

Review Paper on Content Based Image Retrieval Using Hadoop

Simaran

Department of Information Technology, Poornima College of Engineering, India

Abstract: *The Excessive usage of internet causes the explosive growth of digital database. Massive amount of image data is generated due to the high usage of image capturing devices like camera, photo editing software, etc. With this kind of image database, it poses challenges to retrieve image relevant to user query. CBIR system (Content-Based Image Retrieval) is one of the possible solutions to effectively manage image databases. So, fast access to such a huge database requires an efficient computing model. The Hadoop framework is one of the findings based on MapReduce distributed computing model which is widely used for parallel data processing on terabyte or petabyte scales. Therefore the new method to retrieve images called as “CBIR over Hadoop using Map Reduce.”*

Keywords: *CBIR, MapReduce, HDFS, HBASE, Distributed and Parallel Computing, Hadoop.*

I. INTRODUCTION

Content based image retrieval is a solution to search images from huge databases. Nowadays, social networking sites like Twitter, WhatsApp, Facebook, etc. are creating large number of images in digital databases which is difficult to handle using traditional approach. Content Based Image Retrieval over Hadoop using MapReduce helps to find images from large volumes of data. It provides fast access to such a huge database and efficiency to the users. Hadoop is open Java Framework based on the MapReduce computing model. MapReduce is a programming paradigm that allows for massive scalability across hundreds or thousands of servers in a Hadoop cluster. It also provides an associated implementation for processing and generating big data sets with parallel processing on terabyte or petabyte scales.

In the system, the input is a query image and output is all of the similar images in the database. This system involves three steps: Firstly, the features of each image in the database is extracted. Secondly, the feature vector of the query image is extracted. Thirdly, the feature vector of query image is compared to each image in the database.

In this work, we used MapReduce computing model to extract the features of the images that that will later be Stored HBase which is an open source, distributed, column oriented database which runs on the top of the Hadoop.

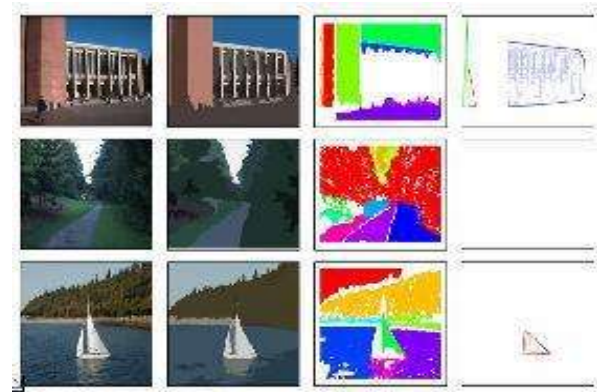


Fig 1. Features Highlighting

II. PREVIOUS WORK

Kannan in 2010 describes that the image mining concept is a better way to extract potential information from the collections. Kun-Che in 2009 suggested pixel wise characteristics extraction and change onto database like tables enables variety of data mining algorithms to make exploration on it.

Earlier, there was a system called text based image retrieval that is also used for image retrieval. In this method, image is retrieved on the basis of tags or keywords assigned to them description assigned to each and every image in the database. It is impossible to assign keyboard or tags to the vast amount of images and it's not even possible to remember all the tags and keywords. Therefore, a new method has been implemented called content based image retrieval which extracts the feature of the image and stores it in the database.

III. PROPOSED WORK

In content based image retrieval, the process of feature extraction is the transformation of the input data into a set of features. It extracts color, texture and shape of an image that can later be used to identify a particular image from the database, after that, similarity matching is done in which it computes the similarity between the query image and images stored in the database.

Similarity between two images is calculated by calculating the distance between two points in an image. This system uses one of the trending Technology Hadoop which is an open-source software framework for storage and large scale processing of data sets on clusters. It uses

one of its tool called MapReduce which has two phases: a map phase and a reduce phase. In map phase, stored split data inputted to map function which will generate key value pairs which will then go through the sorting state which sort these pairs according to the criteria and then they will be sent to reduce phase which accepts these key value pairs as its input and merge all the values with same key.

IV. CONTENT BASED IMAGE RETRIEVAL WORK FLOW

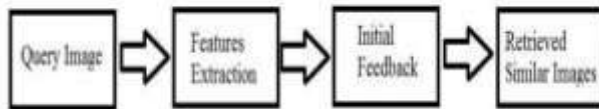


Fig. 2. Steps involved in CBIR Process

In this system, an image will be uploaded by the admin and features of the image are extracted such as color, texture and shape of the image. These features are extracted using MapReduce framework and stored in HDFS in the form of feature vector. When user will enter query image, the features of that image will get extracted and these features are then compared with the images stored in the database. Image with the minimum distance is provided as a result image to user.

Steps followed by admin:

1. Upload all the images to the HDFS.
2. Take one image from HDFS and input it to mapper.
3. Extract the features.
4. Write image and its features in HBASE.
5. Finish all the images in the data base.
6. Collect the output of MapReduce phase.

Steps followed by User:

1. The user queries to the system and image is stored in HDFS.
2. Run a map-reduce job to extract the features and store it into HDFS.
3. The similarity between the query image and the image in database is computed.
4. The Reducer combines all the data from map function and store it into HDFS.
5. Result is sent to the user.

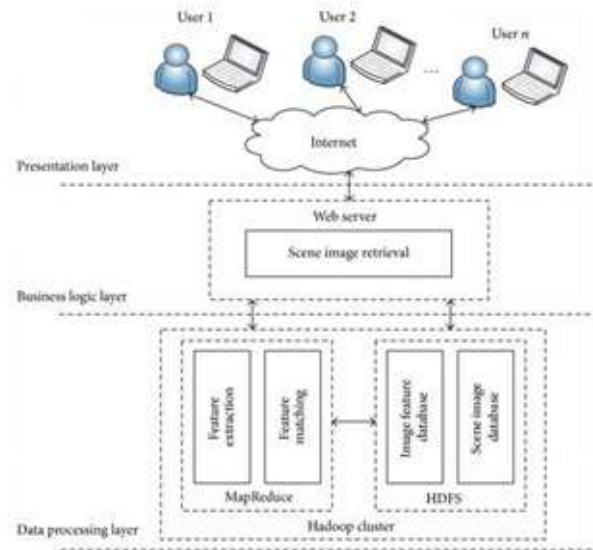


Fig. 3. Working of Map Reduce Model

Fig. 3 shows the working of map reduce model and how Hadoop is being used to retrieve images from database. First is presentation layer which user interaction layer. It is used to input an image or show the result to the user.

Second is Business logic layer in which images and their features are stored into the HDFS.

Third is the data processing layer in which the process of feature extraction and similarity matching is being done and result is sent back to HDFS.

V. CONCLUSION

The abrupt development of internet leads to increase digital database. So, we needed an effective way to handle or analyze the massive data. This paper presents parallel processing of CBIR systems with the help of Hadoop framework to handle large volumes of data. Traditional system of retrieving information has demerits like low efficiency, Poor response time, etc. This system focuses on lowering the transmission speed and time consumption. The development of these systems started with retrieving data based on textual connotations but later advanced to content based retrieval. Since this method reduces processing and response time, it can be used in other image processing areas as well. In this paper, past and current implementation has been reviewed. This work can be further extended to some domain-based applications such as finger print recognition, retina identification, and object detection etc. for large image database.

VI. REFERENCES

- [1] Kusuma B, Megha P Arakeri , Survey on Content Based Image Retrieval Using Mapreduce Over Hadoop, International Journal of Advances in Electronics and Computer Science, Dec. 2015, PP: 05 – 07.
- [2] Dharmendra Pandey, Shivpratap Kushwah, A Review on CBIR with its Advantages and Disadvantages For Low-Level Features, International Journal of Computer Sciences and Engineering, Volume-4, Issue-7, PP: 161-167.
- [3] Nilesh Lohar, Dipankar Chavan, Sanjay Arade, Amol Jadhav, Deepti Chikmurge, Content Based Image Retrieval System over Hadoop Using MapReduce, International Journal of Scientific Research in Science, Engineering and Technology, Volume 2, Issue 1, PP: 123 – 125.