Journal of Systems Architecture 64 (2016) 122–132

Contents lists available at ScienceDirect

Journal of Systems Architecture

journal homepage: www.elsevier.com/locate/sysarc

Real-time continuous feature extraction in large size satellite images

M. Mazhar U Rathore[a], Awais Ahmad[a], Anand Paul[a,*], Jiaji Wu[b]

[a] The School of Computer Science and Engineering, Kyungpook National University, Daegu 702-701, Republic of Korea
[b] School of Electrical Engineering, Xidian University, China

ARTICLE INFO

Article history:
Received 29 July 2015
Revised 15 October 2015
Accepted 10 November 2015
Available online 22 November 2015

Keywords:
Remote sensing
Image processing
Feature extraction

ABSTRACT

Remotely sensed imagery is being increasingly used for the development of the earth observation satellites to investigate human activities, to monitor environmental changes and to update existing geospatial data. The ordinary pictures are difficult to process automatically by computers but can be easily interpreted by humans. The most significant step is how to get anticipated information from the images and how to convert these images into useful data for further studies. The key objective is to satisfy an algorithm claiming to be efficient in large size image processing includ enhanced processing efficiency, finding correlation among data, and extracting continuous features. To achieve these objectives in the setting mentioned above, we propose a real-time approach for continuous feature extraction and detection in remote sensory earth observatory satellite images to find rivers, roads, and main highways. Deep analysis is made on the ENVISAT satellite missions datasets and based on this analysis the algorithm is proposed using statistical measurements, RepTree machine learning classifier, and Euclidean distance. The system is developed using Hadoop ecosystem to improve the efficiency of the system. The designed system consists of various steps including collection, filtration, load balancing, processing, merging, and interpretation. The system is implemented on Apache Hadoop system using MapReduce programming with higher efficiency results in a massive volume of satellite ASAR/ ENVISAT mission datasets.

1. Introduction

Digital image processing is becoming a hot topic these days because of its various applications in security, medical healthcare, agriculture, entertainment and fun, area monitoring, etc. Digital image processing is the use of computer algorithms on digital images to perform image processing. This technology is widely used for the image morphology, feature extraction, segmentation, rendering, and pattern recognition [1–4] and many other digital image operations. Various research also works on image processing aspect of H264/AVC [5–9] such as, in edge detection, deblocking filter, and motion estimation in H264/AVC. Moreover, Feature extraction is the most widely used part of the image processing that can be used for many application such as, security and authentication, object detection, and pattern matching, etc. In practice, two types of feature extraction (feature selection) methods are used, i.e., type I and type II. Type I feature extraction methods mainly focus on the finding of original parameters from the scratch for feature extraction while type II feature extraction method is used to optimize the accuracy of a feature set by removing inconsistent features [10] by given set of features. Also, Type II also used to discover a subset of features associated with optimal identification accuracy [11]. Simpson et al. do well at this in the article Genetic & Evolutionary Type II feature extraction for periocular-based biometric recognition [12].

Remotely sensed imagery is being increasingly used for the development of the earth observation satellites to investigate human activities, to monitor environmental changes and to update existing geospatial data [1,13]. The ordinary pictures are difficult to process automatically by computers but can be easily interpreted by humans. The most significant step is how to get anticipated information from the images and how to convert these images into useful data for different further studies. Moreover, the processing of larger size images or large datasets of thousands of satalite images in an efficient manner is also a key challenge [14].

The continuous features extraction such as roads, river, and highways detection through satellite image is very valuable and efficient for most of the urban planning application. Very few work has been done in the field of continuous natured feature extraction using satellite image processing. The painted lane markings that exist in the most urban roads, in campus sites or in the comparable environments of the theme parks, industrial estates and science parks may not be easily discernible by closed-circuit television (CCTV) cameras because of bad weather conditions, poor lighting and insufficient

* Corresponding author. Tel.: +82 539507547.
E-mail addresses: rathoremazhar@gmail.com (M.M.U. Rathore), awais.ahmad@live.com (A. Ahmad), paul.editor@gmail.com (A. Paul), wujj@mail.xidian.edu.cn (J. Wu).

http://dx.doi.org/10.1016/j.sysarc.2015.11.006
1383-7621/

maintenance. Similar is the case with the river as well. The existence of pavements or curbs is the important feature of roads or rivers on either side defining the boundaries. For the implementation of autonomous navigation or driver assistance systems, the curbs that are parallel to the roads can be harnessed to extort useful features of the roads.

Due to the fact that the use of vision image data is a difficult task for the extraction of the curbs or features of the road edge as curbs are not perceptible in the vision image. Favorable and heuristic lighting and extensive image processing requires to extract the curbs from the camera image. A laser range measurement system is one of the favorable for obstacle detection and depth range measurement under poor lighting, bad weather condition with its best features of the low cast of an alternative to millimeter wave radar system. The significant rise has been observed in the use of laser range measurement system for an autonomous navigation task in the past several years [15–22]. However, the major domain of their use has been in indoor environments [17–22]. Laser range measurements systems have found some of the common tasks of obstacle detection [15,16], map building [21,22], navigation [17,18], and localization [19,20].

However, to keep the properties of rivers in mind, they are long in length and geometrically smooth. These particular attributes can give advantages to most algorithm to construct a river network. River finding, river tracking, and river linking are three typical stages of river extraction. To search the potential river pixels, this methodology is set with in a river window. When creating consecutive river points, the local properties like magnitude and direction are accounted. Continuous and smooth groups of river seeds are linked together to produce different lengths of segments once the river points are found. Finally, the river segments with longer length are selected as a piece of river in the river linkage stage to form a river network.

Therefore, based on the aforementioend needs, this paper presents an efficient mechanism that detects the continuous features in the images (such as river) using statistical computations, Euclidean distance, and machine learning approaches. To gain the more efficiency of the system, the system is implemented on the parallel environment of Hadoop server. The Hadoop has distributed file system, i.e., HDFS and distributed programming language MapReduce, which have the capability to process large size and a large amount of images using parallel tasking on the same dataset. Moreover, the proposed system divides the whole process into various steps to increase the efficiency of the detection mechanism, which includes collection and filtration, segmentation, processing, and merging.

The rest of the paper describes the background and related work in Section 2. Section 3 demonstrates the details of the datasets used for analysis and tested. Section 4 presented the analysis and discussion based on which the proposed system is developed. The proposed system details are given in Section 5. While the evaluation is done in Section 6. Finally, the conclusion is made in Section 7.

2. Background and related work

Remote sensing technology has opened a new way of the data collection era. Automated image processing has reduced human labor and became a desired outcome to increase the efficiency of extracting information. Roads are the one of the most critical components of the landscape while considering continuous feature extraction. That is why automated road extraction from remotely sensed imaginary has become a vigorous research topic.

In the past two decades, a variety of road extractions approaches has been explored in which most of them were developed using panchromatic images. Some reviews are done by Zlotnick et al. [23] and Xiong [24]. There are two fundamental principles: local and global strategies on which the approaches are based. For the local strategy, by means of examining the neighborhood pixels surrounding the target pixel, each pixel is separated by 'road' or 'background' pixel mainly. The edge enhancement techniques are the most popular techniques to find the road edges. For the global strategy, the particular characteristics of the roads are taken into account to filter the results from a local edge enhancement.

The local strategy is the first step that corresponds to the road finding the stage by seeking the road candidates. The road gray values will show a peak or valley shape when roads crossing from one side to another side. The roads are linear features on the image with a certain width. Therefore, morphological operations or particular designed filters, conventional edge filters by Fischler et al. [25] and Geman et al. [26,27] are working to detect potential road candidate.

Though, the roads always systematically appear as the surrounding background, whether it is brighter or darker. Some of the factors may affect the road intensity on the images like atmospheric conditions, sun angle and background structure. To find the road points, the more complex hypothesis of the road attributes along with criteria are considered to improve the road finding procedures as with the only local filters there may produce undesired points or segments. Road seeds are used in the follow on road tracking and the road linking processing after the road finding the step.

Although, some of the research has been done by a various researcher in continous feature detection such as roads and rivers. Gruen et al. used dynamic programming for road extraction from aerial and satellite images [28]. Veit and his research companions [29] used systematic approach to evaluate algorithms for extracting road marking features from images. Similarly, in paper [30], Broggi proposed a system for the road boundaries extraction from various images taken in an out-of-town environment. In addition, some of the limited work is also done with respect to river features [31,32]. Moreover, related concept is also found in [33–36] which also gives a picture of feature extraction in videos as well. However, these mechanisms are not suitable for processing large images such as satalite images or large images dataset containing thousands of images. Moroever, most of the existing workings are not particularly developed for all type of continuous feature extraction, such as River detection. On the other hand, the satellite product or images are an essential part of such type of feature extraction. Therefore, in this paper, we are proposing an efficient mechanism to extract the continuous features, taking Rivers as a use case, to overcome the efficiency limitations of previous feature extraction techniques. Because of the parallel processing nature of Hadoop ecosystem, we use Hadoop server with MapReduce programing to gain efficiency while working on large image or images dataset. The larger size images are efficiently processed by our system by dividing the image into blocks and processing them in parallel. The higher accuracy is achieved by applying statistical methods, machine learning (REPTree classifier), and Eulidean distance. The rest of the section describes the whole details of the system.

3. Dataset and tools used for analysis and evaluation

Datasets are taken from European Space Agency (ESA) [37] for analysis and testing that contain various earth observatory satellite products by monitoring different locations on earth. Two main satellite sensors' data, i.e., Advanced Synthetic Apertures Radar (ASAR) and medium resolution imaging spectrometer (MERIS), of ENVISAT mission, is taken for analysis as shown in Table 1. ENVISAT was working and monitoring Earth from approximately 800 km above the surface [38]. Different types of products that are subjected to the area covered are examined, such as, Sea area, Land area, Ice area, etc. as shown in Fig. 3. ESA monitored products contains satellite image data of various countries, such as, European Countries i.e., Italy, Greece, Spain, Morocco, Poland, Canada, African countries i.e., South Africa, Mauritania, etc. and USA as well. In the mentioned Figure, Product10 covered the area of Ice, Land, and Sea from Canada, Product 7 contains the data from the Sea and Land area in between of Spain and Morocco,

Table 1
Datasets details.

Mode	Mission/sensor	Capturing date	Area overed	Country	Absolute orbit /phase.cycle
Product 1. ASA_APM_1PNPDE20091007_025628_000000432083_00118_39751_9244					
Size: 23MB	**mode:** AP mode medium resolution image	**capturing date:** 07-OCT-2009 2:56:29.3974			
Area covered: Sea and Land	**Country:** Vietnam	**Absolute Orbit /Phase.Cycle:** *39751/2.83*			
Product 2. ASA_APM_1PXPDE20020819_093008_000000622008_00394_02452_0009					
Size: 33MB	**mode:** AP mode brows image	**capturing date:** 8/19/2002 9:30:08			
Area covered: LAND	**Country:** Poland and Germany	**Absolute Orbit /Phase.Cycle:** 2452/*2.08*			
Product 3. ASA_GM1_1PNPDE20100415_224615_000004102088_00345_42483_4425					
Size: 9.4MB	**mode:** GM mode image	**capturing date:** 15-APR-2010 22:46:21.294			
Area covered: Sea and Land (Forest, Desert)	**Country:** Western Sahara, Mauritania				
Absolute Orbit /Phase.Cycle: 42483/*2.88*					
Product 4. ASA_WSM_1PNDPA20050331_075939_000000552036_00035_16121_0775					
Size: 55MB	**mode:** Wide swath mode image	**capturing date:** 31-MAR-2005 07:59:36.4091			
Area covered: Sea and Land	**Country:** Cape town, South Africa				
Absolute Orbit/Phase.Cycle: *16121/2.36*					
Product 5. ASA_WSM_1PXPDE20021117_104431_000000672011_00180_03741_0009					
Size: 67MB	**mode:** Wide swath mode image	**capturing date:** 17-NOV-2002 12:58:52.00			
Area covered: Sea and Land	**Country:** Spain	**absolute orbit /Phase.Cycle:** 3741/*2.11*			
Product 6. ASA_APS_1PXPDE20020819_093043_000000072008_00394_02452_0000					
Size: 497MB	**mode:** AP Mode SLS image	**capturing date:** 19-Aug-2002 09:30:43			
Area covered: Land	**Country:** Poland	**Absolute Orbit /Phase.Cycle:** 2452/*2.8*			
Product 7. MER_FR__1PNUPA20030723_105132_000000982018_00223_07291_0388					
Size: 166MB	**mode:** Not obvious	**capturing date:** 23-July-2003 10:45:43			
Area covered: Land and Sea	**Country:** Spain and Morocco	**Absolute Orbit /Phase.Cycle:** 7291/*2.18*			
Product 8. MER_FRS_1PNPDE20060822_092058_000001972050_00308_23408_0077					
Size: 663MB	**mode:** Not obvious	**capturing date:** 23-April-2010 07:38:17			
Area covered: Land and Sea	**Country:** Tunisia, Libya, Greece, Italy				
Absolute Orbit /Phase.Cycle: 23408/*2.50*					
Product 9. MER_RR__1PNPDK20030813_175754_000026132019_00027_07596_4557					
Size: 42MB	**mode:** Not obvious	**capturing date:** 13-August-2003 18:13:39			
Area covered: Land and Sea	**Country:** USA: 7596/*2.19*	**Absolute Orbit /Phase.Cycle**			
Product 10. MER_RR__1PNRAL20100426_154828_000003662088_00498_42636_0001					
Size: 78MB	**ode:** Not obvious	**capturing date:** 26-April-2010 15:48:28			
Area covered: Land, Ice, Sea	**Country:** Canada	**Absolute Orbit /Phase.Cycle:** 42636/*2.88*			

whereas, Product 9 and product 1 is from USA and Vietnam. Table 1 shows detailed information about the datasets used in our work, such as, the product name, image mode, mission, capturing date and time; the area covered, monitored country, size in MB, the absolute orbit of the satellite, the phase and cycle as well. Finally, all the products of almost 1.7GB size combine to test the system on larger datasets.

EnviView, Beam, and Nest [39] are three popular tools that provide visualization and understanding of ESA Earth Observatory (EO) products. While understanding and performing a basic analysis of products, we use EnviView 2.8.1, Beam VISAT-5.1, and Beam Nest 5.1. For our complex analysis, Hadoop provides an efficient solution through parallel programming and divide-and-conquer facilities [40]. Hadoop 2.3.0 with Map Reduce Java programming is used for algorithm development using divide and conquer mechanism.

We developed and test the proposed algorithm to extract the features of the river on corei5 3.20 GHz $\times$ 4, UBUNTU 14.04 local machine with Hadoop single node setup having 4GB RAM and Gallium 0.4 on AMD OLAND graphics.

4. Image analysis for continues feature extraction

The main focus of the analysis is on ENVISAT/ASAR EO products especially Product1 since ASAR Product1 has more and diverse nature of Rivers as well as diverse covered areas, such as, Sea, and small lakes, city, etc. Initially, the satellite image data is taken from Measurement Dataset (MDS) portion of the product. Keeping in view the continuous behavior of the Rivers, statistical analysis, and pixel value distribution is made for exploring the properties, pattern and behavior of Rivers in the satellite image. We calculated the overall statistical measurements of the products, such as, the mean value of all the pixel values, the diversity and variation in the values to find out the nature and quality of the image data. While exploring overall statistics of various products, we identified that the Product2 has more records since the image quality might be lower due to its overall low mean and standard deviation values. It might also be the case that it covers some of the dark areas, such as Forest.

In the next stage of analysis, we explore pixel view of different image blocks having the different area covered by considering Amplititude_HH band of products. We observed that all the river's pixels look similar and having a continuous nature as shown in continuous black pixels in Fig. 1(c and d). Land image blocks with no river have different pixel view than pixel view of the block with the river as shown in Fig. 1(a) and (c). Since the pixel values of the block, with no river, are quite similar (low S.D), therefore, the difference between the mean value of the pixels of the block and the mean value of the pixels of any sub block of that block is very low.

Pixel value distribution of several image blocks, such as, Land block with no River, Sea block, and Land block with one river, and two blocks that have only one river are also inspected. The distribution among pixel values is quite low for all river pixels as shown in Fig. 2(a) and (b). Only a few pixels in those blocks have the values above 600 and below 400, which results in a minimum mean value. The distribution of image blocks having no River, either Land or Sea block, is entirely dissimilar from River block as shown in Fig. 2(c and d). Pixel values range from 1000 to more than 3000 in case of Land block with no River and from 2700 to 3500 for Sea block. Land blocks that have Rivers are also examined on pixel value distribution. The mean value and pixel distribution of the River portion in the Land block are different from the other part of the block from pixel 61 to 121 as shown in Fig. 2(c). It is also apparent that the pixel values of the River portion have minimum difference among themselves as compare to the other portion of the image.

Considering the fact of greater mean difference between River pixel values and overall block pixel values, the analysis have been performed on mean difference between blocks and sub blocks by considering 10,000 pixels image blocks, which also have one or more River

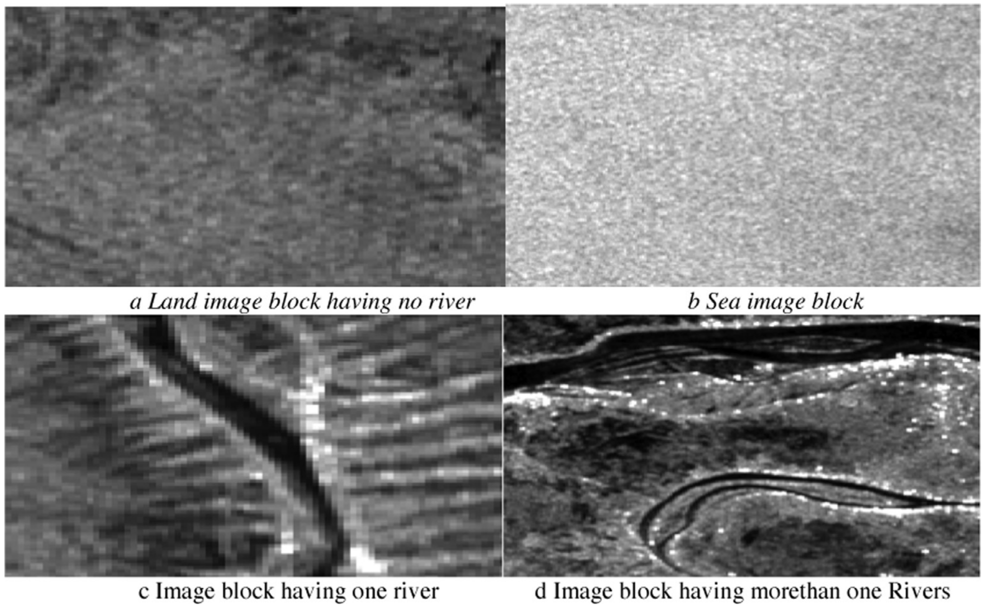


Fig. 1. Pixel view of different image blocks.

Table 2
Statistical measurements of river classified blocks by SVM in Product 1.

Block	Block size (no. of values)	No. of rivers in block	Mean value of rivers pixels Mean_R	Mean value of the overall block Mean_B	|Mean_B−Mean_R|
1	10,000	2	541	2125	1584
2	10,000	1	451	1615	1164
3	10,000	1	3850	2458	1392
4	10,000	1	950	1459	509
5	10,000	2	639	1843	1204

in the blocks. Table 2 clearly shows that the absolute difference between the mean value of river sub block and the mean value of the complete block is very high as shown in the last column of the table. The difference for block 4 is bit lower as compare to other blocks due to river thickness and its existence in the populated city area.

5. Proposed system

Based on the exploration and analysis made in the previous section on earth observatory images, a system is proposed to extract the continuous features, especially rivers exist in the earth observatory satellite images. The proposed system includes the complete implementation model and the algorithm. The proposed system contains various phases including data collection, filtration, segmentation, processing, merging, and the interpretation. The implementation model is depicted in Fig. 3, which has various units for each phase. The first unit is the collection unit that handles collecting and aggregation data. Filtration unit filters the necessary information from datasets and sends it to the Hadoop ecosystem. Next is the Hadoop ecosystem unit, which is the core unit of the system, responsible for the key processing of data. Finally, the interpretation unit uses the results generated by Hadoop ecosystem to extract the continuous features from the satellite image.

5.1. Collection unit

This is the collection point or the ground station of the satellite that receives data from the space station. It does some preprocessing to remove noise from the data that occurs due to the air pollution gases and other interferences in space. Normally the data collected from remote areas are not to be considered in a format for the analysis. Therefore, it also converts the satellite received data into a proper structure form, which can be understandable and can be processed by the normal computer machines. The proper structure of the product captured from Envisat mission is shown in Fig. 4. There are lot sub datasets in a single satellite product. Some of them are:

(1) Main Product Header (MPH): includes the products basis information, i.e., id, measurement and sensing time, orbit, information, etc.
(2) Special Products Head (SPH): contains information specific to each product or product group, i.e., the number of datasets Descriptors (DSD), a directory of remaining datasets in the file, etc.
(3) Annotation Datasets (ADS): contains information of quality; time tagged processing parameters, geo location tie points, solar, angles, etc.
(4) Global Annotation Datasets (GADs): contains calling factors, offsets, and calibration information, etc.
(5) Measurement Dataset (MDS): contains measurements or graphical parameters calculated from the measurement including quality flag and the time tag measurement as well. The image data is also stored in this part and is the main element of our analysis.

5.2. Filtration

The filtration is the second step to drag out the useful information from the fundamental resources to achieve more efficiency. Some of the data might be imprecise and we have to deal with it too that is far away from reality. As we are more interested in image part of the

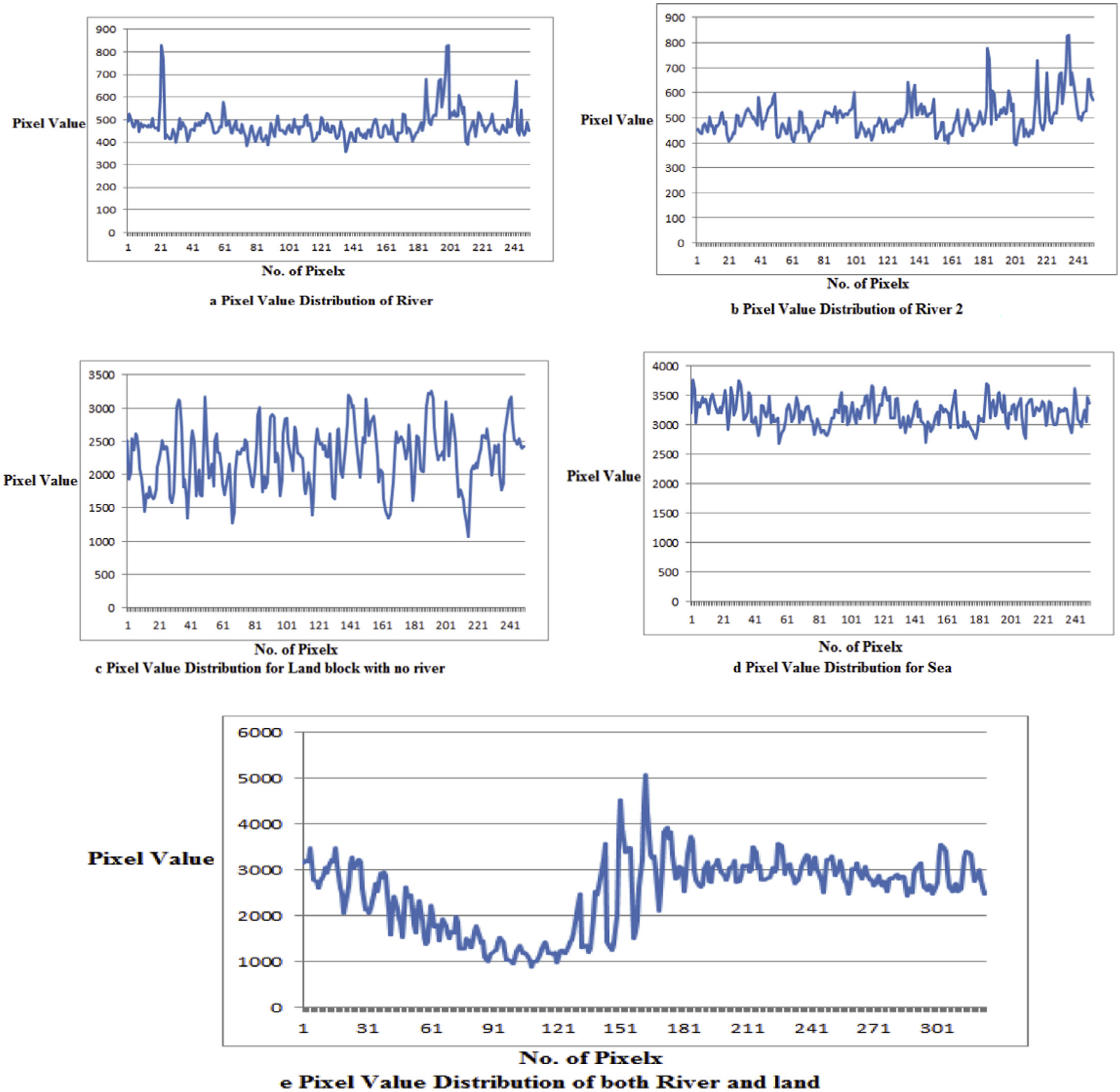


Fig. 2. Pixel value distribution for different image blocks.

satellite product and in Envisat ASAR satellite product, the MDS contains the image data, therefore, filtration unit only filters MDS dataset from the whole product, as highlighted in Fig. 4. You can perceive, a small portion of the whole product is filtered out, all other unnecessary data is discarded, results in increasing efficiency.

5.3. Hadoop ecosystem

Hadoop ecosystem has the ability to process very large datasets in an efficient manner. It has the distributed programming language called MapReduce to process the data stored in distributed Hadoop data nodes. Hadoop also has distributed file system called HDFS, which stored large dataset on to various data nodes by dividing it into chunks to achieve parallelism. This distributed processing and distributed storage make it possible to process very large images in a near real-time manner. Here, in the proposed system, we do all main processing including segmentation, computations, and result in merging by Hadoop. At segmentation step, the whole image is segmented into blocks so that these blocks can be processed in parallel to generate results. In our river detection scenario, the image blocks are simply dividing into fixed size of blocks, as shown in Fig. 5. Each block is represented by the subscripts i and j where i represents the row and j represent the column number of the block. Each segmented block is then processed by computing statistical parameters and machine learning algorithms to extract water features in each block. Fig. 6 shows the image matrix, which is divided into blocks and then each block is divided into sub blocks. The gray color represents the sub block, which has water, means it might be the part of some

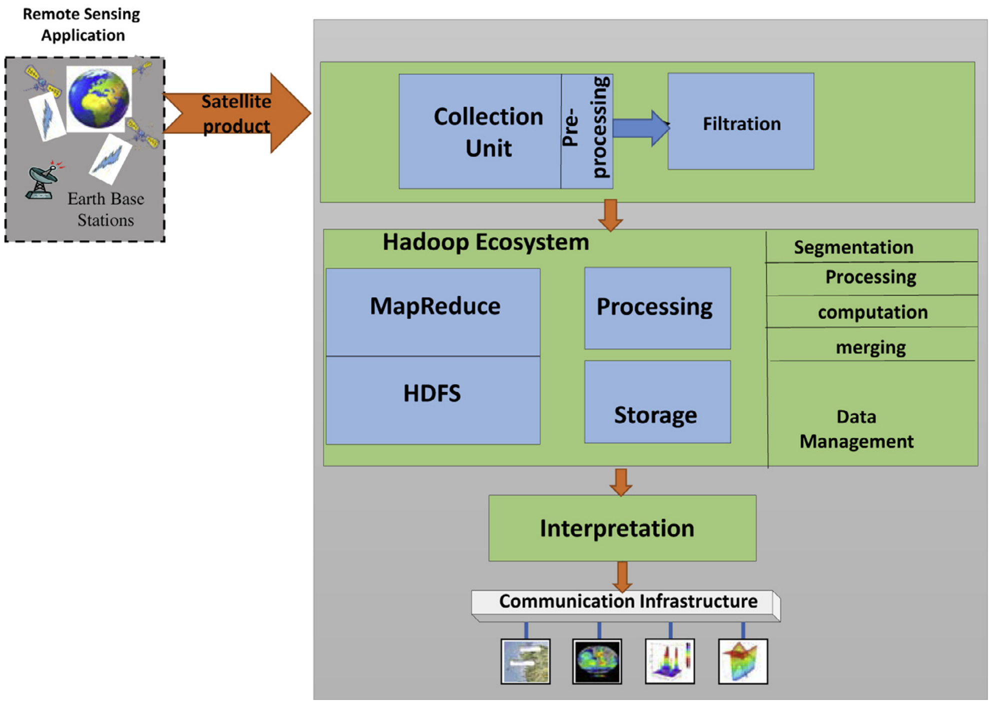


Fig. 3. Implementation model.

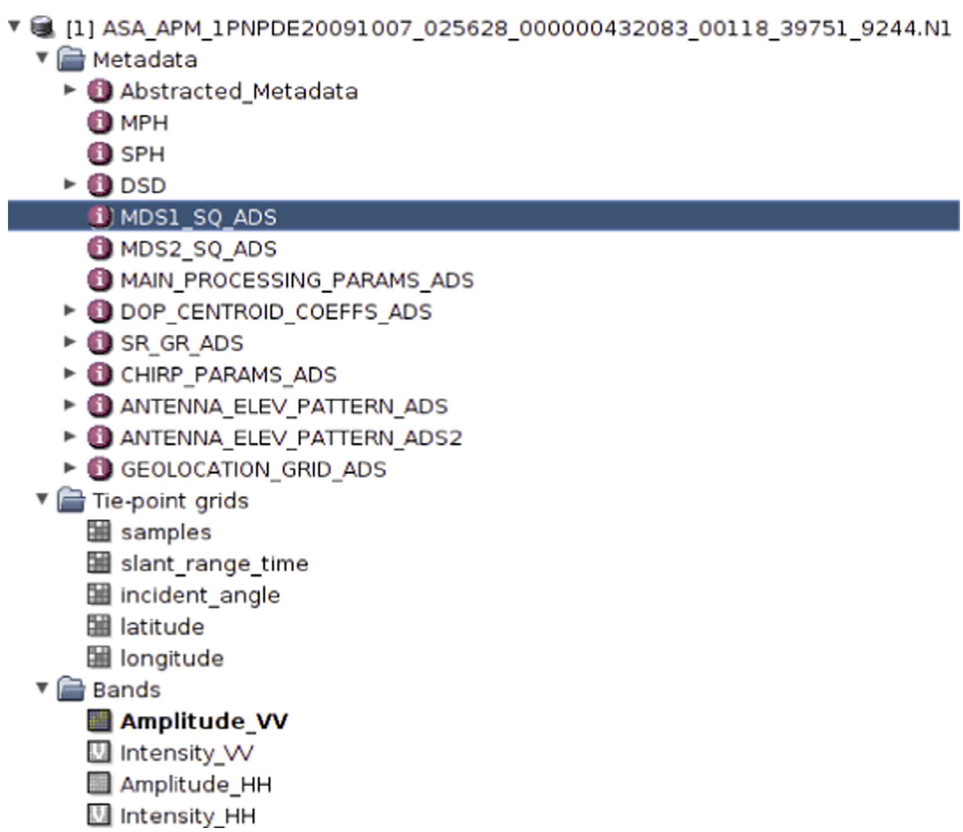


Fig. 4. Envisat generated dataset structure.

river. Later, the blocks are merged by applying proposed technique to combine the water blocks to make one or more river, as shown in the Fig. 6, the river 1 is detected after merging the first block and the block that is beneath that first block. The merging is quite similar to conquer technique in which the blocks are combined and results are conquered to extract the overall river in the images.

5.4. Interpretation

This is the last and optional unit of the implementation model. It takes results from Hadoop ecosystem and visualize them for some user applications, such as marking rivers on the map or drawing some graphs, etc.

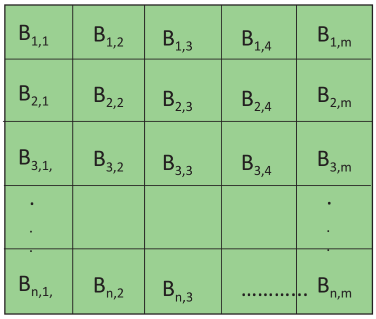


Fig. 5. Image simple segmentation.

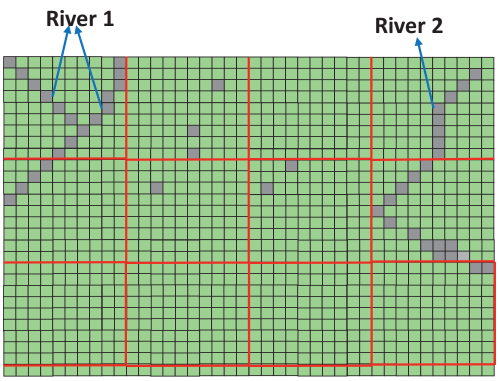


Fig. 6. Merging block results.

5.5. Proposed algorithm for continuous feature extraction as continuous river

Based on the pixel value analysis made in Section 4, the algorithm is proposed, which takes satellite product as an input and detect river or any other continuous feature from that product in the form of image. The algorithm uses various parameters. All symbols and parameters used in the algorithm are described in Table 3. However, most important parameters are mean, variance and Euclidean distance and important operation is the identification of water sub block using machine learning techniques i.e., REPTree. Mean and standard deviation ($\bar{X}_{B_{i,j}}$_ $B_{i,j}$ and S.D_ $B_{i,j}$) of pixel values of block and calculated by Eq. (1) and Eq. (2), respectively. Euclidian distance (ED) between two pixels position is calculated by Eq. (3).

$$\bar{X}_{B_{i,j}} = \frac{\sum \text{Pixl values in block } B}{\text{No. of pixels}} \tag{1}$$

$$S.D_{Bi,j} = \sqrt{\frac{\sum \left(\text{Pixel value}i - \bar{X}_{Bi,j}\right)^2}{\text{No. of pixels}}} \tag{2}$$

$$\text{ED (P1, P2)} = \sqrt{(\boldsymbol{P_1x} - \boldsymbol{P_2x})^2 + (\boldsymbol{P_1y} - \boldsymbol{P_2y})^2} \tag{3}$$

where p_1:pixel1 position = (P_1x,P_1y) and p_2: pixel2 position= (P_2x, P_2y)

Moreover, regression tress logic is used by REPTree to create multiple trees in different iterations. Then its select one of the best three among all generated tree, which will be a representative to all trees. The mean square error on the predictions made by the tree is the key measure that is used to prune the tree. REPTree is a fast decision tree learner, which builds a decision/regression tree using information gain as the splitting criterion and prunes it using Reduced Error Pruning. It only sorts values for numeric attributes once. We preferred RepTree due to its better accuracy and efficiency as compare to other machine learning classifiers.

The algorithms take the Envisat satellite product, which also includes satellite image data. At first level its extract the image data from MDs and discard all other information. Its stores the MDS extracted values into image matrix. In next phase, the whole image matrix is divided into equal blocks. All image blocks are individually processed to find the rivers in each block. Later, each block result is merged and concatenated. Each block is initially tested for the possibility of the river in that block by comparing the pixels variance with the threshold. If the block has the possibility of the river, then it is divided into smaller sub blocks to check whether the each sub block contains water of a river. To check the water sub block, the REPTree machine learning algorithm is sued by supplying four parameters as described in the algorithm1 step 3(iv). Later the water blocks and the rivers from the neighbor block next to the current block are merged by checking the Euclidian distance (ED) between rivers of different blocks or the Euclidian difference between the water sub blocks. Finally, all the results from all blocks are merged based on ED and final set is accomplished called river set. The algorithm 1 is the pseudocode of the prosed algorithm.

6. Evaluation

We evaluate our system with respect to the algorithm computational complexity, accuracy, and the most important, the efficiency with respect to average processing time and throughput.

6.1. Complexity of the algorithm

The complexity of the proposed algorithm is O(mn), calculated in big O notation as follows:

The whole image is divided into m x n number of blocks. Where m is the number of blocks per row of the image and n number of blocks per column of the image

For step 3 (iv), complexity is (m x $\mathbf{n}$) / (10 × 10)
For step 3 (v) the complexity is some number K, where $K < m$ && $K < n$
Similarly, for step 3 (vi), complexity is L, where $L < m$ && $L < n$
For step 4, it is (m x n)
Therefore over all complexity $= C = (mn/100) + K + L + mn = (mn + 100mn + 100K + 100L)/100$
$C = (mn + 100mn + 100N + 100M)/100$
for extreme level as $L < m$ && $L < n$ && $K < m$ && $K < n$
$C = O\,(mn)$

Now, we have understood that the complexity of the proposed algorithm is O (mn), which is too lesser from the complexity O(m^2n^2), which is the complexity of the continuous feature extraction by normal comparison of the pixel without dividing image into blocks and sub blocks.

6.2. Accuracy

The proposed algorithm implementation is based on simple Java programming using iteration mechanism as well as Hadoop (Map Reduce). The algorithms are executed on ASAR and MERIS products for correctness and processing time measurements. The accuracy is measured in terms of number of detected rivers. It detects four rivers from Product 1, two rivers from Product2, two rivers from Product3, three rivers from Product4, and two rivers from Product 5. The detection mechanism could be improved depending upon the satellite image quality and its image taking height. The satellite image of Product1 and the sample river detected image are shown in Fig. 7.

6.3. Efficiency

The continuous feature extraction is implemented using simple Java iteration mechanism and as well as Hadoop divide and merging

Table 3
Parameters and symbols used in algorithm1.

Symbols	Description	Symbols	Description
MDS	Measurement dataset	$B_{i,j}$	image Block *i,j*. Where *i* represents the row and *j* represents the column number.
SP	Satellite product: dataset	$var_{B_{i,j}}$	the variance of pixel values of block $B_{i,j}$
Blk_S	Block size	$\partial_{\mathbf{Min\_RB\_var}}$	threshold define for minimum variation for river block
$\bar{X}_{B_{i,j}}$	Mean of pixel values of block $B_{i,j}$	$River\_Set_{B_{i,j}}$	set which contains the river detected within a block $B_{i,j}$
$\bar{\mathbf{X}}$_Sub_B_k,	Mean of pixel values of sub block B_k	$\partial_{\mathbf{ED\_SB}}$	threshold define for Euclidian distance between water blocks to find continuity for river detection
N_Pix	Number of pixcels	$\partial_{\mathbf{NP\_SB}}$	threshold defined for the number of pixels in river
		var_Sub_B_k,	the variance of pixcel values of sub block B_k

Algorithm 1: Continuous feature extraction in satallite images.

Input: satellite product (Envisat Captured Product) SP
Output: set of rivers in product
Steps:
//Filtration: filter image related data i.e. processed data in MDS. All other unnecessary data will be discarded.
1. Filter (MDS dataset from SP)
// Image segmentation
2. Divide the image into fixed size block i.e. Blk_S = 100 × 100 MDS process_data values, row by row fashion or column by column. Each block will be denoted by $B_{i,j}$ where *i* represents the row and *j* represents the column.
//Processing
3. FOR EACH$_1$ block ($B_{i,j}$) DO
4. Calculate $\bar{X}_{B_{i,j}}$ $var_{B_{i,j}}$
5. IF $var_{B_{i,j}} < \partial_{\mathbf{Min\_RB\_var}}$ THEN // Block does not have any river.
6. $River\_Set_{B_{i,j}} = \Phi$
7. Next $\leftarrow B_{i,j}$ // break
8. END IF
9. Divide $B_{i,j}$ into 10 × 10 sub blocks Sub_B_k, where *k* represents the sub block number in $B_{i,j}$
10. FOR EACH$_2$ sub block: Sub_B_k DO
11. Calculate $\bar{X}$_Sub_B_k, var_ Sub_B_k, |$\bar{X}_{B_{i,j}}$ - $\overline{X}$_Sub_B_k|, |$\bar{X}_{B_{i,j}}$ - var_ Sub_B_k|
12. IF REPTree($\bar{X}$_Sub_B_k, var_ Sub_B_k, |$\bar{X}_{B_{i,j}}$ -$\bar{X}$_Sub_B_k|, |$\bar{X}_{B_{i,j}}$ - var_ Sub_B_k|)== water THEN
13. Water_Set = Water_Set + Sub_B_k
14. END IF
15. END FOR EACH$_2$
16. FOR EACH$_3$ sub block Sub_Bp && Sub_Bq in Water_Set DO
17. IF ED (Sub_Bp, Sub_Bq) < ∂_{ED_SB} THEN
18. MERGE Sub_Bp, Sub_Bq
19. UPDATE Water_Set
20. END IF
21. END FOR EACH$_3$
22. FOR EACH$_4$ sub block Sub_Bz in Water_Set DO
23. IF (N_Pix < ∂_{NP_SB}) THEN
24. DELETE Sub_Bz from Water_Set
25. UPDATE Water_Set
26. END IF
27. END FOR EACH$_4$
28. $River\_Set_{B_{i,j}}$ = Water_Set
29. END FOR EACH$_1$
30. FOR EACH$_5$ $River\_Set_{B_{i,j}}$ DO
31. FOR EACH$_6$ element R in $River\_Set_{B_{i,j}}$ DO
32. Set_NR = EXTRACT_RIVER ($River\_Set_{B_{i,j+1}}$, $River\_Set_{B_{i+1,j+1}}$, $River\_Set_{B_{i+1,j}}$, $River\_Set_{B_{i+1,j-i}}$) //extracting rivers from all neighbor blocks.
33. END FOR EACH$_6$.
34. FOR EACH$_7$ River S in Set_NR and River R in $River\_Set_{B_{i,j}}$ DO
35. IF ED(R, S) < ∂_{ED_BR} THEN
36. MERGE R and S
37. UPDATE $River\_Set_{B_{i,j}}$
38. END FOR EACH$_7$
39. END FOR EACH$_5$

mechanism. The river detection mechanism is implemented without using blocks division-merging mechanism to highlight the efficiency effects while implementing the system using divide-merge mechanism and Hadoop ecosystem. The implementation of the proposed algorithm using Map Reduce divide-and-merging mechanism is more efficient than simple Java iteration implementation due to its divide and conquer nature as shown in Fig. 8. The bars in the graph show average processing time in seconds to process 1MB data from various ESA products. The blue color bars shows the average time consumed in seconds to process 1MB of ESA products by simple Java iteration, while red color bars represents average time taken in seconds to process 1MB of ESA products by Map Reduce implementation on single node Hadoop. Results show that the Map Reduce implementation takes the only half second average time to process 1MB of image data for most of the products. We also observed that simple Java iteration takes more time to process ESA products than Map Reduce implementation. Moreover, ASA-APS products take longer time for its process than other products (i.e., for both simple Java iteration implementation and Map Reduce implementation). ASA-WSM products are processed more efficiently than other APS, APM, FRS, and RR mode products. The processing time varies from product to product due to variation in image bands, and different image modes depending on the product type. ASA-APS product size is larger than other product, so the processing time is exponentially increases due to the multiple nested iteration mechanism in the Java iteration implementation and more division, conquering, and looping in Hadoop implementation.

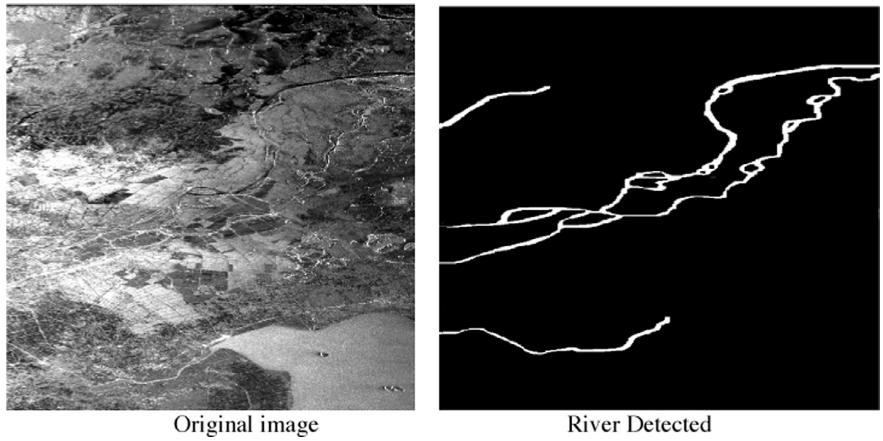


Fig. 7. River detection in product 1.

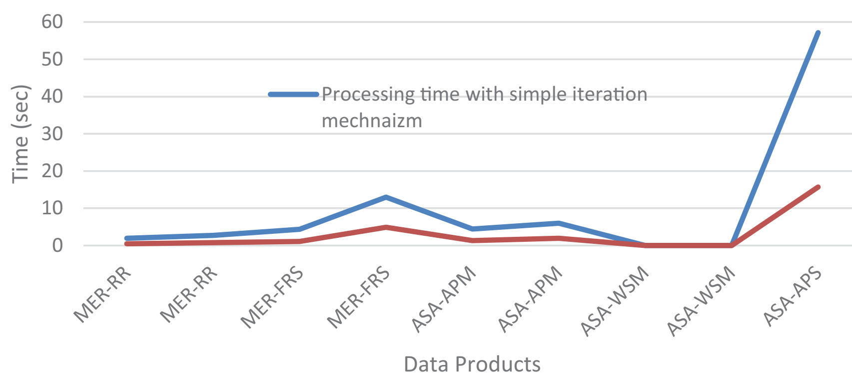


Fig. 8. Processing time taken by the algorithm using various product. (For interpretation of the references to color in this figure, the reader is referred to the web version of this article).

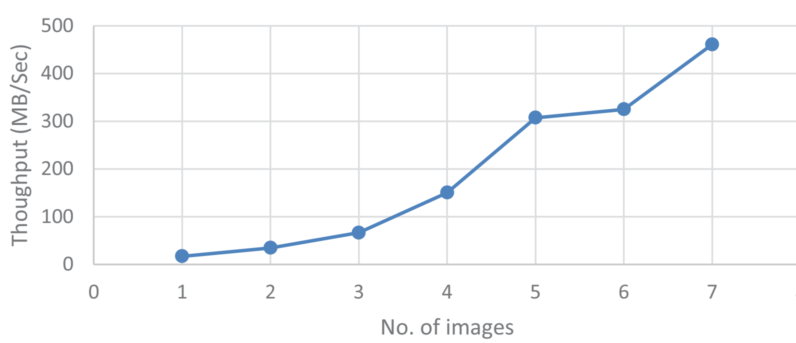


Fig. 9. Throughput of the system by increasing number of satellite images.

The system throughput is also tested by increasing the number of images on Hadoop implementation. Initially, for a single or two images, the throughput is quite less. It is due to the smaller size of the image as the image cannot divided in to chunks to store on multiple data nodes. Hadoop MapReduce programming has a lot of switches between Map and Reduce function. If the size of the image is too short, then it will consume a lot of time in switching as compare to the processing time. Therefore overall throughput is less. However, when we increase the number of images, the throughput tends to increase. For large size datasets, the data is divided into chunks and each chunk is stored on separate data nodes. Each chunk is being processed in parallel because of the Hadoop MapReduce programming nature and HDFS. Resultantly the throughput tends to increase, as shown in Fig. 9.

7. Conclusion

In this paper, continuous features extraction mechanism from satellite images is proposed by taking the river as a continuous feature. The River detection within satellite images is performed by using proposed implementation model. The implementation model has various units and phases including collection, filtration, segmentation, computation, processing, merging and interpretation. The algorithm proposed is based on segmentation, statistical computation, machine learning and the Euclidean distance between pixels. The algorithm proved its efficiency by theoretical testing in terms of complexity as well as by practical implementation using average processing time and system throughput. The algorithm implementation using Map Reduce provides better results as compare to simple Java implementation. Map reduce implementation takes less than 0.5 s average processing time to process 1MB data of most of the ESA satellite products.

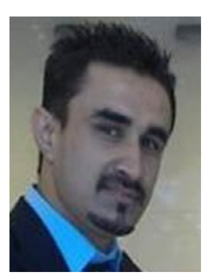
Acknowledgment

This work was supported in part by the Brain Korea 21 Plus Project (SW Human Resource Development Program for Supporting Smart Life) funded by Ministry of Education, School of Computer Science and Engineering, Kyungpook National University, Korea under Grant 21A20131600005, and in part by the Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP). [No. 10041145, Self-Organized Software platform (SoSp) for Welfare Devices]

References

[1] Jiaji Wu, Long Deng, Anand Paul, 3D terrain real-time rendering method based on CUDA-OpenGL interoperability, IETE Technical Review ahead-of-print (2015) 1–8.
[2] Shaohui Liu, et al., A game theory-based block image compression method in encryption domain, The Journal of Supercomputing (2015) 1–20.
[3] Jiao Shi, Jiaji Wu, Anand Paul, Maoguo Gong, A Partition-based active contour model incorporating local information for image segmentation, Sci. World J. 2014 (2014) 19 Article ID 840305.
[4] Jiaji Wu, Anand Paul, Yan Xing, Yong Fang, Jechang Jeong, Licheng Jiao, Guangming Shi, Morphological dilation image coding with context weights prediction, Signal Process. Image Commun. 25 (10) (2010) 717–728 ISSN: 0923-5965.
[5] Anand Paul, High performance adaptive deblocking filter for H.264/AVC, IETE Tech. Rev. 30 (2) (2013) 157–161 ISSN: 0256-4602.
[6] Anand Paul, Jiaji Wu, Jerr-Fa Yang, Jechang Jeong, Gradient-based edge detection for motion estimation in H.264/AVC, IET Image Process. 5 (4) (2011) 323–327 ISSN: 1751-9659.
[7] Anand Paul, K. Bharanitharan, Jhing-Fa Wang, Region similarity based edge detection for motion estimation in H.264/AVC, IEICE Electron. Express 7 (2) (2010) 47–52 ISSN: 1349-2543.
[8] An-Chao Tsai, Anand Paul, Jia-Ching Wang, Jhing-Fa Wang, Intensity gradient technique for efficient intra prediction in H.264/AVC, Proceedings of the IEEE Transactions on Circuits and Systems for Video Technology, 18.5, 2008, pp. 694–698.
[9] Anand Paul, Jhing-Fa Wang, Jia-Ching Wang, An-Chao Tsai, Jang-Ting Chen, Projection based adaptive window size selection for efficient motion estimation in H.264/AVC, IEICE Trans. Fundam. Electron. Commun. Comput. Sci. E89-A (11) (2006) 2970–2976 ISSN: 0916-8508.
[10] P.F.Karen Hollingsworth, Kevin Bowyer, All iris code bits are not created equal, in: Proceedings of IEEE Conference on Biometrics: Theory Applications and Systems, 2008, pp. 1–6.
[11] J. Adams, D. Woodard, G. Dozier, P. Miller, G. Glenn, K. Bryant, GEFE: genetic & evolutionary feature extraction for periocular-based biometric recognition, in: Proceedings of the ACM Southeastern Conference, ACM, 2010, p. 20.
[12] Simpson Lamar, et al., Genetic & evolutionary type II feature extraction for periocular-based biometric recognition, in: Proceedings of IEEE Congress on Evolutionary Computation (CEC), 2010, IEEE, 2010.
[13] Jiao Shi, Jiaji Wu, Anand Paul, Licheng Jiao, Maoguo Gong, Change detection in synthetic aperture radar image based on fuzzy active contour models and genetic algorithms, Math. Prob. Eng. 2014 (2014) 15 Article ID 870936.
[14] N.M.U. Rathore, Anand Paul, Awais Ahmad, Bowei Chen, B. Huang, Wen Ji, Real-time bid data analytical architecture for remote sensing application, IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 8 (7) (2015) 1–12, doi:10.1109/JSTARS.2015.2424683.
[15] T. Dunlay, Obstacle avoidance perception processing for the autonomous land vehicle, in: Proceedings of the IEEE Conference on Robotics and Automation, 2, 1988, pp. 912–917.
[16] E. Krotkov Henriksen, Natural terrain hazard detection with a laser rangefinder, in: Proceedings of the IEEE International Conference on Robotics and Automation, 1997, pp. 968–973.
[17] B. Blanco, B.L. Boada, L. Moreno, M.A. Salichs, Local Mapping from on-line laser voronoi extraction, in: Proceedings of the Conference on Intelligent Robots and Systems, 2000, pp. 103–108.
[18] L. Podsedkowski, J. Nowakowski, M. Idikowiski, I. Visvary, On line navigation of mobile robots using laser scanner, in: Proceedings of the First Workshop on Robot Motion and Control ROMOCO'01, 2013, pp. 241–245.
[19] N. Tomatis Kai, R. Siegwart, Multisensor on-the-fly localization using laser and vision, in: Proceedings of IEEE/RSJ International Conference on Intelligent Robotics and Systems, 2000, pp. 462–467.
[20] Z. Xiaowei, Y.K. Ho, C.S. Chua, Z. Yi, 'The Localization of mobile robot based on laser scanner, in: Proceedings of the Canadian Conference on Electrical and Computer Engineering, 2, 2000, pp. 841–845.
[21] Z. Li, B.K. Ghosh, Geometric feature based 21/2D map building and planning with laser sonar and tactile sensors, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2000, pp. 115–120.
[22] A. Siadat Lallement, M. Dufaut, R. Husson, Laser-vision cooperation for map building and landmarks recognition, in: Proceedings of the IEEE/ISIC/CIRA/ISAS Joint Conference, 1998, pp. 387–392.
[23] Zlotnick, P.D. Camine Jr., Note: finding road seeds in aerial images, CVGIP: Image Underst. 57 (2) (1993) 243–260.
[24] Xiong Demin, White Paper: Automated Road Network Extraction from High Resolution Images, http://www.ncrst.org, 2002.
[25] M.A. Fischler, J.M. Tenenbaum, H.C. Wolf, Detection of roads and linear structures in low resolution aerial imagery using a multisource knowledge integration technique, Comput. Graph. Imaging Process. 15 (3) (1981) 201–223.
[26] D. Geman, B. Jedynak, Detection of roads in satellite images, Proceedings of Geoscience and Remote Sensing Symposium IGARSS '91 on Remote Sensing: Global Monitoring for Earth Management International 4 (1991) 2473–2477.
[27] D Geman, B. Jedynak, An active testing model for tracking roads in satellite images, IEEE Trans. Pattern Anal. Mach. Intell. 18 (1) (1996) 1–14.
[28] Armin Gruen, Haihong Li, Road extraction from aerial and satellite images by dynamic programming, ISPRS J. Photogramm. Remote Sens. 50.4 (1995) 11–20.
[29] Thomas Veit, et al., Evaluation of road marking feature extraction, in: Proceedings of the 11th International IEEE Conference on Intelligent Transportation Systems, ITSC 2008, IEEE, 2008.
[30] Alberto Broggi, Parallel and local feature extraction: a real-time approach to road boundary detection, IEEE Trans. Image Process. 4.2 (1995) 217–223.
[31] G. Priestnall, J. Jaafar, A. Duncan, Extracting urban features from LiDAR digital surface models, Comput. Environ. Urban Syst. 24.2 (2000) 65–78.
[32] Yun Zhang, Detection of urban housing development by fusing multisensor satellite data and performing spatial feature post-classification, Int. J. Remote Sens. 22.17 (2001) 3339–3355.
[33] Anand Paul, Bo-wei Chen, K. Bharanitharan, Jeong Jechang, Jhing-Fa Wang, Video search and indexing with reinforcement agent for interactive multimedia services, ACM Trans. Embed. Comput. Syst. 12 (2) (2013) 1–16 Article no. 25. ISSN: 1539-9087.
[34] Anand Paul, K. Bharanitharan, Jaiji Wu, Algorithm and architecture for adaptive motion estimation in high efficiency video coding, IETE Tech. Rev. 30 (1) (2013) 24–30 ISSN: 0256-4602.
[35] Anand Paul, Yung-Chuan Jiang, Jhing-Fa Wang, Jar-Ferr Yang, Parallel reconfigurable computing-based mapping algorithm for motion estimation in advanced video coding, ACM Trans. Embed. Comput. Syst. 11 (S2) (2012) 1–18 Article no. 40. ISSN: 1539-9087.
[36] Anand Paul, adaptive search window for high efficiency video coding, J. Signal Process. 79 (3) (2015) 257–262 (September 10th 2013, Published online).
[37] [Available online: 14/10/2014, 2312] https://earth.esa.int/ (accessed 15.10.14).
[38] ESA, "ENVISAT Altimetry Level 2 User Manual V1.4, 2011". [Available online: 15/10/2014, 0333] https://earth.esa.int/pub/ESA_DOC/ENVISAT/RA2-MWR/PH_light_1rev4_ESA.pdf
[39] [Available online: 15/10/2014, 0333] http://www.brockmann-consult.de/cms/web/beam/ (accessed 15.10.14).
[40] Mike Olson, Hadoop: scalable, flexible data storage, and analysis, IQT Q. 1.3 (2010) 14–18.

Muhammad Mazhar Ullah Rathore received the Master's degree in computer and communication security from the National University of Sciences and Technology, Islamabad, Pakistan, in 2012, and is currently pursuing the Ph.D. degree at Kyungpook National University, Daegu, Korea. His research interests include Big Data analytics, network traffic analysis and monitoring, intrusion detection, and computer and network security.



Awais Ahmad (S'14) received the B.S. degree (CS) from the University of Peshawar, Peshawar, Pakistan, and the M.S. degree (telecommunication and networking) from Bahria University, Islamabad, Pakistan, in 2008 and 2010, respectively. Currently, he is pursuing the Ph.D. degree at Kyungpook National University, Daegu, Korea. During his research work, he worked on energy efficient congestion control schemes in Mobile Wireless Sensor Networks (WSN). There he got research experience on Big Data analytics, machineto- machine communication, and wireless sensor network. Mr. Ahmad was the recipient of three prestigious awards: (1) Research Award from President of Bahria University Islamabad, Pakistan in 2011 (2) best Paper Nomination Award in WCECS 2011 at UCLA, USA, and (3) best Paper Award in 1st Symposium on CS&E, Moju Resort, Korea, in 2013.

Anand Paul (SM'15) received the Ph.D. degree in electrical engineering from the National Cheng Kung University, Tainan, Taiwan, in 2010. He is currently working as an Associate Professor with the School of Computer Science and Engineering, Kyungpook National University, Daegu, Korea. He is a delegate representing Korea for M2M focus group and for MPEG. His research interests include algorithm and architecture reconfigurable embedded computing. Prof. Paul has Guest Edited various international journals and he is also part of Editorial Team for *Journal of Platform Technology* and *Cyber Physical Systems*. He serves as a Reviewer for various IEEE/IET journals. He is the track Chair for smart human computer interaction in ACMSAC 2015, 2014. He was the recipient of the Outstanding International Student Scholarship Award in 2004–2010, the Best Paper Award in National Computer Symposium, Taipei, Taiwan, in 2009, and International Conference on Softcomputing and Network Security, India, in 2015.

Jiaji Wu is a professor at Xidian University, Xi'an, China. He received the B.S. degree in electrical engineering from Xidian University, Xi'an China, in 1996, the M.S. degree from National Time Service Center (NTSC), the Chinese Academy of Sciences in 2002, and the Ph.D. degree in electrical engineering from Xidian University in 2005.