ISSN:0975-9646

Comparative Evaluation of Cluster based CBIR using different Similarity Measures

Seema Anand Chaurasia[#1], Vaishali Suryawanshi[*2]

1Computer Department, Xavier Institute of Engineering, Mahim
Mumbai University, India

2Computer Department, Thadomal Shahni Engineering College, Bandra
Mumbai University, India

***Abstract*— In this paper, comparative evaluation is made on the result of CBIR system based on different similarity measures like Euclidean distance, City block and Chessboard distances based on cluster based Image retrieval to calculate the deviation from query image. For cluster based Image retrieval LBG algorithm is used. Three experiments are carried out on LBG algorithm by varying number of image categories and by taking Euclidean Distance, City block and Chessboard as similarity measure. K-means algorithm is applied on codebook generated from LBG to create global codebook of query images and global codebook of each image category. Comparisons are made between results obtained from Euclidean Distance, City block and Chessboard similarity measures. In this paper experiments are carried out on COIL database images, first on 720 images having 10 different classes and second on 1440 images of 20 different classes and third on 2160 images on 30 different classes and fourth on 2880 images on 40 different classes by taking 5, 10 and 15 query images from each category based on global codebook created using K-means algorithm on existing LBG codebook..**

***Keywords*— CBIR, QBIC, Precision, Recall, Query Image, Precision, Recall, LBG, Euclidean Distance, Chessboard Distance, City block Distance.**

I. INTRODUCTION

The steady growth of the Internet, the falling prices and easy availability of storage devices, and an increasing pool of available computing power make it necessary and possible to manipulate very large repository of digital information efficiently. Using content based Image retrieval one can search and browse in a large collection of digital image database based on automatically derived image features.

World Wide Web plays an important role in communication, education, industry etc. Size of digital image data is growing rapidly and hence it becomes very important to retrieve and store data efficiently [2, 4]. There are two main approaches for Image retrieval, one based on text based approach and second on content based Image retrieval.

Text based descriptor method is also known as image textual metadata. But there are two main disadvantages related to text-based approach. One is considerable amount of human labour is needed for manual annotation and second one is annotation inaccuracy due to human perception. This method lacks efficiency and simply not practical in databases where number of new images are growing rapidly. The need to locate these images and manage target images in response to user queries has become a very serious problem. Second

approach is Content based Image retrieval (CBIR)[5,6].CBIR is the process of retrieving desired images from a large collection of image database on the basis of color, texture and shape, which can be automatically derived from the images themselves[11,13]. Thus using a Content Based Image Retrieval (CBIR), we can analyze and index images based on their visual contents [12].

One of the main applications of CBIR is category search. In category search user may have a group of images and search is for other additional images of same class or category. In other words, it may be the class that the user has examples and the search is for other elements of the same class. Categories may be derived from the labels or emerge from the database [7, 8].

II. LITERATURE SURVEYED

Search time can be reduced considerably by using clustering [9, 10]. LBG and K-means algorithms are some of the clustering methods which are most widely used. In this section LBG and K-means algorithms are explained.

A. LBG Algorithm

Consider two-dimensional vector space as shown in figure 1.In this algorithm centroid is computed as the first codevector C1 for the training set. Two codevector v1 and v2 are generated as shown in figure 1 by adding constant error to the codevector [1].

Two clusters are formed by Euclidean distance of all training vectors with vectors v1 and v2 based on nearest of v1 and v2. Four clusters are generated by repeating the same procedure for these two clusters and similarly eight clusters are generated by repeating the same procedure for these four clusters. This procedure is repeated for every new cluster until the required size of codebook is reached.

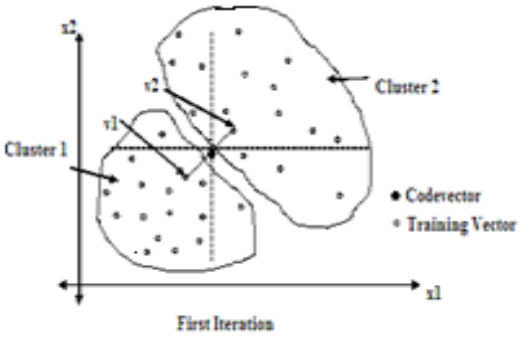

Fig. 1 LBG algorithm for 2-dimensionl class

B. *K-means on LBG Codebook*

K-means algorithm is applied on existing LBG codebook to create global codebook as follows [3]:

- Obtain codebook containing k codevectors using LBG codebook generation algorithm
- Give the above LBG codebook as an input to K-means algorithm by considering first codebook of each category as seed codebook
- Find the squared Euclidean distance of all the training vectors with the k codevectors and k clusters are formed.
- A training vector Xj is put in ith cluster if the squared Euclidean distance of the Xj with ith codevector is minimum. In case the squared Euclidean distance of Xj with codevectors happens to be minimum for more than one codevector then Xj is put in any one of them
- Compute centroid for each cluster
- Replace initial codevectors by the centroids of each cluster respectively
- Repeat the steps 3 to 5 for the respective number of LBG codebooks

III. PROPOSED ALGORITHM

In this section proposed algorithm is explained. In this algorithm common codebook is created for both query images for each category and common codebook for each category containing all images of that category [14]. Following steps shows creation of common codebook by taking Euclidean distance as similarity measure. Same steps need to be carried out by taking City block and Chessboard similarity measures.

Proposed algorithm can be applied as follows:

- Create global codebook of query images from each category by applying K-means algorithm on existing LBG codebook of query images.
- Create global codebook for each category by applying K-means algorithm on existing LBG codebook.
- Find Euclidean distance (similarity measure) ED1, ED2, ED3…ED N between global codebook of query images of each category and global codebook of all categories.
- Sort Euclidean distance array (ED1, ED2, ED3 ….ED N) in ascending order and save corresponding category number. Here N is total number of categories in image database.
- Checking whether category number of global codebook of query images matches with the category with the smallest Euclidean distance and retrieve all the images of that category.

IV. SCHEME OF IMPLEMENTATION

For efficient image indexing and retrieval we are using the LBG algorithm as follows [1]:

- To obtain LBG codebook image is first divided into the non-overlapping windows of size 2x2 pixels. (Each pixel consisting of red, green and blue components)
- These are put in a row to get 12 coordinates per vector. A training set is collection of these vectors. (initial cluster)
- Compute centroid (codevector) of the cluster.
- Split the cluster using LBG
- Repeat the 3,4 till we obtain codebook of required size
- The codebook is stored as the feature vector for the image. Squared Euclidian distance is used as similarity measure

To check the performance of proposed technique two measures are used i.e. precision and Recall [1].

$$\text{Precision} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of relevant images}}$$

$$\text{Recall} = \frac{\text{Number of relevant images retrieved}}{\text{Total number of images in database}}$$

After obtaining codebook using LBG algorithm, K-means algorithm is applied on existing LBG codebook to create global codebook of query images of each category and a single global codebook for all categories. Euclidean distance is calculated between each global codebook of query images and global codebook of all categories. Similarity measure will be based on squared Euclidean distance, City block and Chessboard.

V. SIMILARITY MEASURES

In this paper experiments are carried out by taking three similarity measures .First experiment is carried out by taking Euclidean distance as similarity measure, second on City block and third on chessboard distance.

A. *EUCLIDEAN DISTANCE*

Euclidean distance metric was initially used by Candid. Euclidean distance can be calculated as follows:

If A(x, y) and B(p, q)are two pixels then Euclidean distance can be calculated as follows

$$Distance = \sqrt{\left(|x-p|^2 \oplus |y-q|^2\right)}.$$

B. *CITY BLOCK DISTANCE*

The value of city block distance is always greater than or equal to 0. We will get 0 if there are identical points under

considerations [3]. Value will be high if low similarity is there. City block distance can be calculate as follows:

If A(x,y) and B(p,q) are two pixels then cityblock distance can be given as,

$$Distance = |x - p| \oplus |y - q| .$$

C. *CHESSBOARD DISTANCE*

It is also known as chebyshev distance. It is called chessboard distance, its taken from the chess game. In this game the minimum number of steps taken by a king to move from one square to another square on chessboard equals the Chebyshev distance [3]. Chessboard distance can be calculated as follows: If A(x,y) and B(p,q) are two pixels then chessboard distance can be given as

$$Distance = \max\left(|x - p|, |y - q|\right).$$

Fig.2. Shows a sample database of 100 images by randomly selecting one image from each category. The database has 100 categories, and each category consists of 72 images for a total of 7200 images. The image database used in the experiments is the subset of Columbia Object Image Library (COIL-100).

Fig. 2. Sample database consisting of 7200 Images, the database has 100 categories, 72 images in each category.

The method is implemented in Matlab 7.0 on Intel Core 2 Duo Processor T8100, 2.1 GHz, 2 GB RAM machine to obtain results.

VI. RESULTS

First LBG algorithm is applied on image set to create LBG codebook and retrieval is performed based on LBG codebook. Experiment is carried out on COIL database images, first on 720 images having 10 different classes and second on 1440 images of 20 different classes and third on 2160 images on 30 different classes and fourth on 2880 images on 40 different classes by taking Euclidean distance, City block and Chessboard as similarity measure.

Table 1 shows average Precision/ Recall for different number of image categories using LBG algorithm by taking Euclidean distance as similarity measure. Table 2 shows average Precision/ Recall for different number of image categories using LBG algorithm by taking City block as similarity measure. Table 3 shows average Precision/ Recall for different number of image categories using LBG algorithm by taking Chessboard as similarity measure.

TABLE I

AVERAGE PRECISION/RECALL FOR DIFFERENT NUMBER OF CATEGORIES USING LBG ALGORITHM BY TAKING EUCLIDEAN DISTANCE AS SIMILARITY MEASURE

Number of Categories	Crossover Point of Precision & Recall (%)
10	86.83%
20	80.30%
30	67.90%
40	66.49%

TABLE II

AVERAGE PRECISION/RECALL FOR DIFFERENT NUMBER OF CATEGORIES USING LBG ALGORITHM BY TAKING CITY BLOCK AS SIMILARITY MEASURE

Number of Categories	Crossover Point of Precision & Recall (%)
10	82.14%
20	77.28%
30	65.80%
40	63.49%

TABLE III

AVERAGE PRECISION/RECALL FOR DIFFERENT NUMBER OF CATEGORIES USING LBG ALGORITHM BY TAKING CHESSBOARD AS SIMILARITY MEASURE

Number of Categories	Crossover Point of Precision & Recall (%)
10	84.39%
20	81.05%
30	76.78%
40	69.73%

From the result it is concluded that Euclidean distance and Chessboard similarity measures gives better results as compared to City block. From the result it is also observed that as number of image categories increases Chessboard gives better result as compared to Euclidean distance. Figure 3 shows comparative analysis of LBG algorithm on 10, 20, 30 and 40 Image categories by taking Euclidean distance, City block and Chessboard as similarity measures.

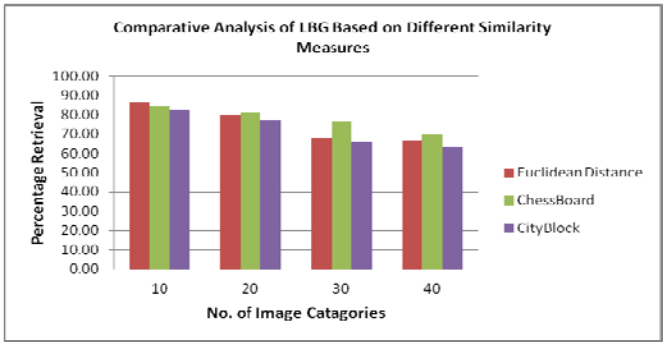

Fig 3. Comparative analysis of LBG algorithm on 10,20,30 and 40 image categories by taking Euclidean distance, City block and chessboard as similarity measures.

After this k-means algorithm is applied on existing LBG codebook to create global codebook of query images and global codebook for each category .Experiments are carried out by increasing number of query images used in creating global codebook of query images and number of image categories in the image set.

Table IV shows Relevance percentage for different number of image categories using proposed algorithm based on Euclidean distance and number of query images. Table V shows Relevance percentage for different number of image categories using proposed algorithm based on City block distance and number of query images. Table VI shows Relevance percentage for different number of image categories using proposed algorithm based on Chessboard distance and number of query images.

TABLE IV
RELEVANCE PERCENTAGE FOR DIFFERENT NUMBER OF IMAGE CATEGORIES AND NUMBER OF QUERY IMAGES USING PROPOSED ALGORITHM BASED ON EUCLIDEAN DISTANCE

No. of categories	No. of Query Images		
	5	10	15
10	80%	90%	90%
20	50%	75%	80%
30	43.33%	53.33%	66.67%
40	42.50%	57.50%	70%

TABLE V

RELEVANCE PERCENTAGE FOR DIFFERENT NUMBER OF IMAGE CATEGORIES AND NUMBER OF QUERY IMAGES USING PROPOSED ALGORITHM BASED ON CITYBLOCK DISTANCE

No. of categories	No. of Query Images		
	5	10	15
10	70%	80%	90%
20	50%	85%	75%
30	40%	53.33%	73.33%
40	38%	46.15%	66.67%

TABLE VI
RELEVANCE PERCENTAGE FOR DIFFERENT NUMBER OF IMAGE CATEGORIES AND NO OF OF QUERY IMAGES USING PROPOSED ALGORITHM BASED ON CHESSBOARD DISTANCE

No. of categories	No. of Query Images		
	5	10	15
10	80%	80%	80%
20	45%	60%	70%
30	40%	63.30%	76.70%
40	42%	55%	70%

VII. CONCLUSIONS

Experiments are carried out by creating codebooks of images using LBG algorithm .Comparative analysis is performed by varying number of image categories and similarity measures. From the result it is concluded that Euclidean distance and Chessboard similarity measures gives better results as compared to City block. From the result it is also observed that as number of image categories increases Chessboard gives better result as compared to Euclidean distance.

Here in proposed method K-means algorithm is applied on existing LBG codebook of images. The efficiency of LBG codebook can be increased by using K-means clustering algorithm. A global codebook of query images is created for each category by randomly selecting 5, 10 and 15 query images from each category. A global codebook is created for each category containing all the images of that category by applying K-means algorithm on existing LBG codebook. Matching is performed between each global codebook of query images and global codebook of all categories. The proposed system uses Euclidean Distance as the similarity measure.

The proposed algorithm effectively reduces overall time complexity. The proposed algorithm number of non relevant effectively minimizes the undesirable results and gives a good relevance percentage by giving minimum images.

The LBG algorithm and proposed algorithm are performed by varying the database to different sizes and by taking different similarity measures. The performance evaluation of LBG algorithm is done by precision and recall and it is observed that, best retrieval results are achieved when size of database is less as compared to large database. The performance evaluation of proposed algorithm is done by observing number of relevant categories and it is observed that best results are achieved when global codebook of query images from each category is made from more number of images.

REFERENCES

[1] Dr. H.B. Kekre, Dr. Tanuja K. Sarode, Sudeep D. Thepade, Vaishali Suryavanshi, “Image retrieval using Texture features extracted from GLCM, LBG and KPE”. International Journal of computer Theory and Engineering, Vol.2, No. 5, October 2010.

[2] Dr. H.B.Kekre ,Dr. Tanuja K. Sarode,”New Clustering Algorithm for Vector quantization using Rotation of error vector”. international Journal Of computer Science and Information Security, Vol.7,no 3, 2010.

[3] Dr. H.B .Kekre, Dr. Tanula K.Sarode,”Vector Quantized Codebook Optimization using K-means” International Journal on Computer Science and Engineering Vol.1(3), 2009, 283-290.

[4] Bang Huang, Linbo Xie “An Improved LBG algorithm for Image Vector 8-1-4244-5540-9,2010 IEEE.

[5] Tejas P. Kokate,”Cluster-based Image Retrieval Techniques” .IJECS Volume 2 Issue 5 May, 2013 Page No. 1474-1478.

[6] Akash Saxena, Sandeep, Saxena, Akanksha Saxena, “Image Retrieval using Clustering Based Algorithm”. International Journal Of Engineering Vol. 1, Issue 3 September 2012.

[7] Hossein Nezamabadi-pour and Saeid Saryazdi, “Indexing and Retrieval in DCT Domain using Clustering Techniques”. World Academy of Science, Engineering and Technology 3 2007.

[8] Arnold W.M. Smeulders, Marcel Worring, Simone Santini, ”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.

[9] Harikrishna Narasimhan, Purushothaman Ramraj,”Contribution-Based Clustering Algorithm for Content-Based Image Retrieval”, Ijarcsse, Vol 2, 2011.

[10] A.Kannan,Dr.V.Mohan,Dr.N.Anbazhagan” Image Clustering and Retrieval using Image Mining Techniques”International Conference on Computational Intelligence and Computing Research 2010 IEEE.

[11] H.B.Kekre, Sudeep D. Thepade, “Rendering Futuristic Image Retrieval System”, National Conference on Enhancements in Computer,Communication and Information Technology, EC2IT-2009, 20-21 March 2009, K.J.Somaiya College of Engineering, Vidyavihar,Mumbai-77.

[12] Dr. H.B.Kekre ,Dr. Tanuja K. Sarode,”New Clustering Algorithm for Vector quantization using Rotation of error vector”. international Journal Of computer Science and Information Security, Vol.7,no 3, 2010.

[13] H.B.Kekre, Sudeep D. Thepade, “Image Retrieval using Augmented Block Truncation Coding Techniques”, ACM Int. Conference on Advances in Computing, Comm. and Control (ICAC3-2009), pp.384-390, 23-24 Jan 2009, Fr. CRCE, Mumbai. Is uploaded at ACM portal.

[14] Seema Anand Chaurasia, Vaishali Suryawanshi,” Hybrid Algorithm for Image Retrieval using LBG and K-means”, International Journal of Computer Applications 94(16):40-4,May 2014.Published by Foundation of Computer Science, New York, USA