*International Journal For Technological Research In Engineering*
*Volume 2, Issue 3, November-2014* *ISSN (Online): 2347 - 4718*

# IMAGE CLUSTERING TECHNIQUE USING COLOUR MOMENT, GRAY LEVEL CO-OCCURENCE MATRIX, ENTROPY AND ENERGY USING K-MEANS CLUSTERING-AN APPROACH

Puspanjali Khatri[1], Joydeep Mukherjee[2]
Faculty of Engineering and Technology, Jadavpur University, Kolkata, India

***ABSTRACT: Advances in data storage and image acquisition technologies have enabled the creation of large datasets. So to deal with these datas, it is important to develop appropriate information systems to efficiently manage these collections. Image searching is one of the most important services that need to be supported by such systems. Here in this paper we propose an image store and clustering method based on colour moments, Gray level co-occurrence Matrix, Entropy and Energy with k-means for clustering the data.***
***Keywords: CBIR, Colour Moments, Gray Level Co-occurrence Matrix, Entropy and Energy.***

## I. INTRODUCTION

In Content Based Image Retrieval (CBIR)system ,image processing algorithms are used to extract Feature vectors that represent image properties such as colour, edge, histogram, texture etc. In this approach It is possible to retrieve images similar to one chosen by the user .One of the advantage of this approach is Possibility of an automatic retrieval process contrasting to the effort needed to annotate the images. Creation of Content Based Image Retrieval (CBIR) Systems, involves research on databases and image Processing, handling problems, such as storage issues to friendly user interfaces [1] as images are complex To manage apart from volumes they occupy retrieval in an application and context dependent task [2] it Requires translations of high-level user perceptions to low-level features known as Semantic Gap Problem. Besides indexing string processing, to index visual features is better to use numerical values for n-feature and
Then represent an image in an n-dimensional space [3].Features of an image are basically group as middle and low level features. Colour and texture comes under Low-Level features and shape, object are grouped as middle-level features. Our initial approach on feature extraction is to extract the colour moment value from an image. The mean,
Variance and standard deviation of an image are known as Colour Moments and stores the value in 1D Array, and calculates the value for all the images used. Secondly the Gray Level Co-occurrence Matrix (GLCM) is calculated. It's a tabulation of how often different Combination of pixel brightness value occur in an image, it's a square matrix. Now the Entropy value of an image is calculated. Entropy basically shows the amount of information of the Image (i) i.e needed for the image compression, as it measures the loss of information of the image as According to the researches made in the past it has been seen that though colour histogram played an Important role in indexing into an image databases, but for very large image databases histogram space With very large dimensions the computational cost of performing distance calculations can be prohibitive Or the reduction of dimensionality of the colour histogram space, and to provide a substantial improvement In retrieval performance entropy is used. Finally Energy is calculated in order to measure the homogeneity, as it gives user the information about the uniformity of the texture. Last but not the least the entire features are merged in a single 1D array of extracted feature for a single Image .K-means clustering algorithm is used here to cluster the data objects putting the similar ones within the same clusters and store the dissimilar objects in the other clusters. The remaining part of the paper is organized as follows, section-2 related work, and section-3 System overview, section-4 Proposed Work, section-5 Experiments and Results, section-6 Analysis, section-7 Conclusions and Future Scope.

## II. RELATED WORK

In the past years some papers has been presented for clustering image database. Some clustering methods are used to cluster similar datas. The image type, color, size and texture characteristics are extracted as a metadata, depending on the stored data some clustering algorithm is used to group the data. Colour represents one of the most widely used visual features in Content Based Image Retrieval. Swain And Ballard [4] in his paper proposed colour histogram intersection and L1 matrices as the similarity measure. In the paper by Sticker and Orengo [5] proposed colour moment and colour histogram, though the result was Not that good. Lin had proposed a system using colour texture and colour histogram. Xiang-yang Wang, yang Lian [6] proposed system for clustering the image with texture and spatial features, further in the paper "Image Clustering using colour moment, histogram, edge and k-means clustering [7] proposed a multi feature Model. Haralick proposed the usage of co-occurrence of matrices for texture feature representation [8].the Problem with the features such as colour histogram was although it is dependent of the object's colour, Ignoring its shape and texture features and also it can be potentially identical for 2 images with different Object content that shares colour information. Problems detected with the canny edge detection are it is very difficult to implement real time response and also its time consuming. So to overcome these drawbacks we have used different feature in our paper including Colour moment, Gray Level Co-occurrence matrix, Entropy and Energy. This multi-features used gives a good Accuracy as compared to the

  

features they had used in the past papers.

## III. SYSTEM OVERVIEW

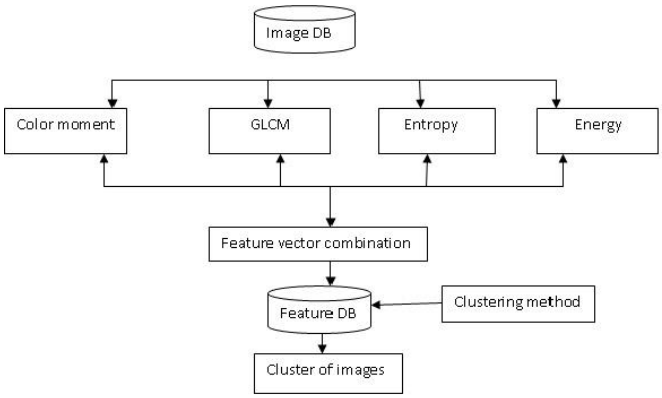


Fig 1: System Overview

*A. Color Moment Analysis Method*

Colour moments are a measure that can be used to differentiate images based on their features of Colour, once calculated these moments provide a measurement for color similarity between images and these values can then be compared to values of images indexed in a database for tasks like image retrieval.

The 3 colour moments are:

Moment1# Mean

Moment2#Standard deviation

Moment3# Skewness

*B. Gray Level Co-occurrence Matrix (GLCM)*

A Gray Level Co-occurrence Matrix is a tabulation of how often different combinations of pixel brightness values (gray-level) occurs in an image. GLCM are created from a gray scale images. It calculates how often a pixel with gray scale values (i) occurs either horizontally/vertically/diagonally to adjacent pixels with value (j).

*C. Entropy*

Entropy shows the amount of information of an image (i),needed for the image compression it measures the loss of image information/message in a transmitted signal. if entropy level is low then the image will have more black area and a perfect image will have 0 entropy.

*D. Energy*

Energy is a measure of local homogeneity and it represents the opposite of entropy basically this feature will tell us how uniform the texture is.

## IV. PROPOSED WORK

Fig 1: System overview of Method-2(Colour Moment, ,Gray Level Co-occurrence Matrix, Entropy, Energy)

Image Color Moment Analysis:

Step 1: Read the image file.

Step 2: We find the mean value using the following function.

$$Ei = \sum_{j=1}^{N} Pij$$

Step 3: We find the Standard Deviation using the following function

$$\sigma \sqrt{\frac{1}{N\left(\sum_{j=1}^{N}[(Pij - Ei)2)\right]}}$$

Step 4: We find the skewness value using the following function

$$\sigma i = \frac{\sqrt{1}}{N}$$

Step 5: Finally store the value of mean, standard deviation and skewness in 1 D array.

Step 6: Step 1-4 is repeated for every image in the database.

Texture Feature Extraction Method Using Gray level Co-occurrence Matrix:

Step 1: Read the image file.

Step 2: Convert the RGB image to grayscale image.

Step 3: Use graycomatrix ( ) to measure gray-level co-occurrence matrix.

The following function is:

Glcm1=graycomatrix (b,'NumLevels',11,'offset',[0 1],'Symmetric', true)

Step 4: Use graycoprops ( ) to calculate energy, contrast and correlation from Gray-levelco-occurrence matrix.

The following function is:

Stats =graycoprops (glcm1,{'contrast','homogeneity','correlation','energy'});

Step 5: Next we calculate the compliment of the grayscale image.

Step 6: step 3-4 is repeated for the complement images.

Step 7: We got energy, contrast, correlation from the complement image.

Step 8: We calculate the mean of the energy of the original image and the energy of the complement image.

Step 9: step 8 is continued for contrast and correlation of the image.

Step 10: Finally store the value of mean value of energy, contrast, correlation in a 1D array.

Step 11: Step1-10 is repeated for every image in the database.

Calculating the distance Using DistMatrix( ):

Step 1: Read the image file.

Step 2: Convert the RGB image to grayscale image.

Step 3: Now we calculate the distance:

C{k}=repmat(A(:,k),1,hB) takes the kth column of A and repeat it to form a matrix with hB columns all the same as that one column hB is the number of rows in B.

Step 4:Now calculate D{k}=repmat(B(:,k),1,hA)Takes the

  

kth column of B and repeat it to form a matrix with hA columns all the same as that one column hA is the number of rows in A.

## V. EXPERIMENTATIONS AND RESULTS

Now here we represent the output of the proposed method and then compare the results of the method with the existing one. The implementation is done using MATLAB7.8.0 (R2009a) image processing tools and statistical tools. In our experiment we consider 4 images of same category and these images create a class. Many images are taken to form a class. After we apply clustering technique on those images it will make different cluster of similar kind of images and how many clusters are created will depend upon the number of image class. Experimental results of our proposed method are shown below: Initially we have taken 16 images of 4 classes each and the proposed method places only images in exact position before clustering. Fig: proposed method using 16 images before clustering.

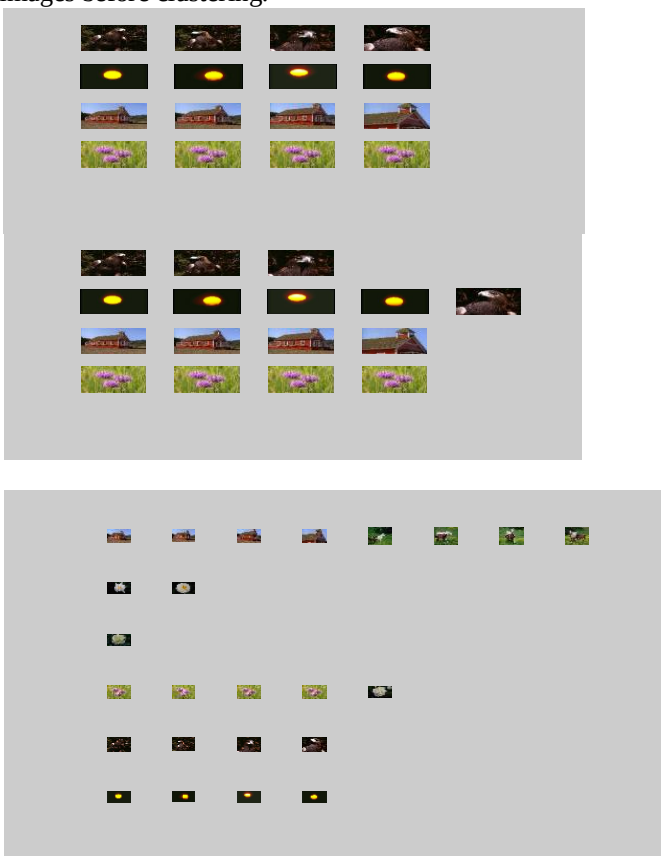

Fig 2: proposed method with 24 images after clustering.

Table 1: Difference between the mother paper and the implemented paper:

| Mother paper | Implemented paper |
|---|---|
| Features used are colourmoment,histogram and edge. | Features used are colour moment,glcm,entropy and energy. |
| Accuracy achieved 79.25% | Accuracy achieved 94.79% |
| Accuracy measure used is Euclidean Distance. | Accuracy measure used is Manhattan Distance. |

## VI. ANALYSIS

Image clustering is the vast field in Multimedia application. Here the clustering efficiency in image database is addressed. Here image clustering is discussed according to the analysis made in the existing paper using colour feature extraction the accuracy is not so good. Performance was found to improve in our proposed method using colour and texture feature extraction. Now from our experiment we see that the accuracy of the proposed method is far better than the existing one. Also comparing in terms of execution our method is faster, efficient in clustering and easy in implementation too. Finally we can conclude that colour feature itself is not sufficient to cluster similar images with better accuracy. But with the combination of colour and texture feature the accuracy can be increased in less amount of time.

Table2: Accuracy table of our proposed methoount of time.

| No.of clusters | No.of images | percentage |
|---|---|---|
| 4 | 16 | 93.75 |
| 6 | 24 | 95.83 |

Overall percentage = 94.79%

## VII. CONCLUSION AND FUTURESCOPE

Image clustering is a very vast field of multimedia application here just an approach to Content Based I mage Retrieval using clustering method is proposed and it has endless scope to work with. Basically the result of clustering depends upon the images used. We can say that our proposed method gives higher accuracy, our system performance is reasonable too according to the accuracy plot given above.
Here Distmatrix function is used to calculate the k-means clustering. In future Chebyshev distance, neural network may be used for further implementation.

## REFERENCES

[1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-Based Image Retrieval at the End of the Years. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12):1349–1380, December 2000

[2] Y. Rui, T. S. Huang, and S. F. Chang. Image Retrieval: Current Techniques, Promising Directions, and Open Issues. Journal of Communications and Image Representation, 10(1):39–62, March 1999

[3] Y. A. Aslandogan and C. T. Yu. Techniques and Systems for Image and Video Retrieval. IEEETransactions on Knowledge and Data Engineering, 11(1):56–63, January/February1999.

[4] M. L. Swain, H H Ballard, ”color indexing”, International journal of Computer Vision,7(1):pp. 11-32,1991.

[5] M Sticker ,M Orengo,”Simlarity of color images”,In:W Niblack,RC Jain eds Pro of Spy storage and Retrieval for image and Video Databases,Vol 2420,san Jose CA,USA:SPIE Press,p381-392,1995.

  

[6] Xiang-Yang Wang, Yong-Jian Yu, Hong-Ying Yang, ”An effective image retrieval scheme using color, texture and ahape features, ”Published in: Journal Computer Standards &Interfaces archive Volume 33 Issue 1,January,2011 Elasevier Science Publishers B. V. Amsterdam, The Netherlands.
[7] Image Clustering using color moments, histogram, edge and k-means clustering, International Journal of science and Research(IJSR),India Online ISSN:2319-7064.by Annesha Malakar and Joydeep Mukherjee.
[8] Haralick RM. Statistical and Structural Approaches Texture [J].Proceeding of the IEEE,1979,67(5):786-804.

  