Retrieval of Images on the Basis of Content: A Survey

[1]Meenakshi Garg, [2]Dr. Manisha Malhotra
[1]Research Scholar, [2]Associate Professor
[1,2]University Institute of Computing, Chandigarh University, Mohali, India

Abstract: With the development of the Internet, and the availability of image capturing devices such as digital cameras, image scanners, the size of digital image collection is increasing rapidly. Efficient image searching, browsing and retrieval tools are required by users from various domains, including remote sensing, fashion, crime prevention, publishing, medicine, architecture, etc. For this purpose, many general purpose image retrieval systems have been developed. In CBIR, images are indexed by their visual content.Content based image retrieval consists of three parts: feature extraction, indexing and retrieval part. The techniques which are used to extract features of an image are called feature extraction techniques. The choice of features plays an important role in image retrieval. Some of the features used are color, texture and shape. Combination of these features provides better performance than single feature. Here we have reviewed some of the work carried out in this field.

I. INTRODUCTION

Commonly, data are shared over the Social networks and other blogging sites on the internet. It mainly consists of multimedia such as images and videos. Apart from this, images have also become a important part of forensic science, medical science, biometrics, etc. Recovery and indexing of this data has become a major issue among many researchers. Content Based Image Classification is a way to classify images into pre-defined set of classes based on their properties. These properties can be based on the colors that make up the image, the edges included in the image or even its temporal features. Feature extraction techniques are used to describe an image in the form of a single feature vector. There are several ways in which features can be extracted from an image. Most commonly used feature extraction techniques include Block Truncation Coding, edge detection, applying transform matrices, etc.

A simple feature extraction techniques based on the Color features of the image is presented in this report. The image is initially divided into 16 equal sized blocks after which the average value of each colour component is calculated. The experiments have been performed in RGB space as well as in LUV space to eliminate impact of color dominance. The extracted features are pre-processed to remove impurities from the data. Various data pre-processing techniques can be used to remove redundancies and imperfections from the data. In this paper, experiments have been performed using Discretization, PKID discretization, Fuzzy fication, Normalization, Numeric to Binary data preprocessing techniques. The preprocessed data is fed to the classifier being used.

Naive Bayesian classifier :

Naïve Bayesian classifier is a simple probabilistic classifier which works by applying the Baye's theorem along with naïve assumptions about feature independence. It assumes value of any feature is independent of values of other features. This assumption is also known as Conditional Independence. Despite the naïve assumption and over simplification, Naïve Bayesian classifiers have proved to be quite useful in complex real world conditions. Naive Bayes classifiers are highly scalable, requiring a number of parameters linear in the number of variables (features/predictors) in a learning problem. It is not a single algorithm for training such classifiers, but a family of algorithms based on a common principle. All naive Bayes classifiers assume that the value of a particular feature is independent of the value of any other feature, given the class variable.

Data pre-processing :

Data pre-processing is a very important step in classification of data. Data pre-processing consists of data cleaning, data integration & transformation, data reduction, normalization, discretization, etc. Discretization pre-processing method discretizes a range of numeric attributes in the given data set into nominal values. When continuous attributes do not follow normal distributions, discretization usually results in a higher classification accuracy. The number of groups (bins) per attribute has to be decided before start of the process.

There are two categories of discretization

1.Supervised .
2.Unsupervised.

Unsupervised discretization ignores class label and where as supervised discretization methods are entropy-based which try to maximize purity of the intervals. This report uses Unsupervised Equal Interval Binning Discretization for all the attributes.

II. LITERATURE SURVEY

Singha, Manimala et al.[1] In this paper, Content Based Image Retrieval has been presented by combining the color and texture features. This proposed approach is termed as Wavelet-Based Color Histogram Image Retrieval (WBCHIR). In terms of accuracy, the results shows that proposed method have perform better as compared to other retrieval methods. Further, steps can be decreased with the use of wavelet transformation.

K. Hemachandran, et al.[2] in this paper, image retrieval technique has been presented. This proposed method is based on the combination of Haar wavelet transformation and lifting scheme which is basically called as lifting wavelet-based color histogram. The color features has translation and rotation invariant. In order to increase the accuracy of the system, the wavelet transformation is used to extract the texture features. And to reduce the processing time, the lifting scheme has been used.

Singh, Sandeep et al. [3] in this paper content based image retrieval system has been presented based on the image recovery system. This novel approach provides precise results among other developed system. Image recovery system is used to estimate the image similarity of every image only in terms of visual features. And then it will return the image with similarity, the feature weight based on the neural network is used to evaluate the efficient feature extraction. In this study, feed forward algorithm has been applied for neural network.

Kushwah, Vinita et al. [4] in this paper an overview of the functionality of content based image retrieval systems. Combining advantages of query based clustering technique for content based image retrieval to find similar and dissimilar image group and data. This system defines automatic retrieval of images from a database, based on the colors and shapes present. In order to evaluate the similar and dissimilar image data, content based image retrieval has been used.

Patel, dileshwer et al. [5] proposed an approach for image retrieval from very large image database. In last decades, various research works has been developed and much technique has been proposed but most of the approaches deals with the accuracy problem. In this paper, novel technique has been presented which uses histogram and color edge of an image with wavelet transform. The proposed algorithm is calculated with two parameters i.e. is precision and images of Wang database.

Mohamadzadeh, Sajad et al. [6] a novel technique depend on the sparse representation and iterative discrete wavelet transform has been presented in this paper. In order to check the applicability of the proposed method, accuracy and average normalized modified retrieval rank has been act as quantitative metrics.in comparison with other approaches, proposed technique gives better performance.

Yue, Jun, Zhenbo Liet al. [7] proposed a novel technique based on CBIR to extract texture feature and color of an image. Initially, a feature vector was formed by extracting the co-occurrence matrix which is relying on color and texture feature. After feature vector was formed, characteristics of various histograms such as local and global color histogram and texture features has been analyzed and compared for content based retrieval image. Depending on the steps, proposed method has been developed by constructing weights of feature vectors. The performance result demonstrated that the fused features retrieval provides better retrieval results.

Murala, Subrahmanyam, R. P. Maheshwariet al. [8] presented a novel techniques using directional local extrema patterns meant for content-based image retrieval application. By comparing gray levels, the LBP makes a relation between reference pixel and its surrounding neighbors. The proposed method can able to extract the directional edge information in 0◦, 45◦, 90◦, and 135◦ directions in an image. The result is compared with existing approaches such as local edge patterns for segmentation, local edge patterns for image retrieval, center-symmetric local binary pattern, block-based LBP. The results after being investigated show a significant improvement in terms of their evaluation measures as compared with other existing methods on respective databases.

Hiremath, P. S., and JagadeeshPujari et al. [9] in this paper, CBIR method has been proposed using color, texture and shape information to attain high efficiency. In this work image may be divided into two equal parts as serve as local descriptors of color and texture. A new method of matching scheme has been presented based on the most similar highest priority (MSHP) principle and the adjacency matrix of a bipartite graph formed using the tiles of query and target image, is provided for matching the images. The combination of the color, texture and shape features provide a robust feature set for image retrieval.

III. METHODOLOGY

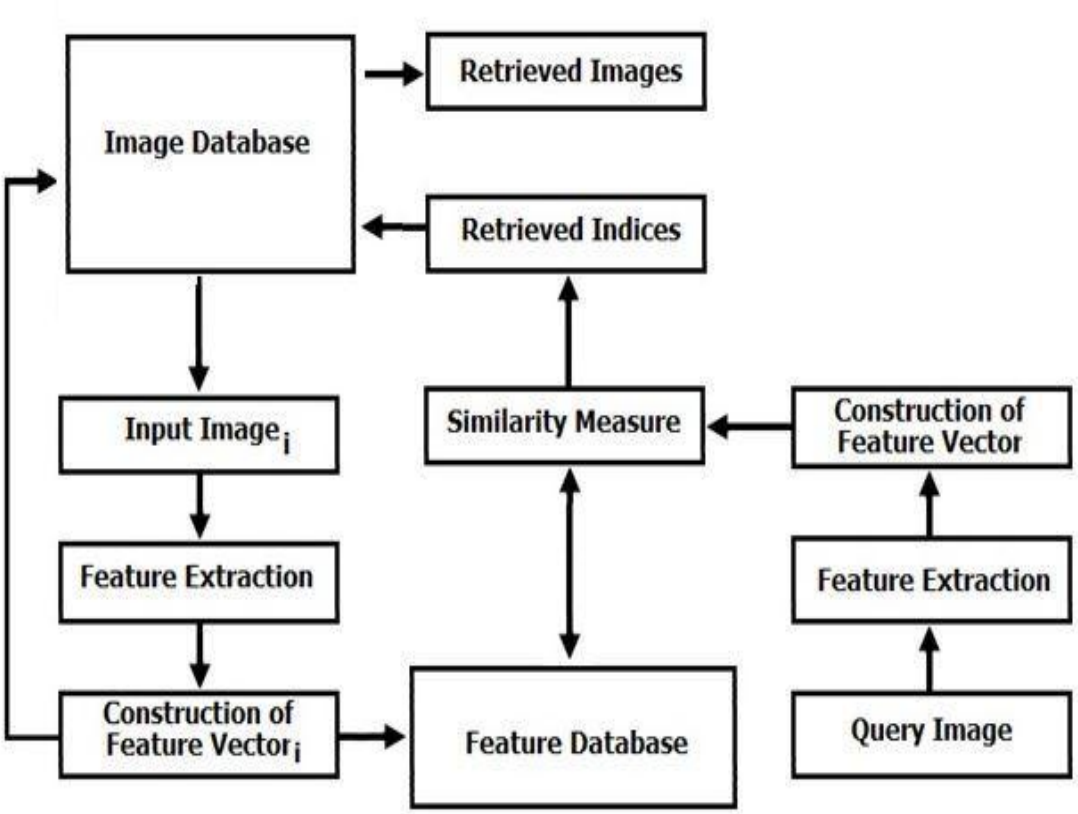

Figure.1 Content-based image retrieval System

Take the image input colour RGB and decompose into its three colour components, i. e. Red(R), Green (G) and Blue B) respectively where each colour component has L intensity levels. For each colour component, compute the probability histogram as follows

$$p(r_i) = \text{Number of Pixels in } r_i / \text{Width} \times \text{height} \qquad V_i \ \varepsilon\ [0, L-1]$$

Where $p(r_i)$ represents the relative frequency or probability of r_i. -th intensity value and the range of intensity value is [0, L -1] .

Compute for each colour component, the standard deviation, skewness and kurtosis from $\text{Bin}_c V$ i ε [0, n -1] where 'C' represents R, G and B colour components respectively.

Construct feature vectors for the query image and the database images using the feature extraction technique. Compute the distance between feature vector of the query image and the feature vectors of the database images for the similarity

measurement. Sort the distance in non-decreasing order and select top N images having minim distances.

IV. CONCLUSION

From extensive literature survey we found that most of the current face recognition system uses structural matching methods to detect structures more reliably e.g. Active shape model. Appearance based methods like PCA, ICA, Gabor-wavelets are used which require small number of features and recognition rate is also high. But these methods require good quality images, large size of database (need training data to learn the classifiers and collection of large amount of data is very expensive) and results are affected by illumination conditions. Whereas, Statistical features are fast to implement and are effective where fast retrieval is required.

REFERENCES

[1] Singha, Manimala, and K. Hemachandran. "Content based image retrieval using color and texture." *Signal & Image Processing: An International Journal (SIPIJ)* 3, no. 1 (2012): 39-57.
[2] Singha, M., K. Hemachandran, and A. Paul. "Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram." *Image Processing, IET* 6, no. 9 (2012): 1221-1226.
[3] Singh, Sandeep, and ErRachna Rajput. "Content Based Image Retrieval using SVM, NN and KNN Classification." *vectors* 4, no. 6 (2015).
[4] Kushwah, Vinita, and ArunAgrawal. "Study on Query Based Clustering Technique for Content Based Image Retrieval."
[5] Patel, Dileshwar, and AmitYerpude. "Content Based Image Retrieval using Color Edge Detection and Haar Wavelet Transform." (2015).
[6] Mohamadzadeh, Sajad, and Hassan Farsi. "Content-based image retrieval system via sparse representation." *IET Computer Vision* 10, no. 1 (2016): 95-102.
[7] Yue, Jun, Zhenbo Li, Lu Liu, and Zetian Fu. "Content-based image retrieval using color and texture fused features." *Mathematical and Computer Modelling* 54, no. 3 (2011): 1121-1127.
[8] Murala, Subrahmanyam, R. P. Maheshwari, and R. Balasubramanian. "Directional local extrema patterns: a new descriptor for content based image retrieval." *International Journal of Multimedia Information Retrieval* 1, no. 3 (2012): 191-203.
[9] Hiremath, P. S., and JagadeeshPujari. "Content based image retrieval using color, texture and shape features." In *Advanced Computing and Communications, 2007.ADCOM 2007. International Conference on*, pp. 780-784. IEEE, 2007.