RESEARCH ARTICLE OPEN ACCESS

An Approach for Content Based Image Retrieval Using Similarity Measures

Ritu A. Mundada [1], Akash D. Waghmare [2]
M.E [1] Student, Assistant Professor [2]
Department of Computer Science, SSBT'S COET College
Jalgaon - India

ABSTRACT
The images are used to identify a person, object and various things. Find the images from a large database are very important today. The image searching, browsing and retrieval can be done using the contents of the image i.e. visual features of an image such as color, shape, texture and spatial layout. This mechanism is called content based image retrieval. For the performance improvement of content based image retrieval, it is very necessary to find effective image feature extraction method and image matching mechanisms. Clusters of image are formed using clustering algorithm and similarity matching is done using feature extraction method. The problem involves in that it compare the structural similarity of the images using the clustering and indexing method. Hence when the more number of image is added, it is unable to retrieve the more relevant image similar to the query image. In proposed system more features of the image will extracted for improving the image retrieval performance using Laplacian filter through mean, median and standard deviation.
Keywords:- Content based image retrieval, wavelet decomposition, K-means clustering algorithm.

I. INTRODUCTION

Military and civilian equipments have been generating giga-bytes of images every day. Interest must be increased in image retrieval because of growth in World Wide Web. Image retrieval has been done using text or metadata such as key word in the 1970s. In 1979 in Florence, a conference was held such as Database Techniques for Pictorial Applications. The old methods retrieve the image based on textual annotation not the visual content. In text based retrieval image can be retrieve by annoting the image manually which is very time consuming and incorrect method. So the retrieving the visual feature is a need of the time So Content Based Image Retrieval (CBIR) is a mechanism is develop in 1980s for the improvement of text based image retrieval mechanism. Retrieve the digital image by using the visual content of the image. Feature extraction can be done using the feature vector. Content Based Image Retrieval (CBIR) is a technique that uses visual contents to search the images from large number of image databases according to the request given by the user in the form of a query image. With the rapid development of multimedia and network technology, image information has been used more widely, for the large image database, how to manage it visually is an urgent problem, flexible, efficient and accurate image retrieval strategy is one of the most essential technologies to solve the problem. Content based image retrieval is based on or retrieve the image using the keyword or tag or string assign to it. Image content consist of color, texture, shape which is the low level feature and the high-level semantic features. The image retrieval characteristic is based on low-level feature that is features can be obtained visually direct from the image. High-level semantic features of image can be retrieve to solve the mathematical model, search algorithm and other issues. But semantic feature possess ambiguity, uncertainty, and dependence on the natural language description. Academic and commercial development community received a lot of attention from Content based image retrieval systems now a days. In such a system user can give the queries such as "retrieve similar images to given image", from a large database. CBIR systems in simple image are developed to retrieve the simple

image from a large collection of images in an effective and efficient manner by development of better techniques and tools.

Section II represent the work related to the image retrieval. Further Section III describes the existing system that is used to retrieve the image from the database. Section IV presents the problem associated with the existing algorithm. Conclusion must be presented in the last section. After which reference is followed.

A. *Color*

Color is the most important and distinguishing visual feature of the image [10]. A color histogram represent the global color distribution in an image which is more frequently used technique for content-based image retrieval because of its efficiency and effectiveness. HSV color space are used for the efficient color histogram matching. The HSV color space is used for the manipulation of hue and saturation (to shift colors or adjust the amount of color) .Color represent the brightness of the image .The same image with different resolution color technique is not correct as difference .

B. *Color Histogram*

The color histogram is an effective representation of the color content of an image if the color pattern is unique and compared with the large number of the data set. The global and local distribution of colors in an image computed effectively characterizing the color histogram. In addition, it is robust to translation and rotation about the view axis and changes only slowly with the scale, orientation and viewing angle. When an image database contains a large number of images, histogram comparison will saturate the discrimination. This problem may be sove using the joint histogram technique.

C. *Color Coherence Vector*

Color coherence vector is a different way of incorporating spatial information into the color histogram. Each histogram bin is partitioned into two types ,first on is coherent, if it belongs to a large uniformly-colored region, or second one is incoherent, if it does not belongs to uniformly color region.

D. *Texture*

Texture is another important feature of images for extracting the image feature. Various texture feature representations have been investigated in pattern recognition and computer vision. Basically, texture feature methods can be classified into two categories: first, structural method and second one is statistical. Structural methods, which including morphological operator and adjancy graph, describe texture by identifying structural primitives and their placement rules. When applied to textures feature they are most effective as they are very regular. Statistical methods, including markove random field and multi resolution technique such as gabor filer and wavelet transform which decomposes the images using the haar wavelet. Texture are characterize by the statistical distribution of the image intensity.

E. *Shape*

shape-based image retrieval consists of measuring the similarity between shapes represented by their features. Some simple geometric features can be used to describe shapes. Usually, the simple geometric features can only discriminate shapes with large differences; therefore, they are usually used as filters to eliminate false hits or combined with other shape descriptors to discriminate shapes. They are not suitable to stand alone shape descriptors. A shape can be described by different aspects. These shape parameters are Mass, Center of gravity (Centroid) , Mean, Variance, Dispersion etc.

II. RELATED WORK

Image shape matching is prime concern in object recognition and identification methods. An image matching is a means of determining the resemblance of one image with the other image. Images are matched based on their shape and texture and it finds variety of applications ranging from image retrieval, object recognition, remote sensing, image classification ,image analysis and so on. In general, image matching techniques are classified into

structure-based and feature-based methods. Structure-based methods compare the shape/structure and the size of the images, whereas the feature- based methods examine the image features like color and texture in addition to size and shape. Therefore, the image shape and size are the most essential component in automatic image matching systems. Moreover, the image comparison can not be done directly on image shapes, it require an image registration process to align the image within the same coordinate space.

Kalavathi, in [8], proposed image comparison method based on Fourier Mellin transformation which accurately compare two images and computes the overlapping similarity and hausdorff distance. For some images, the proposed method has failed to produce accurate result when the spatial difference between the images is high. This may be avoided by modifying the Fourier Mellin (FM) transformation registration. The proposed method has efficiently compared the given images and has produced accurate comparison result. The drawback of this method is that for some images it fails to register the images correctly when the difference in spatial coordinates is high.

Chary, in [6], proposed a Feature extraction method for color image similarity. Retrieval of images within a large image collection based on color projections and different mathematical approaches are introduced and applied for retrieval of images. Images are sub grouping using threshold values ,and RGB color combinations considered for retrieval of images. Proposed system uses combinations of color feature to overcome the problem description.

Li Li [14],proposed algorithm that matches the image based on feature point and special type of descriptors name as DAISY descriptor. SURF descriptors has an problem that it gives the invariant rotation that gives the unstable performance. Rotation of the principal direction obtain according to the DAISY descriptor. This descriptor improves the time than the existing one .DAISY descriptor work in circular fashion because circular neighborhoods gives better feature than the rectangular neighborhoods. Gaussian function are used for selecting the proper key points.It provide more robustness to the image. But this algorithm not suitable for large image scale variation.

P.M.Panchal, in [13], proposed Scale Invariant feature Scale Invariant Feature Transform (SIFT) and Speed Up Robust Features (SURF) local descriptors. Rotation ,orientation , scalling can be done using the SIFT. SIFT produces the key point descriptors. It consist of the four step algorithm. First one is detecting the extrema, second is key point localization ,third is orientation assignment and the last one is key point description. Hessian matrix is used in SURF. It is multi-scale theory. SIFT compare more feature than SURF. SURF has less speed than SIFT.

Shanmugapriya, in [10], proposed method uses three approaches to retrieve the relevant images from the database. Images can be retrieved based on Color, Texture, both Color and texture respectively. The proposed method uses algorithms such as auto color correlogram to retrieve color based images, Gaussian mixture models to retrieve texture based images and Query point movement for relevance feedback. The experimental results conforms that the proposed method gives maximum accuracy when compared to existing work. This method lacks when the structure of object is similar between each other. system can be redesigned to accept semantic in addition to content based queries. To obtain integrated system texture features must be derived from other algorithms.

Elbakry, in [11], proposed Bag-of-Visual Word (BoVW) that can be used successfully in Content-based Image Retrieval .Image retrieval system that uses local feature descriptors such as SURF and SIFT. BoVW retrieve efficiently similar images from standard databases. The proposed system uses SIFT and SURF techniques as local descriptors to produce signatures of the image that are invariant to rotation, orientation and scale. The proposed system is to extract the local descriptors from the processed image and key points are extracted from the image. The system combines the robust techniques, such as SIFT, SURF, and BoVW, to enhance the retrieval process. In the system, we used a k-means algorithm to cluster the feature descriptors in order build a visual vocabulary. But they have many limitation when dealing with

the broad content of image.

III. PROPOSED SOLUTION

1. Input query image.
2. Database images decomposes into different levels using the haar wavelet .
3. A data set contains N elements such as R_0, R_1 ….R_{N-1}, N/2 averages as well as N/2 wavelet coefficient values. The averages are stored in the upper half of the array which is of N element and lower half contains the coefficients. In the wavelet calculation average becomes the input. the system repeated iteration until a single average and a single coefficient are calculated.
4. F-form theory are used for the extraction of image features using the feature vector.
5. In RGB color space direct wavelet decomposition of image is done. After decomposition of image, each resulting sub image form the coefficient matrix using the feature vector. dimension of image feature can be decreases using f-form theory and image matching is perform effectively.
6 Using the segmentation process, clusters of the image is form. In k-means clustering algorithm, The main idea is to define k centroids, one for each cluster. Firstly ,randomly select the 'c' cluster center. calculate distance between each data points and selected cluster center. data point is assign to the cluster center whose distance from the cluster center is minimum or negligible of all the cluster center .recalculate the centroid again. again calculate the distance be data points and new obtain cluster center again. If no data points are reassign then stop the process.
7 The feature vector matrix is stored in index file according to its cluster.
8 Similarity based on the euclidian distance.
9 Find the image similar to the query image.

IV. PROBLEM STATEMENT

The image retrieval method is searching for a tag that would match the descriptive keyword or metadata that describe to the image. This method is called the text based matching of image. The retrieval of images based on their content called Content Based Image Retrieval. The CBIR method gives the results are far more accurate than image indexing and clustering method . The goal of the content-based image retrieval method is to retrieve more relevant image from a large number of dataset that matches to the query image or given image. The features of the images are extracted using feature vector as well as their values and indices are saved in the database. Then all irrelevant items are ideally filter out using the index structure by checking attributes with the query image. The relevant image feature are compared to the feature of the query image based on similarity measure and retrieved items are ranked in order of similarity. The problem involves in that it compare the structural similarity of the images using the clustering and indexing method. But when the more number of image is added ,it is unable to retrieve the more relevant image similar to the query image.

V. CONCLUSION

The proposed content based image retrieval system uses haar wavelet for decomposition of image using wavelet decomposition. After that features are extracted using feature extraction algorithm such as f-norm theory. The clusters of images are form using the k-means clustering algorithm and similarity matching is done using f-norm theory. For more number of images the clustering method has failed to produce accurate result. Instead of comparing only the structural similarity of the image, it needs to extended further to compare the image features.

REFERENCES

[1] Hartigan J.A., Wong, M.A, (1979), "Algorithm AS136: A k-means Clustering algorithm," Applied Statistic, 28, pp. 100-108.
[2] Chen and Y.N.JangY. Chen, J. Z. Wang, " Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval" IEEE Transactions on Pattern

Analysis and Machine Intelligence, Vol. 24, No. 9, pp. 1252–1267, Sept. 2002.

[3] Arnold W.M. Smeulders , Marcel Worring, and Ramesh Jain , "content based image retrieval at the end of early years" , IEEE Transaction on pattern analysis and machine intelligence, VOL. 22, NO. 12. year 2000.

[4] Prashan Premaratne, "New structural similarity measure for image comparison", Emerging Intelligent Computing Technology and Applications,pp.292-297,2012.

[5] Parulpreet singh, "An Approach of Image Retrieval Using Content Based Retrieval System ,"International Journal of Advanced Research in Computer Science and Engineering,Vol.3,February 2013.

[6] R.Venkata Ramana Chary, "Feature Extraction Methods For Color Image Similarity," Advanced Computing an International Journal, Vol.3, No.2, March 2012.

[7] A. M. Patil, "Content Based Image Retrieval Using Color and Shape Features", International

[8] Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering,Vol. 1, Issue 5, November 2012.

[9] P.Kalavathi, "Image Shape Matching Based on FM Transformation Technique,"An Indian Journal of applied reasearch,Vol.4,Issue 10,pp.2249-555, October 2014.

[10] Shankar M. Patil,, "Content Based Image Retrieval Using Color, Texture Shape",International Journal of Computer science Engineering Technology,Vol.3, Issue 9, Sept.2012.

[11] Shanmugapriya, "New Content Based Image Retrieval System Using Gmm And Relevance Feedback",JournalofComputer science,Vol.2,pp.330-340,Nov.2014.

[12] HazemElbakry, "Content-Based Image Retrieval using Local Features Descriptors and Bag-of-VisualWords",International Journal of Advanced ComputerScienceand Applications,Vol.6,pp.212-219, No.9,2015.

[13] Dr. S.K.Singh, " A Survey On: Content Based Image Retrieval Systems Using Clustering Techniques For Large Data sets," International Journal of Managing Information Technology (IJMIT) Vol.3, No.4, November 2011.

[14] P.M.Panchal, "comparision of SIFT and SURF", International Journal of Innovative Research in Computer and Communication Engineering *Vol*. 1, Issue 2, April 2013.

[15] Li Li, "Image Matching Algorithm based on Feature-point and DAISY Descriptor", journal of multimedia, vol. 9, no. 6, june 2014.