Performance of 64 Bins Formed by Partitioning the Histogram using LP and CG Techniques for CBIR

Dr. H. B. Kekre, Kavita Sonawane

Abstract—This paper explores the new technique for CBIR based on bins approach. Bins approach is based on the novel idea of dividing the image histograms into four parts to form 64 bins. Each image is separated into three planes and a histogram of each of them is calculated. Each histogram is divided into four parts with part ids 0, 1, 2 and 3. Three planes, each divided in four parts generates 4^3 bins. Initially these bins have the count of pixels from the image under feature extraction process. These contents are further processed by computing the first four moments namely Mean, Standard deviation, Skewness and Kurtosis. This is generation of four types of feature vectors for each image. Bins formation is experimented with two different techniques. One is linear partitioning (LP) and the other is centre of gravity (CG) partitioning. Based on this multiple feature vector databases for database of 2000 BMP images are prepared as preprocessing work. Query fired to the proposed system is compared with the database images by means of three similarity measures Euclidean distance, Absolute distance and Cosine correlation distance. Performance of proposed approaches is evaluated using three paramemeters namely Precision Recall Cross over Point (PRCP), Longest String (LS) and Length of String to Retrieve all Relevant (LSRR). Performance of both partitioning in 64 bins formation is observed and compared so that the better one can be recommended.

Index Terms— Absolute distance, Bins approach, CBIR, Centre of Gravity, Cosine Correlation distance, Euclidean distance, Linear Partitioning, LS, LSRR, Mean, PRCP, Skewness.

1 INTRODUCTION

THIS paper explores new CBIR technique based on bins approach. The goal of any CBIR system is to make efficient retrieval of a set of images that are similar to a query image from the large image database. The term similarity is in the sense of visually perceived similarity in content as felt by a common user. Explicitly identifying and entering descriptive keywords for each image manually is impractical. Because the creation of text-based queries is time consuming. It is not possible to generate a complete set of keywords for each image where at the same time user has to keep track that it should not catch the description about the other images in the database. It requires the domain experts to observe and analyze the image contents or information so that correct description for them can be created for the indexing process. Many times the content will not be matched if the users of the system and the domain experts are using the different languages for describing the images. Text based retrieval is tedious and time-consuming task which involves this textual annotations. To overcome these drawbacks and limitations vast research is going on this direction. Speed of technology advancement along with the enormous growth of digital information, capturing and storage devices and internet usage are generating the need of automation in the retrieval process for the user satisfaction and time saving. These needs are actually motivating the researcher's to explore some new techniques to provide the automation in this field of image retrieval. This implies that automated Content-Based Image Retrieval (CBIR) systems need to be developed.

The goal of our research is in the same direction, which explores the novel techniques for the image retrieval from very large databases based on image contents. Many CBIR systems have been developed in last few decades like QBIC , NeTra , MARS , Blobworld, PicToSeek and SIMPLIcity, MIT Photobook, Berkeley Chabot Stanford WBIIS Virage System Columbia VisualSEEK and WebSEEK Systems. Each of these systems are exploring different techniques to extract image contents to form the feature vector vectors and overcoming the drawbacks of text based retrieval systems [1-7].

Image contents are classified into two categories local and global. Local contents can be directly extracted from the image that are color, texture and shape information of the image. Global contents can be extracted by performing some computations on local contents or image information so that they can be represented and used efficiently for the images to compare. The global features include signatures, histograms, Eigen images, wavelet representation etc. Various approaches in the CBIR literature have used color and texture information as they are invariant to scaling and rotation.[8-11]. There are different techniques identified by many researchers for CBIR from two different domains, frequency domain [12-15] and spatial domain [16-17]. In frequency domain various transforms and wavelets are used in many different ways to extract the image features. Spatial domain includes the CBIR techniques based on Gaussian, Laplacian Pyramids, color coherence vectors (CCV) and many more. [14-19]. This paper is proposing the approach called "bins approach" for CBIR. Bins

- *Dr. H. B. Kekre Sr. Professor, Computer Engineering Department MPSTME, NMIMS University, Mumbai, India, E-mail: hbkekre@yahoo.com*
- *Ms. Kavita Sonawane is currently pursuing Ph. D in Computer engineering in NMIMS University, Mumbai, India, E-mail: kavitavinaysonawane@gmail.com*

approach explored here is based on partitioning of the R, G and B histograms of the color (BMP) images in the database. It uses two different partitioning methods namely LP and CG partitioning and obtains 64 bins out of 256 bins of each of the R, G and B histograms of the image under process. Feature vector used for comparing the query and database images is none other than these set of 64 bins. Initially these bins are having the count of pixels falling in the specific partitions of R, G and B histograms. These bins data is further processed to extract and represent image contents in various different types. We have computed statistical moments for the intensities being counted into each of these 64 bins. First four statistical moments are generating 4 different types of feature vectors namely Mean, Standard deviation, Skewness and Kurtosis. While computing these moments we have consider each color intensity i.e red, green and blue separately. This way we have prepared multiple feature vector databases. This is repeated for both LP and CG partitioning techniques. Comparison of query and database image feature vector is carried out by means of three similarity measures namely Euclidean distance, absolute distance and cosine correlation distance. Performance of all types of feature vectors with respect to moments, color and partitioning is evaluated and compared using three different parameters Precision Recall Cross over Point (PRCP), Longest String (LS), and Length of String to Retrieve all Relevant (LSRR).

The paper is further organized as follows: Section 2 describes the histogram and its LP and CG Partitioning. Feature extraction process and types of feature vectors are explained in section 3. Similarity measures and performance evaluation parameters are explained in section 4 and 5 respectively. Experimental results and discussions are presented in section 6 that is followed by conclusion in section 7.

2 R, G, B HISTOGRAMS AND PARTITIONING

2.1 R, G and B Histograms

Feature extraction is core phase of the CBIR systems. In this work the feature extraction mainly deals with the histogram of the image. Histogram is representing the frequency of the pixels at each intensity level [19-21].

In this work its starts with the separation of R, G and B planes of image under feature extraction process. Then compute the histogram for each plane. Figure1. Shows the Bus image with its R, G and B planes. Figure 2 is showing the R, G and B histograms computed for the R G and B planes. In each histogram plot x-axis is presenting the color intensity i.e.(R, G and B) from 0 to 255 and y-axis is representing the frequency of each intensity level in terms of counting the number of pixels of each intensity level[21- 25]

2.2 LP and CG Partitioning of R,G and B histograms

In the feature extraction process second important step we execute is bins formation process. Instead of using all 256 bins of each (R, G and B) histograms we are using just 64 bins out of it. It reduces the dimension of the feature vector, which reduces the computational complexity. These 64 bins are not selected randomly; it follows the partitioning method to generate these 64 bins. It includes the partitioning of the histogram. According to linear partitioning (LP) the histograms are partitioned just by taking the equal number of pixels into each partition. LP partitioning into 4 parts. This way three histograms, four parts each will generate 4^3=64 bin addresses. Each of these will hold the count of pixels. Figure 3a shows the LP partitioning of green histogram. LP partitioning takes care that each partition should get equal number of pixels. In this case we found that this idea of LP partitioning is ignoring the pixel intensities, it considers only the pixel count. Pixels intensities are the important contents of an image and to consider the importance of them we have used the other partitioning technique based on centre of gravity (CG). This partitioning divides the pixels in four parts such that each partition will have equal weight in terms of pixel intensities. It may have random no of pixel distribution in each partition 1, 2, 3 or 4. Figure 3b shows the CG partitioning of green histogram. Each partitioning generates 4 parts where each will get a part id 0, 1 , 2 and 3 respectively as shown in Figures 3a and 3b.

This process will be repeated for all three planes of each image and we obtain 64 bins for each image that can be addressed from 000 to 333. Performance of both partitioning techniques is observed, analyzed and compared so that which on is better in 64 bins formation can be proved.

Figure 1. Original Bus Image with its Separated R, G and B planes

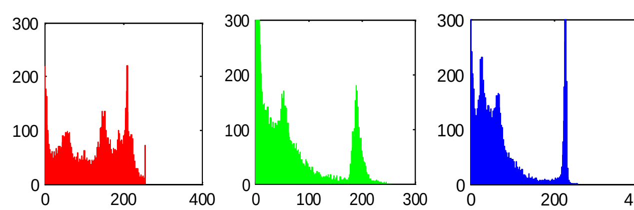

Figure 2. R, G and B Histograms of R, G and B Planes

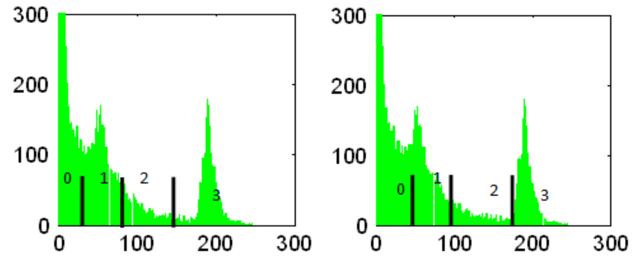

Figure 3a. LP Partitioning 3b. CG Partitioning

3 Moments as Features Extracted to 64 Bins

Feature extraction is core phase of all CBIR systems. This paper proposed a new method of forming 64 bins from R, G and B histogram of the image under feature extraction process. Initially the 64 bins are holding just the count of pixels which is not containing the meaningful information of the image. To represent and to make effeicent use of image contents; statistical properties are computed for the intensities of the pixels counted into each of the 64 bins. Computation of these statistical features is nothing but the calculation of first four moments namely Mean, Standard deviation (STD), skewness (SKEW) and kurtosis (KURTO).

We have obtained variety of features for each image of the database. As preprocessing part we have prepared multiple feature vector databases for all database images by executing the following sequence of steps.

Step1. Read and spilt the image into R, G and B Planes

Step2. Compute R, G and B histograms for R, G and B plane respectively.

Step3. Partition the histogram either using (CG or LP) into four parts.

Sept4. Generate the 64 bins for each image and extract the count of pixels into each bin.

Sept5. Compute the first four moments for the total intensities counted into each bin separately. In this step we compute the moments MEAN, STD, SKEW and KURTO for each color intensity (R, G and B) separately. e.g for Red, green and *blue color intensities the moments calculated are as follows:*

R-color : *Rmean, Rstd, Rskew and Rkurto*
G-Color : *Gmean, Gstd, Gskew and Gkurto*
B-Color : *Bmean, Bstd, Bskew and Bkurto*

Step6: Each feature vector based on different color and moment is stored in separate feature vector database.

The above steps are executed for both partitioning techniques LP and CG and in this way we have prepared multiple feature vector databases for all images in the database.

4. Similarity Matching and Retrieval

Once the preprocessing part is done, the system is ready for the next stage to accept the user query and generate the retrieval results for the same. Whenever query enters into the system , feature vector will be calculated for it in the same manner as it was followed in case of all database images. Now the system will try to match the query feature vectors with feature vectors in the feature database. This matching process is carried out using three similarity measures namely Euclidean distance (ED), absolute distance (AD) and cosine correlation distance (CD) [26-27]. The equations for them are given in equations 1, 2 and 3 as follows:

EUCLIDEAN DISTANCE

$$D_{QI} = \sqrt{\sum_{i=1}^{n} \left|(FQ_i - FI_i)\right|^2} \qquad (1)$$

ABSOLUTE DISTANCE:

$$D_{QI} = \sum_{i=1}^{n} \left|(FQ_i - FI_i)\right| \qquad (2)$$

COSINE CORRELATION DISTANCE

$$\frac{\langle D(n)|Q(n)\rangle}{\sqrt{\left[\langle D(n)|D(n)\rangle\langle Q(n)|Q(n)\rangle\right]}} \qquad (3)$$

where D(n) and Q(n) are Database and Query feature Vectors resp.

The distance between query and database images are calculated and sorted in ascending order for minimum to maximum distance. Then the retrieval set is determined by counting the number of images close to query images (at minimum distance) and retrieve them from the database.

5. Performance Evaluation

Whenever a new algorithm is being proposed and implemented, its performance has to be evaluated so that the efficiency of that can be determined and its effective use can be recommended. Performance of the proposed approaches discussed in this paper is evaluated using three parameters as Precision Recall Cross over Point (PRCP), Longest String (LS), Length of String to Retrieve all Relevant (LSRR). They re define as follows:

$$\text{Precision} = \frac{\text{Number of Relevant Images Retreived}}{\text{Total Number of Images Retreieved}} \qquad (4)$$

$$\text{Recall} = \frac{\text{Number of Relevant Images Retreived}}{\text{Total Number of Relevant Images in Database}} \qquad (5)$$

- **LSRR**: Defines the length of the traversal of sorted distances to recall all relevant images; and
- **Longest String**: Defines the continuous longest string of relevant images collected from sorted distances.

6. Experimental Results and Discussion

CBIR technique explored in this paper is based on bins approach. It generates 64 bins from 256 bins of (three histograms) R, G and B histograms. This approach is based on color contents of the image. Then we are computing some textural properties of the data extracted to 64 bins. These are defined as first four moments representing four types of feature vectors for each image for each color (R, G and B) separately.
Experimentation is carried out with the following details.

6.1 Database and Query Image

Image database used for experimentation is consist of 2000 BMP images including images from 20 different classes. It includes few classes from Wang database. Each class has 100 images. Sample image from each of the 20 classes is shown in Figure 4.

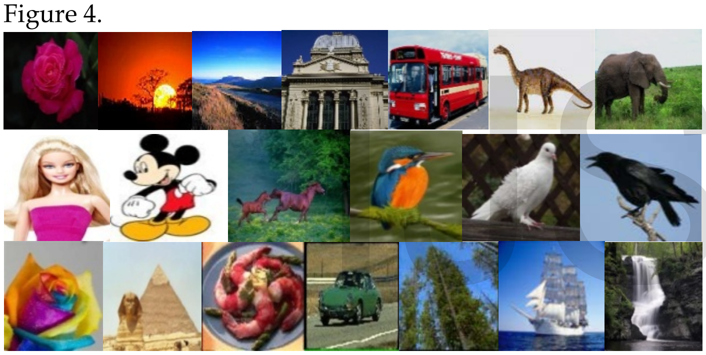

Figure 4. Sample of Image Database having 20 classes.

For all the mentioned database images, system calculates the all types of feature vectors. Each type of feature vector is stored separately and separate feature databases are prepared for all 2000 images in the database. This is the preprocessing part of this CBIR system before the query enters into the system.

Query image: The image can be given as input to the system. There are various approaches to give input to CBIR the system. Some systems are designed to support image contents as input, query by sketch or query by example. This system is designed to support the input as query by example etc. User can give any image as query image to the system. System calculates the feature vector for the same as it calculates for the database images.

To check and evaluate the system performance for all types of feature vectors we have used same set of 200 query images. It includes 10 images selected randomly from each of the 20 classes of image database. Same set of queries are executed for all approaches discussed above and results are presented in following section.

6.2 Query Matching and Retrieval

As discussed in sections discussed above, as preprocessing part multiple feature vector databases are prepared before the query enters into the system. Whenever user fires the query image to the system, all types of feature vectors e.g Rmean, Rstd, Rskew an Rkurto for Red color contents similarly the same moments will be calculated for green and blue colors too. This same process is repeated for each type of histogram partitioning technique, i.e LP and CG. Each type query feature vector will be compared with all respective types of features in the feature vector databases. Comparisons are carried out using three similarity measures ED, AD and CD. All distances are sorted in ascending order and system is taking first 100 images to find out images relevant to query. This count of query relevant images in first 100 images in nothing but the result obtained for parameter PRCP. Here Precision and recall both are same. Further these results are evaluated using two more parameters as mentioned earlier LS and LSRR.

6.3 Results and Discussion

As per the discussion and steps followed in experimental set up; set of 200 queries are fired to the system which are compared by means of ED, AD and CD with all database images and results are presented as follows in terms of three parameters PRCP, LS and LSRR.

6.3.1 PRCP Results

Note: Each numerical value given in table is representing the count of relevant images.

TABLE 1
PRCP: MEAN

	R		G		B	
SM	CG	LP	CG	LP	CG	LP
ED	5103	4949	5747	5260	5578	5460
AD	6430	5930	6520	5720	6343	5819
CD	4712	5066	5549	5596	5410	5795

TABLE 2
PRCP: STD

	R		G		B	
SM	CG	LP	CG	LP	CG	LP
ED	5397	4370	5948	5349	6199	5639
AD	6404	5552	6719	6021	6899	6187
CD	5239	4641	6176	5604	6359	5828

TABLE 3
PRCP: SKEW

	R		G		B	
SM	CG	LP	CG	LP	CG	LP
ED	5053	4202	5345	5028	5769	5419
AD	5877	5279	6109	5618	6411	5977
CD	4773	4348	5353	5204	5802	5600

ISSN 2229-5518

TABLE 4
PRCP: KURTO

	R		G		B	
SM	**CG**	**LP**	**CG**	**LP**	**CG**	**LP**
ED	5649	4548	6234	5601	6442	5855
AD	6535	5802	6024	6186	7014	6389
CD	5547	4849	6313	5799	6526	6001

In all the above results, each entry in the table is total retrieval in terms of total PRCP for execution of 200 query images. Comparing results on partitioning basis we found that CG is performing better as compared to LP in all cases except at 4 places out of 36. Better results are highlighted with yellow color. When we compared these results with ideal PRCP value i.e 20,000 these results are indicating very less retrieval for each color result separately. To improve these results we thought of combining these R, G and B results by using OR criteria so that the results can be improved further.

In following table 5, we have shown the result after OR ing the PRCP results of R, G and B colors. Here we can observe that the results are improved to very good extent as compared to previous. Before applying OR criteria PRCP was around 0.3 as average of 200 query images; Where as it is now crossing 0.5 as average of 200 query images.

TABLE 5
PRCP: 'R' OR 'G' OR 'B'

	ED		AD		CD	
	CG	**LP**	**CG**	**LP**	**CG**	**LP**
MEAN	8422	8088	9418	8597	7853	8093
STD	9408	8355	10110	9207	9243	8505
SKEW	8990	8105	9710	8971	8762	8637
KURTO	9681	8637	**10225**	9458	9496	8778

In above results the consistency that CG is better than LP is maintained and the better results are highlighted with yellow color. Now clearly we can see that the PRCP i.e precision and recall both are same and have crossed 0.5 as an average of 200 query images. This is indicated in the best among all results is highlighted in red color i.e total retrieval is 10225 (for KURTO with AD measure with CG partitioning) out of 20,000 for the execution of 200 query images.

6.3.2 LS Results

Next parameter used for performance evaluation is longest string of i.e. continuous sequence of relevant images in image set sorted according to distances (Min to Max). CBIR users will always look for the maximum retrieval of similar images for the given query. This parameter is trying to address the same for the user's satisfaction.

TABLE 6
MEAN : LONGEST STRING

64_CG ORG	**CD**		**ED**		**AD**	
Flower	24	B	18	B	19	R
Sunset	22	G	14	B	17	B
Mountain	4	G	4	R	4	R
Building	7	B	7	R	10	G
Bus	8	G	10	B	21	B
Diansour	13	B	12	G	51	R
Elephant	7	R	7	R	7	B
Barbie	38	R	37	R	40	R
Mickey	8	G	18	R	30	R
Horses	26	B	28	B	30	B
Kingfisher	11	G	12	G	7	G
Dove	40	G	44	G	55	G
Crow	6	G	6	G	6	R
Rainbowrose	42	G	46	G	52	B
Pyramids	14	R	11	R	15	B
Plates	5	G	5	R	7	B
Car	16	G	16	G	11	G
Trees	10	B	13	B	12	B
Ship	7	G	12	G	13	G
Waterfall	6	R	10	B	7	R
AVG	15.4		16		20.35	
RGBCOUNT	4 11 5		7 7 6		7 5 8	

TABLE 7
MEAN : LONGEST STRING

64_LP ORG	**CD**		**ED**		**AD**	
Flower	24	B	16	B	19	B
Sunset	15	R	23	R	18	R
Mountain	4	R	6	B	5	G
Building	4	R	4	R	5	G
Bus	16	G	13	B	16	R
Diansour	34	G	25	B	44	B
Elephant	8	G	8	R	8	R
Barbie	47	R	44	R	38	R
Mickey	15	B	17	B	20	B
Horses	51	B	32	B	30	B
Kingfisher	18	B	12	B	9	B
Dove	29	G	30	R	28	R
Crow	7	B	8	B	9	B
Rainbowrose	30	B	36	B	37	B
Pyramids	11	R	10	R	22	R
Plates	5	R	8	B	5	G
Car	10	G	8	G	8	G
Trees	7	B	14	B	13	B
Ship	16	G	15	G	15	G
Waterfall	6	R	7	B	7	R
AVG	17.85		16.45		17.45	
RGBCOUNT	7 6 7		6 2 12		7 5 8	

64_CG ORG	TABLE 8 STD : LONGEST STRING					
	CD		ED		AD	
Flower	32	B	29	B	31	B
Sunset	15	B	13	B	24	G
Mountain	4	G	4	G	4	R
Building	7	R	6	G	6	R
Bus	26	G	24	G	28	G
Diansour	54	G	61	B	66	G
Elephant	10	G	10	G	10	G
Barbie	46	R	57	R	60	R
Mickey	27	R	27	R	32	R
Horses	30	G	25	G	34	B
Kingfisher	11	B	8	G	11	B
Dove	53	B	52	B	56	R
Crow	19	G	8	G	10	R
Rainbowrose	46	B	42	G	40	G
Pyramids	27	G	20	G	28	R
Plates	6	G	8	B	8	B
Car	6	R	9	R	8	B
Trees	17	R	20	R	22	R
Ship	11	B	13	R	14	R
Waterfall	11	R	10	R	11	R
AVG	22.35		21.8		24.6	
RGBCOUNT	6 8 6		6 9 5		10 5 5	

64_CG ORG	TABLE 10 SKEW : LONGEST STRING					
	CD		ED		AD	
Flower	32	B	24	B	24	R
Sunset	19	B	16	B	13	R
Mountain	5	B	5	G	5	G
Building	11	R	7	G	9	B
Bus	12	G	17	G	14	B
Diansour	32	G	50	G	66	B
Elephant	9	G	7	G	8	G
Barbie	48	R	62	R	61	R
Mickey	16	R	23	G	24	R
Horses	24	G	20	B	27	B
Kingfisher	7	G	10	B	8	B
Dove	53	R	45	B	51	B
Crow	7	G	6	R	7	R
Rainbowrose	40	G	41	G	32	G
Pyramids	25	G	19	G	22	G
Plates	6	B	10	B	7	B
Car	5	R	7	R	10	B
Trees	16	B	16	G	16	G
Ship	11	B	10	G	10	G
Waterfall	8	G	6	R	8	R
AVG	18.9		19.75		20.7	
RGBCOUNT	5 9 6		4 10 6		6 6 8	

64_LP ORG	TABLE 9 STD : LONGEST STRING					
	CD		ED		AD	
Flower	29	B	34	B	33	B
Sunset	12	G	13	B	19	G
Mountain	5	G	7	R	4	R
Building	7	R	5	R	6	R
Bus	12	B	11	G	13	G
Diansour	39	G	65	G	64	G
Elephant	6	R	6	G	7	B
Barbie	28	R	30	R	28	B
Mickey	9	B	10	R	8	B
Horses	45	B	30	B	43	B
Kingfisher	20	G	16	B	21	B
Dove	28	G	23	G	40	G
Crow	12	B	8	B	12	B
Rainbowrose	28	G	32	G	31	G
Pyramids	18	G	15	G	12	G
Plates	4	R	5	B	6	B
Car	12	B	9	G	8	B
Trees	12	G	10	B	11	B
Ship	8	R	8	R	13	B
Waterfall	5	R	8	R	9	G
AVG	16.95		16.85		18.95	
RGBCOUNT	6 8 6		6 7 6		2 7 11	

64_LP ORG	TABLE 11 SKEW : LONGEST STRING					
	CD		ED		AD	
Flower	23	B	26	B	31	R
Sunset	16	R	14	B	21	G
Mountain	6	B	5	R	4	R
Building	5	G	4	R	5	R
Bus	14	G	11	B	11	G
Diansour	45	G	48	G	59	G
Elephant	9	R	6	R	7	R
Barbie	27	R	32	B	25	R
Mickey	11	B	10	B	8	R
Horses	29	B	34	B	40	B
Kingfisher	10	G	6	G	11	G
Dove	32	G	24	G	50	G
Crow	13	G	12	G	10	G
Rainbowrose	24	G	31	B	34	B
Pyramids	15	G	12	G	13	G
Plates	6	G	7	B	4	B
Car	12	B	9	B	9	B
Trees	16	B	10	B	14	G
Ship	9	B	9	R	10	B
Waterfall	7	R	6	R	8	B
AVG	16.45		15.5		18.3	
RGBCOUNT	4 9 7		5 5 10		6 8 6	

64_CG ORG	TABLE 12 KURTO : LONGEST STRING					
	CD		ED		AD	
Flower	29	B	37	B	35	B
Sunset	17	G	18	G	23	G
Mountain	5	B	4	R	5	B
Building	7	R	7	R	7	R
Bus	29	G	23	G	28	G
Diansour	63	G	61	B	69	R
Elephant	11	G	10	G	10	G
Barbie	42	R	62	R	60	R
Mickey	30	R	29	R	39	R
Horses	35	G	34	B	36	B
Kingfisher	8	B	7	B	12	B
Dove	59	B	56	B	59	B
Crow	11	B	9	B	8	R
Rainbowrose	53	B	46	G	35	G
Pyramids	27	G	22	G	32	R
Plates	7	B	7	B	12	B
Car	9	R	8	R	7	R
Trees	17	G	22	R	23	R
Ship	13	B	11	B	10	R
Waterfall	11	R	11	R	7	R
AVG	23.6		23.65		22.5	
RGBCOUNT	5 7 8		7 5 8		10 4 6	

64_CG ORG	TABLE 14 MEAN : LSRR					
	CD		ED		AD	
Flower	83	B	51	G	38	B
Sunset	45	R	77	G	69	G
Mountain	71	B	60	B	61	G
Building	60	G	58	G	53	G
Bus	65	B	57	B	45	R
Diansour	42	R	52	B	26	B
Elephant	89	R	78	R	53	G
Barbie	34	B	87	R	18	B
Mickey	82	G	69	B	78	G
Horses	68	G	51	B	40	G
Kingfisher	54	G	77	B	73	G
Dove	85	B	52	B	62	G
Crow	99	B	89	G	57	B
Rainbowrose	55	B	62	R	81	B
Pyramids	82	R	78	R	43	G
Plates	75	B	67	G	71	G
Car	57	G	61	B	64	B
Trees	81	G	54	G	39	G
Ship	57	B	62	B	76	G
Waterfall	87	B	79	G	73	R
AVG	68.55		66.05		56	
RGBCOUNT	4 6 10		4 7 9		2 12 6	

64_LP ORG	TABLE 13 KURTO : LONGEST STRING					
	CD		ED		AD	
Flower	35	B	37	B	37	R
Sunset	14	G	16	R	20	G
Mountain	4	R	4	G	7	B
Building	7	R	7	R	6	R
Bus	26	B	14	B	13	G
Diansour	39	G	45	G	54	G
Elephant	10	R	6	G	6	R
Barbie	32	R	32	R	33	R
Mickey	13	B	10	R	14	R
Horses	52	B	43	B	43	B
Kingfisher	14	G	11	G	21	B
Dove	28	G	26	G	43	B
Crow	11	G	10	G	10	G
Rainbowrose	35	G	33	G	34	G
Pyramids	20	G	13	G	13	R
Plates	5	R	5	B	7	B
Car	15	B	11	B	12	B
Trees	10	G	12	B	14	B
Ship	8	G	8	B	13	B
Waterfall	5	R	6	R	8	G
AVG	19.15		17.15		20	
RGBCOUNT	6 9 6		5 8 7		6 6 8	

64_LP ORG	TABLE 15 MEAN : LSRR					
	CD		ED		AD	
Flower	47	B	56	G	71	G
Sunset	66	B	62	B	56	B
Mountain	62	B	66	B	57	B
Building	55	G	56	G	48	G
Bus	51	B	45	B	35	B
Diansour	66	G	56	B	32	B
Elephant	71	B	61	B	46	G
Barbie	19	R	93	R	82	R
Mickey	91	G	95	R	91	B
Horses	28	B	62	G	65	G
Kingfisher	56	B	83	B	85	B
Dove	90	R	49	G	43	G
Crow	100	R	100	R	100	R
Rainbowrose	92	G	90	G	90	G
Pyramids	71	R	67	G	46	R
Plates	63	B	59	G	62	G
Car	65	B	59	B	57	B
Trees	57	B	54	G	51	G
Ship	45	B	45	B	51	G
Waterfall	60	B	68	G	55	R
AVG	62.75		66.3		61.15	
RGBCOUNT	4 4 12		9 8 3		4 9 4	

64_CG ORG	TABLE 16 STD : LSRR					
	CD		ED		AD	
Flower	83	B	51	G	38	B
Sunset	45	R	77	G	69	G
Mountain	71	B	60	B	61	G
Building	60	G	58	G	53	G
Bus	65	B	57	B	45	R
Dinosaur	42	R	52	B	26	B
Elephant	89	R	78	R	53	G
Barbie	34	B	87	R	18	B
Mickey	82	G	69	B	78	G
Horses	68	G	51	B	40	G
Kingfisher	54	G	77	B	73	G
Dove	85	B	52	B	62	G
Crow	99	B	89	G	57	B
Rainbowrose	55	B	62	R	81	B
Pyramids	82	R	78	R	43	G
Plates	75	B	67	G	71	G
Car	57	G	61	B	64	B
Trees	81	G	54	G	39	G
Ship	57	B	62	B	76	G
Waterfall	87	B	79	G	73	R
AVG	68.55		66.05		56	
RGBCOUNT	4 6 10		4 7 9		2 12 6	

64_CG ORG	TABLE 18 SKEW : LSRR					
	CD		ED		AD	
Flower	62	G	36	G	39	B
Sunset	72	G	63	B	69	G
Mountain	89	R	89	R	66	G
Building	90	G	66	G	61	G
Bus	71	G	61	G	48	G
Dinosaur	74	R	86	R	31	B
Elephant	79	G	91	B	64	G
Barbie	63	B	71	B	18	B
Mickey	81	B	80	B	81	G
Horses	50	G	47	G	44	G
Kingfisher	70	B	82	G	76	G
Dove	75	B	62	G	65	G
Crow	77	B	82	B	60	B
Rainbowrose	49	G	48	G	84	B
Pyramids	65	R	83	G	47	G
Plates	86	B	69	G	71	G
Car	85	B	64	B	57	B
Trees	57	B	49	G	38	G
Ship	70	B	80	G	75	G
Waterfall	83	B	80	R	70	R
AVG	72.4		69.45		58.2	
RGBCOUNT	3 7 10		3 11 6		1 13 6	

64_LP ORG	TABLE 17 STD : LSRR					
	CD		ED		AD	
Flower	60	B	53	B	59	B
Sunset	60	G	64	B	71	B
Mountain	76	R	84	G	76	B
Building	78	B	54	G	46	G
Bus	48	B	53	G	49	G
Diansour	84	G	83	G	25	G
Elephant	69	G	66	G	71	G
Barbie	36	B	59	B	16	B
Mickey	84	G	83	G	92	G
Horses	24	G	39	G	34	B
Kingfisher	79	B	83	B	85	B
Dove	67	G	53	G	45	B
Crow	84	G	86	B	51	G
Rainbowrose	91	B	94	G	99	R
Pyramids	65	G	63	G	38	G
Plates	58	G	51	G	57	G
Car	72	B	62	B	63	B
Trees	57	G	44	G	38	G
Ship	58	B	61	B	53	B
Waterfall	73	G	73	G	62	R
AVG	68.55		65.4		56.5	
RGBCOUNT	1 11 8		0 13 7		2 9 9	

64_LP ORG	TABLE 19 SKEW : LSRR					
	CD		ED		AD	
Flower	61	B	56	G	59	B
Sunset	67	R	61	B	69	B
Mountain	69	B	80	B	70	B
Building	80	B	59	G	49	G
Bus	56	G	56	B	53	B
Diansour	92	B	92	G	39	G
Elephant	71	G	69	G	73	G
Barbie	57	R	69	B	24	B
Mickey	85	B	83	G	93	G
Horses	29	G	52	G	42	B
Kingfisher	74	B	83	B	90	B
Dove	67	B	55	G	44	G
Crow	86	G	89	B	51	G
Rainbowrose	90	G	93	G	99	R
Pyramids	65	G	64	G	36	G
Plates	60	G	57	G	61	G
Car	75	B	64	B	60	B
Trees	60	G	48	G	38	G
Ship	55	B	65	B	56	B
Waterfall	79	B	74	G	65	R
AVG	68.9		68.45		58.55	
RGBCOUNT	2 8 10		0 12 8		2 9 9	

64_CG ORG	TABLE 20 KURTO : LSRR					
	CD		ED		AD	
Flower	53	B	39	B	35	B
Sunset	58	G	66	B	59	G
Mountain	81	R	85	B	60	G
Building	84	G	57	G	56	G
Bus	52	G	50	G	41	R
Diansour	64	R	75	B	22	B
Elephant	70	G	85	G	50	G
Barbie	49	B	62	B	10	B
Mickey	68	B	62	B	76	G
Horses	54	G	45	G	38	G
Kingfisher	70	R	79	G	74	G
Dove	52	G	64	G	68	G
Crow	69	B	78	B	57	G
Rainbowrose	47	B	62	G	80	R
Pyramids	58	R	73	G	41	G
Plates	61	G	67	G	73	G
Car	67	B	62	B	62	B
Trees	51	B	41	G	39	G
Ship	59	B	73	G	79	G
Waterfall	77	G	75	G	74	R
AVG	62.2		65		54.7	
RGBCOUNT	4 8 8		0 12 8		3 11 4	

64_LP ORG	TABLE 21 KURTO : LSRR					
	CD		ED		AD	
Flower	61	G	58	G	57	B
Sunset	59	G	63	B	76	B
Mountain	71	B	80	G	73	G
Building	75	B	53	G	43	G
Bus	46	G	53	G	53	G
Diansour	80	B	79	B	24	G
Elephant	64	G	64	G	70	G
Barbie	32	B	51	B	15	B
Mickey	84	B	85	G	93	G
Horses	21	G	40	G	34	B
Kingfisher	74	B	83	B	86	B
Dove	65	G	58	G	47	B
Crow	76	B	79	B	46	G
Rainbowrose	91	G	96	G	99	R
Pyramids	58	G	56	G	37	G
Plates	59	B	56	G	61	G
Car	67	B	63	B	67	B
Trees	59	G	44	G	37	G
Ship	50	B	51	B	56	B
Waterfall	72	G	69	G	61	R
AVG	63.2		64.05		56.75	
RGBCOUNT	0 10 10		0 13 7		2 10 8	

Tables 6 to 13 are representing the results for LS obtained for CG and LP partitioning based histogram bins with moments MEAN, STD, SKEW and KURTO respectively. The results obtained are analyzed with respect to color feature i.e R, G and B, which is performing better, is identified in these results by taking the maximum LS into consideration. We found that green color is performing better among all then red and then blue. All best (maximum among all) are highlighted in green color. In second last row of all tables we are showing the average results of the 20 best results obtained as LS for each type of feature vector. The best values obtained for average LS are 24 for kurtosis with AD measure for CG partitioning and it is 20 for LP partitioning.

6.3.3 LSRR Results

Results obtained for LSRR gives you the length required to be traversed to collect all relevant images from database for the given query. That is why; the ideal LSRR should be as low as possible. Ideal LSRR indicates that the system will take less time to retrieve all relevant images from database. CG and LP partitioning based LSRR results obtained for first four moments MEAN, STD, SKEW and KURTO are shown in tables 14 to 21 respectively. Similar to LS; here also the results are analyzed with respect to color and we found that green is best among all three then blue and red. As for LSRR we are looking for minimum value, we have highlighted these results with green color wherever we obtain best LSRR. The average values recorded in the second last row of these tables are showing that the lowest average obtained amongst all is 54% for CG (kurtosis with AD measure) and 56% for LP based results. LS and LSRR can be considered in % as we have 100 images of each class in the database.

Overall observation of LS and LSRR parameters can be interpreted as maximum retrieval as LS has reached to 69% and the minimum traversal required making recall 1, is just 10% of 2000 images. Means the proposed methods are performing quite well for satisfying the needs of CBIR users.

Based on the work done explained in above sections, following conclusions can be drawn.

7. Conclusion

This paper describes the work done for the core phase of CBIR i.e feature extraction and has addressed the important issues like feature vector dimension reduction and retrieval efficiency. Feature vector size is reduced to just 64 bins instead of using the 256 bins of entire histogram, as it is [28-30].

Bins approach explored here is based on LP and CG partitioning; it has been found that CG is far better than LP in almost all results sets.

Considering the types of feature vectors and their performance, we found even moments are far better than odd moments.

Performance comparison with respect to similarity measure we found AD is best among all and CD is better wherever ED and AD both are doing worst.

Observing the results of performance evaluation parameters we found that average retrieval of 200 query images from the database of 2000 images is crossing 0.5 in terms of precision and recall (PRCP). Average LS and LSRR is 24 % and 54% respectively. Best LS is 69 and best LSRR is 10% are indicating the efficiency of the proposed approaches, which are simple to implement but found effective.
Bins approach explored with 64 bins can be used effectively in various CBIR applications where the computational complexity is reduced to quiet good extent along with the improvement in the relevant retrieval.

REFERENCES

[1] A. Gupta, R. Bach, C. Fuller, A. Hampapur, B.Horowitz, R. Jain, C.F. Shu "The Virage image search engine: an open framework for image management in Storage and Retrieval for Image and Video Databases*", IV, Proc SPIE Vol. 2670,* pp 76-87, 1996.

[2] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker.

[3] "Query by image and video content: The QBIC system," *IEEE Computer, vol. 28, pp. 23-32, 1995*.

[4] C. C. Hsu,W. W. Chu, and R. K. Taira, "A knowledge-based approach for retrieving images by content", *IEEE Trans. On Knowledge and Data Engineering*, 8(4):522–532, 1996.

[5] W. Ma and B. Manjunath, "Netra: A toolbox for navigating large image databases. Multimedia Systems", 7:184–198, 1999.

[6] J.Wang, J. Li, and G. Wiederhold. SIMPLIcity: Semantics-Sensitive Integrated Matching for Picture Libraries. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 23(9):1–17, 2001.

[7] M. Ortega, Y. Rui, K. Chakrabarti, S. Mehrotra, and T. Huang. Supporting similarity queries in MARS. *In 5th Int. Multimedia Conference.ACM*, 1997.

[8] K. S. Thyagarajan, T. Nguyen, and C. Persons, "A maximum likelihood aproach to texture classification using wavelet transform", in *Proc. IEEE Int. Conf. on Image Proc., 1994*.

[9] Ziqiang Feng and David Tien, "Enhancement of Semantics in CBIR", Proceedings of the Third International Conference on Information Technology and Applications (ICITA'05) 0-7695-2316-1/05 © 2005 IEEE

[10] Ying Liua, Dengsheng Zhanga, Guojun Lua,Wei-Ying MabY. Liu et al, "A survey of content-based image retrieval with high-level semantics", *Pattern Recognition,* 40 (2007) 262– 282.

[11] Xiaojun Qi,Yutao Han, "A novel fusion approach to content-based image retrieval", *Pattern Recognition Society. Published by Elsevier Ltd*. 2005.

[12] H.B.Kekre, Sudeep D. Thepade, "Image Retrieval using Color-Texture Features Extracted from Walshlet Pyramid", *ICGST - GVIP Journal, ISSN: 1687-398X, Volume 10, Issue 1*, February 2010.

[13] Arnold W.M. Smeulders, Marcel Worring, Simone Santini, , Amarnath Gupta, Member and Ramesh Jain, , "Content-Based Image Retrieval at the End of the Early Years, *IEEE Transactions on Pattern Analysis And Machine Intelligence, Vol. 22, No. 12, December 2000*

[14] Dr. H. B. Kekre, Kavita Sonawane,"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.

[15] Dr. H. B. Kekre, Kavita Sonawane, "Query based Image Retrieval using Kekre, DCT and Hybrid wavelet Transform over 1st and 2nd Moment", *International Journal of Computer Applications* (0975 – 8887) Volume 32– No.4, October 2011.

[16] M. Singha, K. Hemachandran, A. Paul "Content-based image retrieval using the combination of the fast wavelet transformation and the colour histogram", *Published in IET Image Processing,* Process., 2012, Vol. 6, Iss. 9, pp. 1221–1226.

[17] H. B. Kekre, Kavita Patil , "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.

[18] Xiang Yuan and Chang-Tsun Li, "CBIR Approach To Building Image Retrieval Based On Invariant Characteristics In Hough Domain", ICASSP 2008, 1-4244-1484-9/08/$25.00 ©2008 IEEE.

[19] Avi Kak and Christina Pavlopoulou, "Content-Based Image Retrieval from Large Medical Databases", Purdue University, Proceedings of the First International Symposium on 3D Data Processing Visualization and Transmission, 2002 IEEE

[20] Lin Mei, Gerd Brunner, Lokesh Setia, and Hans Burkhardt, "Kernel Biased Discriminant Analysis Using Histogram Intersection Kernel for Content-Based Image Retrieval" IDEAL 2005, LNCS 3578, pp. 63–70, 2005.C Springer-Verlag Berlin Heidelberg 2005.

[21] Szabolcs Sergy´an, Budapest Tech, "Color Histogram Features Based Image Classification in Content-Based Image Retrieval Systems" SAMI 2008 • 6th International Symposium on Applied Machine Intelligence and Informatics 978-1-4244-2106-0/08/$25.00 © 2008 IEEE.

[22] L. Kotoulas and I. Andreadis , "Colour histogram content-based image retrieval and hardware implementation", IEE Proc.-Circuits Devices Syst., Vol. 150, No. 5, October 2003.

[23] Young-jun Song I, Won-bae Park2, Dong-woo Kiml, "Content-based Image Retrieval using new color histogram", 0-7803-8639-6/04/$20.00 ©2004 IEEE.

[24] S.Suhasini , Dr. K.Sri Rama Krishna, CBIR Using Color Histogram Processing", Journal of Theoretical and Applied Information Technology, 2005 – 2009.

[25] George Gagaudakis and Paul L. Rosin, "Incorporating shape into histograms for CBIR" Pattern Recognition, elsevier Volume 35, Issue 1, January 2002, Pages 81–9.

[26] Miguel Arevalillo-Herráez a, Juan Domingo b, Francesc J. Ferri, "Combining similarity measures in content-based image retrieval", Pattern Recognition Letters 29 (2008) 2174–2181, _ 2008 Elsevier B.V.

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

[28] Wee Kheng Leow, Rui Li,"The analysis and applications of adaptive-binning color histograms", Computer Vision and Image Understanding 94 (2004) 67–91. www.elsevier.com.

[29] Sangoh Jeong, Chee Sun, "Image retrieval using color histograms generated by Gauss mixture vector quantization", Computer Vision and Image Understanding 94 (2004) 44–66. www.elsevier.com

[30] Van Hieu Vu, Quynh Nguyen Huu, Ha Nguyen Thi Thu "Content Based Image Retrieval with Bin of Color Histogram", ICALIP2012, 978-1-4673-0174-9/12/$31.00 ©2012 IEEE

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. at Thadomal Shahani 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. Five 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. 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 23 papers in National/ International conferences / Journals to her credit. She is member of ISTE