INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS AND INNOVATIONS IN TECHNOLOGY

ISSN: 2454-132X
Impact factor: 4.295
(Volume 4, Issue 1)

Available online at www.ijariit.com

Use of Artificial Intelligence Algorithm for Content-Based Image Retrieval System

Sumit Sharma
sharmasumit98@gmail.com
International Institute of Technology and Business, Jhundpur, Haryana

Kavita
kavita.ece87@gmail.com
International Institute of Technology and Business, Jhundpur, Haryana

ABSTRACT

Content-based image retrieval (CBIR) has most important research area in the last couple of years. Image retrieval is the system used for searching out the desired features of a query image. The main purpose of CBIR method is to obtain the accurate and fast results. In this paper, a Query image is retrieved from the database with high accuracy. This going through various steps namely, Pre-processing, feature extraction, Optimization, and matching. In pre-processing noise is removed by using filtering and color conversion is performed on the query image. For extracting features SURF (Speedup Robust Feature) descriptor is used. It is the enhanced form of SIFT (Scale Invariant Feature Transform) and more speedy than it. For optimization Genetic algorithm (GA) is used which improves the features of the extracted image. For matching Euclidian formula is used. It measures the similarity between the extracted feature and the database image. The simulation is carried out in MATLAB simulator tool with an accuracy of approximately 98 %.

Keywords: *Content-Based Image Retrieval (CBIR), SURF (Speedup Robust Feature), Euclidian Formula, Genetic Algorithm (GA).*

1. INTRODUCTION

CBIR is a method that is used to find an image from a set of the large database as per the user requirement. Initially, CBIR method is used to identify text by using text-based method based on the traditional management system. Images are recognized manually which is a costly process. Modern CBIR used the visual content of an image such as color, shape, and texture. Content-Based Image Retrieval [1] i.e. CBIR is a system by which various types of images are retrieved from a large database collection. Several optical features like color, texture, shape and structural arrangement, which are used to develop the databases, are extracted using distinctive techniques. CBIR is also known as query by image content (QBIC) and content-based visual information retrieval (CBVIR) [2]. In CBIR, content-based means the searching of the image on the basis of the actual content of image rather than its metadata.

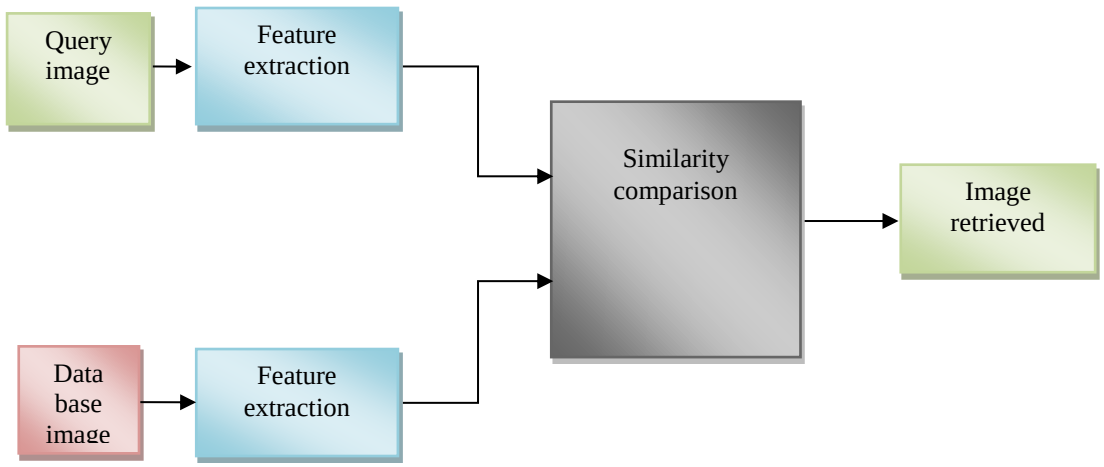

Figure 1: Block Diagram of CBIR

In the proposed work, a number of images of the building are prepared for the test. Test images are stored in the database; therefore, whenever query image is uploaded matching of the query image with the stored images is performed. When the image matched the desired image will be displayed on the screen.

The uploaded image goes through a number of different processes namely Pre-processing, Color conversion, Binarization, and Enhancement [3].

In preprocessing at first stage color conversion is applied on the input image and then channels like red, green and blue colors are extracted from the image. Next step is to convert the image into binary (0,1) form. In the edge detection process, the image is enhanced in the form of an edge, line, and contour. Image enhancement method is used for modifying the digital image to obtained accurate results for further image processing. The main aim of image enhancement is to process a query image for obtaining better results that are suitable for the particular application. Image enhancement sharp images feature such as boundaries, edges or contrast for obtaining a clear image. In the process of building, extracting accurately the features form the original image is a very important step [4].

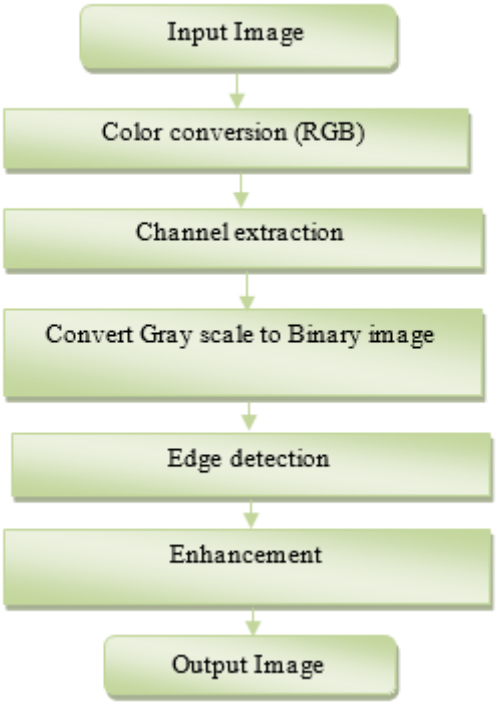

Figure 2: Pre-Processing Stages

2. RELATED WORK

Chaudhari, et al. [5] proposed an algorithm which covers the advantages of different other algorithms to enhance the accuracy and efficiency of the retrieval system. For matching the image color Histogram method has been used. Thus by taking the approximate shape, the retrieval system speed has been increased. The precision rate has been increased up to 72%.

Velmurugan et al. [6] SURF descriptor have been used for enhancing the retrieval accuracy of the system. In the proposed work SURF has been used along with Color Moments as SURF works only with gray scale. For matching the features The Best First search algorithm has been used and finally, Voting method has used to position and retrieved the matched images from the database.

Gnanaraja et al. [7] proposed a graph-based method in CBIR system. A texture based Segmentation algorithm has been used to build the relation between regions and obtained better-enhanced image.

Lai, et al. [8] Represent a user-oriented architecture in the CBIR scheme. For extracting visual features the proposed method acts as a bridge between visual and human perception. The computational parameters like precision, recall performing well.

3. PROPOSED TECHNIQUES

To retrieve the query image successfully with higher accuracy mainly two algorithms are used. The first algorithm named as SURF (Speedup Robust Features) is used for extracting features from the image and second algorithm named as the Genetic algorithm is used for optimizing the features.

3.1 SURF (Speedup Robust Features)

SURF algorithm is the enhanced form of SIFT algorithm. This technique uses a 'quick Multi scale key point detector' to find the key points. A typical version of SURF is faster than the probe, and the technology is better than any other feature extraction techniques. SURF algorithm comprises of three main stages, namely, interest point detection, local neighborhood and matching. For the detection of, this uses a Gaussian square smoothing filter approximation [9, 10].

Algorithm 1st: Proposed SURF Algorithm

For I = 1 to all frames
Detection = Feature points (I)
Points localization= DoG (I)
Where DoG is Difference of Gaussians
If localization need orientation
Orientation=Keypoint_localization (I)
End
SURF_point=All best Feature
Compare the all point using matching and the descriptors obtained from different images is store for the GA
End

Save SURF_points of proposed work for the next phase and we apply the Genetic Algorithm on the SURF_points and find the optimal solution of proposed work for the training and testing.

3.2 Genetic Algorithm

Feature extraction is a more general method in which the original set of features is transformed to provide a new set of features. An attempt at reducing the number of features in an image code using a genetic algorithm is made. It presumes that the potential solution of a problem is an individual and can be represented by a set of parameters. These parameters are regarded as the genes of a chromosome and can be structured by a string of values in binary form. The major components of Genetic Algorithm are named as a crossover, mutation, and a fitness function [11, 12].

Algorithm 2nd: Proposed Genetic Algorithm

Define the population size of the GA (50)
Initialized the GA in MATLAB
Set all initialization parameters
Load SURF feature sets
For I = 1 to all features
For r=1 to all rows (features)
For c=1 to all columns (features)
Define Ft (Threshold) = Average of feature value
Define Fs = SURF_feature (r, c)
Call fitness function
Fit_data= fitness functions (Ft, Fs)
If fitness functions==True
Fit_data= Fs
Else
Fit_data= Ft
Consider as best solution for GA and store as GA_data
End
End
End
End

3.3 EUCLIDIAN DISTANCE (EU)

If $x = (r_1, s_1)$ and y=(r_2, s_2) are the two points of an image the Euclidian distance between these points is given by [13]:
EU (x,y)=$\sqrt{(r_1 - r_2)^2 + (s_1 - s_2)^2}$

On the basis of above-optimized data, we can retrieve the most similar images from the database. The simulation of proposed work is described in next section.

4. SIMULATION RESULT

The simulation of the proposed work is carried out in MATLAB simulator tool. It is a Matrix based software which is designed by linear system package (LINPACK) and Eigen system package (EISPACK). This tool is mostly used by the researchers, designers, and teachers. There are two section 1st is training uploading, 2nd is testing. Both sections have some basic steps like pre-processing, sparse representation, feature extraction, feature optimization and in testing we use one more step which is called matching and for the matching in proposed work we use Euclidean distance technique. After the simulation, we can calculate the performance metrics like Precision rate, Recall rate, F-measure, and Accuracy of proposed work.

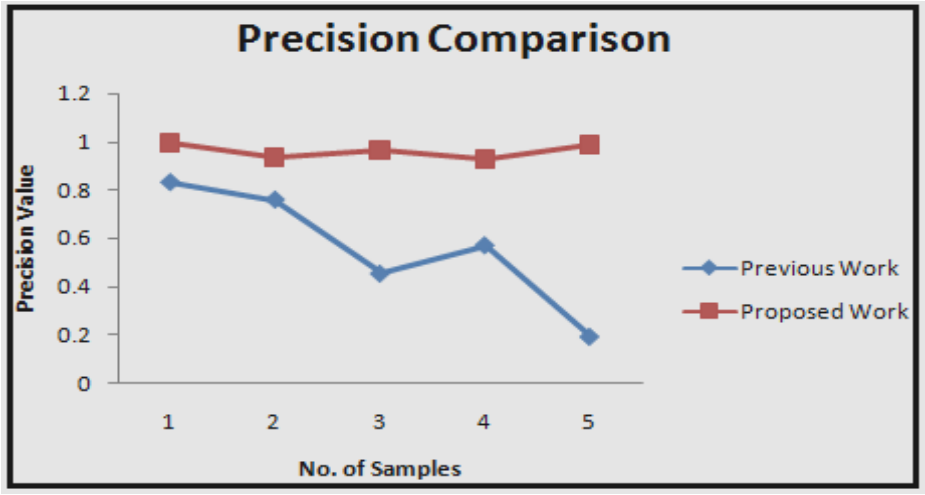

Figure 3: Comparison of Precision Rate

Above figure shows the graphical representation of precision value for proposed work and we compare the proposed precision rate with previous work and founded that the proposed precision is better as compared to the previous work.

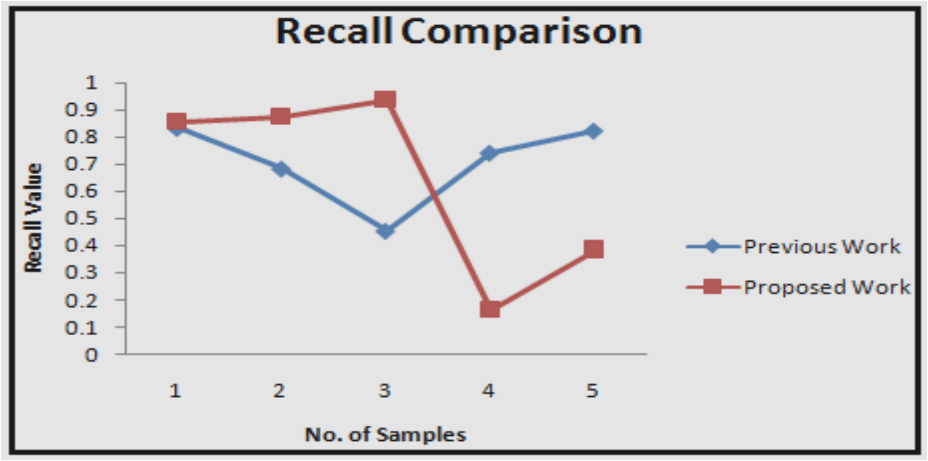

Figure 4: Comparison of Recall Rate

Above figure shows the graphical representation of recall value for proposed work and we compare the proposed recall rate with previous work and founded that the proposed recall rate is better as compared to the previous work by using the SURF descriptor with GA optimization technique.

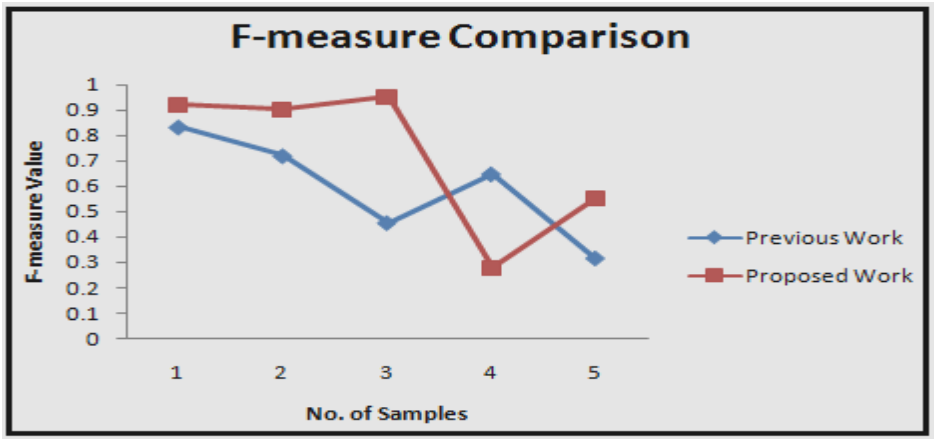

Figure Error! No text of specified style in document.**: Comparison of F-measure Rate**

Above figure show the graphical representation of F-measure value for proposed work and we compare the proposed F-measure with previous work and founded that the proposed F-measure is better as compared to the previous work. If the F-measure is better, then we can say that the retrieving rate of proposed work is better as compared to the previous work.

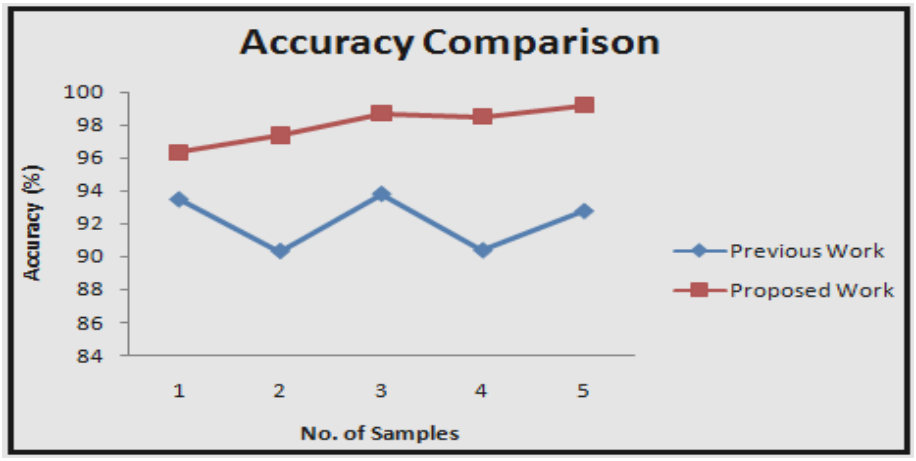

Figure 6: Comparison of Accuracy

Above figure shows the graphical representation of Accuracy value for proposed work and we compare the proposed accuracy with previous work and founded that the proposed accuracy is better as compared to the previous work. The accuracy value shows the retrieving rate by using the SURF feature extraction based on the genetic algorithm for the feature optimization.

5. CONCLUSION

The tense growth in the sizes of images databases has stirred the designing of practical and efficient retrieval systems. The main objective of this research is to retrieve the images from database in a fast and an efficient manner using pre-processing, feature extraction, similarity measures, and classification methods. In the proposed work optimization algorithm (GA) has been used to increase the retrieval accuracy of the images in case of a large dataset. SURF descriptor has been used to extract the features from the pre-processed images. This is one of the speedy descriptors among other existing feature extraction descriptor. Thus by using SURF descriptor the speed of the system has been increased. The performance of the system such as precision, Recall, F-measure, and accuracy has been increased. The average value obtained for precision, Recall, F-measure, and accuracy are 0.563, 0.707, 05934, and 92.144 respectively. In the proposed work by using genetic algorithm the retrieval accuracy of the images is better than the existing work.

6. REFERENCES

[1] Liu, Y., Zhang, D., Lu, G., & Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, *40*(1), 262-282.
[2] Shete, D. S., Chavan, M. S., & Kolhapur, K. (2012). Content-Based Image Retrieval. *InternationalJournal of Emerging Technology and Advanced Engineering*, *2*(9), 85-90.
[3] Lu, Y., Hu, C., Zhu, X., Zhang, H., & Yang, Q. (2000, October). A unified framework for semantics and feature based relevance feedback in image retrieval systems. In *Proceedings of the eighth ACM international conference on Multimedia* (pp. 31-37). ACM.
[4] Cai, W., Feng, D., & Fulton, R. (2000). Content-based retrieval of dynamic PET functional images. *IEEE Transactions on Information Technology in Biomedicine*, *4*(2), 152-158.
[5] Chaudhari, R., & Patil, A. M. (2012). Content-based image retrieval using color and shape features. *International Journal of Advanced Research in Electrical, Electronics and Instrumentation Engineering*, *1*(5), 67-72.
[6] Velmurugan, K., & Baboo, L. D. S. S. (2011). Content-based image retrieval using SURF and color moments. *Global Journal of Computer Science and Technology*.
[7] Gnanaraja, R., Jagadishkumar, B., Premkumar, S. T., & Kumar, B. S. (2014). Content-Based Image Retrieval (CBIR) Using Segmentation Process. *International Journal of Innovative Research in Science, Engineering, and Technology*, *3*, 384-389.
[8] Lai, C. C., & Chen, Y. C. (2011). A user-oriented image retrieval system based on an interactive genetic algorithm. *IEEE transactions on instrumentation and measurement*, *60*(10), 3318-3325.
[9] Shivakumar, B. L., & Baboo, S. (2011). Detection of region duplication forgery in digital images using SURF. *International Journal of computer science Issues*, *8*(4), 199-205.
[10] Panchal, P. M., Panchal, S. R., & Shah, S. K. (2013). A comparison of SIFT and SURF. *International Journal of Innovative Research in Computer and Communication Engineering*, *1*(2), 323-327.
[11] Gali, R., Dewal, M. L., & Anand, R. S. (2012, July). A genetic algorithm for content-based image retrieval. In *Computational Intelligence, Communication Systems and Networks (CICSyN), 2012 Fourth International Conference on* (pp. 243-247). IEEE.
[12] Cho, S. B., & Lee, J. Y. (2002). A human-oriented image retrieval system using the interactive genetic algorithm. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, *32*(3), 452-458.
[13] Bian, W., & Tao, D. (2010). Biased discriminant euclidean embedding for content-based image retrieval. *IEEE transactions on image processing*, *19*(2), 545-554.

