IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727
PP 49-54
www.iosrjournals.org

Satellite Image Mining in Real-time Data Analytical Architecture

Safmina Yousef[1], Smitha Karunan[2]
M Tech Scholar[1], Associate Professor[2]
[1,2]Department of Computer Science Engineering Govt Engineering college, Mananthavady, Wayanad Kerala, India- 670 644

Abstract*—Remote sensors generates the huge amount of data from satellites. Nowadays, there is a large demand for real-time data for remote sensing applications and to extract useful information from the satellite image. This paper examines different mining approaches in various satellite image applications. A two level merging approach is used to extract the sea area. But, the performance is lower due to the lack of feature extraction. The hierarchical region merging approach is used to automatically extract the sea and land area. The merging approach can be better characterized by supervised information combined with the feature extraction. This results in bringing out a real-time analytical architecture to detect land and sea area. The sensors are deployed in forest to feature out the temperature and pressure for the fire detection. Sensors are of high cost. Sensors can be devastated due to climatic changes and also by animals. These issues can be avoided by using satelite image mining application for the fire detection using real-time analytical architecture. The features of the satellite image can be extracted by using the Scale Invariant Feature Transform (SIFT) algorithm. The speed of detecting fire using the satellite image can be increase by the use of Hadoop, a parallel processing framework.*

Keywords: *Hadoop Distributed File System (HDFS), Synthetic Aperture Radar (SAR), Scale Invariant Feature Transform (SIFT), LANDSAT, Remote data.*

I. Introduction

As of late, a lot of enthusiasm for Data mining has risen. Mining techniques can be implemented on new systems as existing methods are upgraded and new products developed. Whereas, mining tools are implemented on high-performance parallel processing systems. Thus, users can analyze large database in minutes. High-speed performance of mining tech-niques makes users analyze large quantities of data.

The advanced technology in mining techniques gives out collecting, managing, analyzing and processing of remote data. Remote sensors are designed to analyze earth observatory system. Many works have been done in the different fields of remote sensing data from the satellite, such as gradient-based edge detection [1] and change detection [2]. Real-time data analytic architecture is focussed on high speed continuous real time and massive offline data.

Nowadays, the world gets changed to the digital world and generating the massive amount of data. From these data sets, to analyze data is at risk by using current technology. The data is generated by streaming data, thus the data will arrive at high speed and the algorithm has to process all the arrived data. Therefore, there is a need for an architecture to analyze both the real-time and offline data sets. The advancement in the computer technology and remote sensing increases the huge amount of sensing data. The earth observatory data from spacecraft is around 1.7GB, this data is collected by single satellite. This massive amount of data has to process to extract the useful information. Thus, there is a need for an architecture for load balancing, data aggregation, and decision analysis.

Remote sensing is the science of obtaining information about areas from satellites. The satellites around the earth are generating varieties of data in every second. Sensing images are in the form of digital images. The image need to be pro-cessed efficiently. Recently, Orfeo toolbox is used to process images in large volumes [3], combined with MapReduce and Hadoop Distributed File System (HDFS). Orfeo results in an efficient and reduced execution time, but shortcomings in high-speed data processing. A problem with the big data analytic was the lack of coordination between databases. Thus, analysts were impeded by a tedious process of exporting data from the database [4]. The analysis of data using Hadoop results in fast parallel processing of data. Many data mining algorithms are migrating towards Hadoop, but the speed-up of the parallel k-means algorithm is not linear. The main reason results in nonlinear were that the communication overhead increases as increase the dataset size. These problems can be solve by using real-time analytical architecture.

Remote sensors collect a large amount of data from the satellite. The collected data has no meaning. The useful data need to get the extract from the collected data. Sometimes, the collected data might be not clear. To resolve such problem, uses an architecture to analyze real-time and offline data

II. Data Mining Techniques

Data Mining is a wide spreading area that offers a great platform for research. The availability of massive amount of data makes complex in knowledge aquiring. Data mining is the discover process. The knowledge discovery process mainly performs selection, integration, transformation and mining of

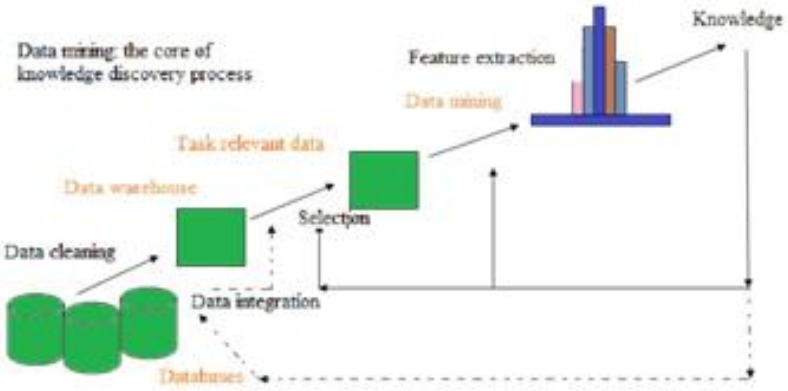

Fig. 1. Data mining process [5]

data. The collected image or data can be used for mining from different databases, data warehouse. Mining techniques can be implemented on existing software platforms as well as hardware platforms, and can be integrated with systems. The various steps in the data mining process are depicted in the figure 1. Data mining is supported by three technologies, namely, massive data collection, multiprocessor computers, and data mining algorithm. The five techniques of data mining are:

Association: It is a simple technique to correlate two or more items.

Classification: Uses decision tree to determine the classi-fication.

Clustering: Clustering is a method of dividing data into groups of similar objects.

Prediction: Prediction method is the combination of trends, classification, and relations.

Feature selection and extraction: It is an attribute reduction process. The data collected from satellites are then extracted to get useful informations. Then the feature extractions are carried out for further data processing. In feature extraction technique simplifies the large amount of resources to explain a huge set of data accurately.

III. Image Mining

Discovering informations from the collected data stored in relational database has been an important work in data mining. The massive collection of image can be mined to discover new information. The main issue of image mining is that it combines the field of content-based retrieval of image, databases and data mining. The image mining process has two phases. The first and the important phase is mining large amount of collected images. The another part is combining the mining of collected data and the corresponding numerical data. Model the image content as a set of blocks, then use any technique to extract the feature. Figure 2 shows the steps in image mining techniques. The collected images are then processed for feature extraction. After extracting the feature from the image mine the content of the image. Evaluate the content of the image with the stored image dataset. The important points to notice on mining images are:

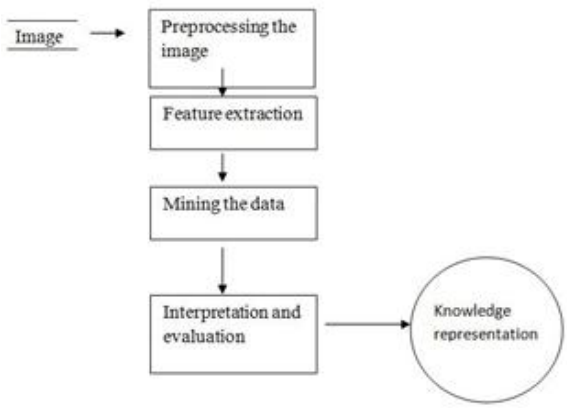

Fig. 2. Image mining [6]

Segment images into identifiable regions.
Compare the segmented image data with the stored dataset. Apply data mining algorithm to generate object of associ- ation rules.

An image is accessed once for the feature extraction. The feature extraction results in image blob and image blob descriptors. The image blobe is stored in a file and the blob descriptors are stored in another file. The blob descriptors are used to identify the object of association rules. Once an image is segmented then it is not necessary to search the image content.

IV. Land-Estimation

An approach to land estimation is to register the Synthetic Aperture Radar (SAR) image acquired by TerraSAR sensor with the geographical map. But these approach is not appli-cable , if map is not available or the updation gets fail due to the natural hazard. Supersolution and Mean-shift algorithm were introduced to solve these issues. K-means is an efficient method used in land mask estimation. Mean-Shift technique is used for optical image pattern recognition. Most of the image filtering techniques used for segmentation were based on Mean-Shift and K-means approach. Because the difference between land and sea are of slight change and the state of sea area is hard to determine [7]. These approach failed in detecting the harbor.

A. Sea-land Segmentation

From the remotely sensed images the land and sea area can be segmented, by using Local Binary Pattern(LBP). LBP is more suitable for remote sensing image processing. For a land pixel, LBP finds out to always zero and to the real land pixel from satellite images, LBP is not zero. It will bring out false alarms [8]. Thus, need a method to reduce the false alarm rate and to segment the land and sea with high precision. The other issues in the existing method to segment land and sea are scalability and the data collected from remote areas are not ready for analysis. LANDSAT 5 satellite images are used for segmenting land and sea area [9]. LANDSAT images do not meet the requirements due to scalability and resolution issues. The data collected from sensors are used to separate the land and sea area. It is more challenging to the coast line extraction and in object detection. The various objects on the ground, make land prediction more complicated than sea area. The hierarchical region merging approach is used to automatically extract the sea and land area [10]. The merging approach can be better characterized by supervised information combined with the feature extraction. This results in bringing out a real-time analytical architecture to detect land and sea area.

B. Data mining technique for sea-land detection

The amount of data rates generated on the digital world is increasing day by day. The existing technologies fails in extracting the sample dataset. It is necessary to design a technique for analyzing both the offline and real-time data. The data mining techniques involves data acquisition, processing and analysis. The collected data are preprocessed. Then the useful data are send to the respective base station for data processing. The filtration and load balancing algorithm filter the collected data and then balance the load. It increases the system efficiency. The sensing satellite uses SPECAN method to convert collected data into image [11]. Filter the useful data from the collected image. In data analysis process calculate mean value and standard deviation of images in each block. Then calculates the maximum sample value. The calculated value for sea area be higher than the land area. The main reason for less mean value is due to the pixel value of land image is lower. Compare the pixel values of collected image with the data already set for land and sea area. It results in the detection of sea area and land area.

Figure 3 shows the remote data process. The remote sensing data unit in data process collects the data from the satellite and then sent to the base station for furthur data process. The data processing are of real-time and offline data process. In offline data process the data are sent to the storage data centre and process the data in future. The storing of incoming real-time data decreases the performance of real-time data process. In real-time data processing the data are directly sent to the filtration and load balancing server. The filtration server only allows the useful information and discarded all unwanted data. The load balancing server divides the filtered data and then sent to the processing server. The load balancing server balances the power of data processing by distributing the real-time data. This server plays an important role in increasing the system efficiency. The processing server process the data and sent result to the aggregration server. Each processing server makes mathematical operations to get the result from the the image segment. The servers do the process in parallel. The aggregration server evaluates the result and then sent to the decision making server. All the servers are supported by algorithms. Finally the data analysis unit in the remote data process provides a decision. The decision making server with the help of decision making algorithm makes the decision and evaluates whether the area is land or sea or forest.

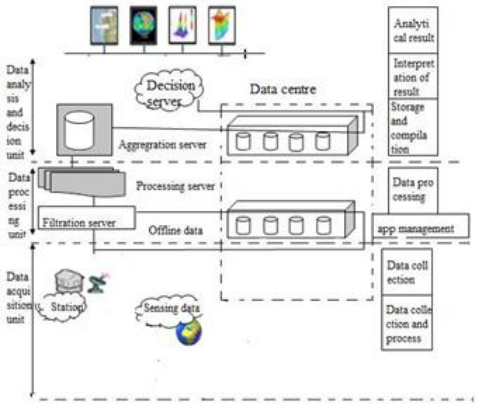

Fig. 3. Remote data process [11]

V. Fire Detection

Forests are an indispensable resource. The occurrence of fire results in increasing damage to the ecosystem. It is necessary to take measures against the forest fire. This requires an efficient method to detect forest fire in real-time monitoring. The Aerial vehicle, ground-based techniques were used for the detection. Moreover, hazardous fire can exploit the life of aerial pilot and the limited range is in ground-based measurement.

A. Routing protocol in fire detection

Researchers find out that the AODV protocol is the best routing protocol used for forest fire detection applications. Energy consumption is the main isssue in AODV protocol even it has best packet deliver ratio. DSDV, LEACH and APTEEN routing protocols are used to overcome the AODV drawback. The cluster node chosen from the sensor nodes for DSDV protocol. Cluster node is chosen dynamically for LEACH and APTEEN routing protocol. Incase of APTEEN cluster node advertise soft threshold value to the sensor nodes. LEACH is not suitable for real-time monitoring. APTEEN(TEEN and LEACH) is enough for monitoring [12]. The APTEEN limits the number of proactive transmissions and for reactive transmission uses threshold values. The main advantage of APTEEN is that it consumes less amount of energy as compared to other routing protocols.

These protocols have lack of security and lack of quality of service in data transmission. Thus, it is necessary to put forward other technique for fire detection.

B. Sensor based fire detection

Temperature and humidity are the main components in fire detection. Set a threshold value for humidity and temperature and check out the sensor readings, if both readings are same generates an alarm indicating the fire is detected. The sensors are deployed to sense the carbon monoxide, humidity, carbon-dioxide and the temperature. The cluster head sensor node collects the data from the corresponding cluster nodes. The control station receives the data from the cluster and the data processing is performed. The main aim of such system is to decrease the fire irrespective of high cost. The slight variations in temperature and dust particles causes the sensors to report fire delay. If a node gets damaged, then the node can interrupt the data process. Hence generates incorrect information.

C. MAC protocol for fire detection

Fire detection applications need high transmission relia-bility. Sensor nodes used for fire detection lacks reliability because of the harsh environment. It is necessary to imple-ment reliable methods. Reliability is described by the number of packets transmitted to the destination during an interval. Implementation of MAC protocols results in limited reliability. Low reliability of MAC protocols were due to the absence of route maintenance. These issues were solved by modifying MAC protocols with the route maintenance. The maintenance are carried out by increasing cost in power consumption and packet delay. Thus, it is necessary to put forward a low- cost and high reliable transmission technique.

D. Visual techniques for fire detection

Visual detection techniques can reduce the fire delay and results in a fire detection accuracy. Visual techniques combine sequential method and color verification technique to make the easy fire detection. Requires a real-time monitoring system to detect the fire with high precision. Satellite images give a real-time information to predict the hazards. Analysis of image by extracting the pixel in real-time results in the prediction of fire detection with the least amount of cost. Camera surveillance systems can be used for the detection of fire. Camera or Optical system can reduce the false information about fire by capturing each movement. Forest-

Watch is a camera system for semiautomatic fire detection. Camera need to install by humans. Capture images within a range, so camera sensor system results in an inefficient detection method. Camera sensors are of high cost. It is necessary to provide a system that detects fire in real-time with less cost.

Visual camera sensors take pictures of forest to detect the fire. A rotating motor is mount on the camera sensor to get ful image of the forest. The collected image are then processed by using MATLAB to detect smoke [13]. The camera sensors need to installl on the communication tower. The tower is placed on a location that faces large areas of the forest.

TABLE I: Comparison Study Of Fire Detection Techniques

Techniques used	Advantage	Limitation
AODV protocol	Best packet ratio for data transmission	Energy consumption
Sensor nodes	Real-time monitoring	Lack of security and reliability
MAC protocol with route maintenanace	Highly scalable	Increase in cost of power consumption
Visual techniques	Real-time monitoring with high precision	Camera sensors are of high cost

The installation of camera sensors on the towers are of high cost. Radio acoustic system is an another method helps in the detection of the forest fire. The radio system consists of radio centre and an acoustic source. The radio system has more sensitivity to change in temperature and provides reliable temperature measurement. The deployment of the radio acoustic sytem in an optimal place is more complex. Table 1 shows the advantages and limitations of various forest fire detection techniques.

Forests can contribute high amounts of economic wealth. So, it is necessary to carry out some approaches to detect and extinguishes fires. Fire causes an environmental damage to human lives. Fire modeling can predict the possible fire behavior without getting burn [14]. Most of the fire detection approaches are based on satellite images. Also, the sensors are deployed in the forest area to check out the temperature. Each sensor nodes are well maintained with a temperature sensor. When there occurs a change in temperature, the sensor nodes send packets to the cluster nodes. When the node gets damaged results in error temperature readings. The image processing techniques are used to overcome the sensor node issues. A rule based color model for fire pixel, which makes use of RGB color space is used to preserve the environment from fire [15]. Since the image processing techniques cannot perform well in the real-time fire monitoring system. The data mining approach uses meteorological data collected from satellite to detect the burned area. Modeled the data with the aim to predict the burned area, may result in the error message or lower accuracy for predicting large fires. The fire creates economical, ecological damage and cause harass to humans. So, it is important to prevent the spreading of fire. The existing method used meteorological data to predict the fire. It results in the prediction of the burned area and is an offline approach [16]. To overcome this prediction of fire, proposed a real-time data analytical architecture for remote sensing. In addition, uses various data mining algorithm for data processing and scale invariant feature transform algorithm (SIFT) for feature extraction of an image.

VI. Conclusion

The data collected from remote sensors are the challenge to the coast-line extraction. The lack of feature extraction in the merging techniques results in the prediction of land and sea area with less precision. Sensor nodes are deployed in the forest and is equipped with a temperature sensor to detect the fire. But when the node gets damaged, it results in error temperature reading and outputs a false alarm. To reduce these issues, used image processing techniques. These techniques may not perform well in real-time monitoring systems. Whereas, the real-time data analytical architecture can analyze both real-time and offline data process. Along with various algorithms, the architecture is also used to detect the land as well as sea area. In addition to real-time data analytical architecture, a scale invariant feature transform algorithm (SIFT) can be used to detect the fire. The occurrence of fire will get notified to the corresponding base station.

References

[1] A. Paul, J. Wu, J.-F. Yang, and J. Jeong,"Gradient-based edge detection for motion estimation in H.264/AVC," , IET Image Processing, vol. 5, no. 4, pp. 323327, Jun. 2011.

[2] 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, April 2014.

[3] Tanuja A, Swetha Ramana , VTU Belgaum, India, "Analyzing Big Data using Hadoop " , International Journal for Innovative Research in Science and Technology, Volume 3, Issue 03, August 2016.

[4] A. Labrinidis and H. V. Jagadish, "Challenges and opportunities with Big Data " , in procedings of 38 th International Conference on Very Large Data Bases Endowment, Turkey, Aug. 2731, 2012, vol. 5, no. 12, pp. 20322033.

[5] https://www.google.co.in/searchbiw=1366bih=662tbm=ischsa=1btnG=Se archqdata+mining+a+KDD+processimgrc=R9E5AVg-KlMNvM.
[6] A. A. Khodaskar; S. A. Ladhake, "Image Mining: An Overview of Current Research" , Fourth International Conference on Communication Systems and Network Technologies, 2014.
[7] J. Martn-de-Nicolas, D. Mata-Moya, P. Jarabo-Amores, N. del Rey-Maestre, J. L. Barcena-Humanes Signal Theory and Communications Department, "Segmentation techniques for land mask estimation in SAR imagery" , International Conference on Computational Intelligence, 2013.
[8] Yu Xia, Shouhong Wan, Peiquan Jin and Lihua Yue, University of Science and Technology of China, "A Novel Sea-Land Segmentation Algorithm Based on Local Binary Patterns for Ship Detection " , International Journal of signal processing and pattern recoginition, 2014.
[9] Paulo Cortez1 and Anibal Morais, University of Minho, Portugal, "A Data Mining Approach to Predict Forest Fires using Meteorological Data " , http://www.dsi.uminho.pt/ pcortez
[10] Dongcai Cheng, Gaofeng Meng and Chunhong Pan, "Sea-land segmen-tation via hierarchical region merge and edge directed graph cut" , IEEE International Conference on Image Processing (ICIP), 2016.
[11] Muhammad Mazhar Ullah Rathore, Anand Paul, Bo-Wei Chen, Bormin Huang, and Wen Ji, "Real-Time Big Data Analytical Architecture for Remote Sensing Application " , IEEE journal of selected topics in applied earth observations and remote sensing, 2015.
[12] V. Devadevan, Bhopal and S. Suresh, SRM University,India, " Energy efficient routing protocol in forest fire detection system " , IEEE 6th International Conference on Advanced Computing, 2016.
[13] Michel Owayjan, George Freiha, Roger Achkar, Elie Abdo, Samy Mallah American University of Science and Technology, "Firxio-forest fire detection and alerting system" , 17 th meditteranean international conference, 2014.
[14] Qasim Siddque "Survey of Forest fire simulation" , Global journal of computer science and technology, 1987, Qasim1987@hotmail.com
[15] Vipin V, Asst. Professor, St.Josephs College of Engineering, Palai, India"Image Processing Based Forest Fire Detection" , International Journal of Emerging Technology and Advanced Engineering, February 2012.
[16] Michael Kostiuk, MA Geography Geospatial Analyst, Canada, "using remote sensing data to detect sea level change" , International conference on ecora 15/Land Satellite Information IV/ISPRS Commission I/FIEOS, 2002.