ISSN 2395-695X (Print)
ISSN 2395-695X (Online)

International Journal of Advanced Research in Biology Engineering Science and Technology (IJARBEST)
Vol. 2, Issue 4, April 2016

# INTERPRETIVE ARCHITECTURE FOR REMOTE SENSING APPLICATION USING REAL TIME BIG DATA

Mrs.B.Rasina Begum
(Anna univ affliated) asst.prof in CSE Department
Mohamed Sathak Engineering college
Kilakkarai,Ramnad dist. TN
*rasinayousuf@gmail.com*

A.Asrin Mahmootha
(Anna univ affliated) ME-CSE Department
Mohamed Sathak Engineering college
Kilakkarai,Ramnad dist. TN
*asrincs@gmail.com*

***Abstract*—The Remote sensing world generates larger volume of real-time data which normally called as Big Data. This contains insight information with significance Today, there is a challenge to a real-time remote sensing Big Data that extract useful information in an efficient manner and to analyze, aggregate and store when they are collected. To overcome these challenges we design a system architecture called real-time Big Data analytical architecture for remote sensing satellite application that allows for both real-time data processing and offline data processing. Our architecture has the capability of dividing, load balancing, and parallel processing of data. Here RSDU acquires data from the satellite and sends this data to the Base Station, where initial processing takes place. And the next unit is DPU, plays a vital role in architecture for efficient processing of real-time Big Data by providing filtration, load balancing, and parallel processing. Final unit is DADU, which is the upper layer unit of the architecture, which is responsible for compilation, storage of the results, and generation of decision based on the results received from DPU. These process results in efficiently analyzing real-time remote sensing Big Data based on earth observatory system.**

***keywords*—Big Data, Data analysis decision unit (DADU),data processing unit (DPU),remote sensing Big Data acquisition unit (RSDU).**

## I. INTRODUCTION

**R**ECENTLY, a great deal of interest in the field of Big Data and its analysis has risen [1]–[3], mainly driven from extensive number of research challenges strappingly related to bonafide applications, such as modeling, processing, querying, mining, and distributing large-scale repositories. The term "Big Data" classifies specific kinds of data sets comprising formless data, which dwell in data layer of technical computing applications [4] and theWeb [5]. The data stored in the underlying layer of all these technical computing application scenarios have some precise individualities in common, such as 1) largescale data, which refers to the size and the data warehouse; 2) scalability issues, which refer to the application's likely to be running on large scale (e.g., Big Data); 3) sustain extraction transformation loading (ETL) method from low, raw data to well thought-out data up to certain extent; and 4) development of uncomplicated interpretable analytical over Big Data warehouses with a view to deliver an intelligent and momentous knowledge for them [8]. Big Data are usually generated by online transaction, video/audio, email, number of clicks, logs, posts, social network data, scientific data, remote access sensory data, mobile phones, and their applications [6], [7]. These data are accumulated in databases that grow extraordinarily and become complicated to confine, form, store, manage, share, process, analyze, and visualize via typical database software tools. Advancement in Big Data sensing and computer technology revolutionizes the way remote data collected, processed, analyzed, and managed [9]–[12]. Particularly, most recently designed sensors used in the earth and planetary observatory system are generating continuous stream of data. Moreover, majority of work have been done in the various fields of remote sensory satellite image data, such as change detection [13], gradient-based edge detection [14], region similaritybased edge detection [15], and intensity gradient technique for efficient intraprediction [16]. In this paper, we referred the highspeed continuous stream of data or high volume offline data to "Big Data," which is leading us to a new world of challenges [17]. Such consequences of transformation of remotely sensed data to the scientific understanding are a critical task. Hence the rate at which volume of the remote access data

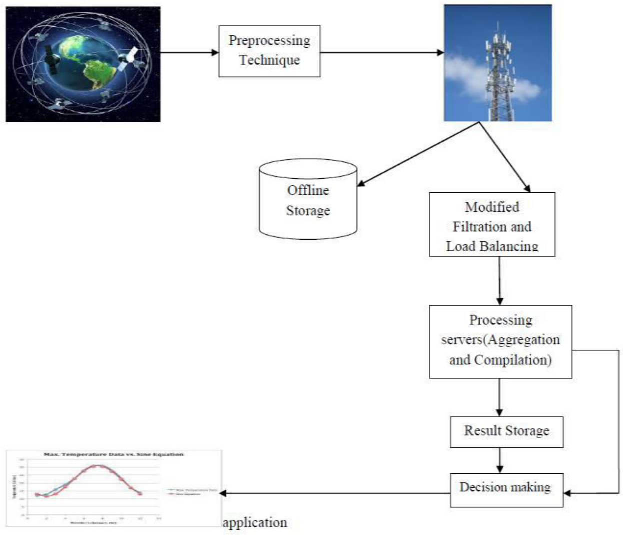
.

## II. MOTIVATION FOR REMOTE SENSING BIG DATA ANALYTICS

The increase in the data rates generated on the digital universe is escalating exponentially. With a view in employing current tools and technologies to analyze and store, a massive volume of data are not up to the mark, since they are unable to extract required sample data sets. Therefore, we must design an architectural platform for analyzing both remote access realtime and offline data. When a business enterprise can pull-out all the useful information obtainable in the Big Data rather than a sample of its data set, in that case, it has an influential benefit over the market competitors. Big Data analytics helps us to gain insight and make better decisions. Therefore, with the intentions of using Big Data, modifications in paradigms are at utmost. To support our motivations, we have described some areas where Big Data can play an important role. Understanding environment requires massive amount of data collected from various sources, such as remote access satellite observing earth characteristics [measurement data set (MDS) of satellite data such as images], sensors monitoring air and water quality, metrological circumstances, and proportion of $CO_2$ and other gases in air, and so on. Through relating all the information drifting such as $CO_2$ emanation, increase or decrease on greenhouse effects and temperature, can be found out. In healthcare scenarios, medical practitioners gather massive volume of data about patients, medical history, medications, and other details. The above-mentioned data are accumulated in drug-manufacturing companies. The nature of these data is very complex, and sometimes the practitioners are unable to show a relationship with other information, which results in missing of important information. With a view in employing advance analytic techniques for organizing and extracting useful information from Big Data results in personalized medication, the advance Big Data analytic techniques give insight into hereditarily causes of the disease.

## III. REMOTE SENSING BIG DATA ANALYTICS ARCHITECTURE

The term Big Data covers diverse technologies same as cloud computing. The input of Big Data comes from social networks (Facebook, Twitter, LinkedIn, etc.), Web servers, satellite imagery, sensory data, banking transactions, etc. Regardless of very recent emergence of Big Data architecture in scientific applications, numerous efforts toward Big Data analytics architecture can already be found in the literature. Among numerous others, we propose remote sensing Big Data architecture to analyze the Big Data in an efficient manner as shown in Fig delineates $n$ number of satellites that obtain the earth observatory Big Data images with sensors or conventional cameras through which sceneries are recorded using radiations. Special techniques are applied to process and interpret remote sensing imagery for the purpose of producing conventional maps, thematic maps, resource surveys, etc.We have divided remote sensing Big Data architecture into three parts, i.e., 1) remote sensing data acquisition unit (RSDU); 2) data processing unit (DPU); and 3) data analysis and decision unit (DADU). The functionalities and working of the said parts are described as below.




### *A. Remote Sensing Big Data Acquisition Unit (RSDU)*

Remote sensing promotes the expansion of earth observatory system as cost-effective parallel data acquisition system to satisfy specific computational requirements. The Earth and Space Science Society originally approved this solution as the standard for parallel processing in this particular context [2]. As satellite instruments for Earth observation integrated more sophisticated qualifications for improved Big Data acquisition, soon it was recognized that traditional data processing technologies could not provide sufficient power for processing such kind of data. Therefore, the need for parallel processing of the massive volume of data was required, which could efficiently analyze the Big Data. For that reason, the proposed RSDU is introduced in the remote sensing Big Data architecture that gathers the data from various satellites around the globe as shown in Fig. It is possible that the received raw data are distorted by scattering and absorption by various atmospheric gasses and dust particles. We assume that the satellite can correct the erroneous data. However, to make the raw data into image format, the remote sensing satellite uses Doppler or SPECAN algorithms [28]. For effective data analysis, remote sensing satellite preprocesses data under many situations to integrate the data from different sources, which not only decreases storage cost, but also improves analysis accuracy. Some relational data preprocessing techniques are data integration, data cleaning, and redundancy elimination. After preprocessing phase, the collected data are transmitted to a ground station using downlink channel. This transmission is directly or via relay satellite with an appropriate tracking antenna and communication link in a wireless atmosphere. The data must be corrected in different methods to remove distortions caused due to the motion of the platform relative to the earth, platform attitude, earth curvature, nonuniformity of illumination, variations in sensor characteristics, etc. The data is then transmitted to Earth Base Station for further processing using direct communication link. We divided the data processing procedure into two steps, such as real-time Big Data processing and offline Big Data processing. In the case of offline data processing, the Earth Base Station transmits the data to the data center for storage. This data is then used for future analyses. However, in real-time data processing, the data are directly transmitted to the filtration and load balancer server (FLBS), since storing of incoming real-time data degrades the performance of real-time processing.

Fig. 2. Remote sensing earth observatory image.

### B. Data Processing Unit

In data processing unit (DPU), the filtration and load balancer server have two basic responsibilities, such as filtration of data and load balancing of processing power. Filtration identifies the useful data for analysis since it only allows useful information, whereas the rest of the data are blocked and are discarded. Hence, it results in enhancing the performance of the whole proposed system. Apparently, the load-balancing part of the server provides the facility of dividing the whole filtered data into parts and assign them to various processing servers. The filtration and load-balancing algorithm varies from analysis to analysis; e.g., if there is only a need for analysis of sea wave and temperature data, the measurement of these described data is filtered out, and is segmented into parts. Each processing server has its algorithm implementation for processing incoming segment of data from FLBS. Each processing server makes statistical calculations, any measurements, and performs other mathematical or logical tasks to generate intermediate results against each segment of data. Since these servers perform tasks independently and in parallel, the performance proposed system is dramatically enhanced, and the results against each segment are generated in real time. The results generated by each server are then sent to the aggregation server for compilation, organization, and storing for further processing.

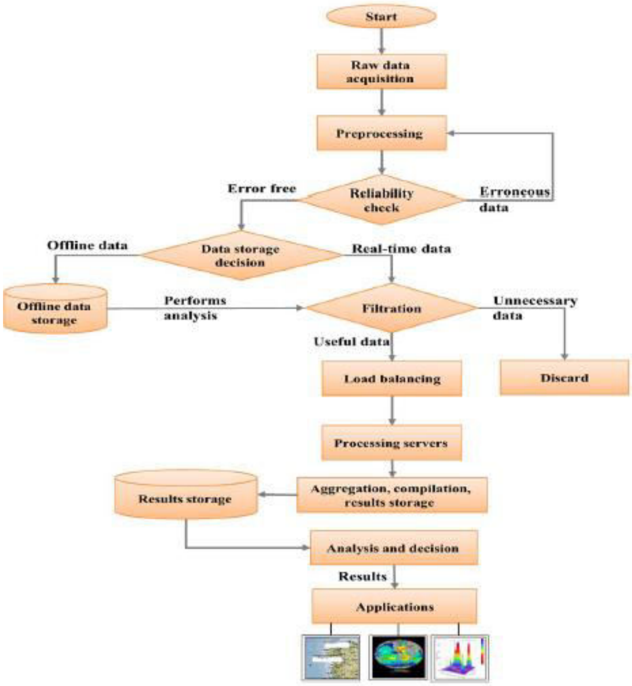



Fig. 3. Flowchart of the remote sensing Big Data architecture.

### C. Data Analysis and Decision Unit (DADU)

DADU contains three major portions, such as aggregation and compilation server, results storage server(s), and decision making server. When the results are ready for compilation, the processing servers in DPU send the partial results to the aggregation and compilation server, since the aggregated results are not in organized and compiled form. Therefore, there is a need to aggregate the related results and organized them into a proper form for further processing and to store them. In the proposed architecture, aggregation and compilation server is supported by various algorithms that compile, organize, store, and transmit the results. Again, the algorithm varies from requirement to requirement and depends on the analysis needs. Aggregation server stores the compiled and organized results into the result's storage with the intention that any server can use it as it can process at any time. The aggregation server also sends the same copy of that result to the decision-making server to process that result for making decision. The decision-making server is supported by the decision algorithm, which inquire Fig. 3. Flowchart of the remote sensing Big Data architecture. different things from the result, and then make various decisions (e.g., in our analysis, we analyze land, sea, and ice, whereas other finding such as fire, storms, Tsunami, earthquake can also be found). The decision algorithm must be strong and correct enough that efficiently produce results to discover hidden things and make decisions. The decision part of the architecture is significant since any small error in decision-making can degrade the efficiency of the whole analysis. DADU finally displays or broadcasts the decisions, so that any application can utilize those decisions at real time to make their development. The applications can be any business software, general purpose community software, or other social networks that need those findings (i.e., decision-making). The self-explanatory flowchart supporting the working of the proposed architecture is depicted in Fig. 3.

## IV. ANALYSIS AND DISCUSSION

Using the proposed architecture for offline as well online traffic, we perform a simple analysis on remote sensing earth observatory data. We assume that the data are big in nature and difficult to handle for a single server. The data are continuously coming from a satellite with high speed. Hence, special algorithms are needed to process, analyze, and make a decision from that Big Data. Here, in this section, we analyze remote sensing data for finding land, sea, or ice area. We have used the proposed architecture to perform analysis and proposed an algorithm for making decision. First, we take satellite-sensed Big Data samples from European satellite Agency (ESA) to analyze land, sea, and ice separately. On the basis of these analyses, we proposed a set of algorithms for handling, processing, analyzing, and decision-making (detecting sea, land, This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination. and ice area) for remote sensing Big Data images using our proposed architecture. In this section, we describe the data sets and tools that are used to perform analysis. Furthermore, we described the analysis findings of the data sets and proposed algorithms.

### *A. Tools, Data Set, and Implementation Environment*

We used BEAM VISAT version 5.0 and EnviView 2.8.1 for simple analysis of the satellite data sets. Beam VISAT and EnviView provide an easy way to understand the structure of ENVISAT mission satellite data sets. They are also useful for simple statistical analysis. For complicated analysis and efficient processing of the Big Data sets, we could not use these tools. Apache Hadoop with MapReduce program using single node setup for sophisticated analysis is used for the implementation of the proposed algorithm, since Hadoop provides the facility of parallel, high-performance computing using a large number of servers. Therefore, it is suitable for analysing a large amount of remote sensory image data. The proposed architecture uses a similar mechanism for load balancing; hence, preference is given to Hadoop for sophisticated analysis, algorithm development, and testing. We analyzed the ENVISAT mission data sets (e.g., products) with advanced synthetic apertures radar (ASAR) and medium resolution imaging spectrometer (MERIS) instruments or sensor. The main focus is given to ENVISAT ASAR data sets because ENVISAT satellite mission has been continuously providing global measurements for the earth including sea, land, ice, and forest since 2002 [28]. It has ten basic instrument data sets for sensing earth. However, we have considered only two instruments, i.e., ASAR and MERIS. More data sets are analyzed from MERIS, ASAR, and few from other ENVISAT sensors, but here we discussed the analytical results of five basic types of ASAR data sets taken from ESA [22].These five remote sensory imaging data sets covered different earth area including sea, desert, forest, beaches, and cities. The data sets included different earth area from Vietnam, Poland, and Germany, Western Sahara and Mauritania and Mali, South Africa, and Spain with different times as shown in Fig. 2. These products are of different types, i.e., altering polarization medium-resolution image (APM), wide swath medium resolution image (WSM), and global monitoring mode image (GMI). The software used for sensing data is ASAR with a different version of 4.02, 3.00S00, 5.03L03, 4.02. A detailed description of these products including their ID/name, product type, sensing time, software version, mission, SPH descriptor and area covered is shown in Table I. The area covered corresponding to each product is also shown graphically using 2-D world map in Fig. 4.

### *B. Findings and Discussion*

On Earth station, the reception of preprocessed and formatted data from satellite contains all or some of the following parts depending on the product.

1) Main product header (MPH): It includes the products basis information, i.e., id, measurement and sensing time, orbit, information, etc.

2) Special products head (SPH): It contains information specific to each product or product group, i.e., number of data sets descriptors (DSD), directory of remaining data sets in the file, etc.

3) Annotation data sets (ADS): It contains information of quality, time tagged processing parameters, geo location tie points, solar, angles, etc.

4) Global annotation data sets (GADs): It contains calling factors, offsets, calibration information, etc.

5) Measurement data set (MDS): It contains measurements or graphical parameters calculated from the measurement including quality flag and the time tag measurement as well. The image data are also stored in this part and are the main element of our analysis.

The MPH and SPH data are in ASCII format, whereas all the other data sets are in binary format. MDS, ADS, and GADs consist of the sequence of records and one or more fields of the data for each record. In our case, the MDS contains number of records, and each record contains a number of fields. Each record of the MDS corresponds to one row of the satellite image, which is our main focus during analysis. In the basic analysis of all these products, we found that the satellite uses the SPECAN algorithm for processing raw data into image data. An image data are normally composed in rows and columns, i.e., matrices. In remote sensory Big Data, the MDS data sets containing the image data are in the form of records. Number of records shows the number of rows in the satellite image and number of sample/line shows the number of field/column in the image. The value against record (i) and sample (j) corresponds to the pixels value of the row (i) and column (j) of the image. Table II shows the algorithm used by the satellite (for processing raw data to image), the number of records in each data sets, number of samples in each record, and mean and standard deviation (SD) of all the MDS process data values corresponding to all records and samples. It is noticed that the mean value of the product is quite lower than all other products mean value. The products 2 only covered the land area; hence, their mean value is quite lower, whereas all other products covered both land and sea area. Since in land satellite images, the color of the land particles is mostly nearer to the block. Therefore, their value is quite lower. Hence, the overall mean value for image particles is lower. In the analysis of the products, we consider 20 random blocks image data, where each block contains 20 000–30 000 sample image-related values from MDS records for both land and sea area.We calculated the mean and SD of all image sample values in each block and calculated the maximum sample value in the block and the normal trend of sample values. We also observed the distribution of MDS image values for land area as well as sea area and then tried to find out the major difference between their data. We assume that the mean value for the land area should be lower as compared to sea data. The land normally has greenery (except deserts), and other objects whose color is nearer to black. Hence, the pixel value is lower, which results in reducing the overall mean value. The pixels SD for the land data is also higher in case of land area. Normally, sea has one color, and there are very few particles on the surface of the sea, which can be ignored. As a result, the color of the sea remains almost same. Therefore, the SD for the sea data is lower and for land, SD is higher as the land has many different things with different colors on its surface. Accordingly, we have considered mean, SD, and the maximum value as the basic parameters for our analysis. Tables III and IV show the overall statisticalanalytical results of all the products with respect to land area and sea area, respectively, in which the minimum and maximum values of mean and SD among all the blocks are presented. The analysis findings are almost according to our assumption except for few abnormalities. It can also be seen that the mean values for land areas are quite lower as compared to sea area and SD values in case of land are quite higher as compared to sea area.

*C. Algorithm Design and Testing*

On the basis of the analysis made in the previous section, a set of algorithms is proposed to process high-speed, large amount of real-time remote sensory image data using our proposed architecture. It works on both DPU and DADU by taking data from satellite as input to identify land and sea area from the data set. The set of algorithms contains four simple algorithms, i.e., algorithm I, algorithm II, algorithm III, and algorithm IV that work on filtrations and load balancer, processing servers, aggregation server, and on decision-making server, respectively. Algorithm I, i.e., Modified filtration and load balancer algorithm (MFLBA) works on filtration and load balancer to filter only the require data by discarding all other information. It also provides load balancing by dividing the data into fixed size blocks and sending them to the processing server, i.e., one or more distinct blocks to each server. This filtration, dividing, and load-balancing task speeds up our performance by neglecting unnecessary data and by providing parallel processing. Algorithm II, i.e., processing and calculation algorithm (PCA) processes filtered data and is implemented on each processing server. It provides various parameter calculations that are used in the decision-making process. The parameters calculations results are then sent to aggregation server for further processing. Algorithm III, i.e., aggregation and compilations algorithm (ACA) stores, compiles, and organizes the results, which can be used by decision-making server for land and sea area detection. Algorithm IV, i.e., decision-making algorithm (DMA) identifies land area and sea area by comparing the parameters results, i.e., from aggregation servers, with threshold values.

## V. RESULTS AND IMPLEMENTATION

We implemented our algorithms in simple java language using Beam-5.0 library [29] as well in Hadoop MapReduce, initially in a single-node environment. In the Hadoop implementation, Map function takes the image block offset as a key and the image block (pixel values) as a value parameter.

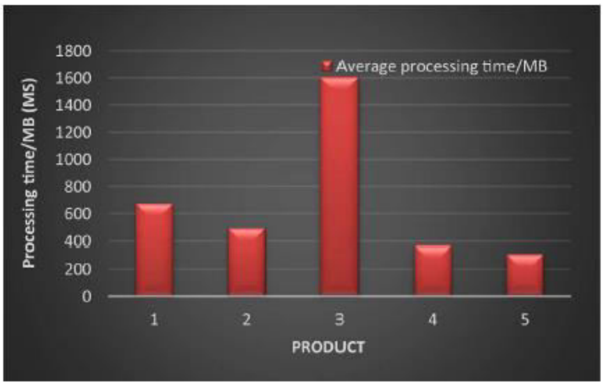



Fig. 8. Average processing time of products using Hadoop implementation.

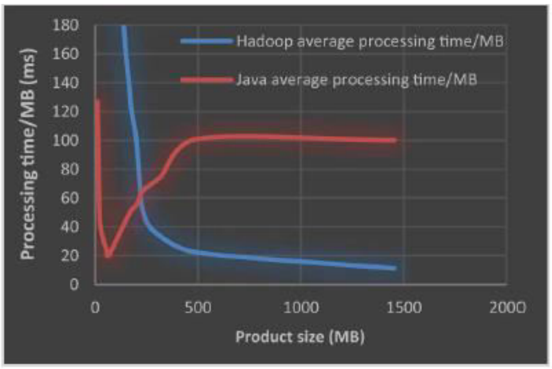



Fig. 9. Efficiency comparison of Hadoop Map Reduce implementation and simple Java implementation.

Since Hadoop MapReduce cannot directly process image blocks, the whole product image data are converted into sequence file to be processed using MapReduce. In such a way, one line of the sequence file contains one image block. Map function performs parameters calculations on incoming block values and finally sends the block number as a key and list of parameters results as a value to the Reduce function. Reduce function uses parameter results for performing decision-making on them. We test and evaluate our algorithms with respect to accuracy and processing time using various ESA products [22]. Accuracy evaluation is done by considering two parameters, such as true positive (TP), which shows the measurements in percentage (%) the land block are correctly identified and false positive (FP) shows the measurements in percentage (%), the sea blocks are incorrectly identified as land blocks. Our proposed system detects different types of area of the world, such as land and sea with the overall accuracy of more than 95% TP and less than 3% FP. Product 2 does not have sea area. Therefore, all blocks are detected as land that results in 100% TP and 0% FP.

## VI. CONCLUSION AND FUTURE WORK

This system has an sarchitecture called Remote Sensing Big Data analytical architecture is used for analyzing real time data or offline data. This architecture comprises of mainly three units i) Remote Sensing Data Acquisition ii) Data Processing Unit iii) Data Analysis and Decision Unit. Initially the data are processed remotely and it is readable by the machines. Then it is transmitted to the Earth Base Station for Data Processing. It allows two type of data processing; they are processing of real time data and offline data, which the data transmitted to offline data storage device. The offline storage data are used for the later usage of data whereas the real time data is directly transmitted to the filtration process and to the load balancer process which improves the system efficiency. Next process is data aggregation unit which is used for the decision and analyzing server. The experimental result shows the efficiency of the analysis.

### FUTURE WORK

In Our Future work, we collect the Big Data which is a remote sensing data. We apply the preprocessing technique for the collected data. Then the data allowed for the data processing unit which allows Modified Filtration and Load Balancing Algorithm for the filtration process and the load balancing process that results in effective analyzing of the data. We use for the real time decision making process such tsunami prediction, earth quake,etc.

## REFERENCES

[1] D. Agrawal, S. Das, and A. E. Abbadi, “Big Data and cloud computing: Current state and future opportunities,” in Proc. Int. Conf. Extending Database Technol. (EDBT), 2011, pp. 530–533.

[2] J. Cohen, B. Dolan, M. Dunlap, J. M. Hellerstein, and C. Welton, “Mad skills: New analysis practices for Big Data,” PVLDB, vol. 2, no. 2, pp. 1481–1492, 2009.

[3] J. Dean and S. Ghemawat, “Mapreduce: Simplified data processing on large clusters,” Commun. ACM, vol. 51, no. 1, pp. 107–113, 2008.

[4] H. Herodotou et al., “Starfish: A self-tuning system for Big Data analytics,” in Proc. 5th Int. Conf. Innovative Data Syst. Res. (CIDR), 2011, pp. 261–272.

[5] K. Michael and K. W. Miller, “Big Data: New opportunities and new challenges [guest editors’ introduction],” IEEE Comput., vol. 46, no. 6, pp. 22–24, Jun. 2013.

[6] C. Eaton, D. Deroos, T. Deutsch, G. Lapis, and P. C. Zikopoulos, Understanding Big Data: Analytics for Enterprise Class Hadoop and Streaming Data. New York, NY, USA: Mc Graw-Hill, 2012.

[7] R. D. Schneider, Hadoop for Dummies Special Edition. Hoboken, NJ, USA: Wiley, 2012.

[8] A. Cuzzocrea, D. Saccà, and J. D. Ullman, “Big Data: A research agenda,” in Proc. Int. Database Eng. Appl. Symp. (IDEAS’13), Barcelona, Spain, Oct. 09–11, 2013.

[9] R. A. Schowengerdt, Remote Sensing: Models and Methods for Image Processing, 2nd ed. New York, NY, USA: Academic Press, 1997.

[10] D. A. Landgrebe, Signal Theory Methods in Multispectral Remote Sensing. Hoboken, NJ, USA: Wiley, 2003.

[11] C.-I. Chang, Hyperspectral Imaging: Techniques for Spectral Detection and Classification. Norwell, MA, USA: Kluwer, 2003.

[12] J. A. Richards and X. Jia, Remote Sensing Digital Image Analysis: An Introduction. New York, NY, USA: Springer, 2006.

[13] J. Shi, J. Wu, A. Paul, L. Jiao, and M. Gong, “Change detection in synthetic aperture radar image based on fuzzy active contour models and genetic algorithms,” Math. Prob. Eng., vol. 2014, 15 pp., Apr.
2014.

[14] A. Paul, J. Wu, J.-F. Yang, and J. Jeong, “Gradient-based edge detection for motion estimation in H.264/AVC,” IET Image Process., vol. 5, no. 4, pp. 323–327, Jun. 2011.

[15] A. Paul, K. Bharanitharan, and J.-F.Wang, “Region similarity based edge detection for motion estimation in H.264/AVC,” IEICE Electron. Express, vol. 7, no. 2, pp. 47–52, Jan. 2010.

[16] A.-C. Tsai, A. Paul, J.-C. Wang, and J.-F. Wang, “Intensity gradient techniquefor efficient intra prediction in H.264/AVC,” IEEE Trans. Circuits Syst. Video Technol., vol. 18, no. 5, pp. 694–698, May 2008.

[17] A. Plaza et al., “Recent advances in techniques for hyperspectral
image processing,” Remote Sens. Environ., vol. 113, pp. 110–122, 2009.

[18] S. Kalluri, Z. Zhang, J. JaJa, S. Liang, and J. Townshend, “Characterizing land surface anisotropy from AVHRR data at a global scale using high performance computing,” Int. J. Remote Sens., vol. 22, pp. 2171–2191, 2001.

[19] A. Labrinidis and H. V. Jagadish, “Challenges and opportunities with Big Data,” in Proc. 38th Int. Conf. Very Large Data Bases Endowment, Istanbul, Turkey, Aug. 27–31, 2012, vol. 5, no. 12, pp. 2032–2033.