RESEARCH ARTICLE OPEN ACCESS

Content Based Image Retrieval Using HADOOP Map Reduce

Hinge Smita[1], Gaikwad Monika[2], Chincholkar Shraddha[3]
Department of Computer Science and Engineering[1,2&3]
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.
Keywords:- 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 these 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. PRILIMINARIES

A. *Content Based Image Retrieval*

The Content Based Image Retrieval is the problem of retrieving the image from the image database by analysing the features of the image rather than its metadata. The features can be primitive features such as color, texture, shape etc.

B. *Color*

Color is the most basic feature which is used in image retrieving process. Colors are generally defined in three dimensional color spaces. These could be RGB(Red, Green, and Blue), HSV(Hue, Saturation and Value), HSB(Hue, Saturation, and Brightness). Most of the image formats such as JPEG, BMP, GIF uses the RGB color space to store information. Color information of the image is stored in the form of color histogram.

C. *Texture*

Texture is nothing but the visual patterns associated with the image. Texture is used to distinguish between the areas of image with same color. There are number of texture features such as degree of contrast, directionality and regularity, coarseness, directionality and randomness etc. The technique that we are using for texture analysis is texture color co-occurrence.

D. *Shape*

Shape is one of the low level features of the image which is used to measure the shape of specific object. The natural objects are generally recognized by their shape. There are different types of shape features such as circularity, convexity, Lake Factor, direction, eccentricity, relative size etc.

E. ***Feature Extraction***

Feature extraction is the process of transformation of the input data into set of features. In feature extraction method, the color, texture and shape features of the image are extracted. MATLAB, SciLab, NumPy these are the software available for feature extraction.

F. Similarity Matching

Similarity matching is the process of approximating the solution based on computation of similarity function between a pair of images. Euclidean distance is the method for similarity matching in which the distance between two points are calculated. According to the distance the images are retrieved. The images with the less distance are provided as an output.

G. ***Hadoop***

Hadoop is open source software framework for storage and large scale processing of datasets on clusters. Hadoop has two subparts first one is MapReduce for computational capabilities and second is HDFS for storage. MapReduce is distributed framework for data processing, especially big data. The MapReduce process of Hadoop completes with two phases Map and Reduce.

In Map phase stored split data inputted to map function which will generate intermediate key pair. Wherever reduce phase accept these key value pair as its input which will merge all intermediate values associated with same intermediate key.

H. HDFS

HDFS stands for Hadoop Distributed File System. In HDFS, data is divided into chunks. HDFS is mainly consist of Namenode and datanode. HDFS uses the master/slave architecture in which Namenode is the master of file system and datanode is the slave of file system. Datanode stores the actual data and namenode stores the metadata for the datanode. namenode and datanode both are replicated to handle the failure and to provide the reliability.

I. ***MapReduce***

MapReduce is a programming language for processing and generating large data sets with parallel and distributed algorithms on cluster. Map reduce program is composed of two basic procedures. Map() procedure that performs filtering and sorting and a Reduce() procedure that performs a summary information.

Map Step: The master node takes the input, divide it into smaller sub-problems, and distribute them to worker nodes. A worker node may do this again in turn, leading to multilevel tree structure. The worker node processes the smaller problems and passes the answer back to the master node.

Reduce Step: The master node then collects the answer to all the sub-problems and combines them in some way to form the output.

III. SYSTEM ARCHITECTURE

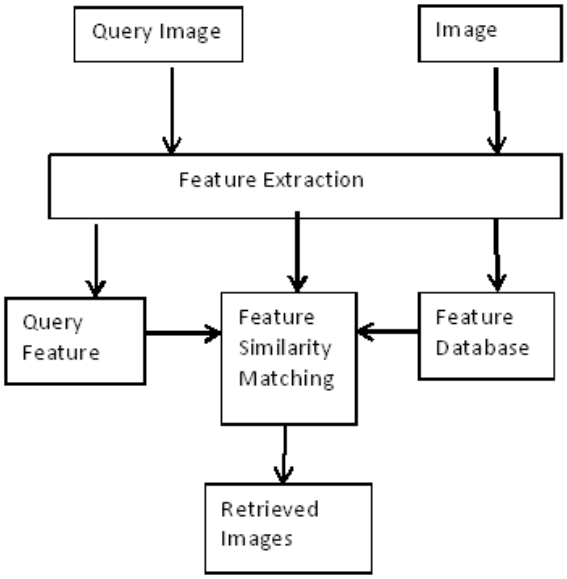

Fig-1: Architecture of proposed system

Fig-1 shows the architecture of proposed system. In this, admin will upload the image. Features of that image are extracted such as color feature, texture feature and shape feature. These features are extracted parallely by using the Hadoop MapReduce framework and are stored in the HDFS in the form of feature vector. Whenever the user provide any query image, the features of the query image is extracted and these features are compared with the feature vector in the database. The images with the less distance are provided as a result to the user.

IV. 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.

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.

REFERENCES

[1] Prof.Deepti Chikmurge"*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] Priyabrat 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