CONTENT BASED IMAGE RETRIEVAL SYSTEM FOR MASSIVE DATA

Hardeep Singh, M.Tech. Research Scholar,
Department of CSE,
SBBS University, Jalandhar
(singhardeep101@gmail.com)

Dr. Vijay Dhir, Professor,
Department of CSE,
SBBS University, Jalandhar
(drvijaydhir@gmail.com)

Dr.Jagdeep Kaur, Associate Professor,Department of CSE
SBBS University, Jalandhar
(sbbsuuietcse@gmail.com)

Abstract:
Improvements in current digital storage media, especially in image capturing devices such as webcams, digital cameras, and the exponential growthin the Internet generates a millions of images per day. This requires an efficient retrieval system to extract useful information from these images. For an example visual information of images is required in different fields such as healthcare, architecture drawing, academic, crime prevention, etc. To this end, many image retrieval systems have been developed. Content-Based Image Retrieval (CBIR) systems as it are now a big demand in society. It is a method of retrieving similar image from the massive dataset. A number of approaches are used to develop an efficient CBIR system, out of which Scale Invariant Feature Transform (SIFT) is very popular. In this paper, we designed a novel CBIR system for massive dataset by taking the advantage of SIFT as feature extraction algorithm. To obtained high quality images, the image features are optimized using Artificial Bee Colony Algorithm (ABC). Then for training and classification Artificial neural Network (ANN) is used. Results are tested by uploading different images in terms of precision, recall, and F-measure. The experiment results show satisfied results compared to past work.

Keywords: **Content-Based Image Retrieval, Scale Invariant Feature Transform, Artificial Bee Colony Algorithm, Artificial neural Network.**

DOI Number: 10.14704/nq.2022.20.8.NQ44232 **NeuroQuantology 2022 ;20(8):2127-2134**

1. Introduction

Technology is changing with each passing day and the demand for the latest devices is increasing. People are more interested in the latest models with built-in cameras. Due to the existence of Internet technology (IT) and the affordable digital image sensors, image databases have proliferated. Due to these image databases, there is an increasing need for efficient image retrieval search algorithms to meet user needs [1]. Efforts have been made to reduce the semantic gap between low-level features and human visual perception to improve Content Based Image Retrieval (CBIR). This system searches images within the databases for the uploaded query images. This image with smallest similarity index is considered as the best matched image [2]. To measure similarity index, initially the feature vector of each images are extracted using some

feature descriptor algorithm. The feature vector and the representation of the image should be discriminative in order to differentiate images. Significantly, it is also expected to remain unchanged for some transformations. According to the image representation, the similarity measure between two images should reflect the semantic relatedness. These two components are crucial to retrieval image [3].

In fact, to retrieve accurate images from large-scale databases is still challenging. The biggest challenge is the semantic gap between the high-level meaning of an image and its low-level visual features. To minimize this semantic gap, efforts have been done by both researchers and scientists [4]. As a result, recent years have seen CBIR make great strides. For example, Google and Baidu are popular search engines that can search for relevant images for any uploaded image [5]. Some e-commerce sites such as Alibaba, Amazon and eBay have similar product search functions. Social platforms like Pinterest have similar content recommendations [6].In this research, to improve retrieval performance of the CBIR system for massive data scale-invariant feature transform (SIFT) with Artificial Bee Colony (ABC) and Artificial neural network (ANN) is used as a feature extraction, feature optimization, and feature classification approaches respectively.

2. Related work

Yadav, and Kaushik (2017)have proposed a CBIR system for medical bigdata. The features like shape of the medical images have been extracted using canny edge detection algorithm. Bayesian classification technique has been used to train and then to classify the images. Shape features of images are applied as an input to the classifier and then indexing has been performed using Euclidean distance method [7]. **Prasanth, and Gunasekaran (2019)**have applied mutual refinement scheme for retrieving big data images using hash tag graph. Initially, the images are pre-processed and then features like frequency and entropy has been extracted. For training, hash tag method was used, which performs dual operations that s training as well as removing the improbability of data.Same process upto feature extraction has been applied on the query image. Then the extracted query features like frequency and entropy has been compared with the hash tag graph. If data is matched then the image is retrieval and visualized. The authors promised to retrieve big data using this technique with high accuracy [8].**Munjal, and Bhatia (2019) have PPLIED** MPEG-7 and Edge Directivity as feature extraction approach. Initially the images are retrieved suing SIMPLE-CEDD approach and the results are satisfactory. The system has sort and filtered the images effectively, which makes the computation time small [9]. **Ramachandran et al. (2020)** have provided a comparative review of multiple image dimension reduction techniques and compare their results. These techniques are principal component analysis (PCA), linear discriminant analysis (LDA), kernel principal component analysis (KPCA), singular value decomposition (SVD), and independent component analysis (ICA). The comparison has been provided on the basis of structure, complexity of multi-dimension dataset, error handline and class label [10].**Hussain,and Surendran (2020)**have proposed a CBIR system for Cluster Based Images. Initially features like colour shape etc is extracted. Axiom 2 is used to extract colour features, which were then stored into the feature database. Then, Bees algorithm has been used to perform local aligned search. Then for retrieving image similarity index among images has been determine. The proposed system provided better result with high accuracy [11].**Muthiah, and Logashanmugam (2021)**have designed CBIR system using conventional neural networks (CNN).The researchers have considered different features of images that include image moment, intensity, image coefficient and many more. Different datasets were used for training and testing. But this makes the system complex. The researchers said that using DWT features better accuracy can be obtained [12].**Zhang et**

al. (2021)have developed a CBIR system for massive images. For retrieving images the concept of Big data image mining has been used. It works in three main steps, segmentation image mining and merging. As per Master-slave operation, query from user has been taken. Then for similarity and matching images, parallel computation process has been performed using Map reduce module. Then the final retrieved image has been provided to the user. The system provides satisfactory results in terms of recall, precision etc [13].

3. Methodology

Figure 1 represents the flowchart for the proposed massive data retrieval using AI approach. Initially, a large set of images was collected and processed to prepare a database. A Graphical User interface (GUI) has been designed for the same.

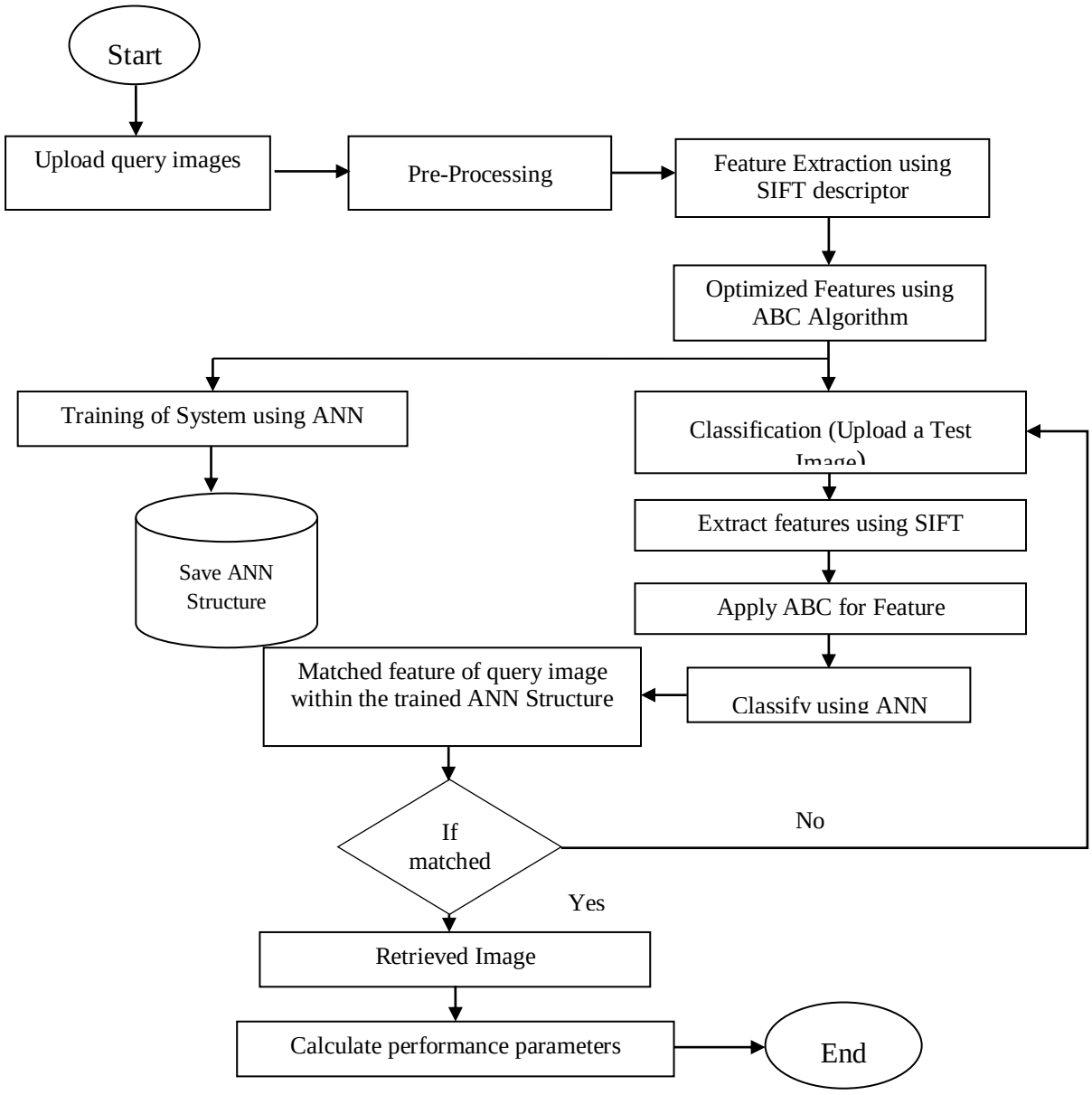

Fig.1 Proposed Work

Step 2: Upload a database of different images to train and validate the system. In both training and testing parts, images are uploaded for processing. In training, the amount of data is larger, but in testing a single query image is uploaded. The uploaded image is shown in Figure 2.

Fig. 2 Query Image

Step 3: Preprocessing is performed to remove noise or background from the image. This step helps to improve the quality of the images. In every image retrieval system, image quality is the main factor that needs to be improved by removing the extra part or information from the image. The color histogram is determined to determine the color composition along with the number of pixels present in the image. The obtained color histogram for the uploaded image is shown in Figure 3.

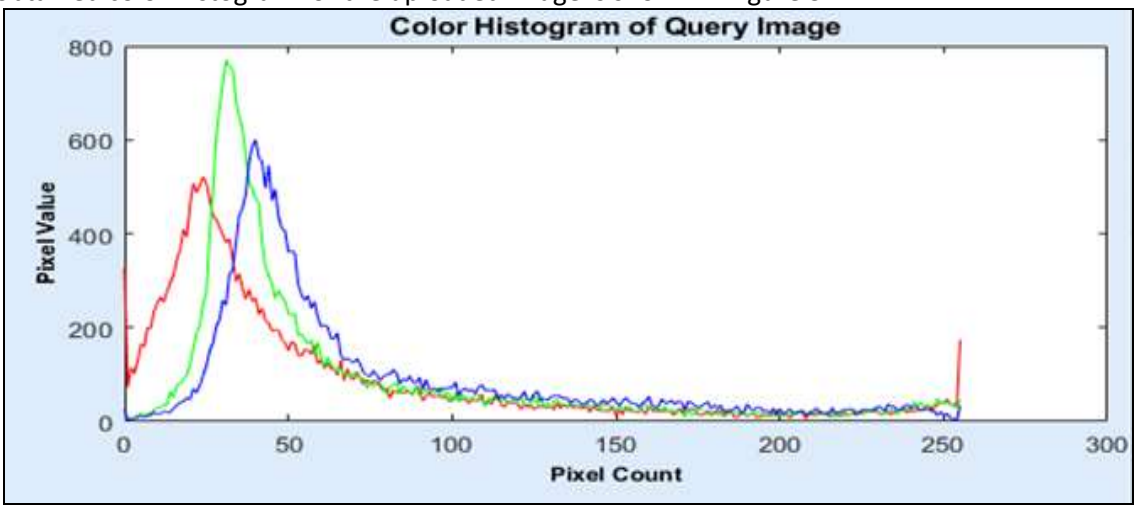

Fig. 3 Color Histogram

Step 4: Next step is to extract features like color, shape from the pre-processed image. This is done using the SIFT feature as descriptor. The concept of using SIFT algorithm is that it is invariant to scale, rotation and translation as well as partially invariant to affine distortion and illumination changes. To efficiently match query images and index images in an image database, all images are first characterized as a set of SIFT features. Every SIFT feature signify a vector of local image measurements in a way that is invariant to image translation, scaling, rotation, and partially invariant to illumination changes and local image deformation. As mentioned earlier, SIFT feature locations are proficiently detected by classifying the maxima and minima of the DoG function in the scale space. Feature vectors are formed by measuring

local image gradients in the region surrounding each location at coordinates relative to the feature's location, scale, and orientation.After SIFT algorithm computation, it returns a set of features based on key points.

Step 5: The keypoints of each image has been optimized by using fitness function of ABC algorithm. It is an optimization algorithm, which provides better and the best images based on the designed fitness function.

Step 6: The artificial neural network is initialized for classification using the feedforward backpropagation concept. After the system is trained, the trained structure is saved and used in the classification part to match the query image with the training dataset. Here, Euclidean distance formula is used to find distance and then used for matching.During the testing phase, upload the query image and repeat steps from 3 to 5. In the classification part, the result is computed if the query image matches the trained ANN structure.

4. Result and Discussion

In this section, the performance of designed CBIR based massive image retrieval system has been presented. Dataset of total 2853 images has been created and is used to make performance evaluation. The analysed result in terms of kappa coefficient and overall accuracy is listed below. The performance of each uploaded image is then evaluated in terms of precision, recall, F-measure and accuracy. The observed values are listed in Table 1.

Table 1 Performance Parameters

No. of Test Images	Precision	Recall	F-measure
1	97.211	98.234	97.7198
2	98.467	97.267	97.8633
3	97.783	96.289	97.0302
4	98.128	97.457	97.7913
5	96.234	98.684	97.4436
6	97.028	99.297	98.1493
7	97.258	96.012	96.6309
8	96.237	97.385	96.8075
9	98.289	97.175	97.7288
10	98.687	96.157	97.4055

Fig. 4 Analysed Parameters

Figure 4 shows the analysed values of precision, recall, and F-measure. The graph is plotted between the number of test images and the parametric values in percentage. The red bar line represents recall values, the blue bar line represents precision value and the green bar line represents F-measure value. For each uploaded query image values are computed. The average value for precision was analysed as 97.53 %. The average value of recall and F-measure was analysed as 97.39 %, and 97.45 % respectively. To show improvement of the proposed work, comparison between proposed and existing work is provided in Figure 5.Hereby, to show improvement of the proposed work using technique of Artificial Neural Network (ANN) and comparison of proposed work with existing work having using of technique Euclidean Distance [13] is provided in Table 5 and Figure 6.

Table Error! No text of specified style in document.**. Comparison of work**

	Precision	**Recall**	**F-measure**
Proposed using (ANN)	97.53	97.39	97.45
Existing using (Euclidean distance)	97	97	97

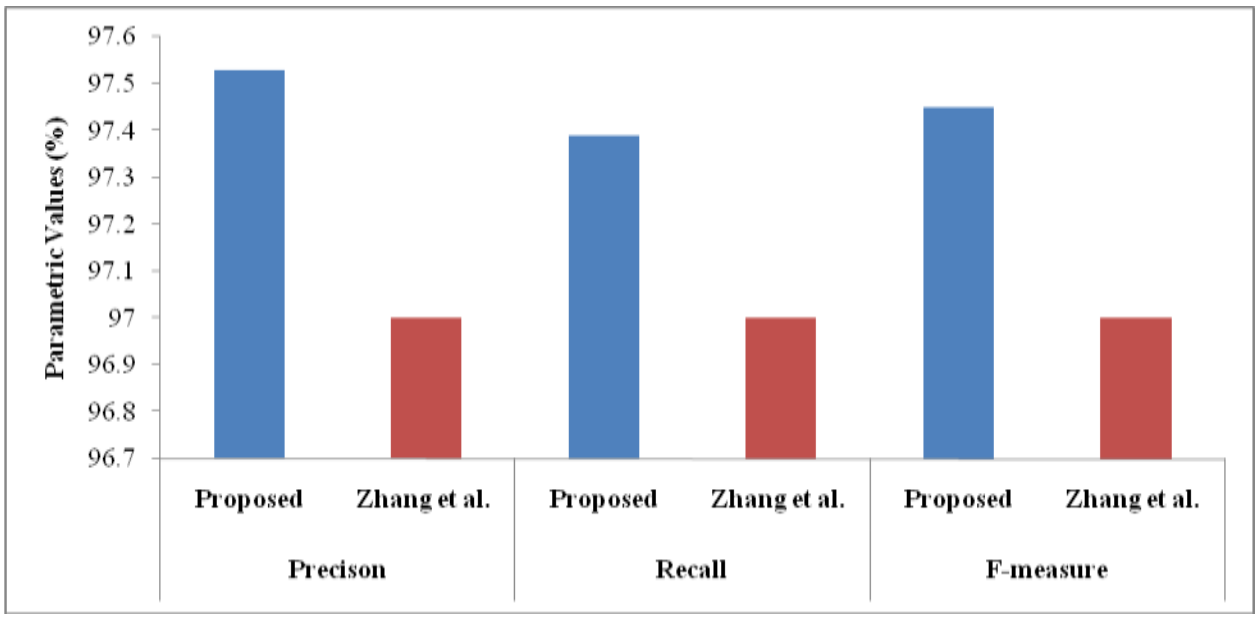

Fig. 6 Comparison of Work

The blue bars show the average values of precision, recall, and F-measure of the proposed work and the red bar indicates the parametric values of the existing work. From the graph it is clearly observed that there is an improvement in the proposed work, when CBIR system for massive data using AI approach has been design. The improvement in the precision, recall, and F-measure are observed as 0.55 %, 0.4 %, and 0.46 % respectively.

5. Conclusion

In this paper, an efficient CBIR system based of SIFT with ABC and ANN algorithm has been presented for massive data retrieval. Size of the extracted features using SIFT algorithm has been optimized by designing an appropriate fitness function of ABC algorithm. Here, ABC algorithm works as a size reduction algorithm to obtained high quality database. This increases the efficacy of the trained ANN structure. It has been inferred that the proposed SIFT with ABC and ANN based CBIR system perform with high accuracy compared to the already exiting methods. Following observations are made:

- The average value for precision was analysed as 97.53 %.
- The average value of recall and F-measure was analysed as 97.39 %, and 97.45 % respectively.
- The improvement in the precision, recall, and F-measure are observed as 0.55 %, 0.4 %, and 0.46 % respectively compared to past work.

References

[1]. Meshram, K. S., & Bhombe, D. L. (2016). Improving Content based Image Retrieval System Using SIFT and SVM.

[2]. Latif, A., Rasheed, A., Sajid, U., Ahmed, J., Ali, N., Ratyal, N. I., ... & Khalil, T. (2019). Content-based image retrieval and feature extraction: a comprehensive review. *Mathematical Problems in Engineering*, *2019*.

[3]. Aziz, M. A. E., Ewees, A. A., & Hassanien, A. E. (2018). Multi-objective whale optimization algorithm for content-based image retrieval. *Multimedia tools and applications*, *77*(19), 26135-26172.

[4]. Kavitha, P. K., & Vidhya Saraswathi, P. (2021). Content based satellite image retrieval system using fuzzy clustering. *Journal of Ambient Intelligence and Humanized Computing*, *12*(5), 5541-5552.

[5]. Li, X., Yang, J., & Ma, J. (2021). Recent developments of content-based image retrieval (CBIR). *Neurocomputing*, *452*, 675-689.

[6]. Alzu'bi, A., Amira, A., & Ramzan, N. (2015). Semantic content-based image retrieval: A comprehensive study. *Journal of Visual Communication and Image Representation*, *32*, 20-54.

[7]. Yadav, V., & Kaushik, V. D. (2017). Indexing and retrieval of medical images form big data. *International Journal of Computer Science and Information Security*, *15*(1), 327.

[8]. Prasanth, T., & Gunasekaran, M. (2019). A mutual refinement technique for big data retrieval using hash tag graph. *Cluster Computing*, *22*(1), 2027-2037.

[9]. Munjal, M. N., & Bhatia, S. (2019, February). A novel technique for effective image gallery search using content based image retrieval system. In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)* (pp. 25-29). IEEE.

[10]. Ramachandran, R., Ravichandran, G., & Raveendran, A. (2020, March). Evaluation of dimensionality reduction techniques for big data. In *2020 Fourth International Conference on Computing Methodologies and Communication (ICCMC)* (pp. 226-231). IEEE.

[11]. Hussain, D. M., & Surendran, D. (2020). Content based image retrieval using bees algorithm and simulated annealing approach in medical big data applications. *Multimedia Tools and Applications*, *79*(5), 3683-3698.

[12]. Muthiah, M. A., & Logashanmugam, E. (2021). Feasibility of conventional neural networks for content-based image retrieval in big data. *International Journal of Public Sector Performance Management*, *8*(3), 253-263.

[13]. Zhang, K., Chen, K., & Fan, B. (2021). Massive picture retrieval system based on big data image mining. *Future Generation Computer Systems*, *121*, 54-58.