Feature Extraction of Image Using Gray-level and KNN based Genetic Algorithm

Hirdesh patidar[1] & Dr Rajendra Gupta[2] & Dr Pratima Gautam[3]
[1]Research Scholar, Computer Science, Rabindranath Tagore University, Raisen.
[2]Associate Professor, Dept. of Computer Science & Information Technology Rabindranath Tagore University.
[3]Professor, Dept. of Computer Science & Information Technology Rabindranath Tagore University, Raisen

Received: July 21, 2018 Accepted: October 08, 2018

ABSTRACT *At the moment, image extraction becomes essential from a large database of images, which reduces storage costs and also offers good image quality. In this work, we propose an approach that extracts the image function based on color, consistency and shape. A color function is selected using probability, entropy and information gain, while grayscale matrices are used for a frame and the module uses a Fourier descriptor (FD). Before extracting these characteristics of the image, we associate the KNNGA-GL algorithm to optimize the result extracted from a large database. The experimental analysis of the proposed approach is simulated on the MATLAB2012a, which includes various functions for simulation. The analysis of our work is done on the basis of the known accuracy. After modeling, it is analyzed that our proposed system works well compared to other approaches*

Keywords: *Content based image retrieval (CBIR), Gain, GLCM, Fourier descriptors, Similarity matching, MATLAB2012A, KNN-Genetic algorithm.*

INTRODUCTION

At present, the use of information systems of images increases significantly with the development of broadband networks, powerful workstations, etc. A huge collection of images can be of interest to the public, from a collection of photos to web pages, otherwise a smooth video database. While the environment requires memory and processing power for processing and storage, it is necessary to have a competent directory and receive visual information when recording or after recording an image. During the current time, image categorization has expanded into an interesting field for investigation in the application. Effective indexing and restoration of huge amounts of color images, categorization plays a significant and stimulating role [1].

When processing images, the color histogram is a demonstration of the highlighting of colors in the image. For digital images, this is basically the number of pixels that have a color in a portion of a fixed list of color ranges that are expanded in the image color space in a set of all possible colors. The color histogram method is very simple and low-level technique and in practice has shown good results [2], specially designed for image indexing and restoration operations, while similar (unnecessary identical) images should be restored together with simple extraction of characteristics. This guarantees the complete transformation and invariance of the revolution in the descriptions of colors under categorical work.

Color is an equivalent three-dimensional vector in the direction of the position in the chromatic space. This allows us to choose a respiratory space of color and quantization phase in this color space. Instead of color, some other image functions are the extraction of textures and shapes, which is also necessary to extract this to reduce the cost of storage and ensure good image quality. In this article, we perform the extraction of all the characteristics of the image: colors, consistencies and shapes. Entropy, probability and aspect of information optimization can be used to extract color images, while gray-matched matrices are used to extract image textures. And for the image form, a Fourier descriptor (D) is used after the function extraction function is needed to reduce or optimize the image from the classification using KNN-GA (near K and the genetic algorithm). The classification of the extracted characteristic is performed on the basis of the similarity coefficient. The phase involved in image processing in the CBIR system is shown in Figure 1.

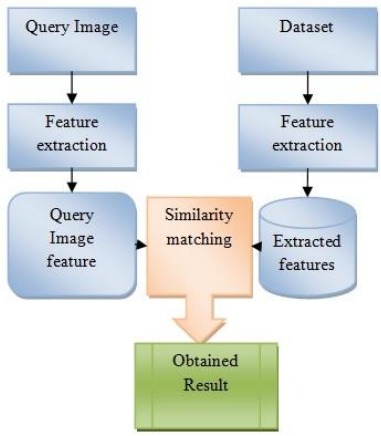

Fig.1 Content-Based Image Retrieval System

RELATED WORK

Extracting the functions from a set of a huge database is a vast area of research, and a lot of work has been done. In this section of this article, we present some of the previous work literature review done in the field of extracting features in the content search engine based on consistency, color and shape.

In this article, [3] the author proposed to predict the offset analysis of your maximum and polunablyudaemy BMMA (semi-BMMA) for the integration of unusual feedback properties and the development of information on unmarked models for SVM-based RF methods. BMMA distinguishes the optimistic feedback from the negative database, based on a local survey, while semi-BMMA can effectively enter information on unlabeled samples, starting from the Laplace regulator BMMA

In this article, [4] the author proposed to introduce a new feedback method (RF) to search for images based on content (CBIR), which uses a mixture of Gaussian (GM), modeled as an image. The GM of each image is acquired as a variation in global GM, which models the probability distribution of aspects of the image database. In each round of RF positive and negative examples are presented to the user of the current round are used for training machines (SVM) Support Vector to distinguish between relevant and irrelevant image based on user preferences. To transfer the similarity between the two images represented as GM approximation used Kullbaka-Leibler (KL), the calculation of which can be accelerated with acceleration, using the benefit of reality that GM images completely eliminated from the general model. The corresponding kernel function based on this distance between the GMs is used for the possible inclusion of GM in the SVM structure. Finally, comparative numerical experiments that showed the merits of the proposed RF system and the advantages presented.

In this article, the author proposed [1] to propose a new method for classifying color images based on content using the Machine Support Machine. Traditional methods of classification do not refer adequately to content based on image classification, one of the reasons for the greater dimensionality of the distinctive feature. The categorization of the color image is performed on the functions derived from the histograms of the color components.

In this article [6], the author has suggested that the modified K-nearest neighbor (MKNN) can be considered as a sort of weighted KNN to facilitate the search for the mark, approximating it by weighing the neighbors of the query. The procedure calculates the proportion of identical neighbors of the labels with respect to the entire number of neighbors. The MKNN classification is based on neighbors tested that include additional information in the evaluation using simple group labels.

In this article [7], the author proposed a well-developed and successful method for indexing and searching for images based on content. The system uses global and local image functions for indexing and fractional distance measurements as a measure of similarity for extraction. The images are quantized before the global functions are extracted. They also presented a new image segmentation method to effectively use the capabilities of the region. R * -The data structure is used to index the characteristics of the region. Experimental results have shown that the proposed system can promote research accuracy and reduce research time.

PROPOSED WORK

The CBIR seem to be extremely useful technique not simply for the management of large numbers of image data, excluding also helpful to a multiplicity of field such as research, clinical medication, education and visual knowledge. CBIR is aimed to retrieve preferred images based on the similarity measurements. The individuality of these similarity measurements includes intensity, color, texture, size, location and so on.

3.1 Feature Extraction Module

The input images, including the instruction and query phase, are everyone process in this section. It is also the most important in image retrieval. While color is the majority admired and instinctive feature base on human visualization, it is applied in the system. In order to get extra dominant features, the CCH method as well as

useful for extracting the imperative feature point. In this module two types of features are describe below:

I. Color Feature Extraction

Input images determination be separated into F*S grids ahead of this stage. Every grid is input to take out the color feature. Initial, the element calculate the average RGB value of the F*F grids. Second, the inside S*S grids in each F*S grids will in addition be input to analyze the average RGB value. The S*S grids' detail RGB information is add on after the F*S grids' color feature information. Everyone those are ready for first K-means clustering. Figure 2 demonstrate the color feature extractions of this segment.

II. CCH feature extraction

The system utilizes CCH (color co-occurrences histogram) to find out the important feature points. All the points are detected for preparing the input data of the neighborhood module and K-means clustering or KNN classifying. The information of CCH feature points, including the 64 dimensions data, combines with the neighborhood module result. Taking it as the input for the second round K-means clustering, the K-means clustering results in a fragment-based database, call the Code book. As the same implementation in query step, K-means is replaced by KNN algorithm. Query data imputed will be classified to improve the training code book, also correct classified result helps for quickly retrieval.

3.2 Grey-level co-occurrence matrix for texture

Grey-Level Co-occurrence Matrix texture dimensions have been the workhorse of image texture as they were anticipated by Haralick. So many image analyst, they are a button you thrust in the software that capitulates a band whose use progresses classification – or not. The foremost works are essentially condensed and mathematical, making the procedure intricate to understand for the student or front-line image analyst. Determine the selected Feature. This computation uses only the values in the GLCM.

I. Contrast
II. Correlation
III. Energy
IV. Homogeneity

These features are planned by distance 1 and angle 0, 45 and 90 degrees.

3.3 Fourier descriptors (FDs)

In general, Fourier descriptors (FDs) are obtained by applying Fourier transform on a shape signature, the normalized Fourier transformed coefficients are called the Fourier descriptors of the shape. The shape signature is any one-dimensional function representing two-dimensional areas. Tree shape signatures are considered in our case, these are centroid distance, complex coordinates, and curvature signature which is derived from shape boundary coordinates.

3.4 K-Means Clustering

K-means clustering is an approach to classify or to group objects based on aspects/features into K number of group [8]. K is positive integer number. The grouping is executed by minimizing the sum of squares of distances among data and the corresponding cluster centroid. Therefore the intention of K-mean clustering is to classify the data.

3.5 K- Nearest Neighbors Algorithm

It is a supervised learning algorithm where the outcomes of original instance query is classified based on majority of K-nearest neighbor group [9]. The principle of this algorithm is to classify a novel object based on attributes and training samples. The classifiers do not employ any model to fit and only based on memory. Specified a query point, we discover K digit of objects (training points) neighboring to the query point. The classification is using majority vote amongst the classification of the K objects. Whichever ties can be broken at random. K Nearest neighbor algorithm used neighborhood classification as the prediction value of the novel query instance.

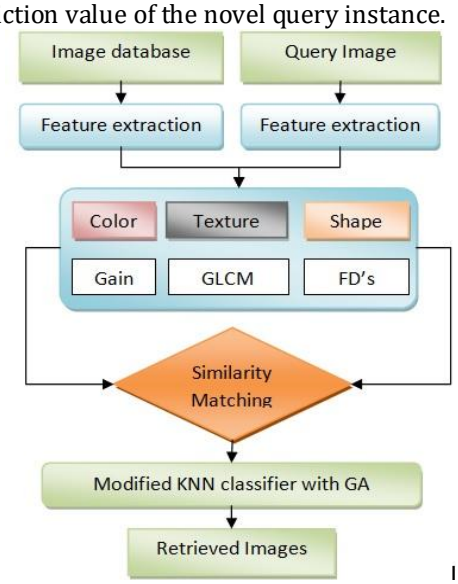

Fig 2: Block diagrams for KNNGA-GL

3.6 PRAPOSED METHOD

Step 1- Verify image database.
Step 2- Read all images from database.
Step 3- Extract image RGB features.
Step 4- Stored all extracted features in a separate single file each.
Step 5- Now read query images.

Step 6- Now applying features extraction for color, texture and shape.

Step 7- Color feature extraction use probability, entropy and gain to find out Majority of color features.

- Probability = no. of occurrences of a target event divided by no. of occurrences + the no. of fail occurs.
- Entropy = measures of uncertainly of a random variable (statistical measure of randomness) and disorder of an image and it achieves its target value.

  $$H(X) = \sum_{k=1}^{m} P(X).\log p(x)$$

- Gain = gain find out the maximum color value and majority of the color.

  Gain = total entropy – independent entropy

Step 8- Texture feature extraction used gray level co-occurrences matrix (GLCM) and edge histogram.

Step 9- Shape feature extraction, extract by Fourier descriptor (FD).

Step 10- After extract all the features of color, texture and shape. Now we will apply similarity matching.

Step 11- Now applying KNN classifier with GA. For classification of images, for best optimized retrieval results.

Step 12- Now used Euclidean's distances formula to calculate the distances between the query images and database.

Euclidean's distances = root $\sum_{i=1}^{k}(xi - yi)$^2

Step 13- Display the result in term of retrieved images.

EXPERIMENTAL RESULTS

To examine the efficiency of the proposed method of content based image retrieval here we used MATLAB 2012A for simulation purpose and done experiment into most popular dataset.

- **Image Data Set [5]**

The coral image data set is very famous image data set for research purpose especially for image retrieval systems. And there are huge amount of images with 10 classes available into that set, but here we used approx 150 images.

Here we have cross verified result with an existing method with our proposed method and found that our implemented method gives the better result to maximum set of images.

Fig. 3: Compared result of existing method with KNNGA-GL method for horse image

Fig. 4: Compared result of existing method with KNNGA-GL method for rose image

Here we have selected only 9 categories of images, and in table 1 shows an individual accuracy of each images and its comparison graph is shown in fig.5. The overall accuracy is shown in table 2 and the compare graph in shown in fig.6. The comparison between existing and proposed method in which the proposed method KNNGA-GL gives the 97.00% in tested images while existing method gives overall accuracy 87.00% on same test cases of the images.

Table 1: Class wise accuracy of different image classes between Existing & KNNGA-GL

Individual accuracy		
Method/Class	**Existing**	**KNNGA-GL**
C1	**100**	**100**
C2	**70**	**90**
C3	**100**	**100**
C4	**100**	**100**
C5	**100**	**100**
C6	**60**	**90**
C7	**80**	**100**
C8	**80**	**90**
C9	**90**	**100**

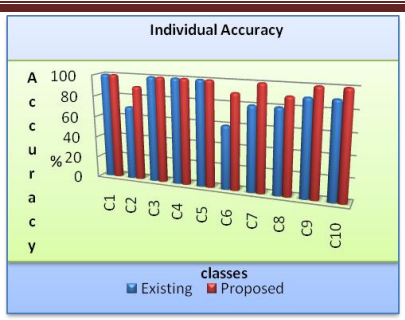

Fig. 5: Comparison of class wise accuracy

Table 2: Overall accuracy

Total accuracy		
Method	**Existing**	**KNNGA-GL**
Accuracy%	87.00%	97.00%

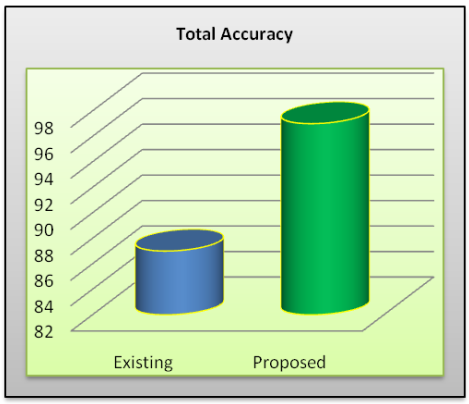

Fig.6: Comparison graph

CONCLUSION

Digital image processing is widely used are area for research for the feature extraction of image, text etc. In this paper we have proposed content based image retrieval using KNN (K-nearest neighbor) and (Genetic Algorithm) GA-GL. The majority of color feature is measured using entropy, probability and information gain while for feature extraction of texture is perform by gray level co-occurrence matrix and edge histogram method which effectively increased the quality of text and shape of image is estimated using Fourier descriptor methodology. After reducing the feature of image on the basis of color, texture and shape we apply hybrid method KNN and GA for similarity matching and optimize the extracted result to augment the quality of the image. The dissertation uses MATLAB 2012a simulation toolbox to simulate proposed approach. The experimental result of proposed approach gives most effective outcomes than existing system. For analyzing the performance of the proposed approach we use performance measurement accuracy in which our proposed approach outperforms. In future work, we need to perform the analysis of our proposed approach is for another performance metrics.

REFERENCES

1. Saurabh Agrawal, Nishchal K Verma, Prateek Tamrakar, Pradip Sircar, "Content Based Color Image Classification using SVM", In International Conference on Information Technology: New Generations, proceeding of IEEE, 978-0-7695-4367, 2011.
2. M. Swain and D. Ballard, "Indexing via color histograms," International Journal of Computer Vision, Vol. 7, pp. 11–32, 1991.
3. Lining Zhang, Lipo Wang ; Weisi Lin, "Semi-supervised Biased Maximum Margin Analysis for Interactive Image Retrieval", Image Processing, IEEE Transactions on (Volume:21 , Page(s):2294 - 2308 ISSN :1057-7149, 2012.
4. Marakakis, A., Siolas, G. ; Galatsanos, N. ; Likas, A. ; Stafylopatis, A., "Relevance feedback approach for image retrieval combining support vector machines and adapted gaussian mixture models", Image Processing, IET (Volume:5 , Issue: 6), Page(s):531 - 540 ISSN :1751-9659, IEEE 2011.
5. Image database: http://wang.ist.psu.edu/docs/related/
6. T. Dharani, I. Laurence Aroquiaraj, "Content Based Image Retrieval System using Feature Classification with Modified KNN Algorithm", Computer Vision and Pattern Recognition, IEEE 2013.
7. Suresh Pabboju, Venu Gopal Reddy, "A Novel Approach for Content-Based Image Indexing and Retrieval System using Global and Region Features", IJCSNS International Journal of Computer Science and Network Security, VOL.9 2009.
8. Rafel C. Gonzalez and Richard E. Woods, "Digital Image Processing", Second Edition, Pearson Education Asia, 2005.
9. Dengsheng Zhang and Guojun Lu, "Review of shape representation and description techniques", Pattern Recognition Society. Published by Elsevier Ltd, Vol.37, pp. 1-19, 2004
10. A. J. M. Traina, A. G. R. Balan, L. M. Bortolotti, and C. Traina Jr., "Content-based Image Retrieval Using Approximate Shape of Objects", Proceedings of the 17th IEEE Symposium on Computer-Based Medical Systems, pp. 91-96, 2004.
11. Aruna Verma, Deepti Sharma, "Content Based Image Retrieval Using Color, Texture and Shape Features", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 4, May 2014.