Interactive KNN and DWT Based Content Based Image Retrieval Mechanism

Pritpal Kaur[1], Sukhvir Kaur[2]
[1]*Research Scholar, Department of Computer Engineering, CTIEMT, Jalandhar*
[2]*Assistant Professor, Department of Computer Engineering, CTIEMT, Jalandhar*

***Abstract*—** Content-based image retrieval has attracted voluminous research in the last decade paving way for development of numerous techniques and systems besides creating interest on fields that support these systems. CBIR indexes the images based on the features obtained from visual content so as to facilitate speedy retrieval. In this work, we present an interactive KNN and DWT based image retrieval system that uses color, contours and texture as visual features to describe the content of an image region. The results demonstrate that each type of feature is effective for a particular type of images according to its semantic contents. Multiple experiments are conducted on different categories of images from the WANG database.

***Keywords*—** *Content based image retrieval (CBIR), color, shape, texture features, DWT, KNN, relevance feedback.*

I. INTRODUCTION

Due to exponential increase of the size of the so-called multimedia files in recent years because of the substantial increase of affordable memory storage on one hand and the wide spread of the World Wide Web (www) on the other hand, the need for efficient tool to retrieve images from large dataset becomes crucial. Problems with traditional methods of image indexing have led to the rise of interest in techniques for retrieving images on the basis of automatically-derived features such as color, texture and shape a technology now generally referred to as Content-Based Image Retrieval (CBIR). After a decade of intensive research, CBIR technology is now beginning to move out of the laboratory and into the marketplace, in the form of commercial products like QBIC. However, the technology still lacks maturity, and is not yet being used on a significant scale. In the absence of hard evidence on the effectiveness of CBIR techniques in practice, opinion is still sharply divided about their usefulness in handling real-life queries in large and diverse image collections. Nor is it yet obvious how and where CBIR techniques can most profitably be used with the advancement in internet and multimedia technologies, a huge amount of multimedia data in the form of audio, video and images has been used in many fields like medical treatment, satellite data, video and still images repositories, digital forensics and surveillance system. This has created an ongoing demand of systems that can store and retrieve multimedia data in an effective way. Many multimedia information storage and retrieval systems have been developed till now for catering these demands.

The most common retrieval systems are Text Based Image Retrieval (TBIR) systems, where the search is based on automatic or manual annotation of images. Retrieval (TBIR) systems, where the search is based on automatic or manual annotation of images. A conventional TBIR searches the database for the similar text surrounding the image as given in the query string. The commonly used TBIR system is Google Images. The text based systems are fast as the string matching is computationally less time consuming process. However, it is sometimes difficult to express the whole visual content of images in words and TBIR may end up in producing irrelevant results. In addition, annotation of images is not always correct and consumes a lot of time. For finding the alternative way of searching and overcoming the limitations imposed by TBIR systems more intuitive and user friendly content based image retrieval systems (CBIR) were developed. A CBIR system uses visual contents of the images described in the form of low level features like color, texture, shape and spatial locations to represent the images in the databases. The system retrieves similar images when an example image or sketch is presented as input to the system. Querying in this way eliminates the need of describing the visual content of images in words and is close to human perception of visual data. A conventional TBIR searches the database for the similar text surrounding the image as given in the query string. The commonly used TBIR system is Google Images. The text based systems are fast as the string matching is computationally less time consuming process. However, it is sometimes difficult to express the whole visual content of images in words and TBIR may end up in producing irrelevant results. In addition, annotation of images is not always correct and consumes a lot of time. For finding the alternative way of searching and overcoming the limitations imposed by TBIR systems more intuitive and user friendly content based image retrieval systems (CBIR) were developed. A CBIR system uses visual contents of the images described in the form of low level features like color, texture, shape and spatial locations to represent the images in the databases. The system retrieves similar images when an example image or sketch is presented as input to the system.

Content-based image retrieval research has produced a number of search engines. The commercial image providers, for the most part, are not using these techniques. The main reason is that most CBIR systems require an example image and then retrieve similar images from their databases. Real users do not have example images; they start with an idea, not an image. Some CBIR systems allows users to draw the sketch of the images wanted. Such systems require the users to have their objectives in mind first and therefore can only be applied in some specific domains, like trademark matching, and painting purchasing. Earlier CBIR systems rely on global image features, such as color histogram and texture statistics. Global features cannot capture object properties, so local features are favored for object class recognition. Similar image elements, like pixels, patches, and lines can be grouped together to form higher-level units, which are more likely to correspond to objects or object parts. Different types of features can be combined to improve the feature discriminability. For example, using color and texture to identify trees is more reliable than using color or texture alone. The context information is also helpful for detecting objects. A boat candidate region more likely corresponds to a boat if it is inside a blue region. While improving the ability of our system by designing higher-level image features and combining individual ones, we should be prepared to apply more and more features since a limited number of features cannot satisfying the requirement of recognizing many different objects in ordinary photographic images. To open our system to new features and to smooth the procedure of combining different features, we propose a new concept called an abstract region; each feature type that can be extracted from an image is represented by a region in the image plus a feature vector acting as a representative for that region. In a typical CBIR system (Figure 1), image low level features like color, texture, shape and spatial locations are represented in the form of a multidimensional feature vector. The feature vectors of images in the database form a feature database. The retrieval process is initiated when a user query the system using an example image or sketch of the object. The query image is converted into the internal representation of feature vector using the same feature extraction routine that was used for building the feature database. The similarity measure is employed to calculate the distance between the feature vectors of query image and those of the target images in the feature database. Finally, the retrieval is performed using an indexing scheme which facilitates the efficient searching of the image database. The idea is that all features will be regions, each with its own set of attributes, but with a common representation. This uniform representation enables our system to handle multiple different feature types and to be extendable to new features at any time.

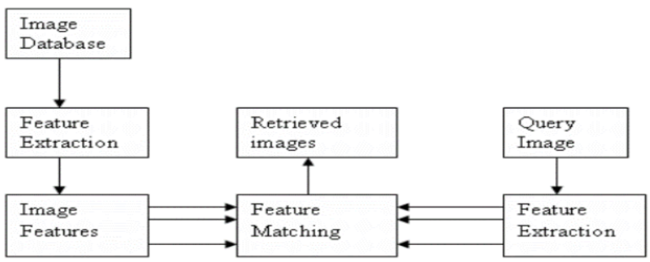

Fig.1. Architecture of a typical CBIR System

II. DWT

The DWT represents the signal in dynamic sub-band decomposition. Generation of the DWT in a wavelet packet allows sub-band analysis without the constraint of dynamic decomposition. The discrete wavelet packet transform (DWPT) performs an adaptive decomposition of frequency axis. The specific decomposition will be selected according to an optimization criterion. Wavelet is an oscillatory function of time or space that is periodic and of finite duration with zero average value. A family of wavelets can be generated by dilating and translating mother wavelet. Wavelet provides time- frequency representation of a signal and is used to analyze non- stationary signals. Multi-resolution technique is used in wavelet transform where different frequencies are analyzed with different resolutions. Big wavelets give an approximate value of a signal, while the smaller wavelets boost up the smaller details. DWT is computed either by using convolution based or lifting based procedures. In both the methods, the output sequence decomposed into low-pass and high-pass sub bands, where each sub bands constituting of half the number of samples of the original sequence.

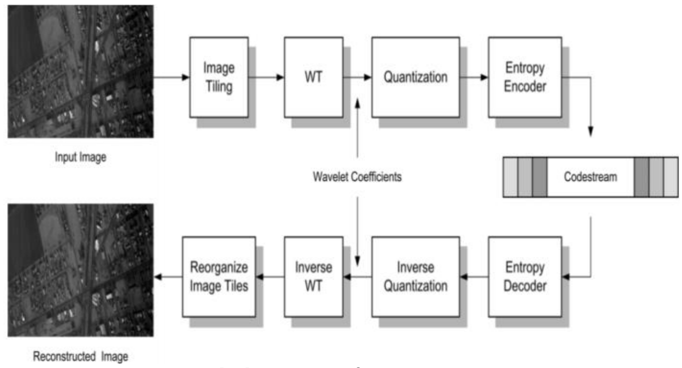

Fig.2. Block Diagram of image processing

The Discrete Wavelet Transform (DWT), based on time-scale representation, provides efficient multi-resolution sub band decomposition of signals. It has become a powerful tool for signal processing and finds numerous applications in various fields such as audio compression, pattern recognition, texture discrimination, computer graphics etc. Specifically the 2-D DWT and its counterpart 2- D Inverse DWT (IDWT) play a significant role in many image/video coding applications. The DWT architecture, the input image is decomposed into high pass and low pass components using HPF and LPF filters giving rise to the first level of hierarchy. The process is continued until multiple hierarchies are obtained. A1 and D1 are the approximation and detail filters.

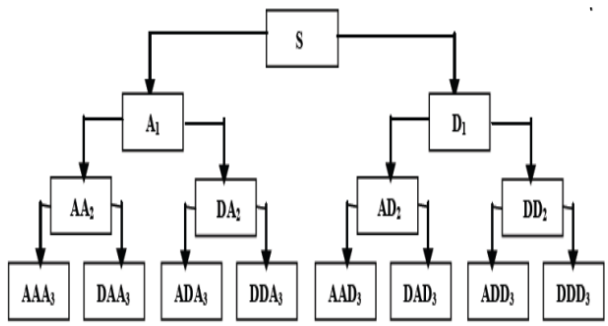

Fig.3. DWT Decomposition

The image is first decomposed into four sub bands of LL, LH, HL and HH. Further the LL sub band is decomposed into four more sub bands as shown in the figure. The LL component has the maximum information content as shown. The other higher order sub bands contain the edges in the vertical, horizontal and diagonal directions. The basic implementation of DWT for images is described as follows. First, an image decomposed into four parts of low, middle and high frequency sub components LL1, LH1, HL1 and HH1 by sampling horizontal and vertical channels using sub band filters. The sub components LH1, HL1 and HH1 represent the first level decomposition. To obtain the next level decomposition the sub component LL1 is further decomposed as shown in figure 5. This process of subsampling is repeated several times based on the requirement. In this work bi-orthogonal wavelets are used to perform watermark embedding and extraction. Bi-orthogonal wavelet generates two basic functions for decomposition and reconstruction.

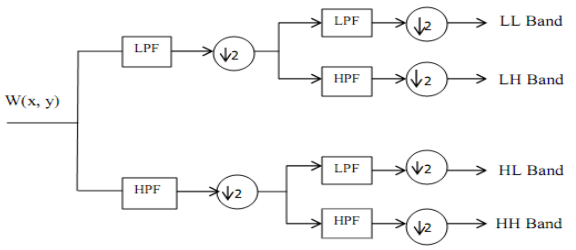

Fig.4. Wavelet decomposition using sub band coding

III. RELEVANCE FEEDBACK

The term "Relevance Feedback" is to take the feedback from the user whether the results are relevant or irrelevant. If the results are satisfied with the user's interest then it's OK, other again user gives the query and same process will follow. This is a feature which having some informational retrieval system. The basic idea behind the relevance feedback is to take the results which are returned from the given query and that information is to use and check whether those results are relevant to perform the new query or not. Further we are able to differentiate the three types of feedback:

1. Explicit Feedback: Basically this feedback gets from the assessors of relevance which indicates the relevance of a document retrieved for a query. This feedback is defined as explicit only when the users of the system know that feedback which is provided that interpreted as relevance judgments.

2. Implicit Feedback: This feedback is inferred from the user's behavior and note about the documents that they do or do not select for viewing, the timing spent on viewing the document and page browsing or scrolling the actions.

3. Blind or 'Pseudo' Feedback: This provides a method for automatic analysis which automates the manual part of the relevance feedback and user gets the better improves retrieval results without ant interaction. This method is to do the normal retrieval to get the initial set of most relevant documents, and to suppose the top 'k' documents are relevant, and then finally do with the relevance feedback as before under this assumption.

IV. KNN

An instance based learning method called the K-Nearest Neighbor or K-NN algorithm has been used in many applications in areas such as data mining, statistical pattern recognition, image processing. Successful applications include recognition of handwriting, satellite image and EKG pattern. In data mining, we often need to compare samples to see how similar they are to each other. For samples whose features have continuous values, it is customary to consider samples to be similar to each other if the distances between them are small. Other than the most popular choice of Euclidean distance, there are of course many other ways to define distance. The k-means clustering algorithm attempts to split a given anonymous data set (a set containing no information as to class identity) into a fixed number (k) of clusters.

Initially k number of so called centroids are chosen. A centroid is a data point (imaginary or real) at the center of a cluster. Each centroid is an existing data point in the given input data set, picked at random, such that all centroids are unique (that is, for all centroids ci and cj, ci ≠ cj). These centroids are used to train a KNN classifier. The resulting classifier is used to classify (using k = 1) the data and thereby produce an initial randomized set of clusters. Each centroid is thereafter set to the arithmetic mean of the cluster it defines. The process of classification and centroid adjustment is repeated until the values of the centroids stabilize. The final centroids will be used to produce the final classification/clustering of the input data, effectively turning the set of initially anonymous data points into a set of data points, each with a class identity.

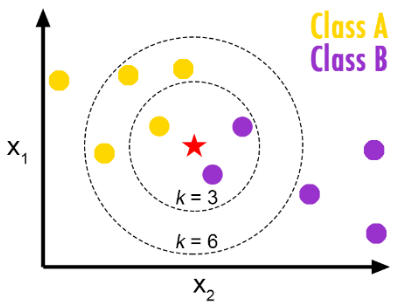

Fig.5. KNN Classification

The training examples are vectors in a multidimensional feature space, each with a class label. The training phase of the algorithm consists only of storing the feature vectors and class labels of the training samples. In the classification phase, k is a user-defined constant, and an unlabeled vector (a query or test point) is classified by assigning the label which is most frequent among the k training samples nearest to that query point.

V. ARCHITECTURE

- Input the query image and database images and divide them into the block of size N * N.
- Wavelet based matching is applied for similarity comparison of database image with the query image.
- After matching of images, system will return the matching images to the query image from the database. User will enter the relevant feedback according to his/her perception.
- Apply the genetic interactive algorithm that uses KNN for further classification of images and returns the top ranked images.

VI. OPENCV

The open source computer vision library, OpenCV, began as a research project at Intel in 1998. It has been available since 2000 under the BSD open source license. OpenCV is aimed at providing the tools needed to solve computer-vision problems. It contains a mix of low-level image-processing functions and high-level algorithms such as face detection, pedestrian detection, feature matching, and tracking. The library has been downloaded more than three million times. In 2010 a new module that provides GPU acceleration was added to OpenCV.

VII. EXPERIMENTAL RESULTS

Multiple number of experiments have been conducted on different categories of images from the WANG dataset. The different categories like buses, buildings, flowers, elephants, mountains, dinosaurs, beaches, human and horses have been used for testing and analysis. A retrieved image is considered to be correct if and only if it is in the same category as the query. The experiments are carried out in a personal computer with Intel Core i5 processor with 8GB RAM. The program is developed using OpenCV libraries and Visual Studio IDE.

Table 1. Results evaluated after testing

Category Name	Relevant After Feedback	Irrelevant After Feedback	Total Retrieved Images After KNN
Human	12	24	6
Beech	17	37	8
Building	17	30	8
Bus	25	29	12
Dinosaur	13	23	6
Elephant	23	37	11
Flower	8	17	4
Horse	14	15	7
Mountain	25	50	12
Food	13	18	6

VIII. PERFORMANCE EVALUATION

The performance of a retrieval system is evaluated based on several criteria. Some of the commonly used performance measures are average precision, average recall, average retrieval rate. All these parameters are computed using precision and recall values computed for each query image. The precision of the retrieval is defined as the fraction of the retrieved images that are indeed relevant for the query. The recall is the fraction of relevant images that is returned by the query. A good retrieval system should have high values for precision and recall. Different set of images are taken in each experiment. Table 1 illustrates the results retrieved by running the multiple number of experiments. All the categories have been taken for testing purposes. The DWT matches the query image with all the database images and after providing the feedback the relevant and irrelevant images are mentioned in column 5 and 6 of table 1. The final retrieved means top ranked images are mentioned in column 7 of table 1. The final retrieved images can have the positive as well as negative images. The positive and the negative in accordance with the query image are written in the column 1 and 2 of table 2. Table 2 also specifies the precision and recall of this work. The computation time is the time taken by the algorithm for fetching the feature vectors. The matching time is the time taken by the algorithm for matching the query image with all of the database images.

Table 2. Precision and Recall Values

Positive In Final Result	**Negative In Final Result**	**Precision**	**Recall**	**Computation Time**	**Matching Time**
2	4	0.3	0.1	1011	2217
1	7	0.1	0.0	1400	2649
1	7	0.1	0.0	1467	4706
8	4	0.7	0.2	1632	4827
4	2	0.7	0.2	517	1261
7	4	0.6	0.2	1626	2883
2	2	0.5	0.1	301	1501
3	4	0.4	0.1	883	2539
2	10	0.1	0.0	1767	3829
4	2	0.7	0.2	982	2973

In the table 2, we have calculated the precision and recall for the same set of images that are used in table 1. The precision is ranging from 0.1 to 0.7 for different set of categories of WANG database which shows a sign of improvement. The overall execution time is computed in milliseconds and needs to be optimized.

IX. CONCLUSIONS

We have reviewed the main components of a content based image retrieval system by applying interactive KNN with DWT, including image feature representation, indexing, query processing, and query-image matching and user's interaction, while highlighting the current state of the art and the key - challenges. It has been acknowledged that it remains much room for potential improvement in the development of content based image retrieval system due to semantic gap between image similarity outcome and user's perception. Contributions of soft-computing approaches and natural language processing methods are especially required to narrow this gap.

X. REFERENCES

[1]. M. Fakheri, T. Sedghi, M. G. Shayesteh1 and M. C. Amirani, "Framework for image retrieval using machine learning and statistical similarity matching techniques," IET Image Process, pp. 1-11, 2013.

[2]. P. MANIPOONCHELVI and K. MUNEESWARAN, "Multi region based image retrieval system," Indian Academy of Sciences, pp. 333-344, 2014.
[3]. H. J´egou, M. Douze, C. Schmid and P. P´erez, "Aggregating local descriptors into a compact image representation," IEEE, pp. 3304-3311, 2010.
[4]. Y. Chen, J. Z. Wang and R. Krovetz, "CLUE: Cluster-Based Retrieval of Images by Unsupervised Learning," IEEE, pp. 1187-1201, 2005.
[5]. R. Fergus, L. Fei-Fei, P. Perona and A. Zisserman, "Learning Object Categories from Google's Image Search," IEEE, 2005.
[6]. Y. Chen, X. Li, A. Dick and A. v. d. Hengel, "Boosting Object Retrieval with Group Queries," IEEE, pp. 765-768, 2012.
[7]. R. Arandjelovi´c and A. Zisserman, "Three things everyone should know to improve object retrieval," IEEE, pp. 2911-2918, 2012.
[8]. M. Perd'och, Chum and J. Matas, IEEE, pp. 9-16, 2009.
[9]. S. A. Chatzichristofis and Y. S. Boutalis, "CEDD: Color and Edge Directivity Descriptor: A Compact Descriptor for Image Indexing and Retrieval," Springer-Verlag Berlin Heidelberg, pp. 313-322, 2008.
[10]. Y. Chen and J. Z. Wang, "A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval," IEEE, pp. 1252-1267, 2002.
[11]. C.-H. Lin, R.-T. Chen and Y.-K. Chan, "A smart content-based image retrieval system based on color and texture feature," ELSEVIER, p. 658–665, 2009.
[12]. Z. Wang, J. Li and G. Wiederhold, "SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture LIbraries," IEEE, pp. 947-963, 2001.
[13]. S. Gandhani, R. Bhujade and A. Sinhal, "AN IMPROVED AND EFFICIENT IMPLEMENTATION OF CBIR SYSTEM BASED ON COMBINED FEATURES," IET, pp. 353-359.
[14]. S. M. H. Khan, A. Hussain and I. F. T. Alshaikhl, "Comparative study on Content Based Image Retrieval (CBIR)," IEEE, pp. 61-66, 2013.
[15]. Sreedevi S. and Shinto Sebastian, "Fast Image Retrieval with Feature Levels," IEEE, 2013.
[16]. S. Ezekiel, Mark G. Alford, David Ferris and Eric Jones,, "Multi-Scale Decomposition Tool for Content Based Image Retrieval," IEEE, 2013.
[17]. K. Juneja, A. Verma , S. Goel and S. Goel , "A Survey on Recent Image Indexing and Retrieval Techniques for Low-level Feature Extraction in CBIR systems," IEEE, pp. 67-72, 2015.
[18]. K. BELATTAR and S. MOSTEFAI, "CBIR using Relevance Feedback: Comparative Analysis and Major Challenges," IEEE, pp. 317-325, 2013.
[19]. D. Jeyabharathi and A. Suruliandi, "Performance Analysis of Feature Extraction and Classification Techniques in CBIR," IEEE, pp. 1211-1214, 2013.
[20]. H. Xie, Y. Ji and Y. Lu, "An Analogy-Relevance Feedback CBIR Method Using Multiple Features," IEEE, pp. 83-86, 2013.
[21]. B. Kaur and S. Jindal, "An implementation of Feature Extraction over medical images on OPEN CV Environment".
[22]. S. Kumar, S. Jain and T. Zaveri, "ARALLEL APPROACH TO EXPEDITE MORPHOLOGICAL FEATURE EXTRACTION OF REMOTE SENSING IMAGES FOR CBIR SYSTEM," IEEE, pp. 2471-2474, 2014.
[23]. K. BELATTAR and S. MOSTEFAI, "CBIR with RF: which Technique for which Image," IEEE, 2013.
[24]. G. Raghuwanshi and V. Tyagi, "Texture image retrieval using adaptive tetrolet transforms," ELSEVIER, pp. 1-8, 2015.