CBIR over HADOOP Map Reduce

Hinge Smita[1] Gaikwad Monika[2] Chincholkar Shraddha[3] Prof. Ajay K. Gupta[4]
[1,2,3,4]Department of Computer Science & Engineering
[1,2,3,4]IOKCOE, Pune-India

Abstract— The development of internet causes the explosive growth of images. It is not possible to handle that amount of images using the conventional method. So there is a need for new method that can handle a huge amount of data and provides the more accurate results to the user. Therefore, we are introducing a new method for retrieving image called as "Content Based Image Retrieval Using Hadoop Map reduce". The main objective of this is to handle a huge amount of data with the principle of parallel processing.

Key words: CBIR, HDFS

I. INTRODUCTION

Before the introduction of CBIR system, there is a method called Text Based Image Retrieval (TBIR) that is used for image retrieval. In this method, images are retrieved on the basis of keywords or tags assigned with the image. The main problem with this method is, we have to assign the keyword to each and every image in the database manually and it is not possible to assign keyword to huge amount of images and it is not even possible to remember all those keywords. And hence the method called content based image retrieval has been introduced. In this method the feature of the image is used as search criteria. These features are automatically obtained features.

Content Based Image Retrieval is the technique for retrieving a specific image from the image database based on the content of the image. Here the content is nothing but the feature associated with that image. There are so many kinds of features like color, texture, shape etc. On the basis of these features, the images are retrieved from the image database according to the users query. And hence the content based image retrieval technique for image retrieval as compared to Text Based Image Retrieval.

In this proposed system we are introducing a novel approach for effective retrieval of images using a MapReduce framework which is generally used for parallel processing and returns the result in shorter time. In the proposed system we are extracting color, shape, texture feature of the image.

II. PROCEDURE

The model proposed in fig1 consists of two phases:

1) Feature extraction phase
2) Similarity matching phase

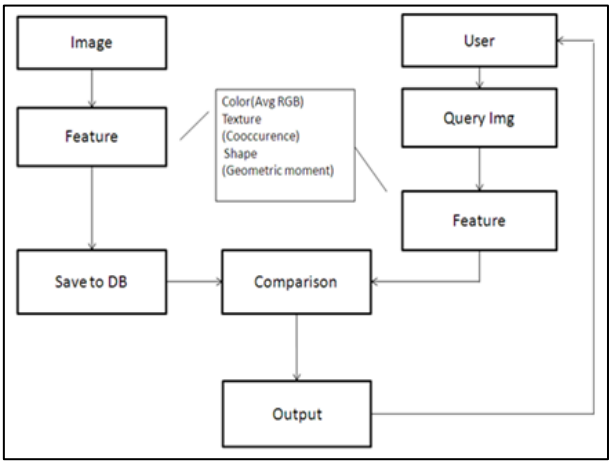

Fig. 1: Architectural Diagram for proposed system

Feature is nothing but the content of image that describes the image. The features are extracted from the image in the form of feature vector. The first phase includes extraction of features from all the images in the database and stores it in the form of feature vector in HDFS. The features extracted in the proposed method are color feature, texture feature and shape feature. These features are extracted parallelly by using hadoop MapReduce. In order to retrieve similar images from the database we must have some similarity measurement technique. The second phase is similarity matching. This phase includes comparison of features of query image with the feature vector in the database. Those images with the less distance are retrieved.

III. MATHEMATICAL MODEL

Sr.no	Description	Observations/Remarks
1	Let S be the System S={S1, S2, S3 }Where, S1- Login module S2- Feature Extraction module S3- Feature Comparison module	S identifies system set
2	S1={Input,Process,Output,Constraint} Input={uid, p} uid=user name P=password Process={uid1,pw1} Uid1=check user name Pw1=check password Output={f,su} F=fail: user is unauthorized su=success: user is authorized Constraint=Username and password must be correct.	Login Module.
3	S2={Input,Process,Output,Constraint} Input={img} img= image	Feature Extraction module

	Process={D1,D2,D3} D1=extract color feature D2= extract texture feature D3=extract shape Output={v} v=feature vector Constraint=Input must an image	
4	S3={Input,Process,Output,Constraint} Input={v,qimg} v=feature vector Qimg=feature vector of query image Process={D1,D2,D3} D1= apply Eucledean Distance algorithm D2= calculate distance D3=comparison Output={Op } Op=retrieved images Constraint=Features of image must be calculated.	Feature Comparison Module.

IV. Result Analysis

Sr.no	Technique	Accuracy	Speed	Efficiency
1	TBIR	30%	20%	40%
2	CBIR	50%	25%	60%
3	Proposed System	80%	60%	80%

Table 1: Existing System and proposed system comparison

A. Graph:

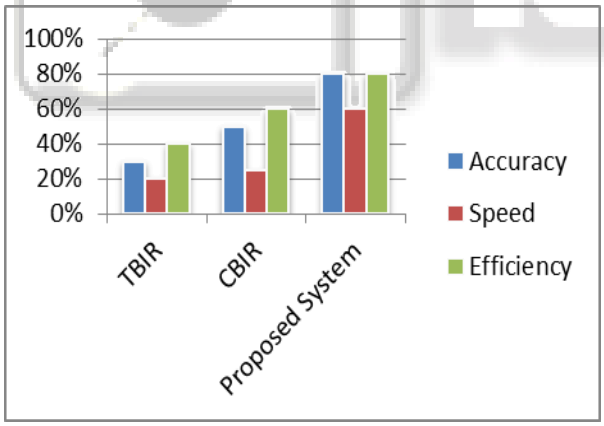

Fig. 1: Graph for comparison of existing system and proposed system

The above graph shows the performance of our system as compared to existing systems. It shows the performance in terms of three parameters i.e. accuracy, speed, efficiency. As we are using three types of features, accuracy is more as compared to existing system. And because of hadoop mapreduce, speed of system is also high.

V. Acknowledgment

We wish to avail this opportunity to acknowledge our profound indebtedness and extend our deep sense of gratitude to our guide Prof. Ajay K. Gupta, Assistant Professor , IOKCOE for his valuable guidance profound advice and encouragement that has feel to the successful completion of this project.

VI. Conclusions

Thousands of images are added to the image database and internet through the various digital devices. So we need an effective approach for storing and handling those images. Therefore we have described a novel approach of Content based Image Retrieval of large dataset by using the hadoop MapReduce for parallel processing.

References

[1] Prof. Deepti Chickmurge "Implementation of CBIR Using MapReduce Over HADOOP"," International Journal of Computer, Information Technology & Bioinformatics (IJCITB)" ISSN: 2278-7593, Volume-2, Issue-2.

[2] Swapnil P. Dravyakar et al"Private Content Based Multimedia Information Retrieval Using Map-Reduce", "International Journal of Computer Science Engineering and Technology (IJCSET) "| April 2014 | Vol 4, Issue 4,125-128,ISSN-2231-0711.

[3] S. Mangijao Singh , K. Hemachandran, "Content Content-Based Image Retrieval using Color Moment and Gabor Texture Feature","IJCSI International Journal of Computer Science Issues", Vol. 9, Issue 5, No 1, September 2012 ISSN (Online): 1694-0814.

[4] Pattnaik ,"A Survey on Text Based Indexing Techniques in Hadoop"," International Journal of Advanced Research in Computer Science and Software Engineering ", Volume 3, Issue 11, November 2013 ,ISSN 2277-128X.

[5] Yong Rui Thomas S Huang and Sharad Mehrotra,"Content Based Image Retrieval with Relevance Feedback in Mars".

[6] Swati V. Sakhare & Vrushali G. Nasre," Design of Feature Extraction in Content Based Image Retrieval (CBIR) using Color and Texture.

[7] Hiremath P.S. and Pujari J., "Content Based Image Retrieval Using Color, Texture and Shape Features," International Conference on Advanced Computing and Communications, ADCOM , pp.780 – 784,2007,

[8] Shankar M. Patil "Content Based Image Retrieval Using Color, Texture and Shape," International Journal of Computer Science & Engineering Technology (IJCSET), Vol. 3, Sept. 2012.

[9] Ryszard S. Choras "Image Feature Extraction Techniques and Their Applications for CBIR and Biometrics Systems" International Journal of Biology And Biomedical Engineering, Vol. 1, 2007

