A Hybrid Approach for Content based Image Retrieval

Preetpal kour, M.Tech Scholar, Department of Computer Science and Engineering, Chaudhary Devi Lal University, Sirsa Haryana, India, preetpal10995@gmail.com

Avininder Singh, Assistant Professor, Department of Computer Science and Engineering, Chaudhary Devi Lal University, Sirsa, Haryana, India, avnindersingh64@gmail.com

Abstract: The image search system processes the visual query and retrieves the relevant visual documents efficiently from the very large scale visual corpus. It will be very difficult to manage this database of images stored at the remote servers. The right tool will be required which can process these images for different operations. These operation include searching etc. It will be difficult to classify the images into groups and then search each class for providing the image as the information against the user request query. The content based image retrieval is the most suitable way to identify the image from the large repository. It will search the image from the large set of images based on contents rather than the image name. It will be having less time to search the image from the large repository when the image is retrieved using content based. In the current research the hybrid approach for content based image retrieval is performed. This proposed procedure will be in the first step perform the classification of the image into multiple classes. The classes are prepared based on the attributes values.

Keywords: content based, QBIC, CBIR, image retrieval, DWT

I.   INTRODUCTION

With the gradual development of internet technology and digital devices billions of people are browsing web and sharing the information in the form of text, images, audio or video. Image Retrieval (IR) is the mechanism by which the images related to query image are searched from the images database and user retrieves the most similar images from the database[2]. The image search system processes the visual query and retrieves the relevant visual documents efficiently from the very large scale visual corpus. This image retrieval system is known as Content Based Information Retrieval system (CBIR) or query by image content retrieval system (QBIC). In technical point of view, there are main three steps in content-based image retrieval system: image representation, image organization, and image similarity measurement. CBIR is to retrieve information from the very big visual database thus organizing the large scale database efficiently and identifying the relevant results of a given query within acceptable time limits and storage demands is the main goal of image retrieval system[3].

Description based search

It is the technique where large amount of information will be stored along to the image. This image description will include the purpose and detailing about the contents of the images. These descriptions are filled by the image generators or may be storage service providers. These information will be used for searching purpose at the later occasion when user query will be submitted. This manual technique of giving description will be having wide number of errors[1].

Wrong description: there are various occasions when the description of the image will be written wrong. This may be intentionally or may be different perception of the author. There is no standard rule that is set for defining the description of the image. This will leads to creating the wrong results for the user query[2].

Incomplete description: there are various occasions when the description of the image will be stored wrong. This means the description written may not be much helpful for searching purpose. This will leads to wrong or incomplete results generation[5].

Language compatibility issue: there are various authors who are having different language which other than the standard language. These languages are difficult to understand and also difficult for processing by the search engine[3].

Types of Information Retrieval Systems

According to the queries content used to retrieve images, the CBIR system may be classified into five categories. 'query by keyword', 'query by example', 'query by sketch', 'query by color map', 'query by concept layout'. User may use keywords as a query to retrieve images. In

query by example; user uses an example image to retrieve similar images. The user may express his intention through a sketch map used as query by sketch. A user specifies the distribution of colors in a given grid pattern generating a color map used as a query to retrieve the images with similar colors. Fusion of multiple keywords used in query formation to retrieve the images related to fused keywords is done in concept by layout[12].
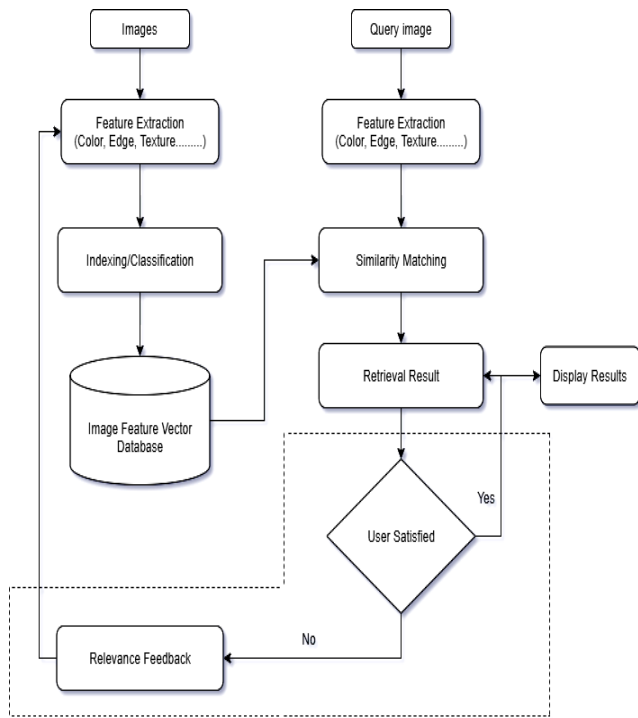
Fig 1: Genaral Framework for Content based Image Retrieval[12]

Hybrid approach

The hybrid approach for the content based image reterival is the efficient way for successful searching of the image from the lareg dataset. There are two steps that are required for searching the image, in the first step there is classification of the images based on various features. The classification will subdivide the larger set into smaller sets, in the second step specific class will be picked for the features extraction and the image identification.

II. LITERATURE SURVEY

[1]Anan Banharnsakun,(2020): author in this paper has worked on the genetic based image retrieval system where the image is recognized based on the content rather than the description. The proposed technique suggested by the author is based on artificial bee colony optimization technique. It identifies the attributes of the image based on the features set or we can say the optimization technique. There are various image features that are read by the technique and identification of the optimized features are performed so that the image with high proportional features are retrieved.   There are various features which are there can be considered as optimized features. The

performance of the optimized process is defined in terms of various performance parameters such that the proposed image is providing better results.

[2]Manpreet Kaur, Sakshi Dhingra (2020): the author in this paper has performed the comparative analysis of the image classification techniques which classify the images into various sub categories. The classification of the images is performed using various classification techniques for example Artificial neural network, KNN, Naïve Bayes, Fuzzy neural network etc. These techniques are supplied with various features identified using statistical method. These features are supplied as the input to the classification technique. This will classify the images based on the statistical features. The results for the classification techniques are compared on the basis of various performance parameters like accuracy, precision, recall, specificity etc. The Naïve Bayes is providing better results in all the aspects such shows higher accuracy and specificity.

[3]Manar Abdulkareem Al-Abaji (2019): the author in this paper has proposed content based image retrieval based on the optimized technique named as Cuckoo search algorithm. The technique first identifies various features based on the statistical measure. These features are processed to generate the classification based on the different classification processes. There are plenty of the techniques where the image retrieval will be required based on these features. It will be comparatively carries less effort to search the image once the small classes of the total image set is prepared. There are plenty of the image procedures that are required which provide the way for searching the image based on the contents and image procedure. The accuracy of the proposed procedure is measured based on different parameters, which shows better performance compared to the other techniques.

[4]M. N. Munjal and S. Bhatia (2019): the author in this paper has proposed a novel technique for the content based image retrieval from the large data repository. The author has proposed a technique where the large images of same purpose are collected into the single repository. The image from the large repository will be searched based on the content based identification. It is the technique basically identifies the features from the large image, these features are statistical features such that the various features are identified from the large image. These features are collected and passed to the novel technique, which identifies the image based on these features. The proposed technique is providing better results in the terms of accuracy and precision.

[5]Zhou Bing and Yang Xin-xin (2010): the author in this paper has proposed a technique based on rotation and flip for the image such that the image matching based on content based image retrieval perform the better results. There traditional techniques are having higher time and

lower accuracy for the image retrieval from the large data repository. There are various premium entities that are there to be working in the light such that some operations are being performed with higher level of accuracy. The current paper provides the technique for the image retrieval based on the content. There are various options that are available such that the operations are performed with the given set of rules. The proposed technique is providing higher accuracy and better operational skills such that the given set of images from the large data repository can be performed well. There are higher number of tools and techniques are available which provide the ability to set the system for image recognition with less time and higher accuracy.

[6]Shao Hong, Cui Wen-cheng and Tang Li(2015): the author in this paper has proposed a scheme for the image retrieval in the field of medical. There are various images which belongs to the different category. These different category images will be having different resolution and intensity values. The images in the category of medical field are having special features. The majority of the images belongs to gray scale and are having different dimensions. These images are having higher scale of intensity overlapping. These images retrieval based on the contents are having better results for the medical field, as these images requires higher level of accuracy   and requires higher skills and are having higher skills for the system management. There are various features which are based on the contents are having higher level ability to integrate the system for the content based image retrieval.

III. ALGORITHM

There are various steps that are to be taken place for the content based image retrieval. There are various images that will be used for the matching process.

Step1: the first step will be to input the image from the user with any of the image source. The inputted image will be processed thereafter.

Step2: the images dataset stored into the memory will be classified into the multiple classes. Each class of the image denotes the images with similar features.

Step3: the features are calculated for the inputted image, these features are various morphological and intensity based features.

Step4: the features of the inputted image will be matched to the specific class of the image based on various features.

Step5: extract the images to which the features are matched to the maximum extent.

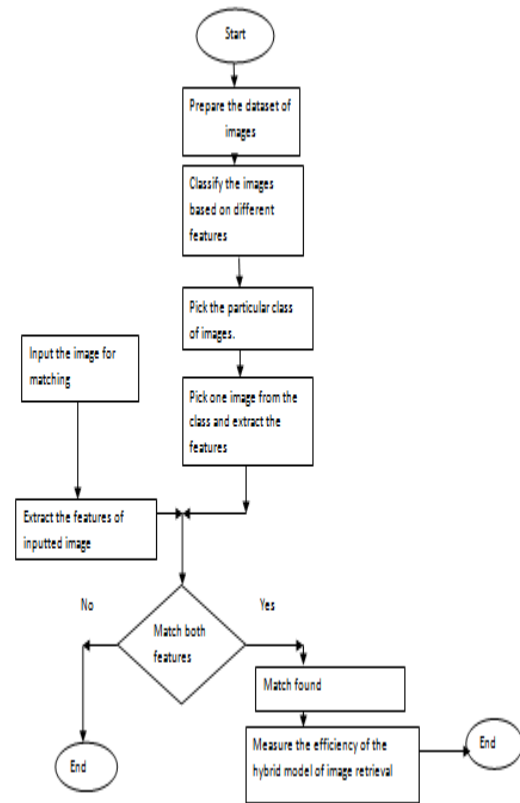Step6: the accuracy of the image matching is calculated.

IV. FLOWCHART

Fig.2 flowchart

V. MOTIVATION

The major research gap lies for reducing machine training time for training large dataset image samples. The existing databases need further refinement for better retrieval results or new datasets are to be generated to foster upcoming demands. The review of literature spreads light on the major issues to be addressed in the field of CBIR. The some of the issues are: the semantic gap in extracting the meaningful image information from image and then transforming into a visual vector in accordance with human perception is the prime most issue. The images belonging to different classes in the database are to be represented by the feature vectors which may be markedly differently from different users.

VI. DATASET

The complete dataset for the results for the hybrid model of content based image retrieval is named as Corel dataset. It has various images in different categories   such that images are classified into multiple classes based on the features selection. The   content based image retrieval performance evaluation is done using various parameters.

VII. RESULTS AND DISCUSSIONS

A Dataset

The complete dataset for the content based recognition is taken from the Kaggle.com. It is the open source

repository from where the open datasets can be downloaded. These datasets are having higher stability for the proposed results generation.

B Comparison of existing and proposed technique

The results on the basis of three parameters are evaluated. The existing technique based on DWT and cuckoo search results are to be compared with the proposed hybrid model.

	DWT and Cookoo search			Hybrid technique		
	Time (s)	Precision (%)	Recall (%)	Time (s)	Precision (%)	Recall (%)
Outdoor	0.38	60.86	56	0.32	67	62
Indoor	0.37	100	92	0.33	100	95
building	0.37	82.6	76	0.31	88.3	82
Nature	0.36	56.62	52	0.28	62.34	56

Table 2 Comparison of results

The whole dataset is classified into four different categories. These four categories are outdoor, indoor, building and nature. The results are evaluated on the basis of various parameters, these parameters are time, precision and recall. The results are evaluated such that the existing and proposed technique are compared with respect to these parameters. The table 1 shows that the proposed hybrid model performance is better with respect to different parameters.
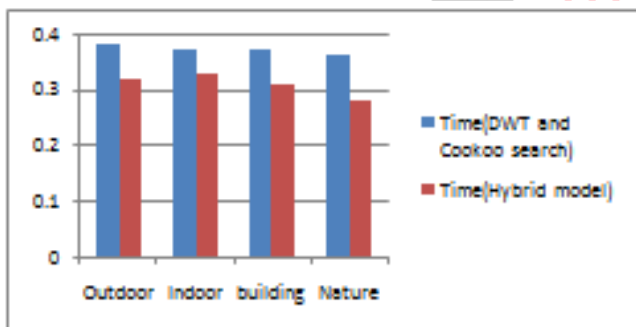
C Comparison of time

Fig. 3 Comparison of time

The graph in fig. 3 shows the comparison of the time for four categories of images using existing and proposed technique. The results in all the four categories are better for the proposed technique.
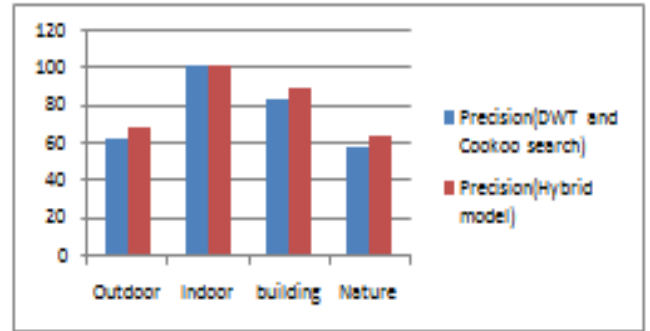
D Comparison of precision

Fig. 4 Comparison of precision

The graph in the figure 10 shows the comparison of the precision of the existing and proposed technique. The results of the proposed technique is better in all the four categories.
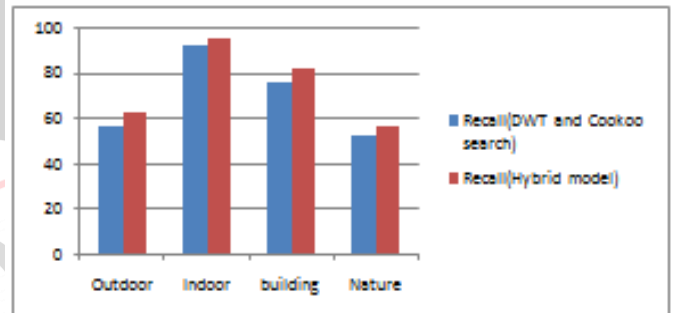
E Comparison of Recall

Fig. 5 comparison of recall

The graph in figure 11 shows the comparison of the recall of the existing and proposed technique. The proposed technique is having recall better compared to the existing technique.

VIII. CONCLUSION

The content based image retrieval is the most suitable way to identify the image from the large repository. It will search the image from the large set of images based on contents rather than the image name. It will be having less time to search the image from the large repository when the image is retrieved using content based. In the current research the hybrid approach for content based image retrieval is performed. This proposed procedure will be in the first step perform the classification of the image into multiple classes. The classes are prepared based on the attributes values. Once the classification is performed the content based image searching will be performed. This searching will requires lower time but provide high accuracy. The proposed approach based on hybrid technique is having better performance in terms of various performance parameters.

IX. FUTURE WORK

The proposed approach is based on hybrid technique, where in the first part there is classification of the images into multiple classes, each class will be having its own set of attributes. The classification based on features will be suitable, but with some level of the error rate. In the future the error rate in the classification can be reduced by involving other improved classification technique. This will further reduces the time for the content based image retrieval.

REFERENCES

[1] Anan Banharnsakun, "Artificial bee colony algorithm for content-based image retrieval", Computational Intelligence,vol. 36,pp: 351–367,2020.

[2] Manpreet Kaur,Sakshi Dhingra," Comparative Analysis of Image Classification Techniques Using Statistical Features in CBIR Systems",IEEE,Vol.3, pp:230-235,2020.

[3] Manar Abdulkareem Al-Abaji," Cuckoo Search Algorithm Based Feature Selection in Image Retrieval System", Journal of Education and Practice ,vol.10, pp:78-88,2019.

[4] M. N. Munjal and S. Bhatia, "A Novel Technique for Effective Image Gallery Search using Content Based Image Retrieval System," 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon), Faridabad, India, 2019, pp. 25-29, doi: 10.1109/COMITCon.2019.8862206.

[5] Zhou Bing and Yang Xin-xin, "A content-based parallel image retrieval system," 2010 International Conference On Computer Design and Applications, Qinhuangdao, 2010, pp. V1-332-V1-336, doi: 10.1109/ICCDA.2010.5540864.

[6] Shao Hong, Cui Wen-cheng and Tang Li, "Medical Image Description in Content-Based Image Retrieval," 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, 2005, pp. 6336-6339, doi: 10.1109/IEMBS.2005.1615946.

[7] Yahiaoui and N. Boujemaa, "Content-based image retrieval in botanical collections for gene expression studies," IEEE International Conference on Image Processing 2005, Genova, 2005, pp. III-1240, doi: 10.1109/ICIP.2005.1530623.

[8] C. Town and D. Sinclair, "Ontological query language for content based image retrieval," Proceedings IEEE Workshop on Content-Based Access of Image and Video Libraries (CBAIVL 2001), Kauai, HI, USA, 2001, pp. 75-80, doi: 10.1109/IVL.2001.990859.

[9] R. Zhang and Z. Zhang, "Effective Image Retrieval Based on Hidden Concept Discovery in Image Database," in IEEE Transactions on Image Processing, vol. 16, no. 2, pp. 562-572, Feb. 2007, doi: 10.1109/TIP.2006.888350.

[10] S. Jang and H. Kwak, "Content-Based Image Retrieval using Shape Information of Central Object," The 9th International Conference on Advanced Communication Technology, Okamoto, Kobe, 2007, pp. 502-505, doi: 10.1109/ICACT.2007.358404.

[11] W. C. Seng and S. H. Mirisaee, "A Content-Based Retrieval System for Blood Cells Images," 2009 International Conference on Future Computer and Communication, Kuala Lumpar, 2009, pp. 412-415, doi: 10.1109/ICFCC.2009.112.

[12] F. Kawanobe, S. Takano and Y. Okada, "Towards Interactive Image Query System for Content-Based Image Retrieval," 2009 Fourth International Workshop on Semantic Media Adaptation and Personalization, San Sebastian, 2009, pp. 56-61, doi: 10.1109/SMAP.2009.22.

[13] S. Deb, "Overview of image segmentation techniques and searching for future directions of research in content-based image retrieval," 2008 First IEEE International Conference on Ubi-Media Computing, Lanzhou, 2008, pp. 184-189, doi: 10.1109/UMEDIA.2008.4570887.