LARGE SCALE SATELLITE IMAGE PROCESSING USING HADOOP FRAMEWORK

Bhagyashri Kamble[1], Namrata Harishchandre[2], Vrushali Kunjir[3]

[1] *Student,Department of Computer Engineering, Sinhgad Academy of Engineering, Maharashtra,India*
[2] *Student,Department of Computer Engineering, Sinhgad Academy of Engineering, Maharashtra,India*
[3] *Student,Department of Computer Engineering, Sinhgad Academy of Engineering, Maharashtra,India*

ABSTRACT

The processing of large scale of images is needed when there are satellite images involved. Now a day's amount of data continues to increase as more information becomes available. And it will increase amount of surface and recognition, segmentation, and event detection in satellite images with a highly scalable system becomes more and more desirable. In this paper, a construction of semantic taxonomy for the land-cover classification of satellite images. Overall system is constructed in a Distributed HADOOP Computing platform. This system is divided into two important part Training and Running classifier. The Training part classifies images one after another such as Vegetation, Building, Pavement, water, snow etc. Large files are distributed and further divided into multiple data nodes. The map processing jobs placed on all nodes are operated on their local copies of the data. It can be observed that the name node stores only the metadata and the log information during the data transfer to and from the HDFS is done through the Hadoop API. Training classifier is constructed using HADOOP MapReduce Framework and is based upon Google Earth. The Running classifier performs zoom-in, zoom-out and calculates the difference between old and new images.

Keyword : - *Hadoop, Image Processing, HDFS, Mapreduce, Remote sensing*

1. INTRODUCTION

In today's world amount of data continues to grow as more information becomes available. With this increasing amount of surface and identification, segmentation, and event detection in satellite images with a highly scalable system becomes more and more desirable. The processing of large amount of images is necessary when there are satellite images involved. Whole system is constructed in a HADOOP Framework. Hadoop is a large-scale distributed batch processing infrastructure. While it can be used on a single machine, its true power lies in its ability to scale to hundreds or thousands of computers, each with many processor cores. Hadoop is also designed to efficiently distribute large amounts of work across a set of machines. The technology of object-oriented classification is now an inevitable trend for the development of High Resolution Remote Sensing Image. Traditional KNN method and clipping- KNN does have their flaws. We are using different techniques for segmentation of image. Hence, we are proposing a system, which will do classification of high resolution remote sensing image using Segmentation of image with HSV and RGB color models. The image will be classified in four sub parts that is Road Land, Water Land, Bare Land and Green Land. Classification and Storage of the Large Image is being done using Hadoop.

2.LITERATURE REVIEW

There Exists some techniques for Satellite Image Classification but in those systems processing of images goes through ordinary sequential ways to accomplish classification. The program loads one after another image and processing each image alone before writing the newly processed Image on a storage device. Therefore, we are in need of a new parallel approach to work effectively on massed image data. In a Hadoop cluster, data is distributed to all the nodes of the cluster as it is being loaded in. The HDFS will split large data files into chunks which are managed by different nodes in the cluster. Data is conceptually record-oriented in the Hadoop programming framework. Individual input files are broken into lines or into other formats specific to the application logic. Every process running on a node in the cluster then processes a subset of these records.

2.1 Kocakulak and Temizel[2]

They have used Hadoop and MapReduce to perform ballistic image analysis which requires large database of images to be combined with an unknown image. This method used correlation technique that imposed a high computational demand. It was shown that processing time was reduced drastically when 14 computational nodes were employed in cluster configuration.

2.2 Li et. Al[3]

He attempted to reduce the computational time taken for executing parallel clustering algorithm on large number of satellite images using Hadoop and MapReduce methodology. The process starts by clustering each pixel with its nearest cluster and then calculates all new cluster centers on basis of every pixel in one cluster set. He proposed another clustering algorithm for processing remote sensing image that uses a parallel K-means approach. In their approach, objects with similar spectral values are clustered together without any formal knowledge.

2.3 Mohamed H. Almeer[1]

He used 112-core highperformance cloud computing system for analyzing Hadoop performance on remote-sensing data. In this analysis the authors show six fold speedup for auto-contrast and eightfold speedup for sharpening algorithm.

2.4 Golpayegani and Halem[4]

They used a Hadoop MapReduce framework to operate on a large number of Aqua satellite images collected by the AIRS instruments. A gridding of satellite data is performed using this parallel approach.

2.5 Banaei, S.M. and Moghaddam, H.K.[5]

They gives a brief idea about the MapReduce which is a distributed data processing model which makes use of open source Hadoop framework for manipulating big volume of data. A very large amount of data in the modern world, espeicially multimedia data, implements new requirements for processing and storage as well. Hadoop allows the processing images on an no of sets of computing nodes with the help of necessary infrastructures.

2.6 Penglin Zhang, Zhiyong Lv, and Wenzhong Shi.[6]

They introduces a structural feature called object correlative index (OCI) which can be used for enhancing the classification of high resolution images. It uses spectral similarity to construct a useful OCI for describing the structural information objectively

2.7 Sarade Shrikant D., Ghule Nilkanth B., Disale Swapnil P., Sasane Sandip R.[7]

This paper mentioned emphasis on using HDFS for retrieval of satellite images also setting up HDFS. It also has few contents about Image retrieval but no sufficient information about Image Processing techniques that can help in Image Segmentation and Feature Extraction.

3. EXISTING SYSTEM

Current processing of images goes through normal sequential ways to complete this job. The program loads image after image, processing each image alone before writing the newly processed image on a storage device. Generally, we use very ordinary tools that can be available in Photoshop. Besides, many ordinary C and Java programs can be downloaded from the Internet or easily developed to perform such image processing tasks. Most of these tools run on a single computer system with a Windows operating system. Although batch processing can be found in these single-processor programs, there will be problems with the processing due to limited efficiency. Therefore, we are in need of a new parallel approach to work effectively on massed image data.

4. PROPOSED SYSTEM

Satellite images play a vital role in today's world for detecting real time events. These events may differ from changing landforms, decreasing glaciers to catastrophic events like earthquakes, tsunamis and sand storms. The numerous amount of changes after these kind of events required to be monitored and capturing satellite images for such event detection can be useful. The idea of this project is to detect the changing landforms in between different vegetation's, store this data, classify it on the basis of certain specified parameters and gain the classified data using properly defined mechanisms.

In a Hadoop cluster, data is distributed to all the nodes of the cluster as it is being loaded in. The HDFS will split large data files into chunks which are managed by different nodes in the cluster. In addition to this each chunk is copied across several machines, so that a single machine failure does not result in any data being unavailable. An active monitoring system then re-replicates the data in return to system failures which can result in partial storage. Even though the file chunks are replicated and distributed across several machines, they form a single namespace, because of that their contents are universally accessible.

4.1 System Architecture

Main Perspective of the system is to make Satellite images classification as:

1. Road land
2. Water land
3. Bare land
4. Green land

Also, how much area it covers must be shown using graph.Application is being designed in such a way that we can give only one image as an input to the cluster, In hadoop we can give multiple image as an input while storing it into the database. Hadoop uses OpenCV library to store images, and for classification we are goinig to use RGB and HSV color models. The advantages of this approach are to store and access images in parallel on a very large scale, and to perform image filtering and other processing effectively.

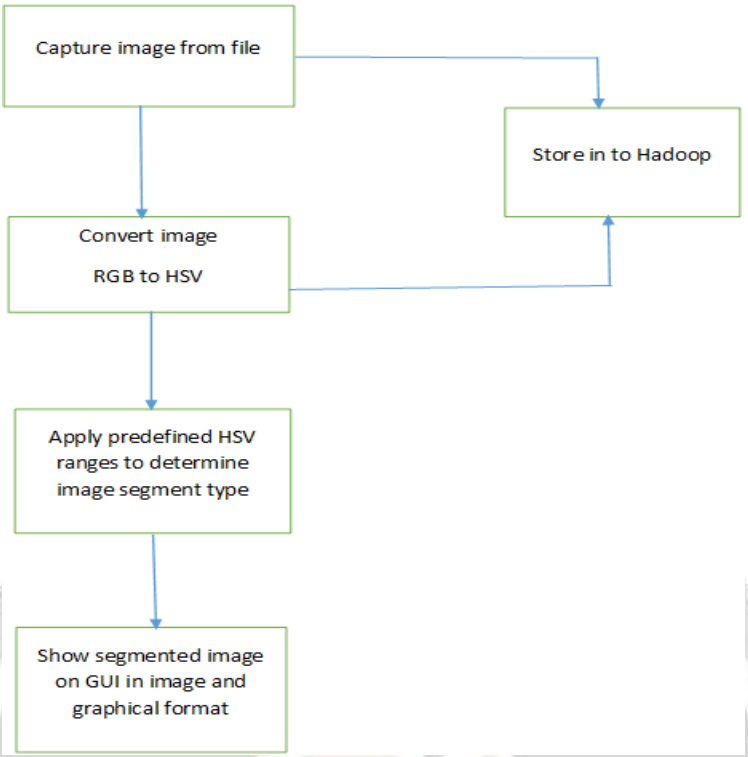

Fig -1: System Architecture

4.2 Algorithm

Step 1: Login with the normal user registered by the application

Step 2: Add Images that has been retrieved from the satellite and store all the images into Hadoop using Hadoop Open CV Library.

Step 2.1: Get U (Images of the data where various types of data like bare land and unused land) as Input to WC.

Step 2.2 : for i=0 to MAX
//MAX = Maximum no of images

Step2.3 : visit i (system application) when logged in by normal user .

Step 2.4 : go to step 2.2 till MAX

Step 2.5 : Output as CP is able to view all images that has been given input to the system.

Step 3: call RC function.

Step 4: call to IE Function

Step 4.1 : Get CP (image) as Input .

Step5: Display Result in the graphical form of the images.

Step 6: Stop.

5. CONCLUSIONS

In this Proposed system, we presented Classification of satellite images in different format by using the Hadoop framework. A large number of images cannot be processed efficiently in the customary sequential manner. We have observed that this Hadoop implementation is better suited for large data classification and storage.

6. ACKNOWLEDGEMENT

We want to thanks to our guide Prof. S.N.Shelake for his important guidance and encouragement in every situation. Suggestions given by him broaden our vision and guided us to succeed in this work. we are also very grateful for his guidance and comments while designing part of our research paper and learnt many things under his leadership.

7. REFERENCES

[1]. Mohamed H. Almeer, "Cloud Hadoop Map Reduce For Remote Sensing Image Analysis" Journal of Emerging Trends in Computing and Information Sciences,Vol. 3, No. 4, April 2012

[2]. H Kocakulak and T T Tamizel, "A Hadoop solution for ballistic image analysis and recognition" International Conference on High Performance Computing and Simulation (HPCS), Istanbul, pp. 836-842, 2011.

[3]. B. Li H Zhao, Z H Lu, "Parallel ISODATA clustering of remote sensing images based on MapReduce", International Conference Cyber enabled distributed computing and Knowledge Discovery (CyberC), Huangshan, pp. 380-383, 2010.

[4]. N. Golpayegani and M. Halem, "Cloud computing for satellite data processing on high end compute clusters," in Proc. 2009 IEEE Int. Conf. Cloud Computing (CLOUD '09), Bangalore, pp. 88–92.

[5] Banaei, S.M. and Moghaddam, H.K. (2014) "Hadoop and Its Role in Modern Image Processing". Open Journal of Marine Science, 4, 239-245. http://dx.doi.org/10.4236/ojms.2014.44022

[6] Penglin Zhang, Zhiyong Lv, and Wenzhong Shi., "Object-Based Spatial Feature for Classification of Very High Resolution Remote Sensing Images", IEEE GEOSCIENCE AND REMOTE SENSING LETTERS, ISSN : 1545-598X, Volume-10, Issue-6, March 2013.

[7] Sarade Shrikant D., Ghule Nilkanth B., Disale Swapnil P., Sasane Sandip R., "LARGE SCALE SATELLITE IMAGE PROCESSING USING HADOOP DISTRIBUTED SYSTEM", IJARCET, ISSN : 2278 – 1323, Volume-3, Issue-3, March 2014.

[8] Vidya Jadhav, Harshada Nazirkar, Sneha Idekar, Prof. P.A. Bandgar, "Recognition of Satellite Images of Large Scale Data Based on Map- Reduce Framework", IJERA, ISSN : 2248–9622, Volume-4, Issue-3, March 2014.

[9] Hadoop Fundamentals &Installation: www.tutorialspoint.com/hadoop/hadoop_tutorial.pdf

[10] J. Dean and S. Ghemawat, "MapReduce: simplified data processing on large clusters," Commun. ACM, vol. 51, no. 1, pp. 107–114, 2008