) of the images. The proposed CBIR system based on RT features is able to improve the accuracy of the retrieval performance and to reduce the computational cost. The retrieval performance is improved further using fuzzy based RFM within 2 to 3 iterations. The proposed mechanism could be tested for video retrieval as future scope of research. [21]
Segmentation and grid module, get combined for feature extraction module, K-means clustering and neighborhood module to build the CBIR system. Furthermore, the concept of neighborhood module which recognizes the side of every grids of image is first contributed . Applying the concept of fragment based code book into the content based image retrieval system also contributes in our system architecture. The experimental results confirm that the proposed CBIR system architecture attains better solution for image retrieval. [22]

REFERENCES

[1]. Ms. K. Arthi 1, Mr. J. Vijayaraghavan 2," Content Based Image Retrieval Algorithm Using Colour Models" International Journal of Advanced Research in Computer and Communication Engineering Vol. 2, Issue 3, March 2013

[2]. Y. A. Aslandogan and C. T. Yu, "Techniques and Systems for Image and Video Retrieval," IEEE Transactions on Knowledge and Data Engineering, Vol. 11, Issue 1, pp. 56-63, Jan/Feb 1999.

[3]. A. W. M. Smeulders, M. Worring, 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, Vol. 22, Issue 12, pp. 1349-1380, December 2000.

[4]. M. J. Swain and D. H. Ballard, „„Color indexing"", International Journal of Computer Vision, vol. 7, Issue 1, pp. 11-32, 1991.

[5]. Irena Valova, Boris Rachev and Michael Vassilakopoulos, "Optimization of the Algorithm for Image Retrieval by Color Features",International Conference on Computer Systems and Technologies- CompSysTech", pp 1-4, 2006.

[6]. Sarfraz and M. Ridha "Content-Based Image Retrieval Using Multiple Shape Descriptors", IEEE/ACS International Conference On Computer Systems and Applications, pp. 730-737, 2007.

[7]. Reshma Chaudhari1, A. M. Patil2," Content Based Image Retrieval Using Color and Shape Features," International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 1, Issue 5, November 2012

[8]. Aura Conci and Everest Mathias M. M. Castro, "Image mining by content" ExpertSystems with Applications, vol. 23, no. 4, pp. 377-383, 2002.

[9]. Chee Sun Won, , Dong Kwon Park, and Soo-Jun Park, "Efficient Use of MPEG-7 Edge Histogram Descriptor" , ETRI Journal, Vol.24, No 1, pp. 23-30, February 2002.

[10]. H.B.Kekre,Sudeep D. Thepade, AkshayMaloo, "Performance Comparison of Image Retrieval Techniques using Wavelet Pyramids of Walsh, Haar and Kekre Transforms" International Journal of Computer Applications (0975 – 8887)Volume 4, No.10, pp.1-8, August 2010.

[11]. Hemalatha M. and Lakshmi Devasena C, "A Hybrid Image Mining Technique using LIMbased Data Mining Algorithm," International Journal of Computer Applications, vol. 25, no 2, pp. 1-5, 2011.

[12]. HiremathP. S and JagadeeshPujari, "Content based Image Retrieval based on Color, Texture and Shape features using Image and its complement", International Journal of Computer Science and Security, Vol.1 : no 4,pp 25-35.

[13]. Murthy V.S.V.S, Vamsidhar E, swarup Kumar J.N.V.R, Sankara Rao P, "Content based Image Retrieval using Hierarchical and KMeans Clustering Techniques", International Journal of Engineering Science and Technology Vol. 2(3), pp. 209-212, 2010.

[14]. N S T Sai and R C Patil, "Image Retrieval using DWT with Row and Column Pixel Distributions of BMP Image" International Journal on Computer Science and Engineering,Vol. 02, No. 08,pp. 2559-2566, 2010.

[15]. Neetu Sharma. S1, PareshRawat . S2 and jaikaranSingh.S,"Efficient CBIR Using Color Histogram Processing", An International Journal(SIPIJ), Vol.2, No.1, pp. 94-112, March 2011.

[16]. Ramamurthy, B. and K.R. Chandran, "Content Based Medical Image Retrieval with Texture ContentUsing Gray Level Cooccurrence Matrix and K-Means Clustering Algorithms"Journal of Computer Science 8 , ISSN 1549-3636, pp. 1070-1076, 2012.

[17]. Rajshree S. Dubey , "Image Mining using Content based Image Retrieval System", IJCSE) International Journal on Computer Science and Engineering Vol. 02, No. 07, pp. 2353-2356, 2010.

[18]. Rajendran P and Madheswaran M, "An Improved Image Mining Technique For Brain Tumour Classification Using Efficient classifier", International Journal of Computer Science and Information Security, vol. 6, no. 3, pp. 107-116, 2009.

[19]. Ediz Saykol, Ali Kemal Sinop, Ugur Güdükbay, ÖzgürUlusoy, "Content-Based Retrieval of Historical Ottoman Documents Stored as Textual Images", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 13, NO. 3, MARCH 2004

[20]. Greg Pass, min Zabih, "Histogram Refinement for Content-Based Image Retrieval " .

[21]. Manish Chowdhury, Sudeb Das, and Malay Kumar Kundu., "Interactive Content Based Image Retrieval Using Ripplet Transform and Fuzzy Relevance Feedback

[22]. Ray –I chang,shu-yu lin,jan-ming ho," A Novel Content Based Image Retrieval System using K-means/KNNwith Feature Extraction"