

# Image Similarity Measurement using Region Props, Color and Texture: An Approach

Satyajit Mondal  
School of Education Technology  
Jadavpur University  
Kolkata

Joydeep Mukherjee  
School of Education Technology  
Jadavpur University  
Kolkata

## ABSTRACT

Image similarity measurement is very important part for image clustering and content based image retrieval. Store the images and searching them with efficiency is the main issue. As the volume of image database increases day by day, efficient searching technique is a challenging job. Here a proposed approach is given for image similarity measurement using regionprops, color, texture and GLCM features.

## General Terms

Image processing, feature extraction, image similarity measurement.

## Keywords

CBIR, Regionprops, Color Moments, GLCM and Energy.

## 1. INTRODUCTION

Image classification is very important part in large number of areas in multimedia applications and other fields. Content Based Image Retrieval plays an important role in the field of medical imaging, business, marketing, social network analysis, computer science, mathematical analysis etc. By Image similarity measurement one can retrieve a range of similar images which are relevant with the query image. When the image database size is high, it is essential to have an efficient image retrieval technique. In content based image retrieval process images can be retrieved depending on their content. Clustering methods help to retrieve similar types of images. Image classification becomes a complex task because of the presence of semantic gap [1]. User has the tendency to translate the high level user perception into low level features. This is known as semantic gap.

For image similarity measurement features are extracted first. Here the features that are extracted from the images are color feature, shape and texture. Shape feature comes under middle level feature. Color and texture comes under low level feature.

Color moment is used for color feature [2] extraction method. Color moment comprises with mean, standard deviation and skewness value of images. When the values are calculated for every image, stored in a 1-D array.

Regionprops is used here for extracting the property of image region as the shape feature. Many region properties are there such as Area, Centroid, Convexlength, Convexareas, Eccentricity etc. Here only Eccentricity is used. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. The value is between 0 and 1. The result is stored as a structure array.

Texture features are extracted using gray level co-occurrence matrix (GLCM) [3]. Gray Level Co-occurrence Matrix 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 energy is calculated. Energy basically measures the textural information in terms of energy whether it is uniform or not. It basically measures texture of gray-scale image represents homogeneity changing, reflecting the distribution of image gray-scale uniformity of weight and texture. It is also known as uniformity or the angular second moment. Here another 1-D array is developed.

Finally the entire three 1D array are merged in a single 1 D array of extracted feature for single image. This process is repeated for every image. Then we apply Manhattan distance method algorithm for similarity measurement.

Remaining portion of the paper organized as follows, section 2- Related Work surveyed by us, section 3- Proposed Work, section 4- System Overview, section 5- Experiments and Results, section 6- Analysis, section-7 Conclusion and Future scope.

## 2. RELATED WORK

In the past, the author Swain and Ballard proposed an approach where image similarity was measured by histogram analysis, intersection and L1 metric. [4]. Color moment based features are more compact and powerful matching parameter which matched more robustly than color histograms [5], actually histograms cannot capture spatial relationship of color regions. Mean, standard deviation and skewness are calculated as color features. Another important feature for similarity measurement is texture feature. A variety of techniques have been developed for measuring texture similarity. For texture feature extraction gray level co-occurrence matrix (GLCM) is used. A proposed approach is discussed in [6] on feature extraction method which deals with color, texture and shape feature extraction method in which the number of regions of interest (ROIs) is determined. As the authors presented in [7], Gray level co-occurrence matrix is used for texture classification. Various types of textural parameters calculated from the gray level co-occurrence matrix which actually describes the details about the overall image content. In the paper [8] authors proposed a modified approach to develop a system for clustering similar images in one cluster and the dissimilar images in other cluster using color moment, histogram, edge and k-means clustering. Again the authors in [9] proposed a method using HSV and lab color space to manipulate images and then compared the result with grey and RGB approach. It shows that the lab color space gave better results than the others. On

the other hand in [10] an approach was proposed and developed based on the HSV color space. In this case the used hue and saturation component of an image is stored in an array and it gave result with higher accuracy level. The recent trends in texture classification have been discussed in [11] where the Gray level co-occurrence matrix method is highlighted to be older and simpler among the statistical approach of representing the image. However, superior performance is achieved by embedding Gray level co-occurrence matrix with the other statistical techniques. Significant drawbacks of texture classification techniques is illustrated, for instance, the structural method can be implemented only on structured textures which are rarely available while the model based technique admits difficulty in implementation due to the complex parameters involved. Gray level co-occurrence matrix is the method of computing the frequency of pixel pairs having the same grey level in the image. The relationship between the reference pixel and the neighboring pixels is calculated to determine the textural features of the image [12].

### 3. PROPOSED APPROACH

Image region property analysis:

- Step 1: Read the image file.
- Step 2: Convert the RGB image into Grayscale image.
- Step 3: Calculate the region property using regionprops function.
- STATS = regionprops(L,properties)
- Step 4 : It is also done for the complement image.

Image Color Moment Analysis:

- Step 1 : Read the image file.
- Step 2 : Mean value calculation using the following function.

$$E_i = \sum_{j=1}^N \frac{1}{N} P_{ij} \dots\dots\dots \text{equation 1}$$

Step 3 : Standard Deviation calculation using the following function.

$$\sigma_i = \sqrt{\frac{1}{N} (\sum_{j=1}^N (P_{ij} - E_i)^2)} \dots\dots\dots \text{equation 2}$$

Step 4 : Skewness value using the following function.

$$S_i = \sqrt[3]{\left(\frac{1}{N} \sum_{j=1}^N (P_{ij} - E_i)^3\right)} \dots\dots\dots \text{equation 3}$$

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

Step 6 : Repeat Step 1-4 for every image.

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

- Step 1: Read the image file.
- Step 2: RGB image converted to grayscale image.
- Step 3: Measure gray-level co-occurrence matrix using graycomatrix( ).
- Glcml=graycomatrix(b,"NumLevels",11,"offset",[0 1],"Symmetric", true)
- Step 4: Calculate energy, contrast and correlation from Gray-level co-occurrence matrix using graycoprops( ).

- Stats =graycoprops(glcml,{,"contrast","homogeneity","correlation","energy"});
- Step 5: Calculate the compliment of the grayscale image.
- Step 6: Repeat step 3-4 for the complement images.
- Step 7: calculate the mean of the energy of the original image and the energy of the complement image.
- Step 8: calculate the mean of the contrast of the original image and the contrast of the complement image.
- Step 9: calculate the mean of the correlation of the original image and the contrast of the correlation image.
- Step10: Store the value of mean value of energy, contrast, correlation in a 1D array.
- Step 11: Repeat Step1-10 for every image.

Similarity measurement using Manhattan distance:

- Step 1: Read the image file.
- Step 2: Convert the RGB image to grayscale image.
- Step 3: Distance calculation:  
Distance is calculated using the following function

$$d(i,j) = |x_{i1} - x_{j1}| + |x_{i2} - x_{j2}| + \dots + |x_{ip} - x_{jp}| \dots\dots\dots \text{equation 4}$$

Step 4: Calculate distance between two points.

### 4. SYSTEM OVERVIEW

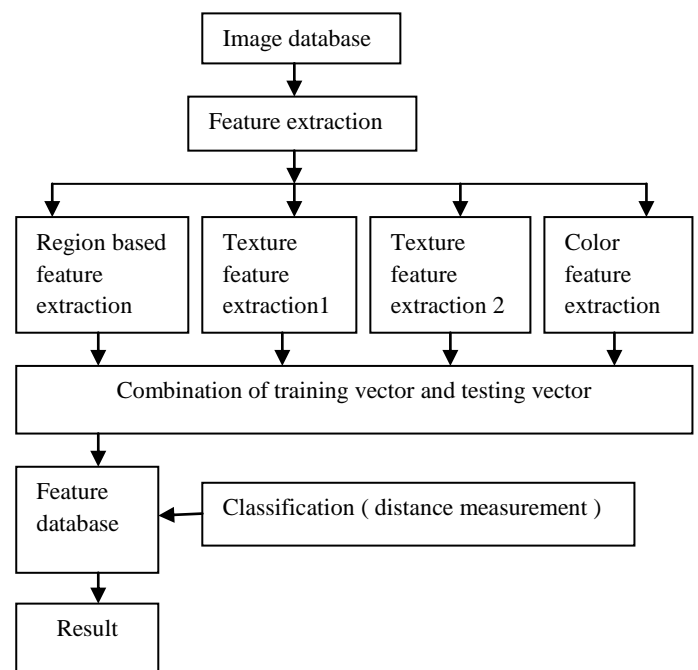
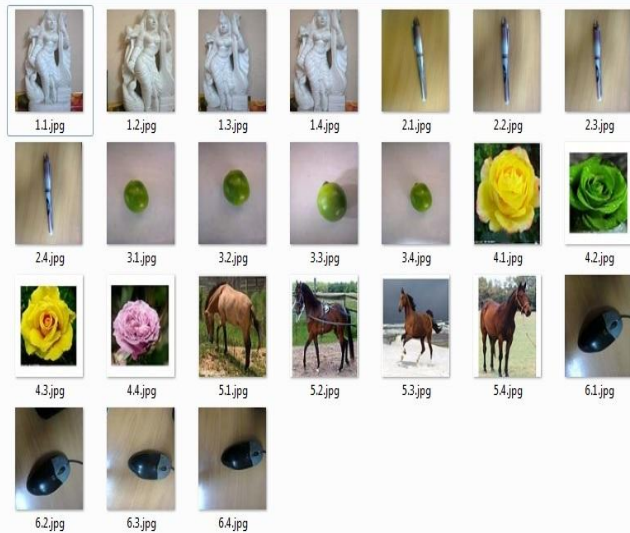


Figure 1 : block diagram of the proposed system

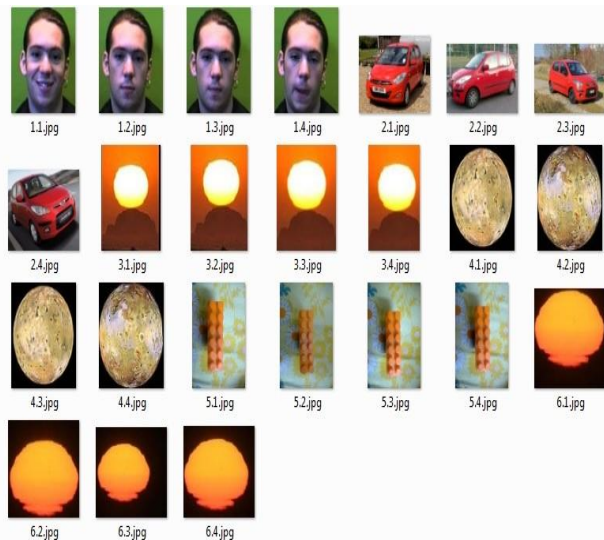
### 5. EXPERIMENT AND RESULT

Now the output of the proposed method is represented and then compares the results of the method with the existing one. The implementation is done using MATLAB7.8.0 (R2009a) image processing tools and statistical tools.

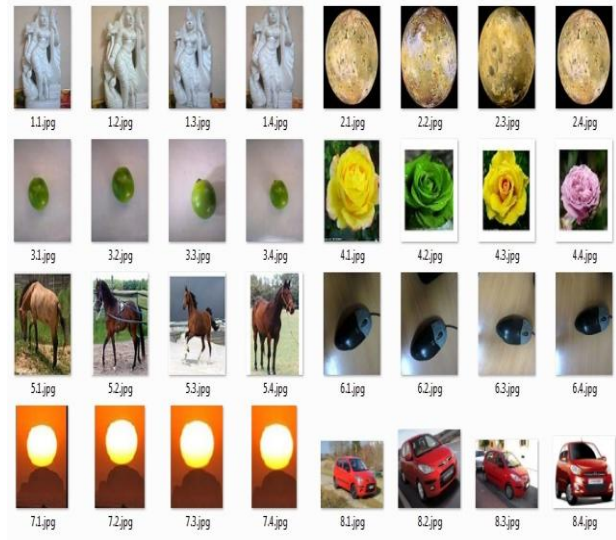
In this experiment 4 images of same category is considered and these images create a class. Many images are taken to form a class. Experimental results of the proposed method are shown below: Initially 24 images of from the 6 classes are taken and the proposed method measure similarity among them.



**Dataset 1 : no. of class 6, total images 24**  
( For this dataset 100% accuracy is achieved )



**Dataset 2 : no. of class 6, total images 24**  
( For this dataset 94.44% accuracy is achieved )



**Dataset 3 : no. of class 8, total images 32**  
( For this dataset 95.83% accuracy is achieved )

**Table 1 : Difference between the existing approach and the proposed approach:**

Existing approach	Proposed approach
Feature used are colour moment, glcm, entropy and energy.	Feature used are regionprops, color moment, glcm and energy.
Accuracy achieved 94.79 %	Accuracy achieved 96.75 %
Accuracy measure by Manhattan Distance.	Accuracy measure by Manhattan Distance.

## 6. ANALYSIS

After the completion of training phase, Manhattan Distance algorithm is applied for similarity measurement, then the accuracy is calculated and it is shown in a table for different classes of images. Finally, an accuracy graph for the proposed approach is shown.

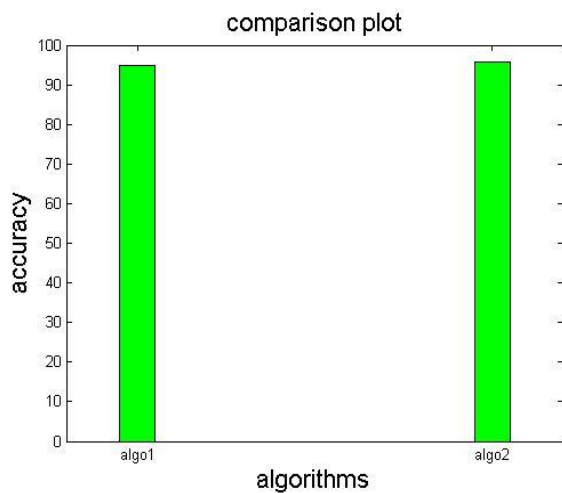
Here an efficient similarity measurement technique is proposed in which regionprops (measure the property of image region) is combined with color and texture. Accuracy is found to be improved in our proposed method using color, texture and regionprops feature extraction. The experiment shows that the accuracy of the proposed method is better than the existing one. It is known that the result of similarity checking depends on the images. Result is good for some images and some image gives poor result.

When the algorithm presented in the existing paper, applied in our dataset the accuracy level fall down. To improve this accuracy level a new feature i.e. regionprops is considered which measure the property of the image region. The proposed system comes with a higher accuracy than the existing one.

**Table 2 : Accuracy table of the proposed method**

No of class	No of images	Percentage
6	24	100
6	24	94.44
8	32	95.83

Overall percentage = 96.75 %



**Figure 2 : Comparison graph of the existing approach and the proposed approach(algo 1 is existing approach and algo 2 is proposed approach)**

## 7. CONCLUSION AND FUTURE SCOPE

Finally it can be concluded that color feature itself is not sufficient for similarity measurement of images with better accuracy. But with the combination of regionprops, color and texture feature the accuracy can be increased in less amount of time.

In future other features can also be extracted for better accuracy and performance.

- Neural network, Chebyshev distance may be used for further implementation.
- Can be apply to a large and different image database.
- Other shape based techniques can be added.

## 8. REFERENCES

- [1] 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.
- [2] P. S. Suhasini, Dr. K. Sriramkrishna, Dr. I. V. Muralikrishna “CBIR using Color histogram processing”, *Journal of theoretical and applied information technology*, pp-116-122, 2009
- [5] D. Ziou and S. Tabbone (1998) "Edge detection techniques: An overview", *International Journal of Pattern Recognition and Image Analysis*, 8(4):537–559, 1998.
- [3] D.S. Guru, Y.H. Sharath & S. Manjunath, *Texture Features and KNN in Classification of Flower Images,, IJCA Special Issue*, 2010.
- [4] J.R.Smith and S.F. Chang, "A fully automated content based image query systems," *Proc.ACM Multimedia 96*, 1996.
- [5] M. Stricker, and M. Orengo, "Similarity of color images", In *SPIE Conference on Storage and Retrieval for Image and Video Databases*, volume 2420, 1995, pp. 381-392, San Jose, USA.
- [6] R.S. Choras, T. Andrysiak, M. Choras, "Integrated color, texture and shape information for content-based image retrieval", *Pattern Anal. Applic.* 10: 333-343, 2007.
- [7] Image clustering technique using colour moment, GLCM, entropy and energy using k-means clustering-an approach, Puspanjali Khatri and Joydeep Mukherjee, *International Journal For Technological Research In Engineering*, volume 2, ISSN (online) : 2347-4718.
- [8] Image Clustering using Color Moments, Histogram, Edge and K-means Clustering, Annesha Malakar and Joydeep Mukerjee *International Journal of Science and Research (IJSR)*, India Online ISSN:2319-7064.
- [9] Jagadeesh Pujari, Pushpalatha S.N. Padmashree D.Desai, "Content-Based Image Retrieval using Color and Shape Descriptors", *Signal and Image Processing (ICSIP)*, 2010 *International Conference on*, pp:239
- [10] Young Deok Chun, Nam Chul Kim and Ick Hoon Jang, "Content Based Image Retrieval using Multi resolution Colour and Texture Features," *Multimedia, IEEE Transactions on* Oct.2008, Volume:10, Issue:6, pgn:1073-1084.
- [11] J.Y. Tou, Y.H. Tay, and P.Y. Lau, Recent trends in texture classification: review, *Symp. Inf. Commun. Technol.* (2009) 64-68.
- [12] M. Sonka, V. Hlavac, and R. Boyle, *Image processing, analysis, and machine vision*, second edition, Albany, NY: Brooks/Cole, (1999) 256-260.