International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 6, June 2017)

Feature Matching Process Using Euclidean Distance of Weighted Block Color Histogram and Color Co-Occurrence Matrix for Content Based Image Retrieval System

J. Vanitha[1], Dr. M. SenthilMurugan[2]

[1]*Research Scholar, Bharathiyar University,Coimbatore.*
[1]*Associate Professor/MCA, EGS Pillay Engineering College, Nagapattinam.*
[2]*Director, AVC College of Engineering, Mayiladuthurai.*

Abstract **- In the present days, Images are widely used everywhere. The Image retrieval are classified into content based image retrieval that is using image contents such as color, texture, shape and spatial information and context based image retrieval that is using annotated text. Content Based Image Retrieval (CBIR) is a well-known technique for effective image retrieval. The fused features are used to retrieve more similar images from the database. Color histogram is the widely used method to extract color features which is liberates translation, rotation and scaling of image and avoid spatial feature. The color co-occurrence matrix of HSV of a pixel extracts spatial feature. The Proposed CBIR system have a fused feature of weighted 3*3 block color histogram and color co-occurrence matrix. The images in the database are indexed with Feature vectors which are used to increase the speed of retrieval. The Feature Matching process is carried out using Euclidean distance of color histogram and color co-occurrence matrix. Further the images are classified which reduce the number of images in the search space and required number of images is to be retrieved.**

Keywords**- Block Color histogram, Color co-occurrence Matrix, Content Based Image Retrieval, Euclidean distance, HSV Color space**

I. INTRODUCTION

Images are widely used nowadays. Image retrieval employs vital role in Military affairs, education, medical science, agriculture etc. Image retrieval can be classified as context based image retrieval and content based image retrieval. Searching of images using keywords and text which is called context based image retrieval, won't give better result instead of image content.

The image contents are color, texture, shape and spatial information. Three stages in Image retrieval are,

1. Feature extraction from image database and query image.
2. Feature Matching Process between query image and image database and index the images in the database.
3. Classified the images in the database and fixed number of images is to be retrieved.

Color Histogram and Color Co-occurrence Matrix are used to extract color and spatial feature from query image and images in image database.

The multidimensional indexing is used after the process of extracting color and spatial feature and stored the values of Hue, Saturation, Value, Color Histogram and Color co-occurrence matrix with the images for increasing retrieval speed.

Further the Feature Matching algorithm is used to sort the similar images. Applying classification is used to reduce the images in the search space. So, the images are classified and fixed numbers of images are to be retrieved.

The rest of the paper is organized as follows. Review of the literature summarized in Section 2. In Section 3, the methods of feature extraction are to be summarized. Multidimensional indexing explained in Section 4. In Section 5, feature matching process is proposed. Conclusions are proposed in Section 6. Then the list of references is provided finally.

II. REVIEW OF LITERATURE

Content Based Image Retrieval (CBIR) System is accessing image in an effective way [1]. The traditional way of an annotated image using text, lacks the automatic and effective description of the image. The main idea of CBIR is to analyze image information by low level features of an image [2], which include color, texture, shape and space relationship of objects etc., CBIR employs in many areas including military affairs, medical science, education, architectural design etc. CBIR system includes QBIC [3], Photobook [4], VisualSEEk [5], Virage [6], Netra [7] and SIMPLIcity [8] etc. Histogram is the most commonly used technique to describe features of image [13]. Shape [10], texture [11] and spatial features [12] etc. were implemented to improve the CBIR. Because of the simplicity and robustness, color is the most effective feature. Color histograms are used to extract color feature [13]. HSV color space are used to represent color for better human visual perception [14,15].

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 6, June 2017)

Color co_occurrence matrix is used for HSV color space [9]. Images can be retrieved quickly and accurately by using fused low-level features [13]. In the Feature matching process, the query image is divided into two blocks, and each image in the database is also divided into two blocks and histograms are also calculated separately, comparison was made separately and finally by considering all local histograms a sorted order of best suitable images were generated, final search results were displayed from the sorted order [17,19]. For Indexing the images in the database, SS-tree indexing is superior for similarity indexing applications Compared with R*-tree indexing [18]. The SR-Tree [24] enhances the disjoint among regions which improves the performance on nearest neighbor queries. The SR-Tree is the best multi dimensional indexing structure among the SS-tree [18,23], the R*-Tree [21,22] and the K-D-B –Tree [20]. K-NN is considered one of the simplest machine learning algorithms [25], to classify the images.

III. Extract The Color And Spatial Information

The image features are extracted using color histogram and color co-occurrence matrix. Before extracting, the image is quantized.

This is the process of reducing number of bins by taking color that are very similar to each other and putting them in the same bin. Maximum bin in Mat lab is 256.

This is the CBIR architecture (Figure 1) of proposed content based image retrieval system.

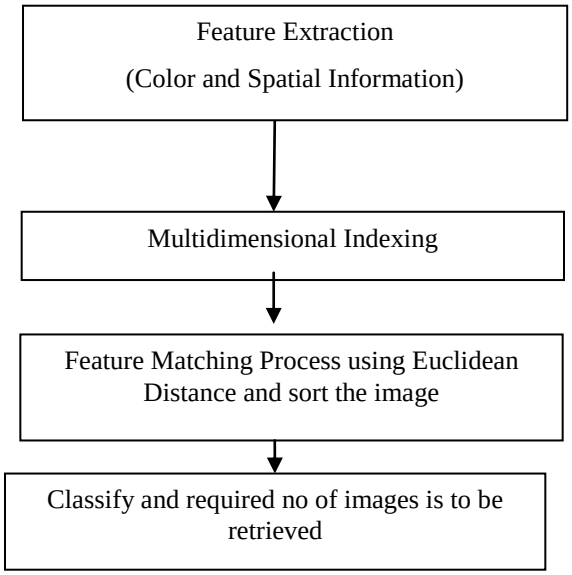

Figure 1: CBIR Architecture

3.1 Color feature extract by Color Histogram

Color histogram defined as a set of bins where each bin denotes the probability of pixels in the image being of a particular color. There are two types of color histograms, Global color histograms and Local color histogram. A Global Color Histogram represents one whole image (Figure 2) with a single-color histogram.

A Local Color Histogram divides an image into fixed blocks and takes the color histogram of each of those blocks. By observation, 3*3 block color histogram is better than global color histogram from the human visual perception [13]. There are two types of division in 3*3 block color histogram.

The 3*3 block color histogram has two types. These are

Type 1: Equally divided and same weight given for each division

Type 2: Unequally divided and double the weight given for center division (Figure 3).

In the proposed System Type 2 is used with weight (Figure 4).

3.2 Spatial feature Extract by Color Co-occurrence Matrix

Color co-occurrence matrix represents three-dimensional matrix where the colors of any pair are along the first and second dimension and the spatial distance between them along the third. In this paper, CCM is simplified to represent the number of color (hue) pairs between adjacent pixels in the image. For each pixel in the image, 4-neighbors (horizontal and vertical neighbors) are accounted [16].

Let I be an N×M image quantized to m colors, and p(x, y) is the color of the image pixel (x, y). Then, the simplified CCM is given by

$$H^{I}(i,j) = \eta((p(x,y), p(N_{x,y})) == (i,j))$$

$$= \alpha \sum_{x=1}^{n} \sum_{y=1}^{M} Ci(x,y) \sum_{(x',y') \in N(x,y)} Cj(x',y') \quad (1)$$

Where η indicates the number of times $(p(x,y),p(N_{(x,y)}))$equals the value of the color indices (i,j) and $N_{(x,y)}$ indicates 4-neighbours of the pixel of (x,y). $C_i(x,y)$ is given by

$$C_i(x,y) = \begin{cases} 1 & if\ p(x,y) = i \\ 0 & Otherwise \end{cases} \quad (2)$$

And the normalization constant α is $\frac{1}{4.N.M}$, for the total number of pixel pairs (p (x, y), p(N(x,y))) is approximately 4*N*M by discounting the difference of boundary pixels.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 6, June 2017)

The simplified CCM is symmetric because the adjacent pixels pairs are neighbors of each other. In this paper, color was quantized to 16 colors, since empirically 16 colors (in hue model) are sufficient for proper color invariant object retrieval. Therefore, the dimension of simplified CCM is 16×16.

IV. Indexing

Indexing is used to reduce the time required for query operation. It will minimize the time of average case and also the worst case. It also supports dynamic insertion and deletion.

For applying this technique to the huge databases, we need to efficiently create multi dimensional index structures, supporting nearest neighbor query. SR-Tree performed most efficiently among other similarity indexing structure. It occupies less volume of space and it supports classification.

The Proposed CBIR system is using SR-Tree algorithm after the process of extracting color and spatial feature and stored the values of Hue, Saturation, Value, Color Histogram and Color co-occurrence matrix with the images.

V. Feature Matching Process

The following steps are used for indexing the images in the image database.

Step 1: Similarity measures between image used as query and the images in database using Euclidean distance of Weighted 3*3 block color histogram (EDwbch(Ii)).

Step 2: Similarity measures between image used as query and the images in database using Euclidean distance of Color Co-occurrence Matrix of HSV of a pixel (EDccm(Ii)).

Step 3: Images stored in the database are sorted using the addition of distance value from step 1 and step 2 (EDwbch(Ii) + EDccm(Ii)).

Algorithm for measuring similarity between query image and the images in database using Euclidean distance of weighted 3*3 block color histogram (EDwbch(Ii))

Step 1: Input Query Image and Image Database.

Step 2: Read the Query Image.

Step 3: The image can be segmented into 3*3 weighted block according to the Proportion 1:2:1.

Figure 2: Original Image

Step 4: Calculate the histogram for each block of the image separately.

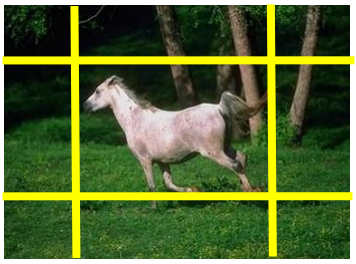

Figure 3: Segmented Image with 1:2:1 Proportion

1	2	1
2	4	2
1	2	1

Figure 4: Weight for each block

Step 5: Read the image from the given database.

Step 6: Do the step 3 – 5 for all the images in the database and stored as an array for future use.

Step 7: Compare the 1st block color histogram of image in database with 1st block color histogram of query image using Euclidean distance $d_i = \sqrt{\sum_{i=1}^{N} (Ii - Qi)^2}$ this process for all blocks and calculate the distance as $dis(Q,I)_{his} = \sum_{i=1}^{9} Wi d_i$ where Wi is the weight of the i^{th} block and d_i is the Euclidean distance of the i^{th} block.

Step 8: The distance marked as EDwbch(Ii) of the image in the database.

Step 9: Continue the step 7 and 8 until all the images are to be read from the image database.

Algorithm for similarity measure using Euclidean distance of Color Co-occurrence Matrix of HSV of a pixel (EDccm(Ii)).

The following steps are used to find the similar images

Step 1: Load the image database .

Step 2: Quantize the images in image database for [16,16].

Step 3: Convert the images from RGB to HSV.

Step 4: Formulate CCM of HSV of a pixel.

Step 5: Load the Query image.

Step 6: Apply the procedure 2-4 to find HSV values of Query image.

Step 7: Determine the Euclidean distance of Query image with database.

Step 8: The distance of Color co-occurrence matrix is marked as EDccm(Ii) of the image in the database.

The images in the database are sorted using the distance value and then the images are classified and fixed no of images are to be retrieved.

VI. Conclusion

In this paper, feature extraction can be carried out on unequally divided and double the weight given for center division of image using 3*3 block color Histogram for color feature and Color Co-occurrence Matrix of HSV of a pixel for spatial feature. The images are indexed by feature vectors like Hue, saturation, value, color histogram and CCM . The feature matching process is using Euclidean distance. Then the images are classified and the fixed number of images is to be retrieved. Also try another feature matching process to give the better result. Therefore, to provide better performance of retrieving more images are needed on the basis of these techniques.

REFERENCES

[1] P. Muneesawang, L. Guan, An interactive approach for CBIR using a network of radial basis functions, IEEE Transactions on Multimedia 6 (2004) 703–716.

[2] R. Datta, D. Joshi, J. Li, J.Z. Wang, Image retrieval: ideas, influences, and trends of the new age, ACM Computing Surveys 40 (2) (2008) 1–60.

[3] W. Niblack, R. Barber, W. Equitz, et al., The QBIC project: querying images by content using color, texture, and shape, in: SPIE 1908, San Jose, CA, 1993,pp. 173–187.

[4] A. Pentland, R.W. Picard, S. Scarloff, Photobook: tools for content-based manipulation of image databases, in: SPIE 2185, San Jose, CA, 1994, pp. 34–47.

[5] S. Mehrotra, Y. Rui, M. Ortega, et al., Supporting content-based queries over images in MARS, in: Proc. of IEEE Int'l Conf. on Multimedia Computing and Systems'97, Ottawa, Ontario, Canada, 1997, pp. 632–633.

[6] J.R. Bach, C. Fuller, A. Gupta, et al., Virage image search engine: an open framework for image management, in: SPIE 2670, 23, San Jose, CA, 1996, pp. 76–87.

[7] J.R. Smith, Integrated spatial and feature image systems: retrieval, analysis and compression, Ph.D. Dissertation, Columbia University, New York City, 1997.

[8] J.Z. Wang, J. Li, G. Wiederhold, SIMPLIcity: semantics-sensitive integrated matching for picture libraries, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (9) (2001) 947–963.

[9] A.Vadivel,ShamikSural,A.K. Majumdar, An Integrated Color and Intensity Co-occurrence Matrix, Pattern Recognition Letters 28 (2007) 974–983.

[10] G. Gagaudakis, P.L. Rosin, Incorporating shape into histograms for CBIR, Pattern Recognition 35 (2002) 81–91.

[11] P.S. Hiremath, Jagadeesh Pujari, Content based image retrieval using color, texture and shape features, in: 15th International Conference on Advanced Computing and Communications, ADCOM 2007, 2007, pp. 780–784.

[12] Y.K. Chan, C.Y. Chen, Image retrieval system based on color-complexity and color-spatial features, Journal of Systems and Software 71 (2004) 65–70.

[13] Jun Yue, Zhenbo Li , Lu Liu , Zetian Fu, Content based image retrieval using color and texture fused features, Mathematical and Computer Modelling 54 (2011) 1121–1127.

[14] Nishant Shrivastava, Vipin Tyagi, An efficient technique for retrieval of color images in large Databases, Computers and Electrical Engineering 46 (2015) 314–327.

[15] Dibya Jyoti Bora, Anil Kumar Gupta, Fayaz Ahmad Khan, Comparing the Performance of L*A*B* and HSV Color Spaces with Respect to Color Image Segmentation, International Journal of Emerging Technology and Advanced Engineering Volume 5, Issue 2, February 2015.

[16] Seong-O Shim, Tae-Sun Choi, *Senior Member, IEEE,* **I**mage indexing by modified color co-occurrence matrix ,IEEE Xplorer.

[17] Chesti Altaff Hussain, Dr. D. Venkata Rao, T. Praveen, Color histogram based image retrieval, International Journal of Advanced Engineering Technology E-ISSN 0976-3945(2013)

[18] David A.White, Ramesh Jain , Similarity Indexing with SS-tree,IEEE,516-523(1996)

[19] Wenqing Huang, Jiazhe Dai and Qiang Wu ,Image Retrieval Based on Weighted Block Color Histogram and Texton Co-occurrence Matrix, International Conference on Information Technology and Management Innovation,848-853(2015).

[20] J. T. Robinson, "The K-D-B-tree: a Search Structure for Large Multidimensional Dynamic Indexes," Proc. ACM SIGMOD, Ann Arbor, US.4, pp 10-18, Apr.1981.

[21] A. Guttman, "R-Trees: a Dynamic Index Structure for Spatial Searching," Proc. ACM SIGMOD, Boston, USA, pp. 47-57, June. 1996.

[22] N. Beckmann, H. P. Kriegel, R. Schneider and B. Seeger," The R*-tree: an Efficient and Robust Access Method for Points and Rectangles," Proc. ACM SIGMOD. Atlantic City,US.4,pp.322-331, May 1990.

[23] D. A. White and R.Jain, " Similarity Indexing: Algorithms and Performance," Proc. SPIE vol. 2670, San Diego, USA , pp. 62-73,Jan 1996.

[24] Norio Katayama and Shin'ichi Satoh , The SR-Tree: An Index Structure for High-Dimensional Nearest Neighbor Queries Proc. ACM SIGMOD Int. Conf. on Management of Data, pp.13-15, May 1997.

[25] Mona Mahrous Mohammeda, Amr Badr, M.B. Abdelhalim Image classification and retrieval using optimized Pulse-Coupled Neural Network, Expert Systems with Applications 42 (2015) 4927–4936.

International Journal of Emerging Technology and Advanced Engineering
Website: www.ijetae.com (ISSN 2250-2459, ISO 9001:2008 Certified Journal, Volume 7, Issue 6, June 2017)

About The Authors

Mrs. J. Vanitha M.C.A., M.Phil is working as associate professor and HOD of MCA Department in EGSPEC, Nagapattinam . Now She is pursuing Ph.D in Bharathiyar University , Coimbatore, Tamilnadu. Her Field of interest is Image Processing.

Dr. M. SenthilMurugan M.C.A., Ph.D., is working as director in AVCCE, Mayiladuthurai. His research interest includes image processing, Biometrics, cloud computing and Computer Networks.