Energy and Buildings 115 (2016) 131–137

Contents lists available at ScienceDirect

Energy and Buildings

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

Automated identification of land cover type using multispectral satellite images

Dragan Stević[a], Igor Hut[b], Nikola Dojčinović[c], Jugoslav Joković[c,*]

[a] University of Priština – Faculty of Technical Science, Str. Kneza Milosa 7, Kosovska Mitrovica, Serbia
[b] University of Belgrade – Faculty of Mechanical Engineering, Blvd. Kralja Aleksandra 73, Belgrade, Serbia
[c] University of Niš – Faculty of Electronic Engineering, Str. Aleksandra Medvedeva 14, Nis, Serbia

A R T I C L E I N F O

Article history:
Received 6 December 2014
Received in revised form 16 April 2015
Accepted 3 June 2015
Available online 9 June 2015

Keywords:
Landscape classification
Remote sensing
Multispectral images
Neural networks
Support vector machines
Logistic regression

A B S T R A C T

Detection of specific terrain features and vegetation, referenced as a landscape classification, is an important component in the management and planning of natural resources. The different land types, man-made materials in natural backgrounds and vegetation cultures can be distinguished by their reflectance. Although remote sensing technology has great potential for acquisition of detailed and accurate information of landscape regions, the determination of land-use data with high accuracy is generally limited by the availability of adequate remote sensing data, in terms of spatial and temporal resolution, and digital image analysis techniques. Therefore, remote sensing with multi-spectral or/and hyper-spectral data derived from various satellites in combination with topographic variables is a valuable tool in landscape type classification. The different methods based on reflectance data from multi-spectral Landsat satellite image sets are used for automatic landscape type recognition. In order to characterize reflectance of landscape types represented in an image, construction of a multi-spectral descriptor, as a vector of acquired reflectance values by wavelength bands, is proposed. The applied algorithms for landscape type classification (artificial neural network, support vector machines and logistic regression) have been analysed and results are compared and discussed in terms of accuracy and time of execution.

1. Introduction

The many global urban and landscape planning applications, as well as applications related to the environment rely on information about the natural environments (vegetation cover on land) and land usage [1–10]. The modern living and working environment combines urban, rural and natural areas, with specific needs in terms of energy efficiency, air-conditioning, aesthetics, etc. [11–16]. Also, since the spatial distribution of vegetation is an important factor in a numerous biological, geographical and chemical cycles related to climate maintenance and change, the information on the environment in context of land cover/vegetation distribution and the changes of greenspace is essential for a better understanding of the sustainability of urban development processes. Therefore, generally, need for distinguishing between urban, rural and natural environment is one of prerequisites of land use planning, as well as estimation of land-use structures and land-cover changes that result from urban growth.

The traditional methods for acquisition of those data rely on data gathering on the field. Such process usually consists of gathering of specific vegetation (or land usage, environment, etc.) data by visiting specific location and fitting them within geo-referential framework. This approach is characterized by high cost of data acquisition by location visit and can last for a long period. Also, depending of terrain, climate and other conditions, sometime is not possible to cover desired locations. Therefore, modern techniques rely on data gathering methods that utilize acquisition from an aircraft/satellite. Recent research has identified a number of different approaches for data acquisition and for landscape characterization and analysis [17–19] which utilize remote sensing imagery as source data in the derivation of spatial data sets with high spatial and temporal resolution. However, methods that use airborne images as input [18–20] requires special aircrafts equipped with specific imagery and/or laser systems to fly over required areas, what increases cost of this methods, thus reducing applicability of those methods. On other hand, majority of satellites that gathers imagery data have already covered large portion of Earth surface or can provide desired data under much lower costs. The series of satellites launched under common name Landsat through it history (1972–present) provided millions of images used in agriculture,

* Corresponding author.
E-mail address: jugoslav.jokovic@elfak.ni.ac.rs (J. Joković).

http://dx.doi.org/10.1016/j.enbuild.2015.06.011
0378-7788/

Table 1
Landsat spectral channels.

Spectral band enumeration	Spectral range (nm)	Ground resolution (m)
1	450–515	30
2	525–605	30
3	630–690	30
4	750–900	30
5	1550–1750	30
6	10,400–12,500	60
7	2090–2350	30
Pan	520–900	15

geology, forestry, regional planning and other related fields. The latest satellite in Landsat series, Landsat 8, was launched on 11th of February 2013. However, most of available images origins from Landsat 7 (Aeronautics) satellite that has been operating since 15th of April 1999. It has polar sun-synchronous orbit on altitude of 705($\pm$5) km. The Landsat satellite equipped with Enhanced Thematic Mapper Plus (ETM+) as the main sensing instrumentation is able to survey the Earth's surface with variable spatial and temporal resolutions, as well as in different spectral bands. In the single acquisition, it captures a multispectral data for the ground within area of $185 \times 185\,km^2$ in the 7 (seven) different spectral bands, including the panchromatic band (Table 1).

The spectral channels width varies from band to band, as shown in Table 1. Additionally, the spatial sampling resolution for 6th spectral band is lower (60 m) in comparison to other bands (30 m) and panchromatic (15 m), meaning that one pixel in images represents area of $60 \times 60\,m^2$ for 6th spectral band, $30 \times 30\,m^2$ for other spectral bands and $15 \times 15\,m^2$ for panchromatic band. However, the 6th channel, that covers wavelengths range belonging to infrared part of spectrum, is not available for civil use.

As stated in Table 1, Landsat can provide more data contained outside of visible part of spectrum, and therefore provide more information than human psycho-visual system can perceived. Although the spectral distribution of reflectance of landscape type can be described, in terms of bare/vegetation regions, as well as different cultures of vegetation, due to variation in acquisition environment, the region overlapping/mixing and nonlinearities caused with varying incident light, it is not possible to uniquely connect resulting spectral distribution with some vegetation culture. Therefore, artificial intelligence and machine learning algorithms have been developed for automatic classification of regions based on spectral distribution of reflected light [8–10]. These methods capture complex pattern of spectral distribution consisting of linear and nonlinear combination of reflectance spectrums of regions.

Starting from importance of the landscape classification as information that can contribute to more detailed mapping of urban areas and towards a more accurate characterization of spatial urban growth pattern, this paper considers a comparative analyses of methods for landscape classification that using information on satellite multi-spectral images. Motivation for using Landsat 7 image set as input dominantly is in their wide availability, since most of Earth is already covered with Landsat 7 data [21].

The application of neural networks in environment classification on bare land, land covered with deciduous, softwood and grasslands has been discussed in our previous paper [22]. Here, we introduce Support vector machines approach as well as two versions of another machine learning algorithm, Logistic regression with linear and nonlinear kernel. The research is focused on comparative analysis of applied approaches for classification of bare regions versus regions covered with vegetation, in terms of accuracy and computational efficiency, in order to provide solution that could be used in real-time application with low deployment costs. Section 2 contains description of the methods applied for landscape classification: artificial neural networks, support vector machines approach and logistic regression, both with linear and nonlinear kernel. Besides the input data availability, the basic reason for using applied algorithms is wide range applicability of the classification methods, where accuracy is better with increasing number of samples. The results are discussed in Section 3, providing comparative analysis of the proposed approaches applied for landscape classification on the chosen example. The survey was conducted in a broader area of Pester plateau, an area of $50 \times 50\,km^2$, located at $43°16'30''$ north latitude and $20°00'00''$ east longitude in the south-western part of Serbia. For illustration, the image of 4th Landsat band covering considered area is presented in Fig. 1. As it can be seen, digital format of Landsat 7 channel is rotated for angle of declination of Landsat 7 orbit respect to Equator, so that northward orientation in each channel is vertically upward. Since digital image formats are strictly rectangular, and original area covered with Landsat 7 sensor is rotated for mentioned angle, in order to match rectangular form data is padded on corners (Fig. 1) with neutral data or black colour.

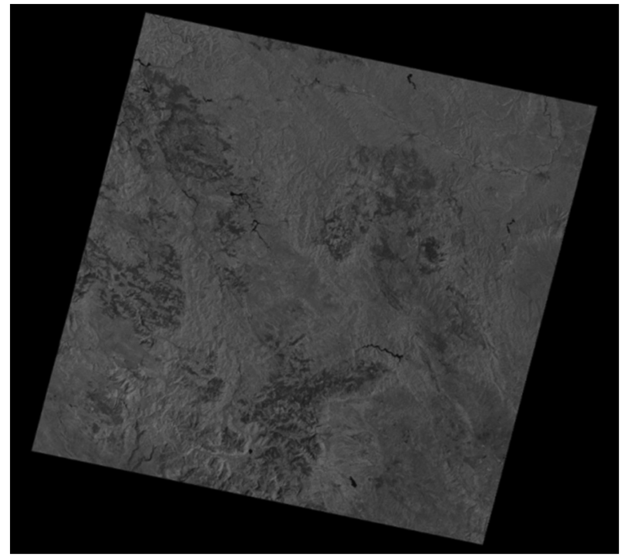

Fig. 1. The 4th Landsat 7 channel of Pešter plateau.

The considered area is characterized by a very sharp relief with large height differences in relief and distinct ridges of the surrounding mountains, the presence of small urban settlements, the diversity of geological substrates and soil, diversity of continental and alpine climates in the lower and higher areas, large temperature differences and high humidity of the air, and a relatively small number of sunny days per year. Since the high level of accuracy for bare/vegetation regions classification (>90%) was obtained, generally, a land-cover classes can be successfully classified using applied low computation complexity algorithms.

2. Methodology

For the purpose of analysis regarding landscape classification data set is formed [22,23], consisting of description vector for a particular pixel as input data and matching bare/vegetation region for that geographic position, as desired response. The landscape data was organized in geo-referent map covering same area as Landsat 7 image set. The geo-referenced map with vegetation culture information overlay can be found in Fig. 2, where different vegetation cultures are represented by corresponding colours. The digital map of vegetation is covered by one image of Landsat 7 channel set, which represents overlay. Since both maps (Landsat 7 channel and vegetation map) have geo-referent data that are available