Volume 5, Issue 12, December 2015 ISSN: 2277 128X
International Journal of Advanced Research in Computer Science and Software Engineering
Research Paper
Available online at: www.ijarcsse.com

A Survey on Medical Image Retrieval Based on Hadoop

A. B. Mamulwar, T. Y. Wandile, S. A. Thorat, G. A. Bele
UG Students, Dept. of Information Technology,
JSCOE, Hadapsar, Pune, India

Abstract— with the use of new technologies in modern hospitals, generation of medical image data increasing drastically day by day. These forces us towards more scalable solution for retrieval of large scale medical image data as the traditional single node systems are slow and unable to handle problem of increasing medical image data. In order to address problem of large amount of medical image data in this paper we are proposing a Hadoop based medical image retrieval system which improves the efficiency of retrieval of medical image. This is achieved by using two key concepts of Hadoop i.e. HDFS and MapReduce. HDFS is used for storing existing feature library and MapReduce framework is used to achieve parallelism. Map is used for matching feature of query image with features in feature library, while Reduce is used to collect result of each Map task [2]. Hadoop based medical image retrieval system can reduce the retrieval time.

Keywords— image analysis, Medical Images, MapReduce, HDFS, Hadoop, Big Data

I. INTRODUCTION

With continuous evolvement in technology, sophisticated medical imaging devices are being used in modern hospitals. Such devices produces large amount of medical data in the form of 2D/3D images on the daily basis. The number of images and their dimensionality increased dramatically [5] during the past years. This causes a problem of storage and retrieval of required image from image database.

With the help of modern image processing techniques, it is possible to help doctors to detect and diagnose complicated diseases. Now a day's doctors tend to refer to the previous cases to be able to detect and diagnose diseases more efficiently. This can be achieved by referring medical image data of previous cases. In this situation the problem is how to store and manage such huge medical image data and how to retrieve required images in stipulated time.

The traditional text-based image retrieval system makes use of the key words to retrieve the marked images [3]. But the main limitation with this approach is that we have to give marking to each image manually, this will increase the workload and contents of the image cannot be completely described by using words. Also each person may understand same image differently, so the understanding of images is different from person to person.

The content-based medical image retrieval (CBMIR)[3] has the advantages of high retrieval speed and high precision. CBMIR technology makes use of visual features of images for image retrieval. The CBMIR algorithm calculates the similarity between the features of query image and feature library. When the number of the features in the library is large, the effectiveness of the single-node retrieval [3] in the traditional browser or server mode (B/S) is difficult to meet the real-time requirements of the images, and the system has a very weak stability and extensibility.

In this paper we are proposing an image retrieval system based on MapReduce to reduce the retrieval of medical images from large scale medical image data. Hadoop is a open-source framework which is capable of handling bigdata. Hadoop uses concept of distributed storage in the form of Hadoop Distributed File System (HDFS) and parallel processing with the help of MapReduce framework [2]. Hadoop provides scalability, fault tolerance, high availability and parallelism [6]. With this features Hadoop is best suited to handle the problem of large scale medical image data.

II. PRILIMINARY

A. Medical Image Data

Medical images are special kind of images that reveal internal structure of the human body which is not visible from outside. Medical images are extensively used by clinicians to diagnose and treat diseases like tumour, cancer etc. Medical images are generated by medical imaging techniques such as X-ray imaging, computer tomography (CT), magnetic resonance imaging (MRI) and ultrasound imaging [7].

B. Hadoop and its Components

Hadoop is an open-source framework allowing distributed computing and is used to handle large amount of data. The key components of Hadoop framework are HDFS and MapReduce [6]. To distribute data across multiples of servers, Hadoop framework uses and manages above two components. However, it is capable of processing huge amounts of data as well as able to process wide range of data. One of the main advantages of Hadoop is that it is able to process not only complex but also unstructured data [2].

1) HDFS:-HDFS is a file system designed for storing files of very large size with streaming data access pattern [6]. HDFS provides write once read many type of data storage which is suitable for storing and processing large files. HDFS is designed to run on clusters of commodity hardware and is highly fault tolerant. HDFS follows master/slave architecture. It includes two main actors which are NameNode and DataNode [8]. DataNode at slave machine stores actual blocks of data but it does not store block location. NameNode at master stores the block locations and metadata for all the files and directories on HDFS.
2) MapReduce:-MapReduce is a software framework for parallel data processing. It is a combination of two processes i.e. Map and Reduce. Mapper is used to divide the large job into small tasks and distribute them among slave nodes in cluster. When all the nodes finish their respective task, Reducer is used to aggregate result of each node and gives combined result. Master node consists JobTracker, which receives job and divides it into small tasks. Then it allocates them to TaskTrackers at slave nodes. TaskTrackers at slave nodes perform actual processing and submits the result to JobTracker.

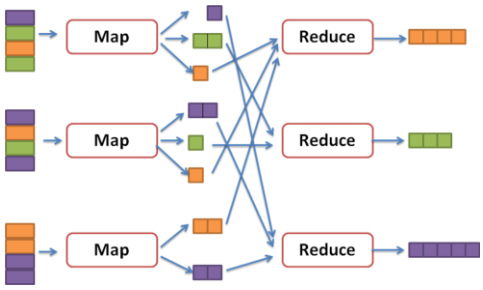

Fig. 1.MapReduce for Color count

III. IMAGE ANALYSIS METHODS

A. Bag of Visual Words[1]

Bag of visual words is a method used for image classification based on local features. In this method image features are treated as "words". Bag of word is a sparse vector of occurrence count of words. Bag of visual word is a vector of occurrence count of local image feature.

Local descriptor is extracted from the set of selected images, then descriptors are clustered into k clusters and centroid of these cluster are used as a visual words.

B. Riesz Miniature[9]

Riesz miniature descriptor represents the image as a single Riesz transform vector, which is nothing but a multidimensional extension of the Hilbert transform. A first step is the down sampling of the image to reduce the dimensionality of the descriptor which uses a linear combination of N–th order Riesz templates at multiple scales. The total weights of the linear combination are extracted from one-versus-all support vector machines. Steerability and multi scale properties of Riesz wavelets allow for scale as well as rotation covariance of the descriptor. By aligning the Riesz templates locally Orientations are normalized and process is carried out analytically. This approach is useful in modelling texture, which shows to outperform state–of–the–art texture attributes in lung classification.

C. Gabor Filter [9]

Gabor filter is a linear filter used for pattern analysis and edge detection. Gabor filters have been known to be used in CBIR for modelling texture. Usually a set of several Gabor filters of different orientations and scales is applied in blocks over the image and histograms of mean filter outputs are used to represent the texture characteristics of the image.

IV. PROPOSED SYSTEM

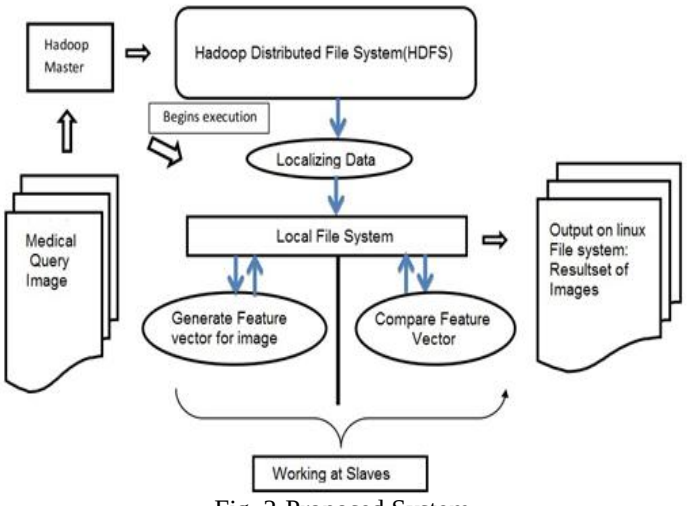

Fig. 2.Proposed System

As shown in proposed architecture diagram user will give an input medical image to the system. This input image is accepted by Hadoop master node and store it in HDFS (Hadoop distributed file system), then Job Tracker in master node will create a MapReduce job for extraction features of this input image. This MapReduce task is given to the task tracker on each node. Task trackers will perform the feature extraction. Resultant features are returned to the master node. After this job tracker will create second MapReduce job to match these features with existing feature vectors in feature library stored on HDFS.

Task trackers perform the feature matching. Resultant matching vectors are returned to the master. Then job tracker creates third and final MapReduce task to retrieve matching images and resultant images are returned to master. Master will return these matched image set to the user. MapReduce jobs are performed by task trackers simultaneously at each slave node, due to this the execution time for this system is expected to be much less as compared to single node system.

V. CONCLUSIONS

In this paper, we have described the overview of Medical image retrieval using MapReduce. The analysis of Big Data, Hadoop, MapReduce, JAVA,HIPI is done and based on these, the application building program has been initiated. A dedicated server operated by master node is implemented to accept query image from the user and three slave nodes to do parallel processing of the image. When user uploads the image to system, Master node will accept image and create a job. After creating the job, the slave will accept the job and will process the query. overall The system will process the query image and will return related images. The system when implemented will be very beneficial for diagnosis at the Hospital.

REFERENCES

[1] Ms.Jyoti S. Patil, ME (Information Technology), Mrs.Sunayana A. Mane (ME (Information Technology) "3-D Image Analysis Using Mapreduce" Pervasive Computing (ICPC), 2015 International Conference on.

[2] Sreedhar C, Dr. D. Kavitha, K. Asha Rani "Big Data and Hadoop" International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) Volume 3 Issue 5, May 2014.

[3] YAO Qing-An 1, ZHENG Hong 1, XU Zhong-Yu 1, WU Qiong 2, LI Zi-Wei 2, and Yun Lifen 3 " Massive Medical Images Retrieval System Based on Hadoop" JOURNAL OF MULTIMEDIA, VOL. 9, NO. 2, FEBRUARY 2014

[4] Chris Sweeney Liu Liu Sean Arietta Jason Lawrence University of Virginia "HIPI: A Hadoop Image Processing Interface for Image-based MapReduce Tasks".

[5] Dimitrios Markonis, Roger Schaer, Ivan Eggel, Henning M¨uller, Adrien Depeursinge University of Applied Sciences Western Switzerland (HES–SO), Business Information Systems, Sierre, Switzerland "Using MapReduce for Large–scale Medical Image Analysis".

[6] Tom white "Hadoop the definitive guide".

[7] www.medicalradiation.com/types-of-medical-imaging.

[8] Prof.Deepti Chikmurge "Implementation of CBIR Using MapReduce Over HADOOP" IEEE Sponsored International Conference On Empowering Emerging Trends In Computer, Information Technology & Bioinformatics International Journal of Computer, Information Technology & Bioinformatics (IJCITB) ISSN:2278-7593,Volume-2, Issue-2.

[9] Dimitrios Markonis , Reńe Donner, Ljiljana Dolamic ,Roger Schaer, Georg Langs, Ćelia Boyer , Henning Muller " Report on and Prototype of final Image Retrieval and Analysis Framework of KHRESMOI"Feb 2014.

