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,