IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

A Study on Median Filter with Big Data

T Phanindra Reddy[1], Vidya Sagar Appaji Setti[2]

CSE Department, MVGR College of Engineering, Vizianagaram, India [1]

Asst. Professor, CSE Department, MVGR College of Engineering, Vizianagaram, India [2]

Abstract: At present global scenario, Image processing plays a vital role in the orientation of research area. There are different image operations performed in different applications, one among is the "image enhancement". As Image enhancement is the problem-oriented procedure, the main aim is to improve the visual appearance of the images, i.e given an input low-quality image that process to get the output with a high-quality image. It is well known that image enhancement techniques are applied on different areas like Aerial imaging Satellite imaging, Medical imaging, Digital camera application, Remote sensing etc.., several methodologies are there to implement image enhancement, and they are: 1.Spatial domain 2.frequency domain. Earlier several research works have been done and proposed in different platforms but the limitation is that they processed one image at once, but in the paper, images are operated at a single time for that the paper introduced a platform named HADOOP [1].

Keywords: Hadoop, Image processing, Image enhancement, spatial domain, frequency domain.

I. INTRODUCTION

The word Image processing is introduced many years ago but the research is going on in this platform. Image processing is the techniques that performs several operations on original image and get benefit from those operations. Each image processing operation having its own advantage. Filtering is one of the image processing techniques and there are many filters which we can apply on an image according to the application. The importance is given to not only the perfection of the output but also the computational. The paper discusses about the recent advancements in distributed computing like MapReduce [2]. The advantage of multimode clustering [3] is also an important parameter in computational speed.

The paper discusses about the advantage of tacking a whole image as a single input instead of tacking image as a matrix or pixels. The load on single system is more while compared to the distributed parallel computing.

We can observe more disk space is require when the image data coming from cloud and should be processed with flash speed.

In the paper one of the image filters, Median filter is to be discussed. Image filters are having some rules. First rule is filter should be uneven. The second rule is the sum of elements of filter should be 1 for the output image is having same brightness as original image, If it is greater than 1 the brightness will increase and if it is less than 1 the output image will be darker than original image.

II. RELATED WORK

Bhat Jasra et al.[4] Proposed a simple method for removing high density salt and pepper noise using neural network and Median filter. There are two phases in this paper they proposed the first phase consisting of neural network application of image the second phase is applying the Median filter.

Vishwanath Gouda et al. [5]discussed about the impulse noise of an image and a new method is using to remove that noise from image. This paper also used the median filter to remove impulse noise. The max-min Impulse detector is used to detect the noise at each pixel and act dynamically to each pixel. Shruti Singh et al.[6] proposed a new method to remove salt and pepper noise in an image, the method consists of trimmed median values and these will apply on each pixel of the image. The only disadvantage of this paper is that it can remove noise only from greyscale images. Eesa Nikahd et al.[7] proposed a new method to use median filter, the method is implemented on hardware an the filter is 1-D. The maximum working clock frequency is independent of median filter window length. The research is going on in the field of Median filtering due to its applications.

III. IMAGE PROCESSING WITHOUT HADOOP

The normal image processing techniques [8] follows sequential methods for computing and processing. There are methods to process an image after image due to these sequential methods the processing time will increase and load on single system will increase. We have simple techniques and process an image in online, but this will take by single system with windows operating system and there is a problem of computational and resource capabilities when the image data is in large scale. Many of us are using programming languages like C and Java to process an image, these are also not sufficient for voluminous amount of data.

IV. IMAGE PROCESSING WITH HADOOP

HDFS (Hadoop Distributed File System)[9] is the file system for hadoop having 64MB or multiples of 64MB of

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

block size. The programming style used is MapReduce. In our approach the Map method is enough to process the images in parallel. The Reduce method is not required for this approach.

The map method will distribute the images to the nodes. The map detects the images using key, value pair. The key input is NullWritable and value is ByteWritable. The final output is written back to the HDFS.

The main thing that is used in implementation is WholeFileRecordReader this is used because the whole image should be read as one record to use the advantage of block size in HDFS. The image input format is also set to WholeFileInputFormat to support the RecordReader. The flow of algorithm is explained below

- 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 BytesArrayInputStream.
- WholeFileRecordReader is the class written for reading the image.
- Get the image height and width.
- In the Map() perform Median filtering for all the images set at the input folder in parallel.
- Set the height and width for the output image.
- Save the final image in jpeg format in HDFS.

V. ENVIRONMENTAL SETUP

The experimentation were performed on cluster equipped with Hadoop. This project has been provisioned with one NameNode and four DataNodes. The NameNode was configured to use two 2.53GHz CPUs, 4 GB of RAM, and 500 GB of storage space. Each DataNode was configured to use two 1.7-GHz CPUs, 4 GB of RAM, and 500 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 I and II shows the master-slave hardware configuration, while Tables III and IV show the cluster hardware and software configurations, respectively.

TABLE I SINGLE NODE CONFIGURATION

PC	Details
Personal Computer	Processor:Intel Core i3, 2x2.53 GHz, RAM: 4 GB, Total Storage: 500 GB. Operating System: Ubuntu 14.04 LTS

TABLE II MASTER-SLAVE CONFIGURATION

Name	Number	Details
Master	1	2 × 2.53 GHz CPU Node, 4 GB RAM, 500 GB disk space
Slaves	3	2 × 1.7 GHz CPU Node, 4 GB RAM, 500 GB disk space

TABLE III EQUIPMENT USED

Name	Number	Details
Name Node	1	2 × 2.53 GHz CPU Node, 4 GB RAM
Data Node	3	2 × 1.7 GHz CPU Node, 4 GB RAM
Network	–	Cisco catalyst 2960 switch connecting all nodes
Storage	4	500 GB

TABLE IV SOFTWARES USED

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

VI. EXPERIMENTAL RESULTS

Figure 1 noise added inputs for median filter

IJARCCE

ISSN (Online) 2278-1021
ISSN (Print) 2319 5940

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified
Vol. 5, Issue 10, October 2016

Figure 2 noise removed outputs with median filter

Figure 1 shows inputs for median filter and Figure 2 shows the outputs for median filter i.e noise removed results respectively.

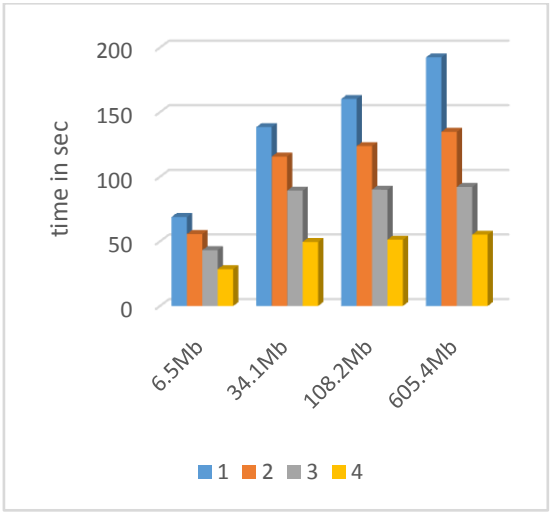

Figure 3 Comparison of processing time for different data and nodes

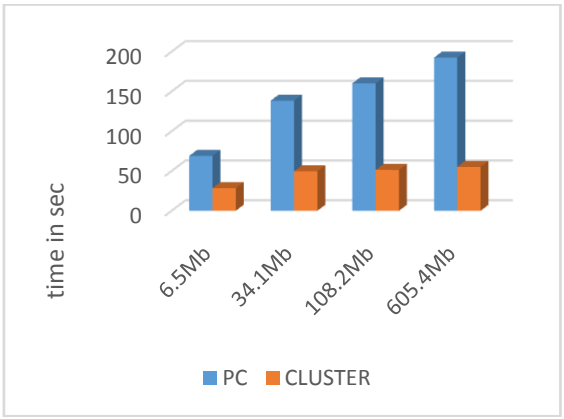

Figure 4 Comparison of processing time for single and 4 nodes cluster

Figure 3 is the bar graph showing the processing time comparison when the same data is processed on single node and 2,3,4 nodes respectively and different data sets also compared on the same graph. Figure 4 compares processing time for single node and 4nodes respectively.

VII. CONCLUSION

The image data coming from different sources increasing rapidly and the data is corrupting i.e adding with noise due to different climatic conditions and interference with other signals. When the added noise is salt and pepper we can remove it by using median filter.
The hadoop application can use where the output is require with flash speed like security cameras. This application can be extended to video data also.

REFERENCES

[1] "Hadoop," Apache, [Online]. Available: http://hadoop.apache.org.
[2] "MapReduce," Apache, [Online]. Available: http://hadoop.apache.org/mapreduce/.
[3] Sridhar Vemula and C. Crick, "Hadoop Image Processing Framework," IEEE International Congress on Big Data, pp. 506-513, 2015.
[4] B. Jasra., Aniqa Yaqoob and S. K. Dubey, "Removal of High Density Salt and Pepper Noise," 6th International Conference - Cloud System and Big Data Engineering, pp. 78-82, 2016.
[5] V. R. Malipatil. and D. G. M. Patil, "Design Modeling and Certain Investigations of," Internation Conference on Microwave, Optical and Communication Engineering, pp. 459-462, 2015.
[6] S. Singh. and A. K. Tiwari, "An Enhanced Decision Based Unsymmetric Trimmed Median Filter For Removal of High Density Salt and Papper Noise," 2015 Fifth National Conference on Computer Vision, Pattern Recognition, Image Processing and Graphics (NCVPRIPG), pp. 1-4, 2015.
[7] Eesa Nikahd, Payman Behnam and R. Sameni, "High-Speed Hardware Implementation of Fixed and Run-time Variable Window Length 1-D Median Filters," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 63, no. 5, pp. 478-482, 2015.
[8] "wikipedia," 12 october 2016. [Online]. Available: https://en.wikipedia.org/wiki/Image_processing.
[9] "MapReduce," Apache, [Online]. Available: http://hadoop.apache.org/hdfs/.