Research Paper
Volume : 3 | Issue : 4 | Special Apr Issue 2014 • ISSN No 2277 - 8179

An Enhanced Content Based Image Retrieval Using Three Region Colour Histogram of The Image

Technology & Innovation

KEYWORDS : CBIR, SIMPLIcity, Histogram, average precision

L.Jayanthi	Research Scholar, Department of Computer Science and Engineering, Periyar Maniammai University
Dr. K.Lakshmi	Assistant Dean, School of Computing Science and Engineering, Periyar Maniammai University

ABSTRACT

Large number of images are created everyday due to the advances in data storage and image acquisition technologies. In order to deal with these data, it is necessary to develop suitable information systems to efficiently manage these collections. Image searching and retrieval of desired images is one of the most important services that support the system with image collections like educational image resources, medical images, finger prints, satellite images, photo collections, museum pictures etc. Content Based Image Retrieval system searches the most similar images of a query image that involves in comparing the feature vectors of all the images in the data base with that of the query image

The color histogram is one of the most important techniques used in Content Based Image Retrieval(CBIR). It's efficient to compute and effective in searching results. Most commercial CBIR systems use color histograms as one of the features. The weakness of the colour histogram based method for measuring image content is that the space information will be missing in a typical colour histogram. It means, the colour histogram will not say where that color was originally present in that image. If the colour feature is blue, we may not be able to identify whether it is sky or sea. In this work, we address CBIR using three region (top, middle and bottom) colour histogram of the image. The combination of these three histograms will indirectly include the space information within it. We have used the Semantic Sensitive Integrated Matching for Picture Libraries database to evaluate the performance of the system. The average precision is used as a metric to measure the performance and compared the performance of our method with four other CBIR methods using Integrated Region Matching(IRM), Fuzzy Integrated Region Matching (F-IRM), Colour Salient Points (CSP) and Ripplet Transform and Fuzzy Relevance Feedback Mechanism (F-RFM) which used the same data set and evaluation method. According to the arrived results, the performance of the proposed three region colour histogram based CBIR system was significantly better than other compared methods.

Introduction

The vast growth of Internet and digital contents of Internet creates much interest on automated image indexing and retrieval techniques. Content Based Image Retrieval systems retrieve images from the image data base according to the user defined query image. It is an automatic retrieval process instead of traditional keyword approach. Histogram based techniques are widely used in the design of a CBIR systems. In a typical colour histogram, the space information will be missing. It means, the colour histogram will not say where that color was originally present in that image; It will not fully explain the nature of colour distribution in the image at various places which actually makes that image. In this paper we have used a combination of histograms for different regions (top, middle and bottom)of the image, which will indirectly include the space information within it.

The rest of this article is organized as follows. Section II describes the existing CBIR systems using Integrated Region Matching (IRM), Fuzzy Integrated Region Matching (F-IRM),Colour Salient points (CSP) and Ripplet Transform and Fuzzy Relevance Feedback mechanism (RT& F-RFM). Section III introduces the block diagram of CBIR using three region colour histogram. Section IV presents the algorithm of CBIR using three region colour histogram. Section V discusses the result of average precision of the system and compared with the other system.

Compared CBIR Systems

A)CBIR using Integrated Region Matching (IRM)

The targeted image retrieval systems are represented by a set of regions corresponding to objects. The objects are characterized by features such as reflecting color, texture, shape and location properties.[1] Each region in the query image is compared with all the regions of the images in the database. After regions are matched, the similarity measure is computed as a weighted sum of the similarity between the region pairs, with weights determined by the matching scheme and the matched images are retrieved.

B)CBIR using Fuzzy Integrated Region Matching (FIRM)

This work proposes a fuzzy logic approach, Unified Feature Matching for region based image retrieval. Image is represented by a set of segmented regions. Each region is characterized by a fuzzy set feature reflecting color ,texture and shape properties. [2] The resemblance of two images is then defined as the overall similarity between two families of fuzzy features and quantified by a similarity measure, UFM measure which integrates properties of all the regions in the images.

C) CBIR using Color Boosted Salient Points and Shape features (CSP)

Salient points are locations in an image where there is a significant variation with respect to a chosen image feature. The set of salient points in an image capture important local characteristics of that image. These salient points form the basis of a good image representation for CBIR. Salient features are determined from the local differential structure of images. [12][14].They focus on shape. Color distinctiveness is also taken into account in addition to shape distinctiveness. The color and texture information around these points of interest serve as the local descriptors of the image.[15] In addition shape information is captured in terms of edge images computed using Gradient Vector Flow Fields. Invariant moments are then used to record shape features. [3][11][13].

D)CBIR using Ripplet Transform and Fuzzy Relevance Feedback Mechanism

Multi scale Geometric Analysis Tool called Ripplet Transform Type I is used in this CBIR to represent the images in the database in different scales and directions . The images in the database are transformed into YCbCr colour space. RT decomposition over the intensity plane (Y) characterizes the texture information. RT decomposition over chromaticity planes (Cb & Cr) characterize colour. Retrieval result is obtained by measuring the Euclidean distance between the RT based features of the query image and the images stored in the database. Feature re-weighting technique utilizes both the relevant and irrelevant result's information to obtain effective result.[9] If the user is satisfied with the retrieved results without using Relevance Feedback Mechanism then the retrieval process is ended in the first pass. If , however, the user is not satisfied, he/she can select top query based relevant based images as positive feedback and the remaining as negative examples for updating the weight parameters and revising the features. Using these feedbacks, Fuzzy feature Evaluation Index (FEI) and weights of features are computed. Thus, Fuzzy Relevance Feedback Mechanism (F-RFM) is used to improve the performance of CBIR systems through continuous learning and interaction with the end users. But in all these cases time complexity per iteration are high and the accuracy of the relevant images are low.[4][8].

The Design of Proposed CBIR System USING THREE REGION COLOUR HISTOGRAM

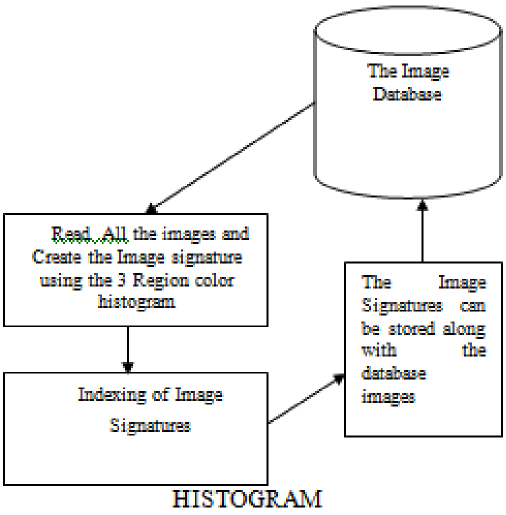

Fig.1. Indexing Phase

Fig.1. above shows the Indexing Phase of the CBIR system. All the images are read. Image signatures using colour information of three regions are created and indexed. The indexed image signatures are stored along with the images of the image data base.

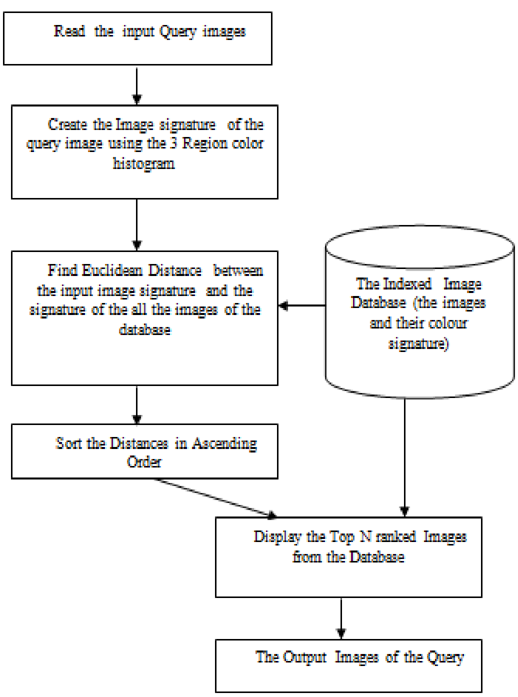

Fig.2. CBIR system using three region Colour Histogram

The block diagram of Three Region Colour histogram based CBIR system is shown in Fig.2.

THE ALGORITHM OF CBIR USING THREE REGION COLOUR HISTOGRAM

Let I be the input query images of size mxn

Apply median filter on the input image

Convert the high colour input image in to a 256 color image

Separate the top, middle and bottom regions R_1, R_2, R_3, of the image

Find the histogram of the regions R_1, R_2, and R_3 as $h_i(r_1, r_2, r_3)$

Use that combined 3 region color histograms as the signature S of input image $S = h_i(r_1, r_2, r_3)$.

Let M be the image signatures of all the images in the dataset.

Find Euclidean Distance between the input image signature and the signature of the all the images of the database

The distance D=M-S

Sort the distance matrix D in ascending order

Find the index of Top N ranked minimum distances of D

Display the Top N ranked Images from the image database using the index

A colour histogram is a type of bar graph, where each bar represents a particular colour of the colour space being used. The bars in a colour histogram are referred to as bins and they represent the x-axis. The number of bins depends on the number of colours there are in an image. The y-axis denotes the number of pixels there are in each bin. In other words how many pixels in an image are of a particular colour. The color histogram is one of the most important techniques in content-based image retrieval. It's efficient to compute and effective in searching results. Most commercial CBIR systems use color histograms as one of the features.

For an m*n image I, the colors in that image are quantized to C_1, C_2, ……, C_k. The color histogram $H(I)=\{h_1, h_2, ……, h_k\}$, where hi represents the number of pixels in color C_i. The color histogram also represents the possibility of any pixel, in image I, that in color C_i.

$$P(C_i) = \frac{h_i}{m * n} \qquad (1)$$

where,

C_i is the i^{th} color index

h_i is the number of pixels with that color index.

m x n is the total number of pixels in the image.

i = 0,1,2,….,L-1. (where L = 256).

Since colour is one of the most prominent perceptual features, in many cases the effect of using histogram in searching and retrieving image is good.

The weakness of the colour histogram based method for measuring image content is that the space information will be missing in a typical colour histogram. It means, the colour histogram will not say where that color was originally present in that image.

For example, in a typical natural image the sky and water may present different shades of blue at different location. Here the location of that color information will play major role to understand the image.

If we prepare different histograms for different regions of the image, then the combination of those histograms will indirectly include the space information within it.

Volume : 3 | Issue : 4 | Special Apr Issue 2014 • ISSN No 2277 - 8179

Let the image is decomposed in to three regions R_1, R_2, R_3.

$$I_1=\{ R_1, R_2, R_3 \} \tag{2}$$

If we see the three histograms of the different regions of the same image, we can find significant difference among them. So, now the image signature created from these three sets of histograms will contain some space information within it

Let h_1 and h_2 represent two combined color histogram of three regions R_1, R_2, R_3, of two images I_1 and I_2.

Now, we can represent the combined histograms of the image I_1 as

$$h_1 = h_1(r_1, r_2, r_3) \tag{3}$$

are then retreived.

The combined histograms of the image I_2 as

$$h_2 = h_2(r_1, r_2, r_3)$$

The Euclidean distance between the color histograms h1 and h2 can be computed as:

$$d^2(h_2, h_2) = \sum_{R_1} \sum_{R_2} \sum_{R_3} \left(h_1(r_1, r_2, r_3) - h_2(r_1, r_2, r_3)\right)^2 \tag{4}$$

If the Euclidean distance between the query image and the image in the database in the database is minimum, the images are said to be the most similar. Images are sorted in the ascending order of the euclidean distance.Top N ranked matched images

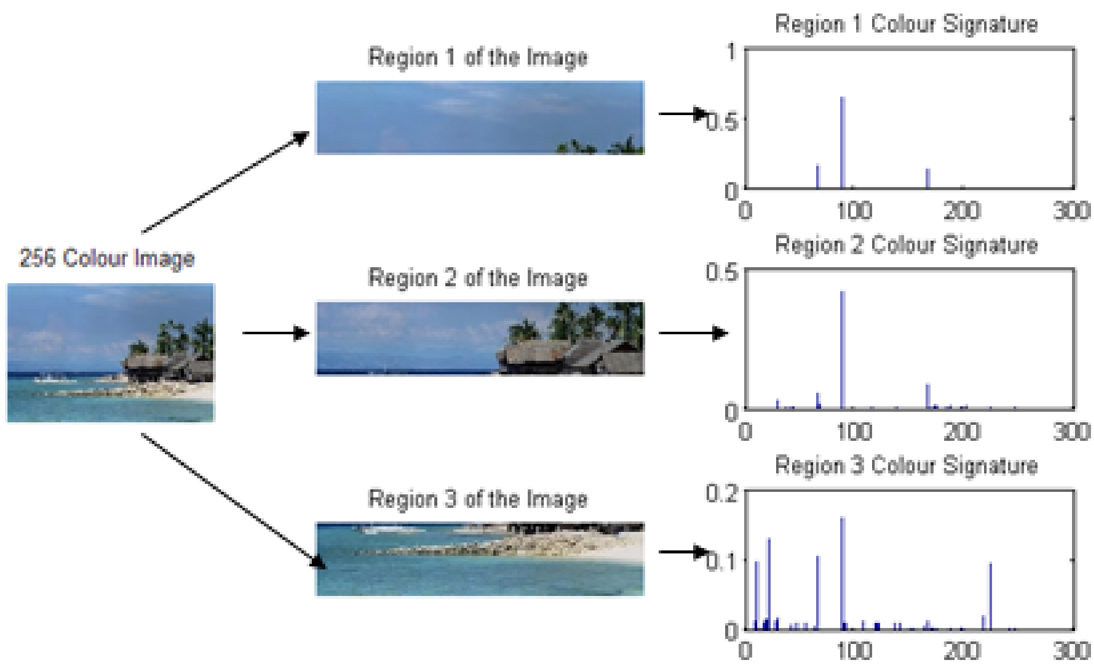

Fig.3. Creating Image Signature from 3 Regions of the Image

Results and Analysis

Repeated experiments were conducted on 1000 images of SIMPLIcity database which contains 10 different categories (African, Ocean, Building, Bus, Dinosaurs, Elephant, Flower, Horse , Mountain, Food) of images to prove the efficiency of the proposed method. The retrieval results obtained using the performance of the proposed three region colour histogram based CBIR system was compared with some of the existing retrieval systems [1,2,3,4] The prototype of the CBIR system was developed using Matlab 7.

Precision was computed using the entire image collection in the SIMPLIcity data set. The each and every images of the dataset was used as query image and the entire dataset is searched for matching images and average precision was calculated for each query. To account for the ranks of matched images, the average of precisions within *k* retrieved images, *k=1,2,..., 100* , is computed, that is,

$$\overline{\rho} = \frac{1}{100} \sum_{k=1}^{100} \frac{n_k}{k} \tag{5}$$

n_k is the number of matches in the first *k* retrieved images.

This average precision is referred to as the ``weighted precision'' because it is equivalent to a weighted percentage of matched images with a larger weight assigned to an image retrieved at a higher rank.

The total number of semantically related images for each query is fixed to be 100. So, we did not considered to display measured recall within the first 100 retrieved images. Because, in this evaluation, recall will be proportional to the precision, since, we are considering 100 top ranked query results and there are 100 images in each category.

The mean of the average precisions of 100 queries on each category is tabulated and given below :

In the following table, the first four columns are the results from 4 reference works. And the last column is the results from this work.

Table 1 . The Average Precision of Different Methods

Category	IRM [1]	FIRM [2]	CSP [3]	F-RFM [4]	3RCS*
Africa	0.48	0.47	0.48	0.49	0.66
Beach	0.32	0.35	0.34	0.40	0.37
Buildings	0.35	0.35	0.33	0.39	0.48

Buses	0.36	0.60	0.52	0.58	0.67
Dinosaurs	0.95	0.95	0.95	0.96	0.98
Elephants	0.38	0.25	0.40	0.50	0.43
Flowers	0.42	0.65	0.60	0.75	0.68
Horses	0.72	0.65	0.70	0.80	0.90
Mountains	0.35	0.30	0.36	0.40	0.44
Food	0.38	0.48	0.46	0.51	0.64
Average	0.471	0.505	0.514	0.578	0.625

3RCS - The Proposed 3 Region Colour Signature Based Method

If we see the above table, the performance of the proposed algorithm was overall good and was little bit lower than F-RFM in the case of Beach, Elephants and Flowers category of images.

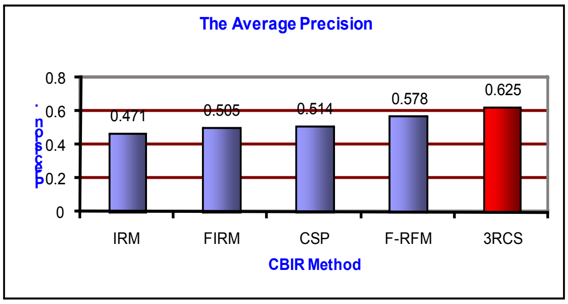

Fig.4. The Average Precision

The above bar chart shows the average performance of the five algorithms in terms of average precision. The performance of the proposed algorithm was significantly better than all other compared algorithms.

Conclusions

The proposed CBIR system has been successfully implemented using MatLab and the performance of the system in terms of precision was validated with images from SIMPLIcity. The arrived results were compared with previous methods IRM, FIRM, CSP and F-RFM and the performance of the proposed three region colour histogram based CBIR system was significantly better than all other compared methods.

In this work, we used three regions from all the images for preparing the colour based image signature by assuming that the top, middle and the bottom portion of image will contain space information which will distinguish the image category in a better manner. Future works may address more efficient and significant ways to separate the image using the colour information which may model the space information of the image in a better manner. In this swork, the three region colour histogram was used as the color signature of the image and the three histograms will increase the storage cost considerably. Future works may address the ways to reduce this size of the color signature by using suitable feature reduction techniques.

REFERENCE

[1] J.Jia Li, James ,Z.Wang and Gio Wiederhold,"IRM: Integrated Region Matching for Image Retrieval," in the Proceedings of the 8th ACM | International Conference on Multimedia, pp.147-156, 2000. | [2] Y.Chen and James Z Wang, "A Region Based Fuzzy feature Matching Approach to Content Based Image Retrieval," in IEEE Transaction | onPAMI, Vol. 24,No.9,pp.1252-1267,2002. | [3] P.S.Hiremath and Jagadeesh Pujari,"Content Based Image Retrieval Using Color Boosted Salient Points and Shape Features of an | Image" in the International Journal of Image Processing (IJIP),Vol.2,No.1, pp.10-17,2008. | [4] Manish Chowdhury,Sudeb Das and Malay Kumar Mundu, " Interactive Content Based Image Retrieval using Ripplet Transform | and Fuzzy Relevance Feedback" PerMIn2012,LNCS 7143, pp.243–251,2012. | [5] Rishav Chakravarti,Xiannong Meng, "A study of Color Histogram Based Image Retrieval" in the sixth International Conference on | Information Technology:New Generations, IEEE Computer Society,pp.1323 – 328, 2009. | [6] A.Ramesh Kumar, D.Saravanan, "Content Based Image Retrieval Using Color Histogram" in the International Journal of Computer | Science and Information Technologies (IJCSIT), Vol.4, Issue 2,pp.242 –245,2013. | [7] S.Neetu Sharma, S.Paresh Rawat and Jaikaran Singh, "Efficient CBIR Using Color Histogram Processing" in Signal and Processing : An | International Journal (SIPIJ), Vol.2,No.1,pp.94- 112, 2001. | [8] Y.Rui,T.S.Huang, S.Mehrotra, "Relevance Feedback: A Power Tool For Interactive CBIR",IEEE Transactions on Circuits and Systems for | Video Technology, Vol.8,pp.644-655 ,1998. | [9] Z.Jin,I.King,X.Li, "CBIR by Relevance Feedback in Laurini. R,VISUAL 2000 ," LNCS, Springer, Heidelberg, Vol.1929, pp.521-529,2000. | [10] K.Suri Babu,V.Nagesh, "Optimal Solution for Image Retrieval Using Integrated Region Matching" in the Proceedings 2nd National | Conference on Challenges and Opportunities in Information Technology (COIT-2008), pp.223-226, 2008. | [11] Mas Rina Mustaffa,Fatima Ahmed,Ramian Mahmod and Shyamal Doraisamy, "CBIR using Color and shape fused features," | Pertanika.J,Science and Technology,Vol.21(1),pp-161-168, 2013. | [12] Ch.Kavitha,B.Prabhakara Rao, A.Govardhan."Image Retrieval based on colour and Texture features of the image sub blocks " in the | International Journal of Computer Applications(0975-8887),Vol.15-No.7, pp.33-37,February 2011. | [13] Ch.Kavitha et al, "Image Retrieval Based on Combined Features of Image Sub-Blocks" in the International Journal on Computer Science | and Engineering (IJSCE),Vol.3,No.4,pp.1429-1438,ISSN:0975-3397. | [14] Chenyang Xu, Jerry L.Prince, "Snakes,Shapes and Gradient Vector Flow",IEEE Transaction on Image Processing,Vol.7,No.3,pp.475- | 488,1998. | [15] B.S Manjunath and W.Y.Ma, "Texture Features for Browsing and Retrieval of Image Data", IEEE Transactions on PAAMI, Volume 18, | No.8,August 1996. |