A Survey on Content based Image Retrieval

Pooja Lather[1]
[1]Department of Computer Science and Engineering
[1]Ambala College of Engg. & Applied Research, Devsthali, Ambala, India

Abstract— Images are rich source of information. With the advancement in web Based search, content-based image retrieval systems (CBIR) are becoming popular to search images from large database with higher accuracy. In this paper we describe survey on content-based image retrieval, approaches used in content-based image retrieval and their limitations.

Key words: CBIR, GA, GLCM, RBIR

I. INTRODUCTION

CBIR research has been started in 1990's.It aims at finding relevant images from the large dataset visualizing its content rather than tags and keywords which we use in text retrieval approach. The requirement to find a desired image from a collection is commonly used by many professional groups, academics, journalists, commerce, design engineers, surveillance and art historians. The necessities of image users can vary significantly, it can be useful to depict image queries into 3-levels of abstraction: primitive features such as color or shape, logical features i.e. the identity of objects exposed and abstract attributes such as the impact of the scenes depicted. While CBIR systems currently work effectively only at the lowest of these levels, most users request higher levels of retrieval.

Difficulties with conventional methods of image indexing [Enser, 1995] have directed to the increase of attention in techniques for retrieving images on the origin of automatically-derived features such as color, texture and shape – this technology now commonly referred to as *Content-Based Image Retrieval* (CBIR). However, the technology still deficiencies maturity, and is not yet being used on a substantial scale. Due to absence of hard proof on the effectiveness of CBIR techniques in exercise, opinion is still abruptly divided about their usefulness in management real-life queries in large and varied image collections. The ideas which are presently used for CBIR system are all underneath research.

II. LITERATURE REVIEW

A. Global Feature based CBIR Systems (GFBIR)

Many image retrieval systems, a query is specified by an image to be matched. We discuss to this indication as "global search" since similarity is created by the overall properties of images. By divergence, there is also "partial search" querying systems called Region-based Image retrieval (RBIR) that retrieve results based on a particular region in an image. Even though RBIR systems enhance the retrieval accuracy, they require high difficult computations to calculate similarity; since these systems must consider every region in the database images resulting in high retrieval response time. Essentially, one of the key points of realizing CBIR is to extract suitable feature vectors to characterize image content correctly. Color, texture and shape features have been used for characterizing image content.

In [5], features such as shapes and texture are mined from the query and database images and are compared by similarity measures of Euclidean distance. The morphological process with spatially-variant structuring component is used for feature extraction. Later the feature extraction process, the feature vectors is calculated by applying Block Truncation coding (BTC) above the feature extracted images. It increases the performance of image retrieval with decrease computational complexity for query execution. Built on HSV color model, a technique of object-based spatial-color feature (OSCF) for color image retrieval is suggested in [3]. Firstly, objects are extracted from color, and then image features are characterized by objects in it. Color and spatial-color feature are assumed for description of objects. The new technique only pays consideration to main central objects. In [6], author offered a novel fuzzy method to classify the color images based on their content, to pose a query in relations of natural language and combine the queries based on neural networks for fast and effective retrieval.

Lin et al. [4] suggested two approaches one color-texture and another color-histogram based image retrieval system (CTCHIR). They suggested (1) three image features, based on color, texture and color distribution, as color co-occurrence matrix (CCM), variance between pixels of scan pattern (DBPSP) and color histogram for K-mean (CHKM) correspondingly and (2) a technique for image retrieval using integrating CCM, DBPSP and CHKM to increase image detection level and shorten computation of image retrieval. From Demonstration they found that, their suggested method overtakes the Jhanwar et al. [16] and Hung and Dai [17] approaches. Raghupathi et al. [7] made a comparative study on image retrieval methods, using different feature extraction methods like color histogram, Gabor Transform, color histogram plus gabor transform, Contour let Transform and color histogram plus contourlet transform. Hiremath and Pujari [18] proposed CBIR system based on the color, texture and shape features by partitioning the image into strips. The features computed on strips serve as local descriptors of color and texture features. The color and texture study are analyzed by using two level grid frameworks and the shape feature is used by using Gradient Vector Flow. The comparison of demonstration result of proposed method with other system [22] [11] [14] [23] reveal that, their suggested retrieval system gives better performance than the others. Rao et al. [15] proposed CTDCIRS (color-texture and dominant color based image retrieval system), they combined three features like Motif co- occurrence matrix (MCM) and variance between pixels of scan pattern (DBPSP) which defines the texture features and dynamic dominant color (DDC) to extract color feature. They matched their results with the work of Jhanwar et al. [16] and Hung and Dai [17] and establish that their method gives better retrieval results than others.

B. Cluster-based Retrieval Systems

Cluster analyses are themself not one precise algorithm, but the common job to be solved. It can be attained by various algorithms that vary considerably in their concept of what constitutes a cluster and how to proficiently find them. Popular concepts of clusters contain groups with small distances between the cluster members, dense areas of the data space, intermissions or particular statistical distributions. Clustering can consequently be expressed as a multi-objective optimization problem. The suitable clustering algorithm and parameter settings (including values such as the distance function to use, a concentration threshold or the amount of expected clusters) depend on the individual data set and proposed use of the results. Cluster analysis as such is not an involuntary task, but an iterative procedure of knowledge discovery or interactive multi-objective optimization that includes trial and failure. It will frequently be necessary to modify data preprocessing and model parameters till the result achieves the desired properties.

In centroid-based clustering, clusters are characterized by a central vector, which may not essentially be a member of the data set. When the number of clusters is permanent to k, k- means clustering gives a formal description as an optimization problem: find the cluster centers and allocate the objects to the close cluster center, such that the squared distances from the cluster are minimized.

In [19], author discussed on the comparative technique used in color histogram based on two major methods used normally in CBIR which are; normal color histogram by GLCM, and color histogram by K-Means. Using Euclidean distance, similarity measures between queried image and the input images are calculated. Research results shows that color histogram with K-Means technique has higher accuracy and precise compared to GLCM.

In Iyengar G. and Lippman A. the authors propose to use clustering method to permit for efficient access to huge image databases [8]. Efficient access is significant, due to the size of large image databases, querying becomes costly even if the images are characterized in a compact manner. With clustering, the task of retrieval is divided into a two stage process. In the first step an suitable cluster is selected and in the second step the top matches from this cluster are returned. They match a clustering technique which uses relative entropy to methods using the Euclidean standard. Kaster T., et all. Suggest touse image clustering methods to allow for quicker searching in image databases. They compare different clustering techniques to find out which fits the task of clustering images best [12]. In Saux B. L. and Boujemaa N. the authors suggested touse image clustering to give a good outline of an image database to help a user find image faster. To cluster this images, they estimate the distribution of image groups and search the best descriptive for each cluster [20]. They characterize images by a high-dimensional feature vector and suggested a new clustering algorithm which they match to other clustering methods. In [2], [9], and [10] give overall information about clustering of data and the estimation of results. In [13] a new clustering algorithm based on the EM algorithm is proposed and a method to avoid the problematic of finding an initial partition by iterative piercing of an initial Gaussian relating all data points is presented.

C. Genetic Algorithm for CBIR Systems

Genetic algorithm (GA) is part of artificial intelligence rely on the theory of natural selection and evolution. It is an competently global searching algorithm built on the norm of "survival of the fittest" and used for optimization problem and searching problems. Content-based image indexing and retrieval (CBIR) implementation using only one content feature doesn't give sufficient retrieval accuracy. Therefore to overcome this problem, any different model for the content based image retrieval system must be based upon association of multiple features for the image like color, texture, and shape. GA has been widely used for Optimization problem these centroids must be optimized using any search optimization procedure like genetic algorithm (GA) for increasing average recall and average precision of image retrieval and accuracy.

Shrikant. et.al.,[21] have introduced a way of using different feature descriptors such as, color ,texture and shape descriptors to signify low level features of image. There are the techniques called atrous wavelet transform (AWT) and Julesz's texton elements are used to produce the texton image. Also the multi texton histogram (MTH) is one of the techniques for these tasks. They integrate the advantages of co- occurrence matrix and histogram by signifying the attributes of co-occurrence matrix using histogram. User directed mechanism for CBIR using an interactive genetic algorithm (IGA) is proposed and implemented. The color characteristics such as the mean value, standard deviation and image bitmap of a color image are used as features for retrieval.

Arvind Nagathan et al., [1] introduced a cbir system which make use of feed-forward back propagation neural network.at first step neural network is trained about the features of images. The image features reflected here are color histogram as color descriptor, GLCM (gray level co-occurrence matrix) as texture descriptor and edge histogram as edge descriptor. The training is done using back propagation algorithm. This trained when presented with a query image retrieves and shows the images which are relevant and similar to query from the database.

III. CBIR ISSUES

Notable observations in the related works are as follows:

1) Image is rich source of information Therefore, perfect descriptions for semantic features is still not possible. This is due to the variety of visual features, which broadly occurs in many real applications of image retrieval.
2) Content-based image indexing and retrieval (CBIR) implementation using only one content feature doesn't give satisfactory retrieval exactness because image itself contain many feature therefore considering only one feature affect precision and accuracy.
3) For speeding-up image retrieval and to reduce the search domain it will be beneficial to group the images into clusters.it will improve accuracy by reducing search domain rather than using large databases.
4) There are many of approaches used to define the similarity between different objects, like, Dice

coefficient, Euclidean distance, cosine coefficient and distance-based measurements. Therefore selecting appropriate distance measurement is also essential.

5) The current CBIR systems use basically global features, or region based features to characterize the content of an image. Even though RBIR systems can increase the retrieval accuracy, RBIR systems has a high retrieval response time and they high complex computations to calculate similarity; since these systems must consider each region in the database images especially and color images have large dimensions and the computations are quite time consuming,
6) Most commonly faced problem when designing CBIR system is to make a system general-purpose .All current CBIR systems suffer from inadequate generalization performance and accuracy. Therefore combining some methods like content-based, using global features, multi-features extraction, clustering technique classification and artificial intelligence to construct the CBIR system, we can get more and closer to generalization.

IV. Conclusion

Images are full with information and features that can be used for image retrieval. We must select and extract the most significant features that lead to a good retrieval result. A review of given literature shows that image search is usually done considering only one feature which can't provide good results. So it's important to consider multiple feature such as color, Shape, texture .Searching the whole database will be time consuming therefore, images with similar features are grouped into related clusters. Images can be clustered using K-means or any other clustering approach .Clustering of images will be done to query processing so that to answer a query, the system doesn't require to search the entire image database. This method saves major query processing time and computation load without compromising the retrieval precision in large database.

References

[1] Arvind Nagathan, Manimozhi, Jitendranath Mungara. "Content-Based Image Retrieval System using Feed-Forward Backpropagation Neural Network" IJCSNS International Journal of Computer Science and Network Security, VOL.14 No.6, June 2014.
[2] Berkhin P. "Survey of Clustering Data Mining Techniques". Technical report, Accrue Software, San Jose, CA, 2002.
[3] Chaobing Huang, Yarong Han, Yu Zhang. "A Method for Object-based Color Image Retrieval", Fuzzy Systems and Knowledge Discovery (FSKD), 9th International Conference on, pp: 1659-1663,2012
[4] C.H. Lin, R.T. Chen and Y.K. Chan, "A smart content-based image retrieval system based on color and texture feature", Image and Vision Computing vol.27, pp.658–665, 2009.
[5] Daisy, M.M.H., TamilSelvi, S. and Prinza, L. "Gray Scale Morphological Operations for Image Retrieval", International Conference on Computing, Electronics and Electrical Technologies [ICCEET], pp. 571-575,2012.
[6] Fernando, R. and Kulkarni, S., 'Hybrid Technique for Colour Image Classification and Efficient Retrieval based on Fuzzy Logic and Neural Networks', Neural Networks (IJCNN), The 2012 International Joint Conference on, pp:1-6.
[7] G. Raghupathi, R.S. Anand, and M.L Dewal, "Color and Texture Features for content Based image retrieval", Second International conference on multimedia and content based image retrieval, July-21-23, 2010.
[8] Iyengar G., Lippman A. "Clustering Images Using Relative Entropy for Efficient Retrieval". Proc. Workshop on Very Low bitrate Video Coding, Urbana, IL. 1998.
[9] Jain, A.K., Dubes R.C., "Algorithms for Clustering Data". Prentice Hall Inc., Englewood Cliffs, New Jersey, ISBN: 0-13-022278-X, pp: 320, 1988.
[10] Jain A. K., Murty M. N., Flynn P. J. "Data Clustering: A Review. ACM Computing Surveys", Vol. 31, No. 3, pp. 264–323, 1999.
[11] J. Li, J.Z. Wang and G. Wiederhold, "IRM: Integrated Region Matching for Image Retrieval", In Proceeding of the 8th ACM International Conference on Multimedia, pp- 147-156, Oct. 2000.
[12] K¨aster T., Wendt V., G. Sagerer. "Comparing Clustering Methods for Database Categorization in Image Retrieval". Proc. DAGM 2003, Pattern Recognition, 25th DAGM Symposium, Vol. 2781 of Lecture Notes in Computer Science, pp. 228–235, Magdeburg, Germany, 2003.
[13] Linde Y., Buzo A., Gray R.. "An Algorithm for Vector Quantization Design". Proc. IEEE Transaction Communications, Vol.28, pp.84–95, 1980.
[14] M. Banerjee, M. K. Kundu and P. K. Das, "Image Retrieval with Visually Prominent Features using Fuzzy set theoretic Evaluation", ICVGIP, 2004.
[15] M. B. Rao, B. P. Rao, and A. Govardhan, "CTDCIRS: Content based Image Retrieval System based on Dominant Color and Texture Features", International Journal of Computer Applications, Vol. 18– No.6, pp-0975-8887, 2011.
[16] N. Jhanwar, S. Chaudhurib, G. Seetharamanc and B. Zavidovique, "Content based image retrieval using motif co-occurrence matrix", Image and Vision Computing, Vol.22, pp-1211–1220, 2004.
[17] P.W. Huang and S.K. Dai, "Image retrieval by texture similarity", Pattern Recognition, Vol. 36, pp- 665–679, 2003.
[18] P. S. Hiremath and J. Pujari, "Content Based Image Retrieval based on Color, Texture and Shape features using Image and its complement", 15th International Conference on Advance Computing and Communications. IEEE. 2007.
[19] Rasli, R.M.Muda, T.Z.T., Yusof, Y.,Bakar, J.A. 'Comparative Analysis of Content Based Image Retrieval Technique using Color Histogram. A Case Study of GLCM and K-Means Clustering', Intelligent Systems, Modelling and Simulation (ISMS), Third International Conference on, pp: 283 – 286,2012
[20] Saux B. L., Boujemaa N. "Unsupervised Robust Clustering for Image Database Categorization". Proc.

International Conference on Pattern Recognition, Vol. 1, pp. 259–263, 2002

[21] Shrikant Chavate , Prof. Vikas Gupta , " An approach used for user Oriented Content Based Image Retrieval using Interactive Genetic Algorithm," International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 11, November 2013

[22] Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content Based Image Retrieval", IEEE Transactions on Pattern Analysis and Machine Intelligence. Vol. 24, No.9, pp. 52-1267, 2002

[23] Y. Rubner, L. J. Guibas and C. Tomasi, "The earth mover's distance, multidimensional scaling, and color-based image retrieval", Proceedings of DARPA Image understanding Workshop. Pp. 661-668, 1997

