www.ijecs.in
International Journal Of Engineering And Computer Science ISSN: 2319-7242
Volume 5 Issue 1 January 2016, Page No. 15457-15462

Efficient Content Based Image Retrieval Using Combination Of Dominant-Color, Shape And Texture Features And K-Means Clustering

Bhagyashri Amrutkar[1], Lokesh Singh[2]

[1]Department of Computer Science & Engineering , RGPV University, Bhopal, India.
amrutkar.bhagyashri@gmail.com
&
[2]Department of Computer Science & Engineering , RGPV University, Bhopal, India.
lokesingh@gmail.com

Abstract: There is a huge demand for the efficient content based image retrieval system because of the availability of large image databases. In this paper we have present an efficient CBIR framework by extracting the Dominant-color, Texture, edge features and by clustering feature database. We have applied the dominant color extraction using color-quantization technique. Initially the image is divided into some partitions using the color quantization algorithm, here we are dividing into eight partitions and the eight dominant colors are obtained from that partition. Next for shape feature extraction sobel color edge detection technique is used. And local binary pattern (LBP) is performed on gray scale image to extract the texture feature. Then all features discussed above of image are combined to form a single feature vector. K-means clustering is applied over combined feature vector of database images. Finally, to retrieve similar images from database similarity matching is performed by Euclidian distance which compares feature vector of clustered database images with that of query image. The result of this proposed approach provides efficient, more accurate result.

Keywords: Dominant color, K-means clustering, Local Binary Pattern (LBP), Euclidian Distances.

1. Introduction

An image may be defined as a two dimensional function,
f (x,y), where x and y are spatial(plane) coordinates, and the amplitude of f at any pair of coordinates (x,y) is called the intensity or gray level of the image at that point. When x, y, and the intensity values of f are all finite, discrete quantities, we call the image a digital image [17].

Image retrieval system is a computer system for browsing, searching and retrieving images from a large database of digital libraries. The use of Metadata such as captioning, keywords or descriptions to the images stored in the database along with the images or the low level feature extracted from the image like shape, color, texture etc have been used till now for the image retrieval from the existing search engine. A user formulating a query usually has in mind just one topic, while the results produced to satisfy this query may belong to different topics. Therefore only parts of the search results are relevant for a user [17].

In this age of technology, for all intents and purposes all circles of human life including trade, government, scholastics, healing facilities, wrongdoing aversion, reconnaissance, designing, structural engineering, news-casting, style and visual computerization, and authentic examination use pictures for effective administrations. An expansive gathering of pictures is alluded to as picture database. A picture database is a framework where picture information are coordinated and put away [2].Image information incorporate the crude pictures and data separated from pictures via computerized or PC helped picture investigation.

Content based Image Retrieval (CBIR) techniques are also working with the same lay [2].A component

vector is removed from each photo in the database and the course of action of all highlight vectors is formed as a database record . At request time, a part vector is removed from the inquiry picture furthermore; it is facilitated against the component vectors in the record . The key difference between the distinctive systems lies in the parts that they remove and in the counts that are utilized to take a gander at highlight vectors [1-7].

Due to the characteristics of invariant to image size and orientation, Color is one of the most widely used low level features for CBIR. Various color models are available for image retrieval which is RGB, HSV, and CMY etc. The RGB model uses three primary colors, red, green and blue, which are used in different combinations to reproduce other colors. This model has the advantage of being easy to extract. A true-color RGB image's bit depth is 24 bit that means each pixel will have a red, green and blue value ranging from 0 to 255. Various techniques are available to extract the color feature. Coherence vector, color histogram, color correlogram are the main techniques [1, 3, 4, 5].

Texture is the natural property of all surfaces, which describes visual patterns. Like colors in the image, the textural characteristics are also effective ways of describing visual content. Texture features have also been widely used in CBIR applications. This is a feature that describes the distinctive physical composition of a surface[20].Various methods are available which are used to describe texture feature such as gray-level cooccurrence matrix and Haar discrete wavelet transform etc[1,7].

Shape description is an important task in content-based image retrieval. It is important in CBIR because it corresponds to region of interests in images [21].Various methods to extract shape feature are edge histogram descriptor (EHD) and Sobel descriptor etc[2,5].

In addition to these various techniques of feature retrieval of image, clustering techniques also used in content based image retrieval on the retrieved feature vector of database images. Clustering means dividing given image feature vectors into clusters so that feature vector in same cluster are most similar.The advantage of application of clustering techniques in CBIR is that similar image retrieval process becomes less time consuming because query image feature vector is compared with the feature vector of center of clusters only.

2. Literature Review

In 2010, Jun Yue, Zhenbo Li , Lu Liu , Zetian Fu,. [1] Propose a method for content-based image retrieval (CBIR) by extracting color and texture features of an image quickly. Here RGB color space is converted to HSV color space, which is then quantified rationally. Color histogram technique is used for color feature extraction and texture features retrieved by a gray-level co-occurrence matrix. These features are fused by constructing weights of feature vectors. The experiments of image retrieval show that the fused features retrieval brings better visual feeling than the single feature retrieval.

In 2012, Neha et al. [7] introduced a co-histogram based picture recovery which is the blend of shading, composition and shape. They have utilized Wang database for result examination. They have contrasted their strategy and the past calculation and observed to be better regarding exactness and review.

Clustering means dividing given image feature vectors into clusters so that feature vector in same cluster are most similar [19] . In 2013,Umesh Kumar presented a impact of clustering in CBIR [8].Here in this thesis work, before searching similar images, clusters are computed based on color similarity then some low level features as texture, shape and spatial of query image are extracted which are compared to only clustered images features only. The result of applying clustering is improved speed with accurate result. Here hierarchical clustering is used to form clusters of similar color features for fast retrieval and then k-means clustering is applied on those clusters for more accurate image retrieval .

In 2012, Nishant et al. [15] proposed another picture recovery method. It recovers comparative pictures in diverse stages. The photos are at first recuperated in perspective of their shading segment similarity. The significance of the recouped pictures is then further improved by planning their surface and shape highlights independently. Generally CBIR contrast request pictures highlight vector and each and every other picture in the database. This decreases the precision of the system as the interest incorporates the whole database which contains a wide grouping of pictures. Moreover accomplishment of shape

develop CBIR depends as for exactness of Segmentation framework used. Deplorably it has been exhibited that correct division is still an open issue. Present procedure wipes out the dependence over precise division framework to some degree by narrowing down the chase range at each stage.
In 2013, Ramesh Kumar proposed a Content-based image retrieval (CBIR) using color histogram technique [14]. And a color histogram is a representation of the distribution of colors in an image. For digital images, a color histogram represents the number of pixels that have colors in each of a fixed list of color ranges, that span the image's color space, the set of all possible colors.

In 2014, Choudhary et al. [16] proposed a substance based picture recuperation composed technique which isolates both the (shading) and surface (composition) segment. To partition the shading segment, shading minute (CM) is used on shading pictures and to remove the surface component, neighborhood twofold illustration (LBP) is performed on the Grayscale picture. By then both shading and organization highlight of picture are joined to shape a singular part vector. Finally likeness organizing is performed by Euclidian division which differentiations highlight vector of database pictures and question pictures. LBP essentially used for face affirmation. Regardless, we are going to use LBP for basic pictures. This joined strategy gives definite, profitable, less bewildering recuperation structure.

3. Proposed Method

Proposed method uses low level features of image such as dominant color, Texture, shape for more accurate image retrieval result, and preprocessing of the image feature database using K-means Clustering for efficient image retrieval system.

Color quantization algorithm is used to extract the dominant colors from the image. Actually, in a given color image, the number of actual colors only occupies a small proportion of the total number of colors in the whole color space, and some dominant colors covers a majority of pixels, So it will not influence the understanding of image content though the quality of image is reduced if we use the dominant colors to represent image and which helps for efficient image retrieval on similarity matching. The RGB color space is used. First, the RGB color space is divided equally into 8 partitions. And for each block the average value of color say Xi find out. Then Xi(R, G, B) is considered as Dominant color. Thus we calculate total 8 Dominant colors for a image.

Here the Local Binary Pattern is applied for texture feature extraction. First RGB color space image is converted into gray scale image. Then gray scale image is subdivided into 3by 3 blocks. gray level pixel values is calculated for every block. from that LBP value is calculated. From LBP value LBP image is obtained and LBP histogram is calculated. The histogram of LBP features extracted from each block is concatenated into single histogram. This extracted feature histogram represents the texture of image.

For shape feature extraction, improved color edge detection approach used here. Color-edge detection in RGB space means identifying edges in an RGB image which are clearly perspective to human viewer. First RGB image is converted into three separate R, G, and B component image. Then 4 gradient vector are used to identify horizontal, vertical, diagonal, antidiagonal edges for each pixel in each component image. The property of Gradient vector is it points in the direction of the maximum rate of change of f at coordinates (x, y).Out of above 4 mentioned gradient vector, the gradient is having maximum value is taken as dominant gradient for each pixel. And these maximal gradients identify edge pixels. This technique is repeated for each component color image and result is combined.

Then all features discussed above of image are combined to form a single feature vector. K-means clustering is applied over combined feature vector of database image. Clustering causes efficient similar image retrieval since similarity measure is applied between combined feature vector of clustered database image and that of query image.

For similitude examination, we have utilized Euclidean separation, d utilizing mathematical statement

$$d = \sqrt{\left(F_Q[i] - F_{DB}[i]\right)^2} \qquad (1)$$

3.1 Algorithm

Input: Query Image

Output: Most Similar Images

Step 1: Input image is processed.

Step 2: Apply Dominant color feature extraction using uniform color quantization algorithm.

Step 3: Apply Texture feature extraction using Local Binary Pattern (LBP) technique.

Step 4: Apply Edge feature extraction using improved color-edge detection method.

Step 5: Combine all feature vectors of image to form a single feature vector.

Step 6: Similarity measure is then found by Euclidean distance.

$$\sqrt{\sum_{i=1}^{N}(xi - yi)^2}$$

It is applied on combined feature vector of query image and that of clustered feature database images.

Step 7: Topmost 20 similar images are retrieved.

3.2 Architecture

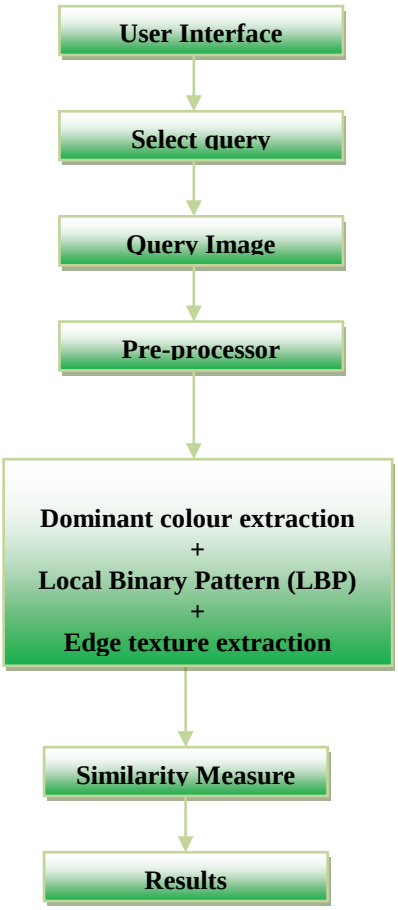


Figure 1: Architecture

4. Result Analysis

We have considered the famous database created by Wang .It is a collection of 1000 images. There are 10 classes in the database. We consider every class one by one for examination. Following is the demonstration of Result images on a image query of different classes. When we apply our calculation in the above database, then we watch that our outcome is better in correlation to the conventional methodology.

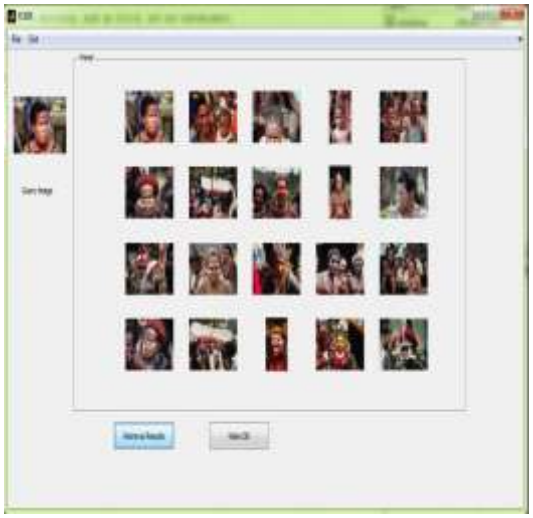
Assessment of a data recovery framework for the most part spotlights on how important are the recovered results (exactness/precison)

Precion can be calculated by the following formula:

$$\text{Precision}=\frac{Tp}{Tp+Fp} \quad (2)$$

Where Tp = True Positive
Fp= False Positive

A decent approach to describe the execution of a classifier is to take a gander at how accuracy and review change as you change the limit. A decent classifier will be great at positioning genuine plane pictures close to the highest priority on the rundown, and have the capacity to recover a considerable measure of plane pictures before recovering any geese: its exactness will stay high as review increments. A poor classifier will need to take an extensive hit in accuracy to get higher review. As a rule, a production will introduce an accuracy review bend to demonstrate how this tradeoff searches for their classifier. Figure 2 to 5 are the results obtained by our methodology.



DOI: 10.18535/Ijecs/v5i1.10

Figure 2: Retrieval Result images on a query of African man

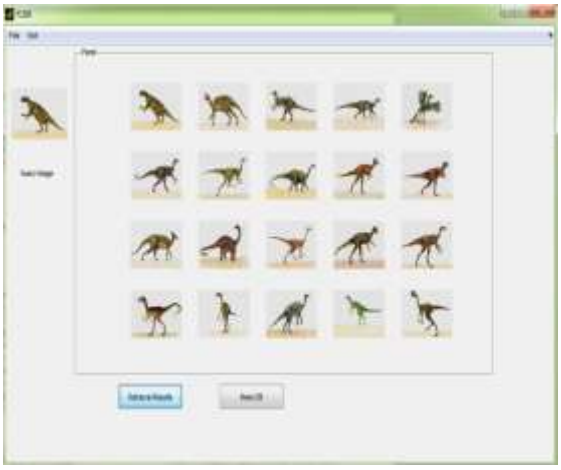


Figure 3: Retrieval Result images on a query of Dinosaur

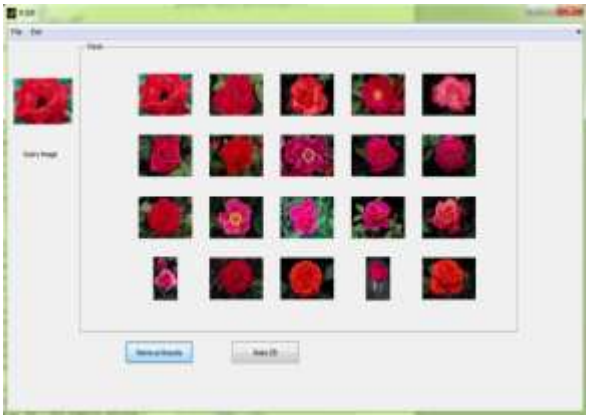


Figure 4: Retrieval Result images on a query of Rose

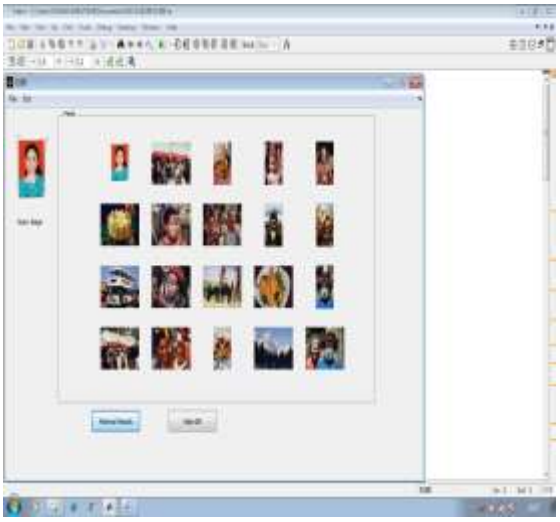


Figure 5: Retrieval Result images on a query of man

The accuracy compariosn is shown in Table 1 , Figure 6 and 7 which shows the significant improvement from the previous approach.

Table 1: Comparison Table

S.NO	*Category*	*Proposed Approach*	*Previous Approach[16]*
1	Africa	1	NA
2	Beaches	0.8	0.5
3	Buildings	0.8	0.7
4	Buses	1	0.98
5	Dianosaurs	1	1
6	Elephants	0.6	0.6
7	Flowers	1	0.89
8	Horses	0.95	0.8
9	Mountains	0.9	0.7
10	Food	0.55	0.6
Overall accuracy		0.86	0.75

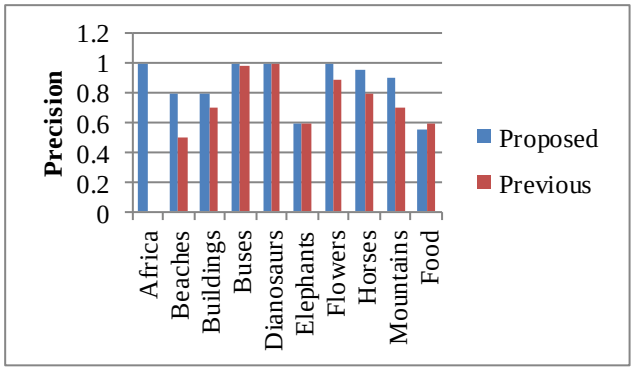


Figure 6: Precision Comparison

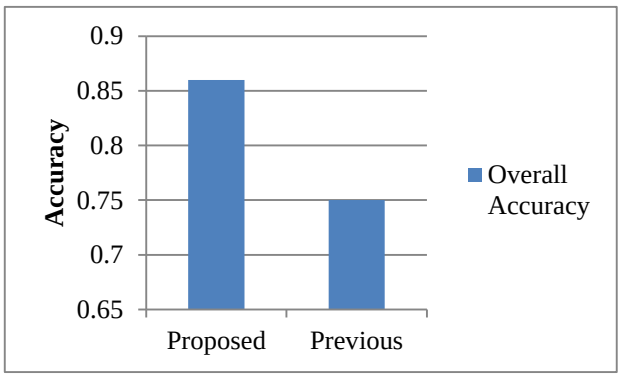


Figure 7: Accuracy Comparison

5. Conclusion and Future Work

In this paper we have presented an efficient and more accurate CBIR system .More Accuracy is achieved by extracting the three features Dominant Color, Shape and Texture and Efficiency is achieved by utilizing K-means clustering on featured database. It can accomplish higher recovery effectiveness due to use of dominant color features and K-means clustering of feature database and edge texture feature.

In future, to reduce the effect of image background on image matching decision, dominant color extraction on weighted basis can be implemented. Also in future, the implementation of clustered as well as indexed featured database can be possible so that image retrieval will be more efficient.

References

[1] Jun Yue ,Zhenbo Li , Lu Liu , Zetian Fu, "Content-based image retrieval using color and texture-fused feature," www.elsevier.com/locate/mcm, Mathematical and Computer Modelling 54,pp. 1121–1127,2010.

[2] Swapnalini Pattanaik, Prof.D.G.Bhalke, "Beginners to Content Based Image Retrieval", International Journal of Scientific Research Engineering &Technology (IJSRET), Volume 1, Issue2, pp .040-044, May 2012.

[3] Reshma Chaudhari, A. M. Patil, " Content Based Image Retrieval Using Color and Shape Features", International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering), Vol. 1, Issue 5, November 2012.

[4] YShankar M. Patil, "Content Based Image Retrieval Using Color, Texture & Shape", International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 3, No. 9, Sep 2012.

[5]Poulami Haldar, Joydeep Mukherjee, "Content based Image Retrieval using Histogram, Color and Edge", International Journal of Computer Applications (0975 – 888),Volume 48,No.11, pp.25-31, June 2012.

[6] S.Pavani , Shivani, T.Venkat Narayana Rao , Deva Shekar ,"Similarity Analysis Of Images Using Content Based Image Retrieval System" ,International Journal Of Engineering And Computer Science , Volume 2 ,Issue 1 , pp. 251-258,Jan 2013.

[7] Neha Jain,Sumit Jain,and Ravi Mohan Sairam, "Content-based Image Retrieval using Combination of Colour, Shape, and Texture features" , International Journal of Computer Research,Vol.3, No.1, Issue 8, pp.70-76, March 2013.

[8] Priyanka Gupta, Umesh Kumar and Ankur Chawla, "Impact of clustering in image retrieval techniques", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 3, Issue 4, pp.322-324, April 2013.

[9] Kiranjeet Kaur, Sheenam Malhotra, "A Survey on Edge Detection Using Different Techniques",International Journal Of Application or Innovation in Engineering and Management (IJAIEM), Volume 2, Issue 4, pp.496-500, April 2013.

[10] G.Ohashi and Y.Shimodaira "Edge Based Feature Extraction Method and it's Application to Image Retrieval", A Report to Depart. of Electrical and Electronics Engineering, Shizuoka University.

[11] P. A. Hemalatha, "Image Retrieval by Content Using Descriptors", International Journal of Engineering Sciences &
Research Technology, Vol.2, Issue 10, pp. 27803-2809, October 2013.

[12] R.Malini, C.Vasanthanayaki, "An Enhanced Content Based Image Retrieval System using Color Features", International Journal Of Engineering And Computer Science, Volume 2, Issue 12, pp. 3465-3471, Dec 2013.

[13] Monika Rathore,MR. Ravindra Gupta MR.Gaurav Shrivastava, " A Effective Content Based Image Retrieval System Based on Features &Techniques", International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) ,Volume 2, Issue 12, pp 3195-3201, December 2013.

[14] A.Ramesh Kumar, D.Saravanan " Content Based Image Retrieval Using Color Histogram", International Journal of Computer Science and Engineering (IJCSIT),Volume 4,Issue-2, pp. 242-245, 2013.

[15] Nishant Shrivastava and Vipin Tyagi "Multistage Content-Based Image Retrieval ", A Report to Dept. Of CSE, Jaypee University of Engineering & Technology

[16] Roshi Choudhary, Ruchika, et al. "An Integrated Approach to Content Based Image Retrieval." Advances in Computing, Communications and Informatics (ICACCI, 2014 International Conference on. IEEE, 2014.

[17] Rafael C. Gonzalez, Richard E. Woods, "Digital Image Processing", Textbook.

[18] https://en.wikipedia.org/wiki/Color_histogram.

[19] https://en.wikipedia.org/wiki/K-means clustering.

[20] https://en.wikipedia.org/wiki/Image_texture.

[21] https://en.wikipedia.org/wiki/Feature_extraction.

[22] Nai-Chung Yang, Wei-Han Chang, Chung-Ming Kuo, Tsia-Hsing Li, A fast MPEG-7 dominant color extraction with new similarity measure for image retrieval,Journal of Visual Communication and Image Representation 19 (2) , pp.92–105, 2008.