International Journal of Engineering Research and Development
***e**-ISSN: 2278-067X, **p**-ISSN: 2278-800X, www.ijerd.com*
Volume 4, Issue 4 (October 2012), PP. 73-85

Linear Equation in Parts as Histogram Specification for CBIR Using Bins Approach

Dr. H. B. Kekre[1], Kavita Sonawane[2]
[1]MPSTME, NMIMS University, Sr. Professor, Mumbai, India.
[2]MPSTME, NMIMS University, Ph.D Research Scholar, Mumbai, India.

Abstract:—This paper explores the use of linear equations as histogram specification for CBIR system based on bins approach. This specification modifies the histogram of R, G and B planes of image. It actually pushes the intensities to little higher level which gives enhanced image where in the image details can be seen clearly. This modified image contents are utilized effectively to form the feature vectors of the image. This system is producing good results as compared to results obtained using the contents from the original histogram. To extract these image features each image plane modified using the histogram specification goes through the following few steps. Modified image planes are partitioned into two parts by computing the center of gravity. Each partition will get an id assigned to it, which leads to the bins formation process. This CG based partitioning of three planes generates the 23 = 8bin addresses. Each of these addresses is indicating the region of specific range of intensities from R, G and B planes. Image contents are extracted to these eight bins and computed in different formats. These eight bins holding the image contents in various formats are representing the image feature vectors of dimension eight. Three similarity measures Euclidean distance, Absolute distance and Cosine Correlation distance are used to compare the query and database image feature vectors. Performance of the proposed system is evaluated using three parameters Precision Recall Cross over Point (PRCP), LSRR (Length of String to Retrieve all Relevant) and Longest String. Experimentation of this system is carried out with database of 2000 BMP images having 20 classes. Set of 200 query images is used to check the response of the system which includes images selected randomly from all 20 classes in database.

Keywords:—Linear Equation, Histogram Specification, Center of Gravity, Bins, Euclidean distance, Absolute distance and Cosine Correlation distance, PRCP, Longest String, LSRR

I. INTRODUCTION

CBIR system explained in this paper is based on the bins approach. Main idea implemented in this paper is that histogram specification can be utilized very effectively and efficiently in core phase of the feature extraction for many applications based on CBIR technique. Feature extraction is very important phase of any CBIR system because it generates the compact feature vectors from huge image contents [1-3]. These feature vectors representing the image are expected to be of very small size as compared to entire image size. There are various contents present in the image and can be used to form the feature vector. Based on the information present in the image; features are broadly classified into two types local and global image features. Based on the properties of the contents extracted to form the feature vector, they can further be classified into variant and invariant feature vectors. Variant feature vectors are extracted from the image contents which are varying with translation and rotation applied to image. Image contents which are independent of these variations or transformations applied to image are called invariant image features. Three prominent image contents subjective to human perception are Color, Texture and Shape descriptors.[4-9]. In this work we are using the color information of the image to form the feature vector. We are working in R, G and B color space. We first separate the image into R, G and B planes and a color histogram is obtained for each of them. Feature extraction process followed in this work is actually based on the modification of the image histogram using the histogram specification. Histogram specification is defined through the simple linear equations i.e 'y=mx+c' in parts. Previous work with the color histograms are based on histogram shifting, equalization of histogram, modification of histogram using polynomials, original histogram with CG partitioning etc.[10-17]. Modification functions used in these techniques have some constraint while pushing the intensities towards higher side. This work has tried to remove this limitation by making use of linear equations. Histogram modified using the linear equations are then partitioned using CG into two parts which is the origin of the bins formation process. Three planes partitioned into two parts generating eight bin addresses. The eight bins are designed to extract the image information in different formats namely Count of pixels falling in specific partition, 'Total of intensities' for the pixels counted in each bin and 'Average of intensities' for the pixels counted in each bin. These are used as feature vectors of dimension eight. According to literature survey, we found that CBIR systems designed with histogram bins are time consuming and have high computational complexity as compared to our proposed system. This is because many systems are comparing the images bin by bin for the histogram of size 256 bins which takes more time and increases the complexity [18-22]. We could save our execution time and reduce the complexity because of the formation of eight bins formed by partitioning the histogram in two parts. Time require to compare the query and database image is also depends on one more important factor i.e. similarity measure. Commonly used similarity measure in many CBIR system is Euclidean distance, minkowiski distance.[22-24]. We are using three similarity measures namely Euclidean distance, absolute distance and cosine correlation distance. We found better performance using the other two measures mentioned as compared to Euclidean distance. Performance of the CBIR system can be evaluated in terms the accuracy in relevance of images retrieved as well as on the completeness of retrieval.

Completeness indicates that all images relevant to query could be retrieved from database. To evaluate performance of proposed system in terms accuracy and completeness we are using parameter PRCP (Precision Recall Cross over Point) along with Longest String and LSRR (Length of String to Retrieve all Relevant). Database of 2000 BMP images of size 128 x 128 from 20 different classes each having 100 images is used for the experimentation. Organization for rest of the paper is as follows. Section II gives the detail description of the linear equations used as histogram Specification. Section III describes the Feature extraction process with implementation details. Section IV discusses the role of similarity measures followed by performance evaluation in Section V. Section VI presents the Results along with their discussion. Finally Section VII highlights the observations and conclusions.

These equations are worked out here with three variations of delta (Δ). This change in histogram has brought positive change in the image as well as in the results to some extend and after certain value of (Δ) it started falling down. This (Δ) is the increment given for the original intensities of image to go high. Information about final paper submission is available from the website.

II. HISTOGRAM SPECIFICATION WITH LINEAR EQUATIONS

A. Histogram Specification

Image histogram acts as a graphical representation of the tonal distribution in a digital image. In simple words image histogram is just a bar graph of pixel intensities. Pixel intensities are plotted along with the x-axis and number of occurrences for each intensity is plotted across y axis. Histograms are used in many CBIR systems because of their efficiency, and insensitivity to small changes in camera viewpoint [25-26]. Histogram specification is special tool used for image enhancement and normalization. It is a class of transformation which changes the original histogram of an image into desired one. In this work we have proposed a new specification to modify the histogram using linear equations in parts. It modifies the histogram such that each intensity level will be pushed high with some increment. This is described in detail with help of equations and graphical representation of the same in Figure1. We can observe in Figure 1. The new specification used to modify the histogram is shown by green color. This is linear increment achieved in intensity levels in two parts. First part is line joining two points P1 and P2. Where P1 is (0,0), and P2 is ((0.5-Δ), (0.5+Δ)).

Let,

$$y = mx + c \tag{1}$$

Where $m = \frac{0.5 - \Delta}{0.5 + \Delta}$ and $c = 0$

Where Δ is the variable as shown in Figure1 use to give increment to the original intensity. In this case we have tried the three value of Δ as 0.05, 0.1 and 0.15 to increase the intensities.

Let, replace m and c in equation (1) we get the line joining the point P1 and P2 i. e

$$y = \left(\frac{0.5 - \Delta}{0.5 + \Delta}\right) * x \tag{2}$$

Where x varies from 0 to 0.5-Δ and so y in range 0 to 0.5+Δ.

Now, equation for line joining two points P2 to P3 is

$$y = mx + c \tag{3}$$

Where $m = \left(\frac{1-(0.5-\Delta)}{1-(0.5+\Delta)}\right)$ and $C = \frac{-2\Delta}{0.5-\Delta}$

Using three different Δ, we get three different lines each in two parts (P1, P2) and (P2, P3) using the equations 2 and 3 respectively. These histogram specifications obtained using above equations with three Δ are shown in Figure 2.

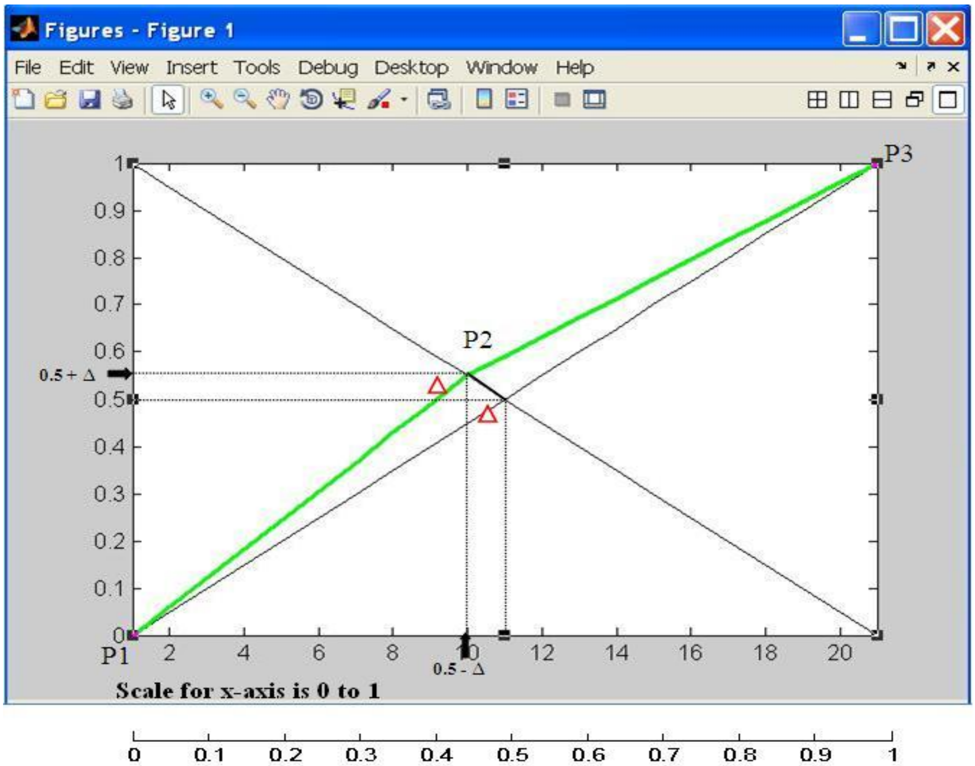


Figure 1. Histogram Specification Linear Equation for P1P2 and P2 P3

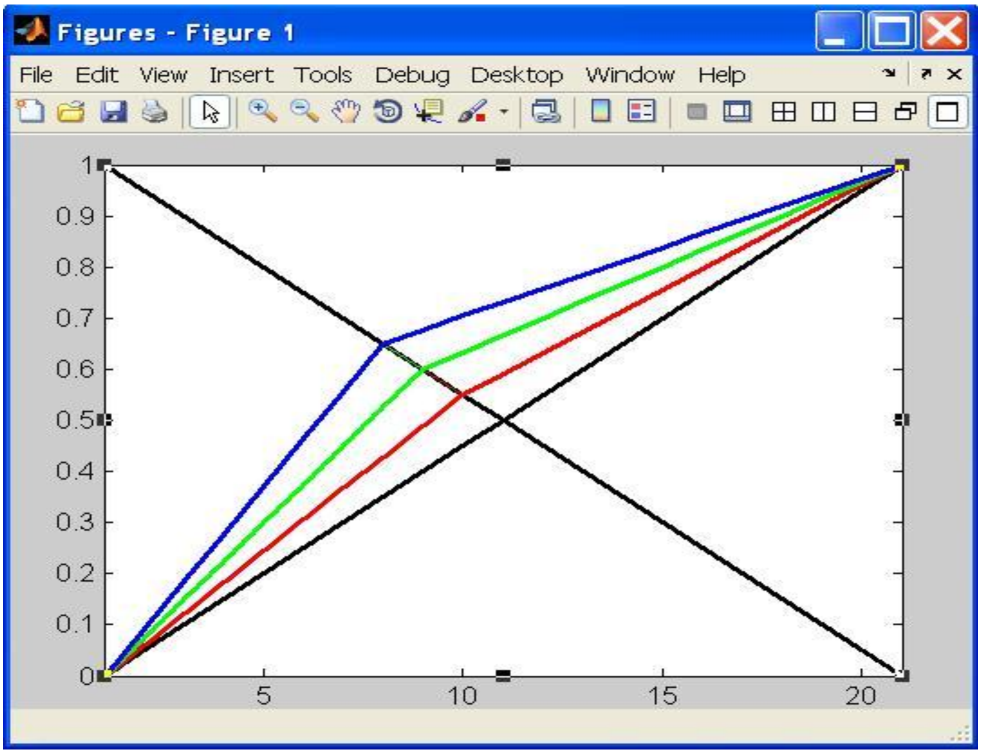


Figure 2. Histogram Specification with respect to three different Δ replaced in Eq.2 and 4.

Fig. 1 shows the histogram specification function curves obtained in parts using the linear equations 2 and 3. Fig. 2 shows the three specifications in red, green and blue colors obtained to modify the histogram with three different Δ i.e. 0.05, 0.1 and 0.15 respectively.

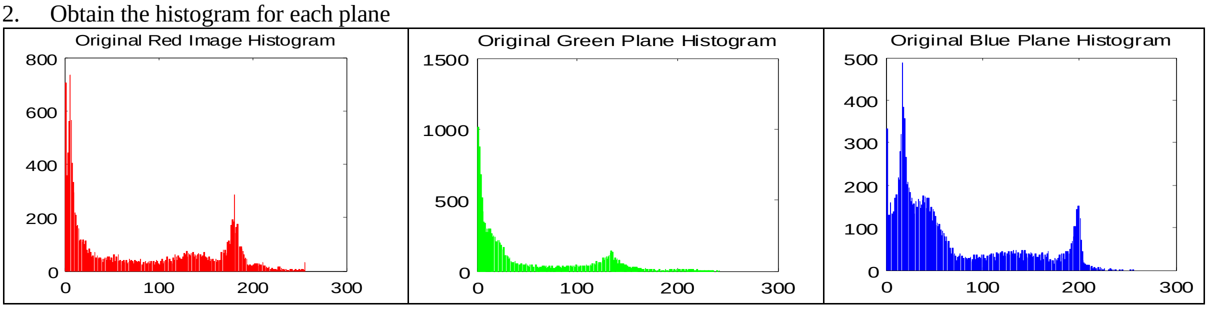
III. FEATURE EXTRACTION PROCESS

Following sequence of steps given with implementation details are followed to extract the feature vectors of the image under process.

1. Separate the image into R, G and B planes. As shown in Fig 3 .

Figure 3. Bus Image with its R, G and B planes

2. Obtain the histogram for each plane



3. Modify the histograms using the Linear equations 2 and 3. Here we have shown in Fig.4, the modification of red plane histogram using all three values of Δ i.e 0.05, 0.1 and 0.15.

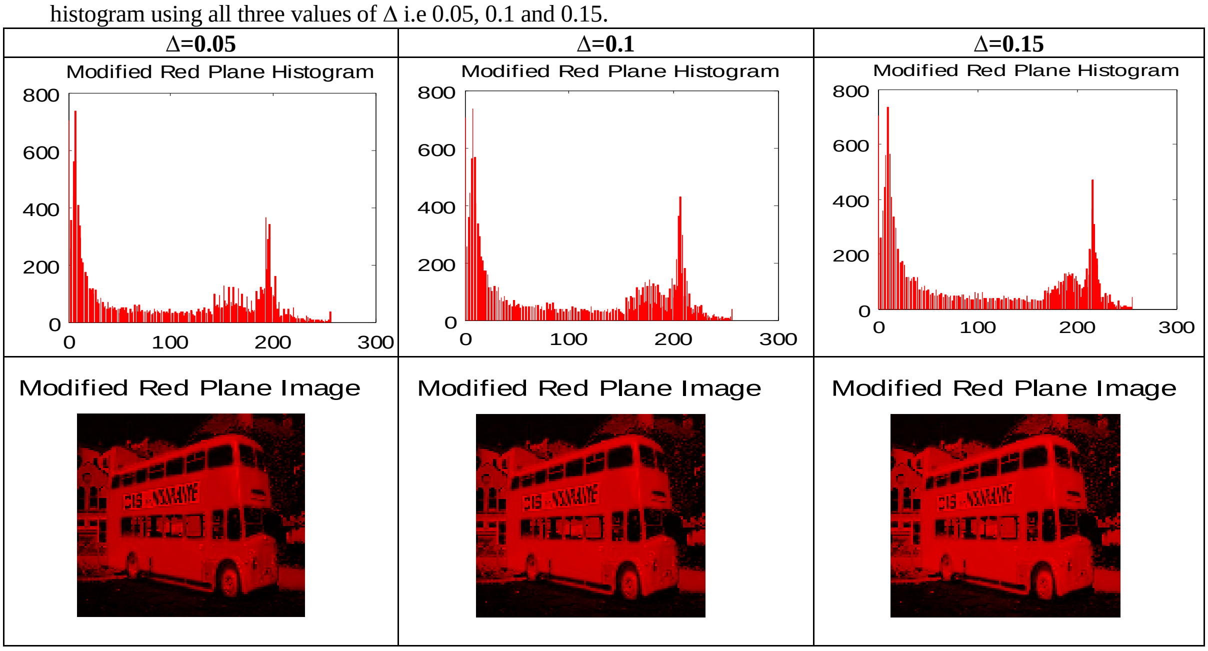


Figure 4. Application of histogram specification to modify the original histogram (for Red plane of bus image)

4. Modified histograms are then partitioned into two parts by computing the Center of Gravity. We get two partitions with id '0' for below CG and '1' for above CG as shown in Fig. 5.

$$CG = \left[\frac{\left(L_1 W_1 + L_2 W_2 + \ldots.+ L_n W_n\right)}{\sum_{i=1}^{n} W_i} \right] \quad (4)$$

Where Li is intensity Level and Wi is no of pixels at Li

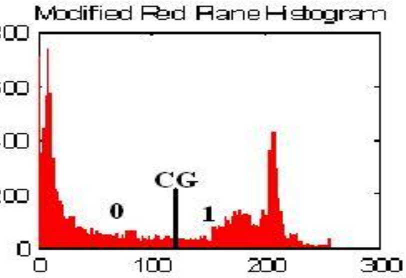


Figure 5. CG Based Partitioning of modified Histogram

5. As shown above once the modified histograms are partitioned using CG we start with the pixel by pixel processing of image under feature extraction process.

Pick up the pixel Pi(x,y) from image with its R, G and B intensities named as Ri, Gi and Bi respectively. Check where these intensities fall in either 0 or 1 partition of the respective color histogram.
Let, (Ri<=CG), (Gi>CG) and (Bi<CG)

Based on above conditions the flag bits assigned to this pixel are '0', '1' and '0' i.e. "010". It indicates that the pixels will be counted in Bin 2.
Same process is applied to all image pixels and the count of pixels is made available into eight bins from bin 000 to111.

6. Once the count of pixels is obtained, this information is used to represent the image contents in two different formats as follows.

R_Total is →
$$R_T = \sum_{i=1}^{N} R_i \quad (5)$$

R_Avg is →
$$\overline{R} = \frac{1}{N}\sum_{i=1}^{N} R_i = \frac{R_T}{N} \quad (6)$$

Where N is the Count of pixels in each bin.

Count of pixels obtained in eight bins itself can be used as feature vector to represent the image and compare it. Along with this, two more feature vectors used here to represent the image. One is sum of intensities of each color is computed separately for the pixel count of each bin. Equation (5) shows the sum of red intensities as R_Total. Similarly the G_Total and B_Total is obtained for each image as feature vectors.
Second type of feature vector is that, instead of taking the sum we are taking the average of intensities into consideration. It is calculated using equation (6), which is showing the average for red intensities as 'R_avg'. Same process is applied to other two color intensities i.e Green and Blue feature vectors G_Avg and B_Avg are obtained.

7. Multiple feature vector databases are prepared for all variations with respect to all factors used for computing the feature vector. The different factors used are three specification functions used to modify the histogram, count, total and average information and three colors.

3 colors x 3 Linear Equation based histogram specifications x two variations to represent the image contents (i.e Total and Average of intensities) =18 feature databases In addition to it count of pixels into eight bins for each of the three histogram specifications gives three feature databases.
i.e now 18 + 3 =21 feature vector databases are prepared for all 2000 BMP images in the database. The seven steps explained above are followed to extract the features of the image and form the feature vector databases.

IV. APPLICATION OF SIMILARITY MEASURE

After the preprocessing work i.e. Preparation of the feature vector databases for all database images system is ready to face the query image and generate the retrieval results. It compares the query feature vector with database feature vectors. This comparison process is carried in our system by means of three similarity measures namely Euclidean distance (ED), Absolute distance (AD) and Cosine correlation distance (CD).

Euclidean Distance	Absolute Distance:	Cosine Correlation Distance
$D_{QI} = \sqrt{\sum_{i=1}^{n} \left|(FQ_i - FI_i)\right|^2}$ (7)	$D_{QI} = \sum_{i=1}^{n} \left|(FQ_i - FI_i)\right|$ (8)	$\frac{\langle D(n) \mid Q(n) \rangle}{\sqrt{[\langle D(n) \mid D(n) \rangle \langle Q(n) \mid Q(n) \rangle]}}$ (9) Where D(n) and Q(n) are Database and Query feature Vectors resp.

Equations 7 and 8 are measures that are calculating the distance between the two feature vectors that are being compared. Equation. 9 measures the similarity in terms of cosine angle between two feature vectors. The distance and angle obtained between the query and database image feature vectors will be arranged in ascending order from min to max distance between them.

V. PERFORMANCE EVALUATION OF THE PROPOSED SYSTEM

Performance of the system can be tested by firing queries from different classes. We check the time required to extract the query image feature vector and comparing it with the database image feature vectors and generate the retrieval results. After testing the system its performance should be evaluated so that we can determine its strength and accuracy [27-28]. We are using three parameters to evaluate the performance of the system designed with Bins approach based on partitioning of the modified histogram with three specifications. The parameters used are defined below.

A. PRCP: Precision Recall Cross over Point

This parameter is designed on the analysis of two traditional CBIR parameters that are Precision and Recall [10-17], [27-28]. Using PRCP two major things can be measured about the system that is accuracy (Precision) and completeness (recall) as given in equations (10) and (11).

$$\text{Precision} = \frac{\text{Relevant Retreived Images}}{\text{All Retrieved Images}} \tag{10}$$

$$Recall = \frac{Relevant\ \ Retreived\ \ Images}{All\ \ Relevant\ \ In\ \ Database} \tag{11}$$

PRCP=1; Accuracy tells that the CBIR system under evaluation process is generating the retrieval for the given query such that the resultant set does not contain the irrelevant images. Completeness indicates that system is able to recall or retrieve all the images available in the database which are relevant to query. It indicates the best case performance of the system which proves that how ideal the system is.

PRCP =0; It indicates the worst case performance of the system. It tells that retrieval result generated by the system contains only irrelevant images; it does contain a single relevant image. Now if not a single relevant image is being retrieved from database recall becomes 0. Means the system is neither accurate nor complete. In short we can say PRCP values between 0 to1 tells that how far we are from the ideal system

B. Longest String

In most of the CBIR systems we found that only initial string of images relevant to query are taken into consideration for final retrieval result. This may contain very few images where we get recall very low. It can be noticed that continuous set of large no of relevant images appear at a place other than beginning. To take the advantage of this fact in retrieval process this parameter is designed, used and termed as 'Longest String'.

C. LSRR: Length of String of Retrieve all Relevant

This parameter measures the length and in turn time required to retrieve all images from database relevant to query. Input taken to measure this parameter is images arranged in sorted order at minimum to maximum distance from query image. It goes on collecting images relevant to query till recall becomes 1. It is advantageous to have this parameter as low as possible. Lowest value obtained for LSRR indicates that all images relevant to query available in database are retrieved with minimum length of traversal and minimum time.

VI. RESULTS AND DISCUSSION

A. Database and Query formulation

Database: To test the performance of the system with respect to the all small variations used in feature extraction process we have used database of 2000 BMP images. It includes 100 images from each of the 20 different classes. Sample image is taken from each of the 20 classes to represent that class and are shown in following Figure 6.

Figure 6. 20 Sample Images from database of 2000 BMP images from 20 classes

Query Formulation*:* Query formulation part of this system is based on query by example image [29]. Whenever user wishes to search for the particular set of images, user has to give a sample query image as input to the system. For this query then system computes the feature vector same way as it has computed for all database images.

This experimentation is carried out with set of 200 query images. This includes 10 images selected randomly from each of the 20 classes.

B. Result, Analysis and Discussion

Once the query image is fired to the proposed system it computes the feature vector and proceeds for the comparison process using three similarity measures ED, AD and CD.Results obtained for each parameter for all 21 feature vector databases are shown separately.

Note: Results shown in all tables and charts are using following nomenclature

Histograms modified with three Δ values Δ= 0.05, Δ = 0.1 and Δ=.15 are termed as MOD1, MOD2 and MOD3 respectively., Original Histogram: ORG and 'Count of Pixels' is COUNT, sum of intensities for R, G and B colors are 'TOTAL' and for Average of intensities it 'AVG'.

C. PRCP

Total PRCP for 200 query image for Count of Pixel

SM	PRCP			
	ORG	***MOD 1***	***MOD 2***	***MOD 3***
ED	4991	4976	4976	4877
AD	5319	5376	5393	5318
CD	4877	4920	4918	4812

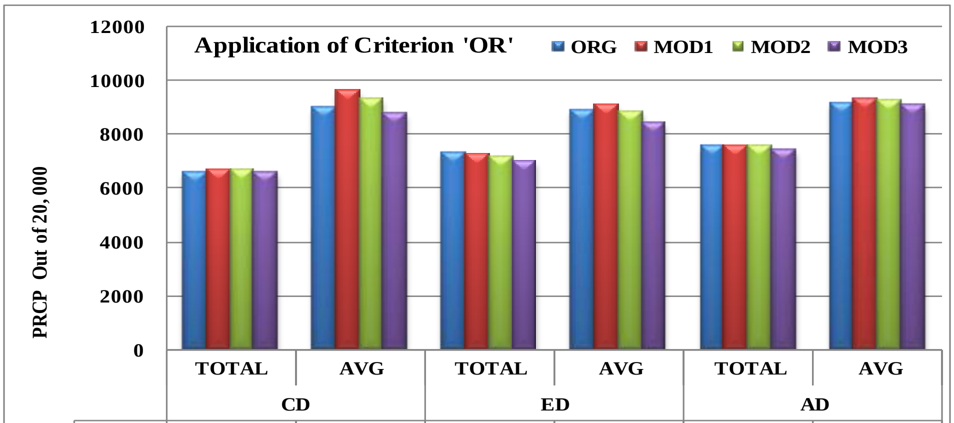
Table II. Total PRCP for 200 query image for 'Total of R, G and B Intensities'

PRCP : TOTAL												
SM	R				G					B		
	ORG	***MOD1***	***MOD2***	***MOD3***	***ORG***	***MOD1***	***MOD2***	***MOD3***	***ORG***	***MOD1***	***MOD2***	***MOD3***
ED	5080	5240	5311	5378	4739	4746	4792	4869	4546	4575	4516	4409
AD	5448	5624	5758	5813	5188	5191	5260	5346	4951	4941	4921	4816
CD	4716	4740	4837	4872	4379	4373	4380	4408	3933	4195	4182	4120

Table III. Total PRCP for 200 query image for 'Average of R, G and B Intensities'

PRCP : AVG												
SM	**R**				**G**					**B**		
	ORG	***MOD1***	***MOD2***	***MOD3***	***ORG***	***MOD1***	***MOD2***	***MOD3***	***ORG***	***MOD1***	***MOD2***	***MOD3***
ED	5573	5792	5720	5346	5448	5833	5675	5491	5309	5997	5874	5632
AD	5749	5987	6036	5767	5480	6002	5937	5728	5421	6014	6050	5944
CD	5722	5874	5759	5353	5802	6392	6161	5826	5236	6352	6211	5899

Chart1. PRCP for Total and Average after application of Criterion ‘OR’ to R, G and B Results



	CD TOTAL	CD AVG	ED TOTAL	ED AVG	AD TOTAL	AD AVG
ORG	6565	8978	7316	8911	7582	9141
MOD1	6683	9617	7273	9083	7586	9302
MOD2	6688	9307	7152	8815	7564	9282
MOD3	6572	8784	7004	8398	7419	9098

In above Tables I, II and III results obtained for parameter PRCP are displayed. Here each entry in the table corresponds to the total PRCP value obtained for all the results of 200 query images from 20 different classes and that is why each value in the table is out 20,000. In Table I , we can observe that for count of pixels MOD1 and 2 are performing equally better as compared to ORG. Table II represents the PRCP results obtained for total of intensities, where we found that modifying the histogram using linear equation with Δ =0.15 is giving better performance as compared to other histograms for red and green color. For blue color MOD1 is better among all. In Table III, observing results of Average of intensities, we have noticed that performance of MOD1 is on top among all other histogram based bins and after that MOD 2 is better. Among all three tables we observe that the best value obtained for PRCP is 6392 for green color with feature vector type average of intensities extracted from the histogram modified using the Δ=0.05. It indicates that Precision and recall are reached to 0.3. About this performance we can say that although modified histograms are performing better than original histogram, we still have scope to improve this result. To give direction to this thought we have come up with the refinement to be made to this result by applying the OR operation termed as Criterion OR over the separate results obtained for R, G and B colors. After the application of Criterion OR over R, G and B color results of total and average of intensities we could really improve the result. These results are shown in Chart 1. In Chart 1, we can notice that the best value obtained for PRCP is 9617 for MOD1 with CD measure for feature vector average of intensities. Means the precision and recall is reached to 0.5, and this is very good improvement than the previous results obtained without the application of Criterion OR. Among the results shown in Chart 1, we found that Average of intensities are performing better as compared to total of intensities for all CD ED and AD distance measures. It can also be noticed that among different histogram specification used in feature extraction MOD1 and MOD2 are performing better in all results sets.

D. Longest String

The next parameter used to evaluate the system performance is longest string. As discussed earlier in definition of longest string, the ideal value for this parameter should as long as possible. Each of the 200 query images are executed to obtain longest string with respect to all similarity measures for all types of feature vectors. When the results obtained for each color (R, G and B) separately for the feature vector types total and average of intensities, we have taken the maximum longest string obtained among 10 queries result for three colors. Because of this we could analyze the performance of each color and we have found that for total of intensities the better performance obtained for three colors is in order Red, Green and Blue. Similarly for average of intensities we found it as Green, Red and Blue. Means red and green are playing significant role in this CBIR system. Results obtained for longest string are shown in following charts 2, 3 and 4 for Count of pixels, total of intensities and average of intensities respectively. In Chart 2, we found that few classes dinosaur, dove, horses, rainbow rose, sunset have achieved very good results for longest string. They got values in range between 20 to 45. The minimum value we have obtained among all results is 3 and maximum is 48 as longest string. In chart 3, for feature vector total of intensities we found that Dinosaur, Barbie, Dove, Sunset, Horses, Bus, Rainbow rose and Pyramids are performing better as compared to other classes in the database. These values are in range from 20 to 35. The best value for this result set is 36 for class dinosaur and the lowest result is 3 for two quires form mountain and car class. In Chart 4 for average of intensities we found that classes Barbie, Dove, Rainbow rose, Sunset, Dinosaur, Bus and Mickey are performing better as compared other classes. They achieved good results for longest string in range 20 to 95. Best value obtained here is for class Barbie i.e. 95 and minimum is 4 for Classes Mountain and building. This is very good result we could obtain for the feature vector type average of intensities. Observations written below each chart is highlighting the performance of each histogram specification used to modify the histogram to improve the CBIR results. Overall results are when analyzed we can say that, although individual results for queries from few classes are not good but average performance for all the classes is checked and shown in the last columns of all charts named as AVG. The overall average values for 20 queries from 20 classes we have got the average longest string as 13, 16 and 23 for count of pixels, total of intensities, and average of intensities respectively.

E. LSRR

Results obtained for LSRR are shown in Charts 5, 6 and 7 with data tables for count of pixels, total of intensities and average of intensities respectively. Ideal value for parameter LSRR should be as low as possible as it decides the time and length of the traversal required to collect all images of query class from database to make recall 1. Similar to the longest string parameter result, results obtained for LSRR are also analyzed for R, G and B colors. Values plotted in all charts are average values obtained for 20 queries from 20 classes of database. We have taken the minimum % LSRR obtained among 10 queries result for three colors. The values plotted in these charts are those which are minimum among all 10 queries and three color results set. In Chart 5 we can easily notice that performance given by modified histogram based feature vectors means MOD1 and MOD2 are better as compared to other results. The best minimum values we obtained for LSRR is with CD measure i.e just 58% traversal gives us 100% recall. In Chart 6 we can observe that there not much difference in the LSRR % obtained for all cases. But still looking at the fine derails we found that MOD1 and 2 are better as compared to others. The best LSRR obtained here is 59% for AD measure with MOD2. In Chart 7 we can observe that again MOD1 and MOD2 are performing better as compared to other histogram based feature vectors. The best LSRR value obtained is 59 % with AD measure for MOD1. We have analyzed the LSRR results obtained for all approaches discussed based on the variation in the feature extraction process. We found that LSRR values shown in all graphs are giving the average LSRR, means we can say that that for all most all approaches with respect to all factors LSRR is 60%. This indicates that this system gives recall value 1with just 60% of traversal.

Chart 2. Longest string for Count of Pixels

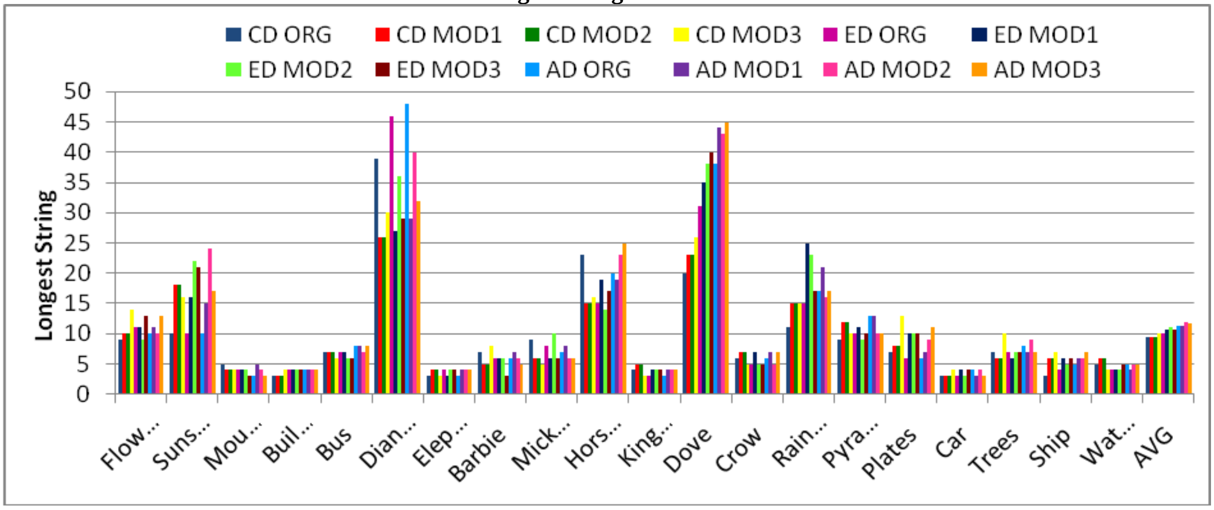


Chart 3. Longest string for Total of Intensities

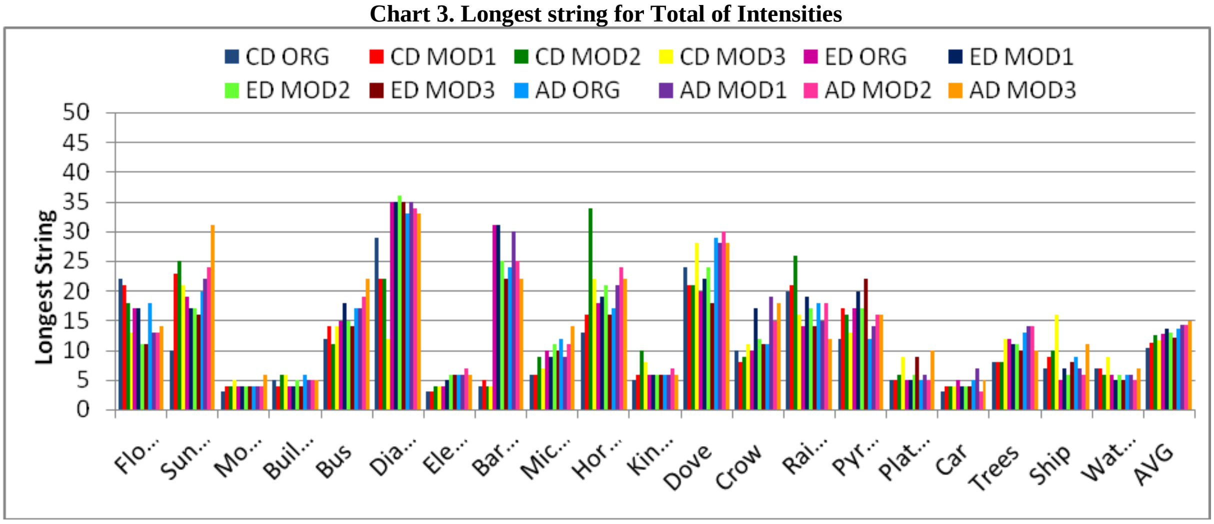


Chart 4. Longest string for Average of Intensities

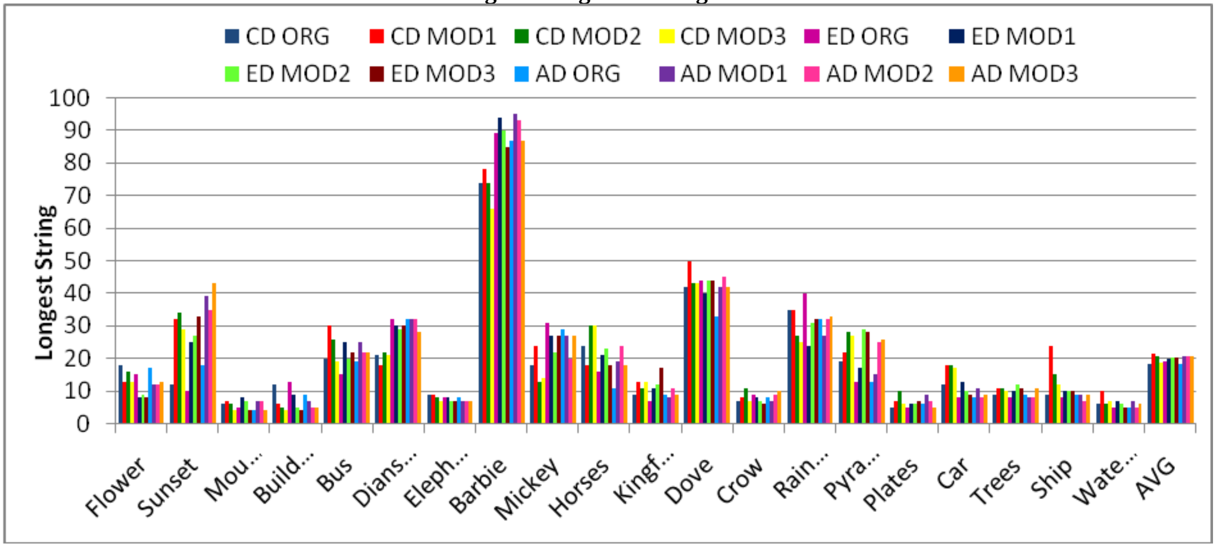
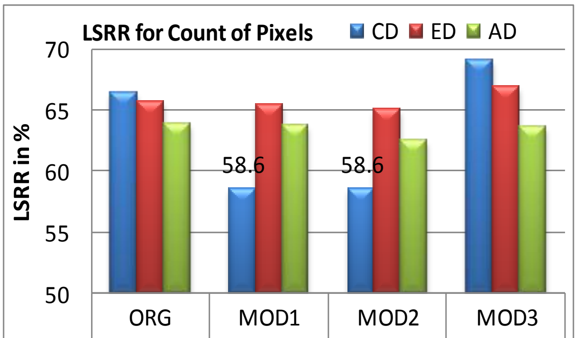
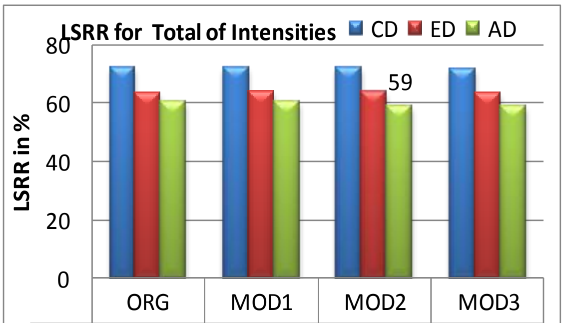


Chart 5. Percent LSRR for Count of Pixels



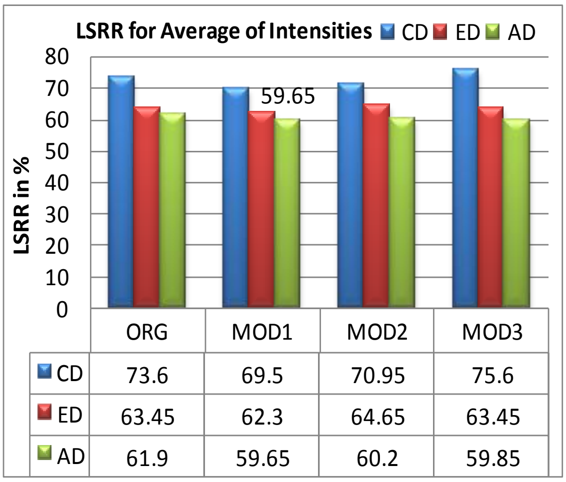
	ORG	MOD1	MOD2	MOD3
CD	66.45	58.6	58.6	69.15
ED	65.75	65.55	65.15	67
AD	63.9	63.75	62.6	63.7

Chart 6. Percent LSRR for Total of Intensities



	ORG	MOD1	MOD2	MOD3
CD	72.6	72.3	72.6	71.7
ED	63.85	64.15	64.1	63.85
AD	60.8	60.5	59	59.4

Chart 7. Percent LSRR for Average of Intensities



Various conclusions can be drawn from the analysis done for the system with respect to different factors such as linear equation with three different Δ values.(MOD1, 2 and 3), similarity measures , type of contents extracted to bins(count of pixels, total of intensities and average of intensities), R, G and B colors and the evaluation parameters etc. Conclusions are given as follows.

VII. CONCLUSION

The CBIR system is designed with the Bins approach which mainly based upon the histogram specifications used to modify the histogram and also on the CG based partitioning of the histogram to form eight bins out of 256 bins. Such CBIR system design suggests the use of linear equations to modify the histogram such that the intensities can be pushed up by some amount. This modification brings positive change in the features extracted and so in the retrieval of relevant images. Important point extracted from this research is that bins designed with original histogram and designed with the modified histogram if compared, it clearly tells that histograms modified with linear equations (MOD1, MOD2) are giving far better performance than the original histogram based bins. Instead of taking the 256 bins of histogram as feature vector, we could reduce the dimension of the feature to just eight bins by CG based partitioning of the modified histogram. This greatly reduces the time and computation complexity. System's evaluation with parameter PRCP indicates that this system gives very good precision and recall value i.e**0.5** which is obtained for **average of 200** query images. This indicates the better performance of this CBIR system as compared to others. Parameters longest string gives best result for MOD1 and MOD2 histogram specifications with CD and AD measures we have got this best values for class Barbie for all 10 queries between 90 to 95. Parameter LSRR has proven the strength of the system by giving the average LSRR just 60% for all other factors considered. Which tells that 60% traversal of the sorted string of database images (according to the distance sorted form min to max) collects all query relevant images from database. Among Similarity measure we found absolute and cosine correlations distance are producing far better results as compared to ED which is used in many other CBIR systems. Means we can suggest the other researcher to make use of the two measures to check and improve their system performance. Analysis done for three colors tells that red and green are dominating as compared blue color contents of the images in many cases.

REFERENCES

[1]. Hualu Wang, Ajay Divakaran, Anthony Vetro "Survey of compressed-domain features used in audio-visual indexing and analysis", 2003 Elsevier Science, www.elsevier.com/locate/yjvci.

[2]. Yong Rui and Thomas S. Huang, Shih-Fu Chang, "Image Retrieval: Current Techniques, Promising Directions, and Open Issues", Journal of Visual Communication and Image Representation 10, 39–62 (1999).

[3]. Min-Hee Jang1, Si-Yong Lee1, Sang-WookKim,"On Extracting Perception-based Features for Effective Similar ShaderRetreival", 2011 35th IEEE Annual Computer Software and Applications Conference.0730-3157/11 © 2011 IEEEDOI 10.1109/COMPSAC.2011.21

[4]. S. Nandagopalan, Dr. B. S. Adiga, and N. Deepak, "A Universal Model for Content-Based Image Retrieval" World Academy of Science, Engineering and Technology 46 2008.

[5]. Seongtaek Jo, Kyhyun Um, "A Signature Representation and Indexing Scheme of Color-Spatial Information for Similar Image Retrieval" 0-7695-0577-5/00 $10.00 0 2000 IEEE.

[6]. Ki-Hyun Kang, Yong-In Yoon, Jong-Soo Choi, "Additive Texture Information Extraction Using Color Coherence Vector" Proceedings of the 7th WSEAS International Conference on Multimedia Systems & Signal Processing, Hangzhou, China, April 15-17, 2007.

[7]. SubrahmanyamMurala, Anil BalajiGonde, R. P. Maheshwari, "Color and Texture Features for Image Indexing and Retrieval", 2009 IEEE International Advance Computing Conference (IACC 2009), Patiala, India, 6-7 March 2009.

[8]. R.VijayaArjunan, Dr.V.Vijaya Kumar, "Image Classification InCBIR Systems With Colour Histogram Features" 978-0-7695-3845-7/09 © 2009 IEEE DOI 10.1109/ARTCom.2009.233.

[9]. Wasim Khan, Shiv Kumar. Neetesh Gupta, "A Proposed Method for Image Retrieval using Histogram values and Texture Descriptor Analysis", International Journal of Soft Computing and Engineering (IJSCE), ISSN: 2231-2307, Volume-I Issue-II, May 2011.

[10]. DinuColtuc, Philippe Bolon Jean-Marc Chassery "Exact Histogram Specification", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 15, NO. 5, MAY 2006.

[11]. Dr. H. B. Kekre, KavitaSonawane,"Bins Formation using CG based Partitioning of Histogram Modified Using Proposed Polynomial Transform 'Y=2X-X2'for CBIR", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 3, No. 5, 2012.

[12]. DebashisSen,Sankar K. Pal, "Automatic Exact Histogram Specification for Contrast Enhancement and Visual System Based Quantitative Evaluation", IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL. 20, NO. 5, MAY 2011.

[13]. Dr. H. B. Kekre, KavitaSonawane,"Feature Extraction In Bins Using Global And Local Thresholding Of Images For CBIR", (IJ-CA-ETS)ISSN: 0974-3596, October '09 – March '10, Volume 2 : Issue 2.

[14]. Dr. H. B. Kekre, KavitaSonawane,"Bin Pixel Count, Mean And Total Of Intensities Extracted From Partitioned Equalized Histogram For CBIR", IJEST ISSN : 0975-5462 Vol. 4 No.03 March 2012.

[15]. Dr. H. B. Kekre, KavitaSonawane,"Bins Approach To Image Retrieval Using Statistical Parameters Based On Histogram Partitioning of R, G, B Planes", International Journal of Advances in Engineering & Technology, Vol. 2, Issue 1, pp. 649-659Jan 2012. ©IJAET ISSN: 2231-1963.

[16]. Dr. H. B. Kekre, KavitaSonawane,"Feature Extraction in the form of Statistical Moments Extracted to Bins formed using Partitioned Equalized Histogram for CBIR", International Journal of Engineering and Advanced Technology (IJEAT) ISSN: 2249 – 8958, Volume-1, Issue-3, February 2012.

[17]. Gi-HyoungYoo, BeobKyun Kim, "Content-Based Image Retrieval Using Shifted Histogram" ICCS 2007, Part III, LNCS 4489, pp. 894–897, 2007© Springer-Verlag Berlin Heidelberg 2007.

[18]. Nam Yee Kim, Kang Soo You, Gi-HyoungYoo "An Efficient Histogram Algorithm for Retrieval from Lighting Changed-Images", 2008 Second International Conference on Future Generation Communication and Networking Symposia.

[19]. NST Sai , RavindraC.Patil "Image Retrieval using Equalized HistogramImage Bins Moments", International J. of Recent Trends in Engineering and Technology, Vol. 3, No. 1, May 2010.

[20]. Dr. Thomas Ottmann "Feature Histograms for Content-Based Image Retrieval", Dissertation work at the Albert-Ludwigs-Universit¨at Freiburg, Germany in 2002.

[21]. Wei-Min Zheng, Zhe-Ming Lu, Color Image Retrieval Schemes Using Index Histograms Based On Various Spatial-Domain Vector Quantizers" International Journal of Innovative Computing, Information and Control ICIC International °c 2006 ISSN 1349-4198, Volume 2, Number 6, December 2006.

[22]. V. Vijaya Kumar1, N. GnaneswaraRao, IHBM: Integrated Histogram Bin Matching For Similarity Measures of Color Image Retrieval, International Journal of Signal Processing, Image Processing and Pattern Recognition, Vol. 2, No.3, September 2009.

[23]. Dr. H. B. Kekre, KavitaSonawane, "Effect of Similarity Measures for CBIR Using Bins Approach", International Journal of Image Processing (IJIP), Volume (6) : Issue (3) : 2012.

[24]. Dr. H. B. Kekre, KavitaSonawane, "Retrieval of Images Using DCT and DCT Wavelet Over Image Blocks", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 2, No. 10, 2011.

[25]. Greg Pass RaminZabih, "Histogram Refinement for Content-Based Image Retrieval" 0-8186-7620-5/96 , 1996 IEEE. http:/www.cs.cornell.edu/home/rdz/refinement.html,

[26]. P.S.Suhasini , Dr. K.Sri Rama Krishna, "CBIR Using Color Histogram Processing", Journal of Theoretical and Applied Information Technology, www.jatit.org, © 2005 – 2009.

[27]. MarjoMarkkula, Marius Tico, "A Test Collection for the Evaluation of Content-Based Image Retrieval Algorithms "A User and Task-Based Approach"

[28]. Julia Vogela,,BerntSchiele, "Performance evaluation and optimizationfor content-based image retrieval", 0031-3203/$30.00 _ 2005 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

[29]. Lucia Ballerini1, Xiang Li, "A Query-by-Example Content-Based Image Retrieval System of Non-Melanoma Skin Lesions",homepages.inf.ed.ac.uk/rbf/PAPERS/MCBR09.pdf

Dr. H. B. Kekre has received B.E. (Hons.) in Telecomm. Engg.from Jabalpur University in 1958,M.Tech (Industrial Electronics) from IIT Bombay in 1960, M.S. Engg. (Electrical Engg.)from University of Ottawa in 1965 and Ph.D. (System Identification) from IIT Bombay in 1970. He has worked Over 35 years as Faculty of Electrical Engineering and then HOD Computer Science and Engg. at IIT Bombay. For last 13 years worked as a Professor in Department of Computer Engg. atThadomalShahani Engineering College, Mumbai. He is currently Senior Professor working with Mukesh Patel School of Technology Management and Engineering, SVKM's NMIMS University, Vile Parle(w), Mumbai, INDIA. He has guided 17 Ph.D.s, 150 M.E./M.Tech Projects and several B.E./B.Tech Projects. His areas of interest are Digital Signal processing, Image Processing and Computer Networks. He has more than 450 papers in National / International Conferences / Journals to his credit. Recently twelve students working under his guidance have received best paper awards. Seven of his students have been awarded Ph. D. of NMIMS University. Currently he is guiding eight Ph.D. students. He is member of ISTE and IETE.

Ms. Kavita V. Sonawane has received M.E (Computer Engineering) degree from Mumbai University in 2008, currently Pursuing Ph.D. from Mukesh Patel School of Technology, Management and Engg, SVKM's NMIMS University, Vile-Parle (w), Mumbai, INDIA. She has more than 10 years of experience in teaching. Currently working as a Assistant professor in Department of Computer Engineering at St. Francis Institute of Technology Mumbai. Her area of interest is Image Processing, Data structures and Computer Architecture. She has 14 papers in National/ International conferences / Journals to her credit. She is member of ISTE.