Image based Remote Sensing Analysis using Hadoop Platform

Doddamani Basavaraj[1], Shivabasayya Kulkarni[2], and Siddalingeshwar Patil[3]
[1,2] *Senior Scale Lecturer, Dept of CSE, Government Polytechnic for Women, Hubli*
[3] *Lecturer, Dept of CSE, Government Polytechnic for Women, Hubli*

***Abstract*—Image processing algorithms related to multiple large images have been tested and utilized on the Hadoop MapReduce in parallel. Although there has been considerable research utilizing the Hadoop platform for image processing rather than for its original purpose of text processing, it had never been proved that Hadoop can be successfully utilized for high-volume image files. Hence, the successful utilization of Hadoop for image processing has been researched using three different practical image processing algorithms. We extend the file approach in Hadoop to regard the whole JPG image file as a unit by expanding the file format that Hadoop uses. Experiments have shown that the method is scalable and efficient in processing multiple large images, and the difference between the non hadoop runtime and the Hadoop runtime is clearly noticeable.**

***Index Terms*—Hadoop, MapReduce, Remote Sensing, Image Processing.**

I. INTRODUCTION

The challenge of processing large number of images in parallel and to derive customized products, several efforts have been made in the past few years towards incorporation of high-performance computing models. This study analyzes the recent advancements in distributed computing technologies as embodied in the MapReduce programming model and extends that for image processing of collected images.

We make two contributions, which are the alleviation of the scarcity of parallel algorithms for processing large numbers of collected images using the parallel Hadoop MapReduce framework and the implementation for large images. Our research has been conducted to find an efficient programming method for customized processing within the Hadoop MapReduce framework and to determine how this can be implemented.

Performance tests for processing large archives of collected images were performed with Hadoop. The following three image analysis algorithms were used: Sobel filtering, auto contrasting, image sharpening. The findings demonstrate that MapReduce has a potential for processing large-scale images.Current approaches to processing images depend on processing a small number of images having a sequential processing nature. These processing loads can almost fit on a single computer equipped with a relatively small memory. Still, we can observe that more disk space is needed to store the large-scale images. The large-scale JPG images taken and are used in this image processing approach. The developed approach can also be used for other formats.

II. PROBLEM STATEMENT

The project takes the images in JPG format and applies color sharpening and contract enhancement on the input image to get the better quality image.We will compare the performance in terms of number of images processed against the serial processing and Hadoop parallel processing and draw performance graphs.

In order to process a large number of images effectively, we use the Hadoop HDFS to store a large amount of images, and we use MapReduce to process these in parallel. The main advantage of this project is distributed processing of large images using Map reduce model, due to this the efficiency of system increases to process more images.

EXISTING SYSTEM

The current processing of images goes through ordinary sequential ways to accomplish 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 found in Photoshop, for example. 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 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

capabilities. Therefore, we are in need of a new parallel approach to work effectively on massed image data.

PROPOSED SYSTEM

In order to process a large number of images effectively, we use the Hadoop HDFS to store a large amount of images, and we use MapReduce to process these in parallel. The advantages of this approach are three abilities: 1) to store and access images in parallel on a very large scale, 2) to perform image filtering and other processing effectively. The main advantage of this project is distributed processing of large images using Map reduce model, due to this the efficiency of system increases to process more images.

III. LITERATURE SURVEY

Nowadays, with the wide support for spatial data sharing, more and more images are becoming publicly available where users can download them freely. Analyzing of these images requires high speed network connection for downloading them and powerful storage and computing resources for local processing. Therefore it might be more efficient to process images remotely, preferably on a high-performance computer server that are close to the data servers.

HADOOP - HDFS – MAPREDUCE

Hadoop is an Apache software project that includes challenging subprojects such as the MapReduce implementation and the Hadoop distributed file system (HDFS) that is similar to the main Google file system implementation. In our study and in others as well, the MapReduce programming model will be actively working with this distributed file system. HDFS is characterized as a highly fault-tolerant distributed file system that can store a large number of very large files on cluster nodes. MapReduce is built on top of the HDFS file system but is independent. HDFS is an open source implementation of the Google file system (GFS). Although it appears as an ordinary file system, its storage is actually distributed among different data nodes in different clusters.

MapReduce is a parallel programming model for processing large amounts of metadata on cluster computers with unreliable and weak communication links. MapReduce is based on the scale-out principle, which involves clustering a large number of desktop computers. The main point of using MapReduce is to move computations to data nodes, rather than bring data to computation nodes, and thus fully utilize the advantage of data locality. The code that divides work, exerts control, and merges output in MapReduce is entirely hidden from the application user inside the framework.

HADOOP APPROACH TO IMAGE PROCESSING

In order to process a large number of images effectively, we use the Hadoop HDFS to store a large amount of stored images and we use MapReduce to process these in parallel. The advantages of this approach are three abilities: 1) to store and access images in parallel on a very large scale, 2) to perform image filtering and other processing effectively, and 3) to customize MapReduce to support image formats like JPG. The file is visited pixel by pixel but accessed whole as one record. The main attraction of this project will be the distributed processing of large images by using a MapReduce model to solve the problem in parallel.

Image processing basically includes the following three steps.

1. Importing the image with optical scanner or by digital photography
2. Analyzing and manipulating the image which includes data compression and image enhancement.
3. Output is the last stage in which result can be altered image or report that is based on image analysis.

Purpose of Image processing: The purpose of image processing is divided into 5 groups. They are:

1. Visualization - Observe the objects that are not visible.
2. Image sharpening and restoration - To create a better image.
3. Image retrieval - Seek for the image of interest.
4. Measurement of pattern – Measures various objects in an image.
5. Image Recognition – Distinguish the objects in an image.

HADOOP APPROACH

In order to process a large number of images effectively, we use the Hadoop HDFS to store a large amount of images data, and we use MapReduce to process these in parallel. The advantages of this approach are three abilities: 1) to store and access images in parallel on a very large scale, 2) to perform image filtering and other processing effectively.The file is visited pixel by pixel but accessed whole as one

record. The main attraction of this project will be the distributed processing of large images by using a MapReduce model to solve the problem in parallel.

IV. IMPLEMENTATION

The following Modules are:

1. Hadoop Approach Color Sharpening: Image sharpening is used to enhance both the intensity and the edge of the image in order to obtain the perceived image. In order to process a large number of images effectively, we use the Hadoop HDFS to store a large amount of stored images data, and we use MapReduce to process these in parallel.
2. Hadoop Approach Sobel Filter: A large number of images cannot be processed efficiently in the customary sequential manner. Although originally designed for text processing, Hadoop MapReduce proved suitable to process JPG format images in large quantities. Sobel edge filter is used to detect edges and is based on applying horizontal and vertical filters in sequence.
3. Hadoop Approach Auto-Contrast: Hadoop MapReduce installed in a parallel cluster proved suitable to process JPG format images in large quantities companies that own massive images and need to process those in various ways, our implementation proved efficient and successful. Auto-contrast is a point operation whose task is to modify the pixel intensity with respect to the range of values in an image.

V. SNAPSHOTS AND GRAPHS

The project takes the input images in jpg format

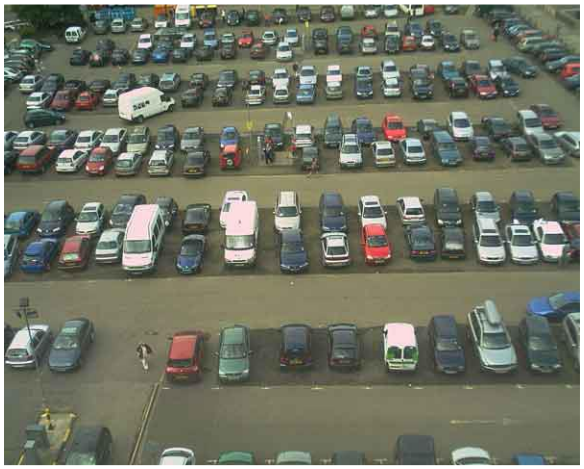


Image 1

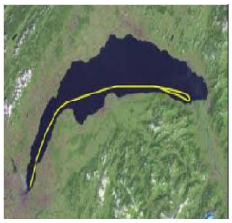


Image 2

Image 3

Image 4

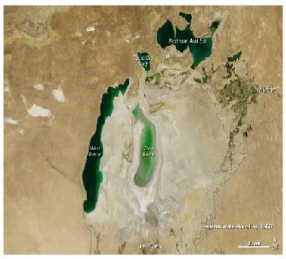


Image 5

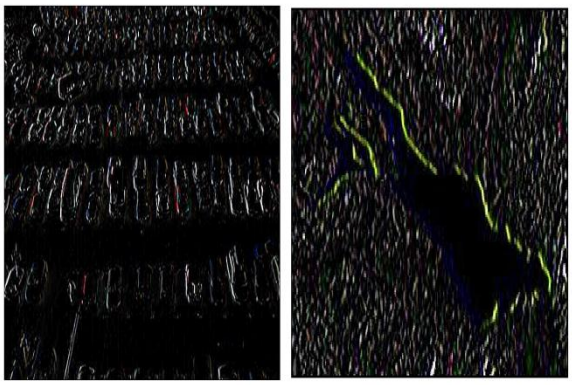
Hadoop Approach to Image Processing: In order to process a large number of images effectively, we use the Hadoop HDFS to store a large amount of stored imagesdata, and we use MapReduce to process these in parallel.

Hadoop Approach Color Sharpening: Image sharpening is used to enhance both the intensity and the edge of the image in order to obtain the perceived image.

Figure: Output for color sharpening

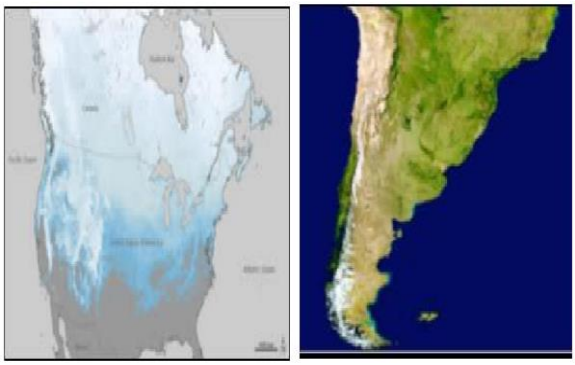
Hadoop Approach Sobel Filter: Sobel edge filter is used to detect edges and is based on applying horizontal and vertical filters in sequence.

Figure: Output for Sobel filtering



Hadoop Approach Auto-Contrast: Auto-contrast is a point operation whose task is to modify the pixel intensity with respect to the range of values in an image.

Figure: Output for Auto-Contrast



OUTPUT GRAPHS

Figure: Output graph for color sharpening

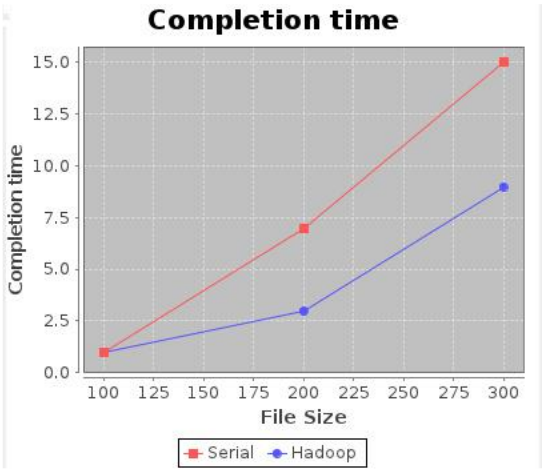


Figure: Output graph for Sobel filtering

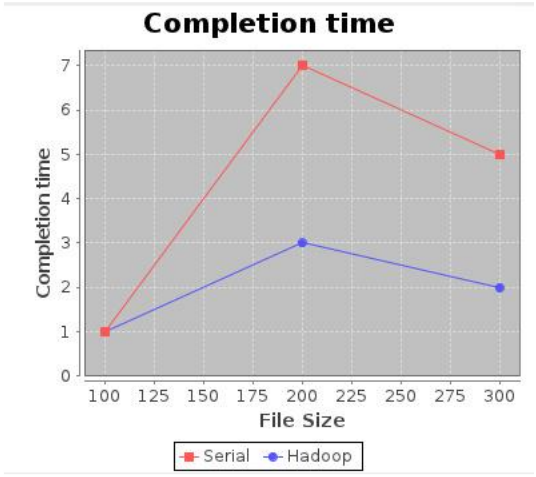
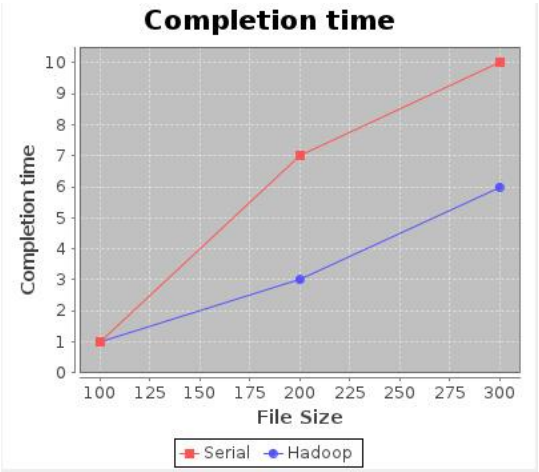


Figure: Output graph for Auto-Contrast



VI. CONCLUSION

In this paper, we presented a case study for implementing parallel processing of large stored images in jpg format by using the Hadoop MapReduce framework. The experimental results have shown that the typical image processing algorithms can be effectively parallelized with acceptable run times when applied to large images. A large number of images cannot be processed efficiently in the customary sequential manner. Although originally designed for text processing,Hadoop MapReduce installed in a parallel proved suitable to process jpg format images in large quantities. For companies that own massive images and need to process those in various ways, our implementation is efficient and successful.

VII. FUTURE ENHANCEMENT

We have observed that this parallel Hadoop implementation is better suited for large data sizes than for when a computationally intensive application is required. In the future, we might focus

on using different image sources with different algorithms that can have a computationally intensive nature.We can customize MapReduce to support image formats like TIFF.

The large-scale TIFF images taken from satellites can also be used in this image processing approach.Other image formats such as BMP, GIF, and JPEG can also be handled by the Hadoop Java programming model developed.

Finally, we can apply this to other image formats such as the JPEG, BMP, and GIF formats.This method is scalable and efficient in processing multiple large images used mostly for remote sensing applications.

REFERENCES

[1] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters", Commun. ACM, vol. 51, no. 1, pp. 107-113, Jan. 2008

[2] Y. Li, D. Crandall and D. Huttenlocher, "Landmark classification in large-scale image collections", Computer Vision 2009 IEEE 12th International Conference on, pp. 1957-1964, Sept 2009.

[3] 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.

[4] 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

[5] T. Blaschke, Object based image analysis for remote sensing ISPRS Journal of Photogrammetry and Remote Sensing, vol. 65, pp. 216, 2010.

[6] Li Bo et al., "Parallel ISODATA clustering of remote sensing images based on MapReduce", Proc. CyberC 2010 Huangshan, pp. 380-383, 2010.

[7] C. Lee, S. D. Gasster et al., "Recent developments in high performance computing for remote sensing: A review", Selected Topics in Applied Earth Observations and Remote Sensing IEEE Journal of, vol. 4, no. 3, pp. 508-527, 2011.

[8] He B, Fang W, Luo Q, et al. Mars: a MapReduce framework on graphics processors[C]//Proceedings of the 17th international conference on Parallel architectures and compilation techniques. ACM, 2008: 260-269.

[9] W. Lai et al., "towards a framework for large-scale multimedia data storage and processing on Hadoop platform", The Journal of Supercomputing, pp. 1-20, 2013.