

An Efficient Framework for Image Analysis using Mapreduce

S Vidya Sagar Appaji¹, P.V.Lakshmi² and P.Srinivasa Rao³

¹ CSE Department, MVGR College of Engineering, Vizianagaram

² IT Department, GITAM, Visakhapatnam

³ CSE Department, MVGR College of Engineering, Vizianagaram

Abstract—In this present modern era the general image collections cannot be handled efficiently on one device due to the fact that the image size being large, high computational costs and performance overhead of modern image processing algorithms. Therefore, image processing often requires disseminated (distributed) computing. But practically, distributed computing is a very complicated subject that demands strong technical knowledge in that area and often cannot be used by researchers who develop this image processing algorithms. A framework is needed that allows the researchers to concentrate on the image processing tasks and hides them from the complicated details of distributed computing. Additionally, the framework should provide the researchers with the familiar image processing tools. The paper describes the usage of MapReduce paradigm framework that provides the ability to divide the image into clusters for distributed processing. The basis allows the improvement of image processing by using traditional methods. The purpose for selecting Hadoop is the technology and its timely delivery and cost-effectiveness, which is now a key constituent for success in many government endeavors, businesses, scientific and engineering disciplines. The Hadoop software stack that was used consists of an extensible MapReduce execution engine, along with pluggable distributed storage engines, and a range of procedures for declarative interfaces—is a popular choice in many industries. Tests have shown that this method is ascendable and efficient in handling multiple large images used mostly for remote sensing applications, and the variance between the single PC runtime and the Hadoop runtime is clearly appreciable.

Keywords: Hadoop, MapReduce, Image Processing Tools, Framework, Distributed Computing.

1. INTRODUCTION

Image Analysis is a vast area that needs thorough analysis and it's also very difficult to process large image data sets. In the paper we have focused on the image stitching algorithm on some basic images. Conventional image processing algorithms often require more time and effort and in the paper we have proposed a novel way of processing images and have also seen how the platform will outperform the time as the image size increases with increase in the number of nodes. Image processing is playing a very important tasks in many areas of research such as remote sensing [1], medical imaging [2], Internet analysis, astronomy, etc. The industry also very widely uses image processing. Presently, image processing often requires distributed computing. Generally the sizes of modern image collections, are large (terabytes and

petabytes of data); such huge collections of data cannot be stored and processed efficiently on a single machine. In addition, current image processing algorithms are becoming very challenging and, hence, computationally intensive.

There are many challenges in processing large and complex image datasets to derive modified products, and several efforts are being made in the recent years towards incorporating of high-performance computing models. The following study analyzes the recent developments in distributed computing technologies which is embodied in the programming model of MapReduce and extends that for image processing for complex images.

Generally in the existing system the processing of images goes through certain ordinary sequential steps. The program will be loading images one after the other, processes each image alone before writing the updated processed image on to a storage device. In general, we use very ordinary and simple tools that can be found in Photoshop. Also, many ordinary C and Java programs are available or can easily be developed to perform some basic image processing tasks. Majority of these tools will run on a single computer along with Windows operating system. Although we can process the images batch wise in these single-processor systems, there will be problems with such processing due to limited capabilities. Hence forth, we are in need of a new approach that can process parallel and also to work effectively on massed image data.

As per the literature the importance of HADOOP framework [3] and advantages of this technology to process large images have been identified in the paper from which we have generically framed some steps that actually match with the original conventional programming models.

1. Loading of Images
2. HADOOP Distributed File System [4]
3. MapReduce [5] Programming
4. Output Image

Loading of Image: As we have worked on a Single Image, we have considered that image in the hadoop distributed file system database.

HADOOP Distributed File System: To process a large set of image efficiently this set of images are to be fed to HADOOP distributed file system. And also it will be necessary to divide the higher resolution images into multiple smaller segments and assign each image segment to various slave machines to efficiently compare the images. This can be normally done in distributed environment.

MapReduce Programming: MapReduce is the powerful programming model that can be used in Hadoop framework. MapReduce programming is very efficient because it will divide the work among slave nodes and will perform the work in parallel.

Output Image: The output images will save on the HDFS and we can copy those to our local file system and compare the results.

Image stitching is the algorithm developed in the paper. Image stitching is having applications like high quality image mosaics for satellite imaging, video stitching, object insertion, medical imaging. The stitching algorithm consists of following steps:

Algorithm 1:

- I. Detect and Describe Interest point features
- II. Associate features together
- III. Robust fitting to find transform
- IV. Render the combined image

The above steps are the normal stitching algorithm steps. We can use different techniques at each step to find speed and accurate result. We are integrating these steps with Hadoop processing steps to get the results even faster.

Detect and Describe Interest point features: In this step the image features can be found from all the images. Image features can be extracted using different methods like edges, corners, blobs/regions, ridges. In the paper blob/regions detection is used. We use Fast Hessian detector region detector to detect the features of images. This will locate the random points in images and extract the features of the points selected. These features are calculated depends on the Hessian matrix.

Associate features together: The features in all the images are compared together and find the similarities in all the images. In this step with the original points the duplicate points may be recognized those are having similar features.

Robust fitting to find transform: In this step the fundamental matrix computation will use. The fundamental matrix will compute the necessary points that are matched between pair of images the duplicate points will be removed in this step.

Render the combined image: This is the last step in image Stitching. The images that are having the matched features will be overlapped on each other and the final result will be the Stitched panorama image.

Image stitching algorithm steps are common for normal stitching without Hadoop but Hadoop steps also included for reducing time and processing the job in parallel. The steps of image stitching is included in MapReduce programming and then compare the results. The image stitching with hadoop flow will work as in Algorithm 2.

Algorithm 2:

- Load images into HDFS.
- Create input and output paths.
- Read images in Buffered image format as a single file.
- To read image as single file set the input format to WholeFileInputFormat class.
- Mapper will read image storage bytes as ByteArrayInputStream.
- WholeFileRecordReader is the class written for reading the image.
- Get the image height and width.
- In the Map() perform image stitching steps from I to IV in Algorithm 1.
- Stitching algorithm will combine two images at each DATANODE at a time.
- The Map() output will be submitted to the Reduce().
- Reduce() will perform the same function of Map() with the output of Map() images.
- Set the height and width for the output image.
- Save the final image in jpeg format in HDFS.

The rest of the paper is structured as follows. In section 2 Related work about hadoop and stitching, section 3 Environmental setup, In section 4 Results and finally in section 5 conclusion and future work are explained.

2. RELATED WORK

Matthew Brown and David G. Lowe [6] did the panoramic image stitching and discussed about 1D and 2D stitching and related problems and they discussed multi-image matching problem and invariant local features are used to detect features. Cheng-Ming Huang, Shu-Wei Lin, and Jyun-Hong Chen [7] used stitching the sequence of images with recurring patterns and implementing the same Image stitching operation and they performed Image stitching and object insertion in the image and used homography matrix with respect to mosaicked image that is used for fast stitching. Fan Yang, Yang He, Zhen Sheng, Ang Yan [8] automated image stitching for x-ray images and discussed about down sampling to decrease the computation and the automated stitching of full spine and lower limb was proposed. Somaya Adwan, Iqbal Alsaleh, Rasha Majed [9] have discussed about stitching of medical images using dynamic time wrapping technique and compare the results with the normal stitching algorithms and perform dimensional reduction to reduce the computational complexity.

3. ENVIRONMENTAL SETUP

The experimentation were performed on cluster [10] equipped with Hadoop. This project has been provisioned with one NameNode and four DataNodes. The NameNode was configured to use two 2.5-GHz CPUs, 2 GB of RAM, and 100 GB of storage space. Each DataNode was configured to use two 2.5-GHz CPUs, 1 GB of RAM, and 100 GB of disk storage. Besides this, all the computing nodes were connected by a gigabit switch. Ubuntu 14.04 LTS, Hadoop 2.7.1, and Java 1.7.0_78 were installed on both the NameNode and the DataNodes. Table 1 shows the master-slave hardware configuration, while Tables 2 and 3 show the cluster hardware and software configurations, respectively.

Table 1. Master and Slave Specifications

Name	Number	Details
Master	1	2 × 2.5 GHz CPU Node, 2 GB RAM, 100 GB disk space
Slaves	3	2 × 2.5 GHz CPU Node, 2 GB RAM, 100 GB disk space

Table 2. Hardware Specifications

Name	Number	Details
Name Node	1	2 × 2.5 GHz CPU Node, 2 GB RAM
Data Node	3	2 × 2.5 GHz CPU Node, 2 GB RAM
Network	–	1 Gbit switch connecting all nodes
Storage	4	100 GB

Table 3. Software Specifications

Name	Version	Details
Hadoop	2.7.1	Installed on each node of the computer
Ubuntu	14.04	Pre-configured with Java and the Hadoop.
Java Openjdk	7	For Programming Image Processing

4. RESULTS



Figure 1



Figure 2



Figure 3



Figure 4



Figure 5 output Stitched image

The above Figure 1 to Figure 4 shows the input images that are given to the stitching job and figure 5 shows the stitched output image. This is the application is used when the camera cannot capture the entire location and when zooming decreases the quality so take the normal view of images and then get the high quality panorama image.

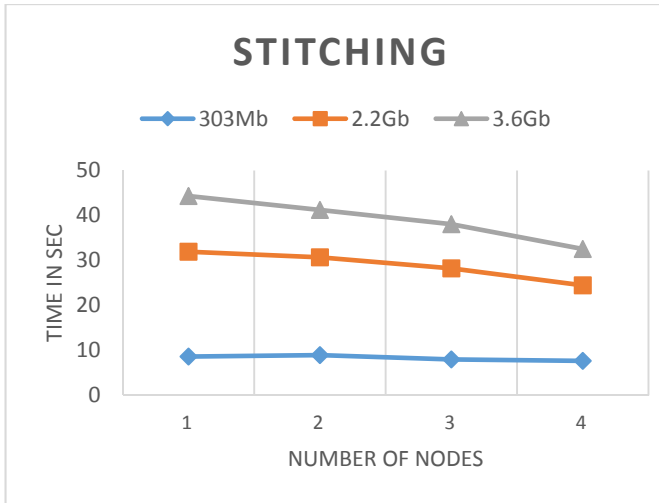


Figure 6

The above Figure 6 shows the results compared for stitching job for different size images and run each on single node, 2 nodes, 3 nodes, 4 nodes.



Figure 7



Figure 8



Figure 9



Figure 10-output stitched image

The above example i.e., figure 7 , 8&9 are input images the figure 10 is the output image that is stitched output of all the Input images. This shows that the thief face is not clear in all the input images but the output image has the clear face of the thief.

5. CONCLUSION AND FUTURE SCOPE

In this paper we have attempted to process small scale images in which we analyzed Image stitching algorithm.

Image stitching is the process which combine more images which are having intersection area. This application can be implemented to the security cameras to detect thieves when the face is not clear in one camera or capture but by using the stitching from adjacent cameras we can detect the face. The high quality video processing can be achieved from low resolution cameras.

The future scope of the paper is to develop algorithms for stitching when the security cameras in different locations and the face of thief is not clear in all the cameras then the face should be stitched.

Hence we can apply these techniques by which we can reduce the load and improve the performance of this algorithms.

This novel approach will reduce the load and will improve the performance with respect to the present algorithms.

REFERENCES

- [1] Z. Lv, Y. Hu, H. Zhong, J. Wu, B. Li and a. H. Zhao, "Parallel K-means clustering of remote sensing images based on MapReduce," *Web Information Systems and Mining (WISM '10)*, p. 162–170, 2010.
- [2] G. Fox, X. Qiu, S. Beason, J. Choi, J. Ekanayake, T. Gunarathne, M. Rho, H. Tang, N. Devadasan and a. G. Liu, "Biomedical case studies in data intensive computing," *Springer-Verlag*, vol. 5931, pp. 2-18, 2009.
- [3] "Hadoop," apache, [Online]. Available: <http://hadoop.apache.org/>.
- [4] "HDFS," apache, [Online]. Available: <http://hadoop.apache.org/hdfs/>.
- [5] "MapReduce," apache, [Online]. Available: <http://hadoop.apache.org/mapreduce/>.
- [6] Lowe, Matthew Brown and a. D. G., "Automatic Panoramic Image Stitching using Invariant Features," *International Journal of Computer Vision*, 2007.
- [7] Cheng-Ming Huang, Shu-Wei Lin and a. J.-H. Chen, "Efficient Image Stitching of Continuous Image," *IEEE*, vol. 15, no. 10, pp. 5910-5918, 2015.
- [8] Fan Yang, Yang He, Zhen Sheng and A. Yan, "Improvement of automated image stitching system for DR X-ray images," *elsevier*,

pp. 108-114, 2016.

- [9] Somaya Adwan, Iqbal Alsaleh and R. Majed, "A new approach for image stitching technique using Dynamic Time Warping (DTW) algorithm towards scoliosis X-ray diagnosis," *ELSEVIER*, pp. 32-46, 2016.
- [10] Ghemawat, J. Dean and a. S., "MapReduce: simplified data processing on large clusters," *Commun. ACM*, vol. 51, p. 107-114, 2008.

AUTHORS

Mr S. Vidya Sagar Appaji currently working as an Assistant Professor in the Department of CSE of MVGR College of Engineering. He is having over 11 Years of teaching experience. His Research includes Algorithms, Image Processing etc...



Dr P.V.Lakshmi is Professor and BOS chairman in the Department of Information Technology, GITAM University, Visakhapatnam. She has authored several books, several book chapters and research papers in International Conferences and Journals. Her research areas include Cryptography, Algorithms in Bio Informatics, Network Security and other areas.



Dr P. Srinivasa Rao currently working as an Associate Professor in the Department of CSE of MVGR College of Engineering. He is having over 11 Years of teaching experience. His Research includes Data Warehousing and Mining, Distributed Computing, Image Processing etc..,

