



Content Based Image Retrieval of Corel Images Along With Face Recognition

Deepa Joseph¹, Ajai Mathew²

¹Student, Department of Electronics and Communication Engineering
Amal Jyothi College of Engineering, Kanjirappally, India

²Assistant Professor, Department of Electronics and Communication Engineering
Amal Jyothi College of Engineering, Kanjirappally, India

¹deepajoseph@ec.ajce.in; ²ajaimathew@amaljyothi.ac.in

Abstract—: *Image Retrieval basically deals with identification of similar images from a large database. Image retrieval based on rich content of the image is known as Content Based Image Retrieval (CBIR). The visual content of an image is analysed in terms of different features extracted from the image. The efficiency of CBIR techniques depends on the database selected. The database considered is Corel database. Content Based Image Retrieval based on color, shape and texture alone fails in face recognition. To analyse the entire Corel database, Principal Component Analysis (PCA) is introduced in the system. PCA works best in situations where relationship among pixels is linear. Thus face recognition is done with PCA which maintains data reduction. The idea of PCA fails in situations where linearity among pixels does not hold. This means that PCA alone cannot be used for image retrieval. To analyse the entire Corel database, PCA is incorporated in to the basic feature extraction methods. The proposed system reduces computation cost and time. The accuracy of the retrieval system is increased.*

Keywords— *Face recognition, Corel database, PCA, Basic feature extraction*

I. INTRODUCTION

The retrieval of an image from a large database has turn into an important task due to frequent manipulation of digital images. The application of image retrieval includes both judiciary and fashion technology. The challenge in image retrieval lies in similarity between the query image and the recognized image. The optimal solution is achieved when the computational cost and time is minimized and highest retrieval performance is obtained. Content Based Image Retrieval (CBIR) has already proven over concept based image retrieval (text based) in the last decade. Now the research mainly focuses on refining the CBIR[1] system. The term ‘Content based’ pointers to the analysis of an image based on its contents other than using metadata which includes keywords, tags or descriptions. CBIR extracts the rich content of an image in terms of its features. The basic feature extraction includes color, shape, and texture extraction.

Now the CBIR system looks into user friendly interfaces. The CBIR systems are designed and trained to extract the different features of an image automatically. CBIR based on basic feature extraction turns to be irrelevant in domain specific image retrieval tasks. Here the gap between query image and recognized image is much greater. In order to perform domain specific image retrieval, domain specific features are to be considered. The domain selected here is face recognition. The proposed system deals with Corel images focusing on face

recognition. The algorithm implements Principal Component Analysis (PCA) along with extraction of color, shape and texture. PCA deals with face recognition where, color, texture and shape features are used for basic image matching. The distance between images is calculated using Euclidean distance. The algorithm works smoother without much overhead. i.e., the computational time is reduced.

II. CONTENT BASED IMAGE RETRIEVAL

Image retrieval is the task of searching for similarity in images from an image database. An image can be described in terms of its visual contents such as color, shape, texture and spatial layout. Content Based Image Retrieval (CBIR) indexes an image in terms of its visual contents which is extracted and described by feature vectors. The available feature vectors constitute a feature database. In order to carry out the image retrieval, user provides a query image. The query image is analysed and described as a feature vector. The distance between feature vectors of query image and database images are calculated and sorted out based on minimum distance. The block diagram of CBIR system is shown in Fig 1. The selected database for image retrieval is the Corel database. It is widely used in the field of CBIR..

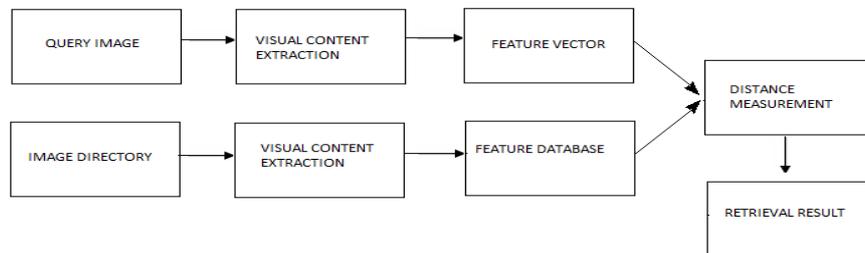


Fig 1. Block diagram of Content Based Image Retrieval System

A. Visual Content Description

The basic visual content of an image include color, shape, texture and spatial relations. Visual content also include domain specific visual content like face recognition. Visual content is generally invariant to image alterations. The following are the parameters used in the proposed system for the extraction of visual content.

HSV Histogram

Hue is invariant to illumination and camera direction. Thus HSV histogram shown in Fig 2 is used to identify such variations in an image.

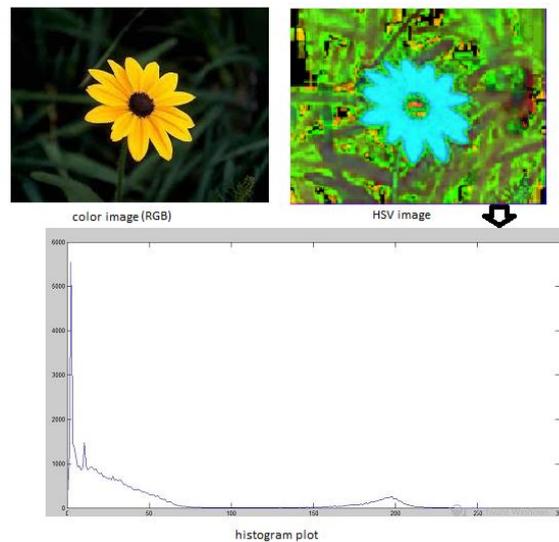


Fig.2. HSV histogram

Color Moment

Color is an important feature of images. It is invariant to image scaling, translation, and rotation. Color space widely used are RGB (Red Green and Blue) and HSV (Hue Saturation and Value). Color moments [6] are measures that characterize color distribution in an image in the same way that central moments uniquely describe a probability distribution. The first order (mean), the second (variance) and the third order (skewness) color moments are effective in representing color distributions of images. Each comparison between images results in a similarity score, and the lower this score is the more identical the two images.

Color Autocorrelogram

A color autocorrelogram gives the spatial information of pixels in an image. This improves the information of color content.

Canny Edge Detection

Shape[3] can generally be defined as the description of an object regardless of its position, orientation, and size. The edges are detected through canny edge detection as shown in Fig 3.

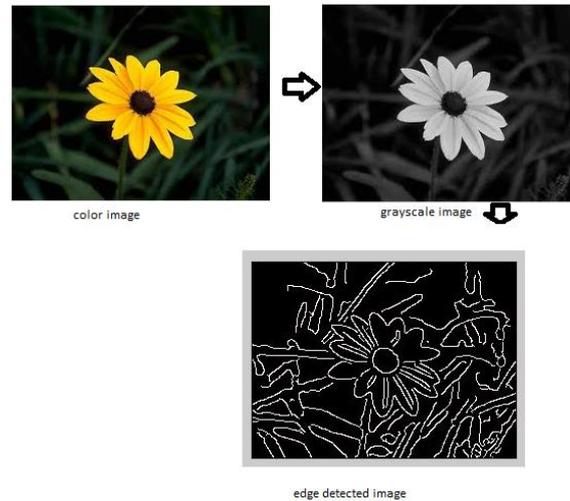


Fig.3. Canny edge detection

Wavelet Transform

Texture [2] is a description of the spatial arrangement of color or intensities in an image. Texture refers to visual patterns with properties of homogeneity. Texture features typically consist of contrast, uniformity, coarseness, and density. Image patterns can be recognized by texture content through structural and statistical observations. The statistical analysis includes wavelet transform. It transforms the image into a representation with both spatial and frequency characteristics.

Domain Specific Visual Content

Principal component [3] of images is the domain specific visual content associated with face recognition. Principal Component Analysis (PCA) is a statistical procedure concerned with elucidating the covariance structure of a set of variables. In particular it allows identifying the principal directions in which the data varies. In computational terms the principal components are found by calculating the Eigen vectors and Eigen values of the data covariance matrix.



Fig.4. Principal components of database

PCA algorithm [5] is applied to an image database after re sizing each image into equal size. Consider an image of $N \times N$ size. The image is converted to vector of size $N^2 \times 1$. Each image in the database is replaced by the corresponding image vector. Thus created database matrix where each column corresponds to the image

vector. The data reduction is obtained by removing the redundant data from the database matrix. This is done by subtracting the mean of the database matrix from the original database matrix. The covariance matrix of the reduced database matrix is calculated in the obtained lower order.

The Eigen vectors and corresponding Eigen values are obtained from the covariance matrix. The set of Eigen vectors whose Eigen values are greater than one, are filtered out and used to create the basis of database matrix. The basis of database matrix is formed by multiplying the reduced database matrix with the set of Eigen vectors filtered out. The query image is projected to the basis to compute the Euclidean distance. The database with principal components alone is shown in Fig.4. Probably the most common approach to compare images directly is the Euclidean distance. To be able to compare images using a Euclidean distance, the images have to be of the same size which can be achieved easily with scaling algorithms.

III. SYSTEM DESIGN

A. CBIR System

The proposed system is shown in Fig. 5. The system consists of three stages. First stage includes creation of database from the selected directory of images. Second stage consists of providing the query image. Third stage performs image retrieval.

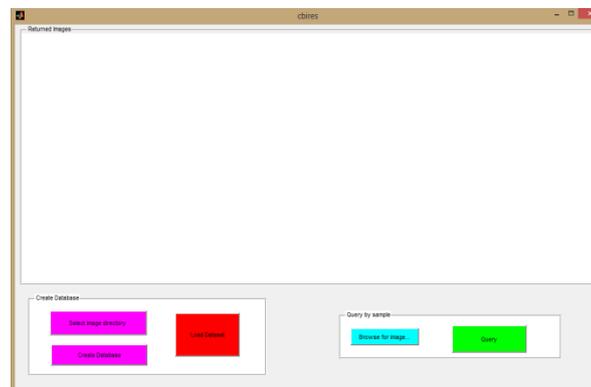


Fig.5. CBIR system

Creation of database begins with selecting the directory of images. Thus the database created is saved in the user defined folder in the specified name. This indicates the successful completion of first stage in the CBIR system. The created database is stored and can be utilized for future retrieval operations. Thus database needs to be created for a single time for the same set of images. The second stage of CBIR system deals with query image. Query image is given as the user defined input. The final stage of CBIR system is image retrieval. This is performed by clicking on the Query button. The returned images are obtained in allocated space.

A. Algorithm for image retrieval

Consider the color image $I_i \in D$ where D is the directory of images. The image I is processed in such a manner to generate feature vector F_i . The feature vector F is given by

$$F_i = [a_1 \ a_2 \ a_3 \ a_4 \ a_5 \ a_6] \quad (1)$$

where $1 \leq i \leq N$. N is the total number of images. a_1 is the HSV histogram value. a_2 is the color moment value. a_3 is the color autocorrelogram value. a_4 is the canny edge detector value. a_5 is the wavelet transform values and a_6 is the PCA value. All F_i s constitute the dataset.

Consider I_0 is the query image. The Euclidean distance E is calculated as

$$E = \sqrt{F_{I_i}^2 - F_{I_0}^2} \quad (2)$$

where F represent the corresponding feature vector. The images are sorted out based on the minimum E. Thus the results are obtained and displayed by the system. The flow chart of the algorithm is shown in Fig. 6.

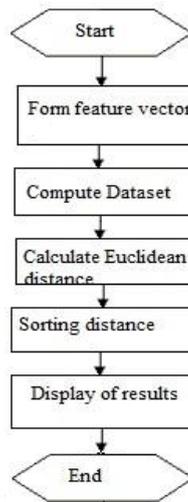


Fig.6. Flow chart of the algorithm

IV. OBSERVATIONS AND RESULTS

The major problem in performance evaluation of content-based image retrieval systems is that neither a standard test database nor a standard performance measure is available. Thus in early reports of content-based image retrieval systems, the results are often restricted to the presentation of retrieval results of one or more example queries, which is easily used to give a positive impression of the abilities of a system.

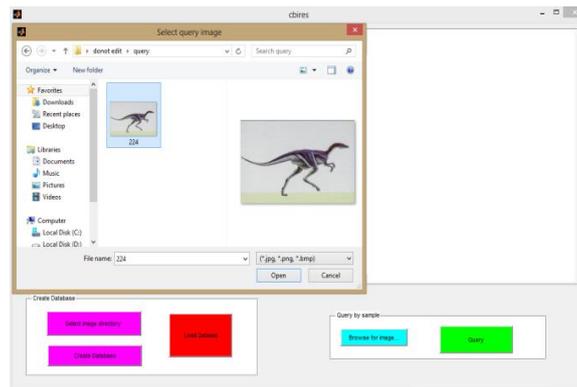


Fig 7: Query example 1

A typical Corel image is given to the system as input as shown in Fig. 7. Color texture and shape features matters in this kind of image retrieval. The returned images shown in Fig. 8. are obtained with less computational time and better accuracy.

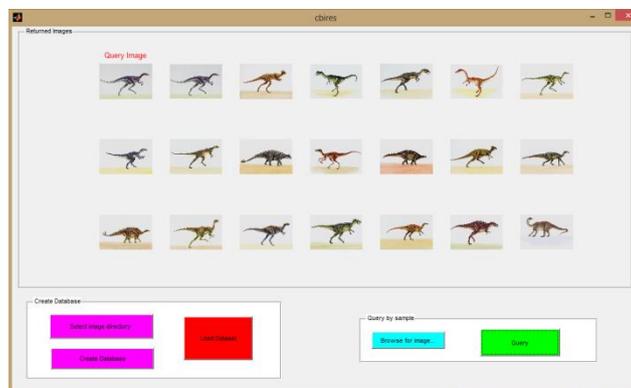


Fig. 8: Recognized images based on the proposed system for query example 1

The proposed system is experimented with facial images. Principal component is the key factor in face recognition. Facial image is given as the user defined input as shown in Fig.9. The returned images are shown in Fig.10.

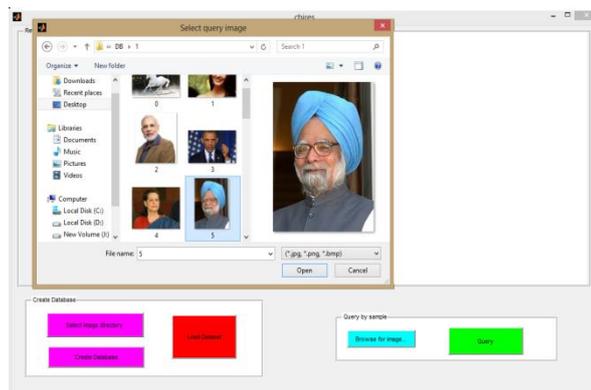


Fig.9. Query example 2

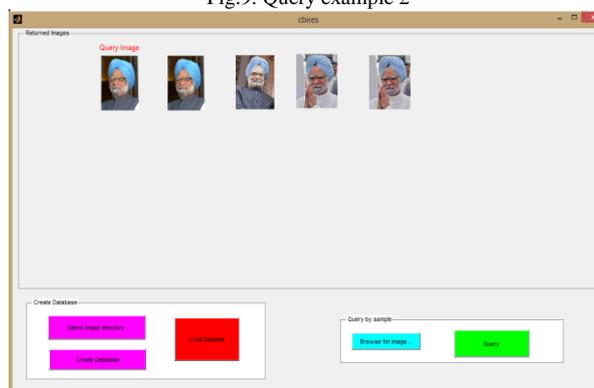


Fig.10. Recognized images based on the proposed system for query example 2

Thus the proposed system has proven accuracy over entire Corel database along with face recognition. The system performs with less computational time.

V. CONCLUSION

Content Based Image Retrieval (CBIR) system for the entire Corel database is proposed. The proposed method incorporates basic feature extraction with PCA. The basic feature extraction which includes color, shape and texture analysis aids in the recognition of basic elements in the image where pixels hold nonlinear relation. Color moments, Color autocorrelogram and HSV histogram methods are used for color feature extraction, canny edge detection method is used for shape feature extraction and Wavelet transform is used for texture feature extraction. Facial elements hold linear relation among pixels. Thus PCA is used for face recognition. PCA does not add much computation due to the data reduction techniques adopted. Thus the performance of the proposed system in the analysis of entire Corel database along with face recognition is much greater. The Content Based Image Retrieval (CBIR) system for the entire Corel database is proposed. The proposed method incorporates basic feature extraction with PCA. The basic feature extraction which includes color, shape and texture analysis aids in the recognition of basic elements in the image where pixels hold nonlinear relation. Color moments, Color autocorrelogram and HSV histogram methods are used for color feature extraction, canny edge detection method is used for shape feature extraction and Wavelet transform is used for texture feature extraction. Facial elements hold linear relation among pixels. Thus PCA is used for face recognition. PCA does not add much computation due to the data reduction techniques adopted. Thus the performance of the proposed system in the analysis of entire Corel database along with face recognition is much greater.

REFERENCES

- [1] Prof. C. S. Gode, Ms. A. N. Ganar, "Image Retrieval by Using Colour, Texture and Shape Features", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering Vol. 3, Issue 4, April 2014.
- [2] Jun Yue, Zhenbo Li, Lu Liu, Zetian Fu., "Content Based Image Retrieval using Color and Texture Fused Features", 1121-1127, 2011, Elsevier.
- [3] Muzammil Abdulrahman, Tajuddeen R. Gwadabe, Fahad J. Abdu, Alaa Eleyan, "Gabor Wavelet Transform Based Facial Expression Recognition using PCA and LBP", IEEE 22nd Signal Processing and Communications Applications Conference 2014.

- [4] MShadma Parveen, Shweta Yadav, and Neelu Chauhan "Sketch4Match - Content-Based Image Retrieval System Using Sketches", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, April 2013.
- [5] R. C. Thomaz and G. Giraldi, "A New Ranking Method for Principal Components Analysis and Its Application to Face Image Analysis", Image and Vision Computing, vol. 28, no. 6, pp. 902-913, 2010.
- [6] Hi-Chun Huang, Patrick P. K. Chan, Wing W. Y. Ng, Daniel S. Yeung, "Content-Based Image Retrieval Using Color Moment and Gabor Texture Feature", Proceedings of the Ninth International Conference on Machine Learning and Cybernetics, July 2010.
- [7] Swati Agarwal, A. K. Verma, "Content Based Image Retrieval using Color Edge Detection and Discrete Wavelet Transform", IEEE International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014