Review Paper for Content based Image Retrieval Methods
Miss. Megha Jain¹ Mrs. Pushpa Singh²
¹M.Tech. Scholar ²Assistant Professor
1,2Department of Electronics & Communication Engineering
1,2Millennium Institute of Technology, (MIT), Bhopal (M.P.)

Abstract— The paper presents a review on numerous ways in which of content based image retrieval. Content based image retrieval may be a system by that numerous images are retrieved from an outsized database assortment. These databases are ready using numerous visual features like color, texture, shape and spatial layout that are extracted using completely different techniques. As image database volume is increasing rapidly, researchers are probing for a much more robust mechanism to retrieve images and to get more accurate results. Therefore the analysis focus has been shifted from low-level feature extraction algorithms to the high-level more accurate results. Therefore the analysis focus has been shifted from low-level feature extraction algorithms to the high-level feature extraction mechanism. To develop better content based image retrieval system, it is necessary to boost numerous processes concerned in retrieval like feature extraction, image segmentation, image decomposition and similarity matching techniques. During this paper we tend to discuss the basic aspects, visual features and techniques for quick and retrieval images from the database. In this paper, completely different well-known options for image retrieval are quantitatively compared and their correlation is analyzed. We compare the options for two different image retrieval tasks (color a clear and medical radiograph) and a transparent distinction in performance is determined, which can be used as a basis for an appropriate selection of features. In the medical field, images, and especially digital images are produced in ever increasing quantities and used for diagnostic sand therapy.

Key words: Image Retrieval, Medical Image, Content based image

I. INTRODUCTION

This section provides an introduction to content based image retrieval systems (CBIRs) and also the technologies employed in them. Image retrieval has been an extremely active research area over the last 10 years, but first review articles on access methods in image databases appeared already in the early 80’s. As the technology growing throughout the society, the digital images, multimedia files, visual objects are also increasing. This huge amount of images requires novel methods to search and access the images. Advances in medical and other technologies have provided extensive image generation, its storage and transmission capabilities. Due to the increase in the usage of these digital images in various fields, researchers are focusing on new ways by which images can be easily, quickly and accurately retrieved and accessed from large databases. Image retrieval generally and content-based image retrieval (CBIR) particularly are well-known fields of analysis in data management within which an outsized number of strategies have been projected and investigated however within which still no satisfying general solutions exist.

The need for adequate solutions is growing because of the increasing quantity of digitally produced images in areas like journalism, medicine, and personal life, requiring new ways in which of accessing images. As an example, medical doctors have to be compelled to access massive amounts of images daily, home-users usually have image databases of thousands of images, and journalists additionally need to search for images by numerous criteria.

In literature, the term content based image retrieval (CBIR) has been used for the first time by Kato et. al. To explain his experiments into automatic retrieval of images from a database by color and shape feature. The typical CBIR system performs two major tasks. The first one is feature extraction (FE), where a set of features, called feature vector, is generated to accurately represent the content of each image in the database. The second task is similarity measurement (SM), where a distance between the query image and each image in the database using their feature vectors is used to retrieve the “closest” images. For CBIR feature extraction the two main approaches are feature extraction in spatial domain and feature extraction in transform domain. The feature extraction in spatial domain includes the CBIR techniques based on histograms, BTC, VQ. The transform domain methods are widely used in image compression, as they give high energy compaction in transformed image. So it is obvious to use images in transformed domain for feature extraction in CBIR. Transform domain results in energy compaction in few elements, so large number of the coefficients of transformed image can be neglected to reduce the size of feature vector. Reducing the size feature vector using fractional coefficients of transformed image and till getting the improvement in performance of image retrieval is the theme of the work presented here. Many current CBIR systems use average Euclidean distance on the extracted feature set as a similarity measure.

The two methods which are used for image retrieval are Text based image retrieval and Content based image retrieval. Text based image retrieval system also known as concept based image retrieval system. In concept based image retrieval user poses the query using natural language text, subject heading, keywords or annotations of the image. These systems do not actually understand the actual content of the images. Metadata is used for image indexing in concept based system. There is various limitation of concept based image retrieval system. There are number of ways to say the same thing. Annotation of images is never complete and is time consuming process because human perceptivity can lead to a number of errors. A new method for image retrieval is needed where the human factor would be relieved from the annotation task and doing it automatically. Most web based image search engines rely only on metadata and this produces a lot of garbage in the results. In these search engines, humans have
to enter the keywords manually and it is inefficient and expensive way to find images in a large database [7].

Content-based image retrieval is the modern image retrieval system. The Content based image retrieval systems are used to extract image features, index those using appropriate structures and efficiently process user queries providing the required answers. The query processing includes segments and features extraction and search in the feature space for similar images. In Content based image retrieval system various techniques are bought together effectively for the same purpose as image processing, information retrieval and database communities. It is also called query-by-image content and content-based visual information retrieval [9].

II. LITERATURE SURVEY

In the past decades various methodologies have been proposed content-based image retrieval. A brief introduction of these methodologies is as follows:

Content-based image retrieval using moment-preserving edge detection: In 2003, Shyi-Chyi Cheng has been planned A content-based image retrieval algorithm supported a new edge detection technique. Both the question and database images are divided into non-overlapping square blocks and coded by the mean in every uniform block and by edge data in non-uniform block. The coded blocks of a query image are then used to find matches from an image database. The edge feature in a given block is detected by applying the moment-preserving principle to the image data. The edge directions are approximated by multiples of 45° to speed up the matching method while not introducing obvious distortion. For a larger database, a selective filtering strategy supported the visual-pattern histograms are additionally delineated to further speed up the retrieval method. The solution to the edge detection problem during a given block is also analytic. This algorithm may be performed in no time for large database applications with any need for special hardware. Supported the edge detector, an image may be coded by visual-pattern codes block by block in real time. An identical strategy based on the visual-pattern codes join together makes the retrieval method robust and accurate, and therefore the methodology using the visual-pattern type histograms to separate out dissimilar images dramatically shortens the time needed to retrieve images, while not sacrificing the retrieval accuracy. The speed of retrievals may also be increased while not affecting the robustness of the system by clustering the type histograms of a database earlier. Future work will deal with linking edge patterns into objects to the planned system, and increasing the database size (Algorithm) [1].

Kernel-based distance metric learning for content-based image retrieval: In 2007, Hong Chang, Dit-Yan Yeung planned for a particular set of features chosen for representing images, the performance of a content-based image retrieval (CBIR) system depends critically on the similarity or dissimilarity measure used. Rather than manually selecting a distance operates in advance, a more promising approach is to be told a decent distance function from data automatically. During this paper, authors propose a kernel approach to enhance the retrieval performance of CBIR systems by learning a distance metric based on pairwise constraints between images as supervisory information. In contrast to most existing metric learning strategies that learn a Mahalanobis metric similar to performing linear transformation within the original image space, we tend to outline the transformation within the kernel-induced feature space that is non-linearly associated with the image space. Experiments performed on two real-world image databases show that planned technique not solely improves the retrieval performance of Euclidean distance while not distance learning, however it additionally outperforms different distance learning strategies considerably due to its higher flexibility in metric learning [2].

Wavelet optimization for content-based image retrieval in medical databases: In 2010, G. Quellec, M. Lamard, G. Cazuguel, B. Cochener, C. roux planned a content-based image retrieval (CBIR) technique for diagnosis aid in medical fields. Within the planned system, images are indexed in a very generic fashion, without extracting domain-specific features: a signature is built for every image from its wavelet transform. These image signatures characterize the distribution of wavelet coefficients in every sub band of the decomposition. A distance measure is then outlined to check two image signatures and so retrieve the foremost similar images in a database once a query image is submitted by a physician. To retrieve relevant images from a medical database, the signatures and also the distance measure should be related to the medical interpretation of images. As a consequence, authors introduce many degrees of freedom within the system so it may be tuned to any pathology and image modality. Especially, they propose to adopt the wavelet basis, inside the lifting scheme framework, and to use a custom decomposition scheme. Weights are introduced between sub bands. All these parameters are tuned by an improvement procedure, using the medical grading of every image within the database to outline a performance measure. The system is assessed on two medical image databases: one for diabetic retinopathy follow up and one for screening mammography, further as a general purpose database. Results are promising: a mean exactitude of 56.50%, 70.91% and 96.10% is archived for these three databases, when five images are returned by the system. In this article, two kinds of generic wavelet-based image signatures, with associated distance measures, have been evaluated in a CBIR system. They take advantage of the flexibility inherent in the wavelet transform framework to adapt the system to any specialized database. In particular, a way to adapt the wavelet transform to a high-level criterion, within the lifting scheme framework, is proposed in this paper. It makes it possible to generate any wavelet transform, respecting the orthogonality relations, and with a desired number of vanishing moments. A controlled random search, based on a genetic algorithm, is then performed in the predict and update filter space in order to find the optimal wavelet transform, and a similar search is performed in the distance weight vector space in order to maximize the precision of the system. The adapted wavelet transforms could also be used to store images in a compressed format, such as JPEG 2000, as their signature is computed [3].

Content based Image Retrieval for Medical Images using Canny Edge Detection Algorithm: In 2011, B Ramamurthy and K.R. Chandran have been planned the
enlargement of digital data content has led to the necessity for rich descriptions and efficient retrieval tool. To develop this, content based image retrieval method has contend a vital role within the field of image retrieval. This paper aims to produce efficient medical image data retrieval from a large content of medical database victimization one of the images content like image shape, because efficient content-based image retrieval within the medical domain is still a difficult problem. The main objective of this paper is to produce an efficient tool that is used for efficient medical retrieval from a large content of database and that is employed for further medical diagnosis purposes. In this work, we have used canny edge detection algorithm for extracting the shape features for the medical images. After extracting the shape feature, which are given as input to Support Vector Machine (SVM) for classifying the images and then classified images are indexed and labeled for making easy for applying retrieval algorithm in order to retrieve the relevant images from the database. This work retrieves the images from the huge medical database as required by the doctors for their further evaluation about their patients’ diseases. So that doctors can give proper treatment to their patients in proper time. Due to the rapid expansion of medical image database, unfortunately, the scalability of the medical image database is also increased in day to day medical treatment activities. When the scalability of the medical image database is increased, then the performance of the retrieval process is decreased. Hence the performance of the retrieval process is needed to be increased more along with the large scalability of the database. To overcome this issue, the canny edge detection algorithm, Support Vector Machine (SVM), and spatial indexing method were used. This will give retrieval accuracy, speed, and clarity of the medical images to predict exact disease situation of the patients. So that doctors can get exact idea about disease in order to give the treatments for their patients [4].

A Novel Approach for Content Based Image Retrieval: In 2012, Nidhi Singh, Kanchan Singh, Ashok K. Sinha have planned the problem of content based image retrieval in dynamic surroundings is addressed. It is not possible for systems that analyze images in real-time wherever the images are stored or added on an ongoing basis. During this paper, the authors propose a framework that is able to select the foremost appropriate features to analyze recently received images thereby improving retrieval accuracy and efficiency. An improved algorithm is planned here. The algorithm is includes of planning feature vectors segmentation which can be utilized in similarity comparison between query image and database images. The framework is trained for various images within the information. The planned algorithm has been tested on numerous real images and its performance is found to be quite satisfactory in comparison with the performance of standard strategies of content based image retrieval [5].

A weighted dominant color descriptor for content-based image retrieval: In 2013, Ahmed Talib, Massoudi Mahmuddin, Husniza Husni, Loay E. George have been proposed a new method in which, color has been extensively utilized in the method of image retrieval. The dominant color descriptor (DCD) that was projected by MPEG-7 may be a famous case in purpose. It is based on compactly describing the distinguished colors of an image or a neighbourhood. However, this method suffers from some shortcomings; particularly with regard to the object-based image retrieval. During this paper, a new semantic feature extracted from dominant colors (weight for every DC) is planned. The newly projected technique helps reduce the result of image background on image matching decision wherever an object’s colors receive much more focus. Additionally, a modification to DC-based similarity measure is additionally planned. Experimental results demonstrate that the planned descriptor with the similarity measure modification performs higher than the present descriptor in content-based image retrieval application. The planned descriptor considers as step forward to the object based image retrieval [6].

An Experimental Study on Content Based Image Retrieval Based On Number of Clusters Using Hierarchical Clustering Algorithm: In 2014, Monika Jain and Dr. S.K.Singh introduced a new algorithm. According to this, nowadays the content based image retrieval (CBIR) is becoming a source of actual and fast retrieval. CBIR presents challenges in indexing, accessing of image data and how end systems are evaluated. Data clustering is an unsupervised methodology for extraction hidden pattern from huge data sets. Several clustering and segmentation algorithms each suffer from the limitation of the quantity of clusters specific by a human user. It is usually impractical to expect an individual’s with comfortable domain data to be available to select the quantity of clusters (NC) to return. This paper discusses the image retrieval based on NC which is evaluated using hierarchical agglomerative clustering algorithm (HAC). In this paper, authors determine the optimal number of clusters using HAC applied on RGB images and validate them using some validity indices. Based on number of clusters, we retrieve set of images. These cluster values can be further used for divide and conquer technology and indexing for large image dataset. An experimental study is presented on real data sets. An experimental study on cluster oriented image retrieval with color features is shown in this paper. The selected color features of the image database mainly of same category and the image query are then clustered using HAC algorithm for similarity measurement purpose. The approach is examined in the experimental study with Corel image dataset. However for very large dataset, it can be used as preprocessing and applied on other unsupervised algorithms taking information regarding NC. The results of experiment can be used for indexing and be enhanced for divide and conquer technology. It will become good approach for efficient content based image retrieval system for very large dataset [10].

Content Based Image Retrieval: Classification Using Neural Networks: In 2014, Shereena V.B.and Julie M. David proposed a new algorithm. In a content-based image retrieval system (CBIR), the main issue is to extract the image features that effectively represent the image contents in an exceedingly database. Such an extraction needs a detailed analysis of retrieval performance of image features. This paper presents a review of elementary aspects of content based image retrieval as well as feature extraction of color and texture features. Normally used color features as well as color moments, color histogram and color
correlogram and Gabor texture are compared. The paper reviews the rise in efficiency of image retrieval once the color and texture options are combined. The similarity measures based on that matches are created and images are retrieved is mentioned. For effective indexing quick looking of images supported visual features, neural network based pattern learning may be accustomed achieve effective classification. This paper investigated numerous feature extraction algorithms in CBIR. A Study of various color and texture options for image retrieval in CBIR is performed. Various strategies are accessible for feature extraction in CBIR. They are also known and studied to know the image retrieval method within the CBIR systems. Studies created on experiment results show that the strategy based on hybrid combination of color and texture options has higher retrieval accuracy than the opposite strategies supported signal feature extraction. Color moments, color histograms, color correlogram and Gabor texture are thought about for retrieval. It is difficult to say that one feature is superior to others. The performance depends on the color distribution of images. The combination of color descriptors produces higher retrieval rate compared to individual color descriptors. Color moments and color histogram features are often combined to get better results. Color histograms and correlograms may be combined retentive advantages of histograms with special layout. Similarly, texture features may be combined with color moments or color histogram to induce correct results for image retrieval. From the studies, it is found that only one color feature or texture feature is not enough to explain an image. There is wide increase in retrieval efficiency when each color and texture features are combined. Additionally authors have reviewed numerous papers associated with totally different classification strategies for the advance of image retrieval in CBIR. Among totally different classification strategies, Neural Network classification is an efficient technique for image retrieval. It takes into consideration the characteristics of relevant and irrelevant images. Neural Network classification has significantly improved the recall rate and additionally retrieval time, as a result of its extremely efficient and accurate classification capability [11].

Low-Level Features for Image Retrieval Based on Extraction of Directional Binary Patterns and its Oriented Gradients Histogram: In 2015, Nagaraja S. and Prabhankar C.J. presented a complete unique approach for image retrieval based on extraction of low level features exploitation techniques like Directional Binary Code (DBC), Haar Wavelet Transform and Histogram of Oriented Gradients (HOG). The DBC texture descriptor captures the spatial relationship between any combine of neighbouring pixels in an exceedingly native region on a given direction, whereas Local Binary Patterns (LBP) descriptor considers the connection between a given pixel and its surrounding neighbors. Therefore, DBC captures a lot of spatial information than LBP and its variants; conjointly it will extract a lot of edge information than LBP. Hence, we use DBC technique so as to extract grey level texture features (texture map) from every RGB channels on an individual basis computed texture maps are any combined that represented color texture options (color texture map) of an image. Then, we decomposed the extracted color texture map and original image exploitation Haar Wavelet Transform. Finally, we encode the form and native options of and local transformed images exploitation Histogram of Oriented Gradients (HOG) for content based image retrieval. The performance of planned methodology is compared with existing ways on two databases like Wang’s Corel image Caltech 256. The analysis results show that our approach outperforms the present ways for image retrieval [12].

III. CONCLUSION

In this paper we have introduced various methods of content-based image retrieval. Each method is different from each other. With the development of technology, more and more images come into view as and have become a part of our daily existence. There is a wide range of applications which require image processing tools. Some of its examples are crime prevention medicine, Fashion and graphic design, Architectural and engineering design, Publishing and advertising, research, law etc. Thus, there are many technologies have been developed to meet the requirement [8].

REFERENCES


[13] Assistant Professor
[14] Millennium Institute of Technology
[15] Bhopal, India
[16] pushpabir25@gmail.com