DETECTION OF FERTILE SOILS BASED ON SATELLITE IMAGERY PROCESSING

V. Grishkin[1,a], E. Zhivulin[1], A. Khokhriakova[1], S. Karimov[1]

[1] *Saint Petersburg State University, 7–9 Universitetskaya nab., Saint Petersburg, 199034, Russia*

E-mail: [a] valery-grishkin@yandex.ru

The paper proposes a method for detecting fertile soils based on the processing of satellite images. As a result of its application, a map of the location of fertile and infertile soils for a given region of the earth's surface is formed and the corresponding areas are calculated. The method for detecting fertile soils is based on the fact that fertile soil includes areas covered with vegetation in the spring-summer period. Therefore, by measuring the spectral characteristics of these areas in the late autumn period, when there is no vegetation on them, it is possible to obtain objective parameters of fertile soils. For detection, a number of classifiers are being built that recognize two classes - fertile soil and sand, which is especially important when monitoring areas prone to desertification. The feature vector used for classification is a set of indices similar to the well-known NDVI index. This set of indices is calculated for each pixel of the image by its values in different spectral channels. Classifiers are implemented using CUDA parallel computing technology on a GPU. Based on the results of the experimental study, a classifier is selected that has shown the best characteristics of the recognition quality.

Keywords: satellite image segmentation, multispectral images, cene classification

Valery Grishkin, Evgeniy Zhivulin, Anastasiia Khokhriakova, Sardor Karimov

1. Introduction

Currently, there is a problem - land desertification. To effectively combat this, it is necessary to know the situation in a particular region at the current moment in time, as well as to be able to track the dynamics of desertification. One way to do this is to monitor the area using satellites. This method allows you to cover the entire territory, while providing sufficient accuracy for such a scale. In addition, it does not require large financial costs and allows you to collect and process information about the state of the soil over a long period of time.

When monitoring the state of land, it is necessary to distinguish between fertile and barren soil. Since land desertification occurs mainly due to the onset of sand, this paper considers a method that detects areas occupied by fertile soils and sandy areas on satellite images. We propose a method for detecting fertile soils in specified regions based on processing multispectral satellite images. These images from various satellites are publicly available on their respective websites. In this work, we use open data obtained by the Sentinel satellites within the European Copernicus program [1].

2. Data acquisition and preprocessing

Data of multispectral imagery from Sentinel-2 satellites and the results of their preprocessing are used as input images. These data are freely available on the Copernicus Open Access Hub [2]. Images of the earth's surface for the region of interest are taken at intervals of 5 days. The surface survey from satellites is carried out in various spectral ranges with a resolution of at least 20 by 20 meters in one image pixel. Table 1 shows a list of spectral and additional channels used to detect fertile and sandy soils.

Table 1. Wave range and maximum resolution of the channels used.

Channel	Description	Wavelength	Resolution
B02	Blue	492.4 nm	10 m
B03	Green	559.8 nm	10 m
B04	Red	664.6 nm	10 m
B05	Vegetation red edge	704.1 nm	20 m
B06	Vegetation red edge	740.5 nm	20 m
B07	Vegetation red edge	782.8 nm	20 m
B08	NIR	832.8 nm	20 m
B08A	Narrow NIR	864.7 nm	20 m
B09	Water vapor	945.1 nm	20 m
B11	SWIR	1613.7 nm	20 m
SNW	Snow probability	-	20 m
CLD	Cloud probability	-	20 m
SCL	Scene classification data	-	20 m

Additional channels contain the results of preprocessing images captured in the Sentinel data center. It should be noted that for the detection of fertile and sandy soils, we use only 13 channels out of 23 available.

Channels B02, B03, B04 represent reflected light in the optical range. Channels B05, B06, and B07 represent infrared radiation. Two channels B08 and B08A belong to the near infrared range. The next spectral range B09 is associated with reflections from water vapor. Then three channels follow,

reflecting the results of data processing by the Sentinel-hub service itself. The first one (SNW) is the probability that a given pixel in the image is part of a snow-covered area, and the second one (CLD) is the probability that it belongs to a cloud cover. The last of these channels (SCL) is an image mask describing the belonging of each pixel to one of the 12 classes. Among these classes are the following: vegetation, bare soil, water, clouds, cloud shadows, snow.

Interaction with the Copernicus Open Access Hub website is carried out using requests that are supported by the corresponding Sentinel-hub service. The request specifies the date range (the service provides the latest snapshot from the specified range), the coordinates of the area of interest, the size of the resulting image in pixels, the list of channels from which the data is requested, and the view in which it will be loaded. The size of the requested area is chosen equal to 10 x 10 km, which corresponds to the dimensions of the resulting image equaling to 512 x 512 pixels when using a channel resolution of 20 meters.

Before loading data, it is possible to pre-process it directly on the site by performing arithmetic operations on it. These operations allow for normalizing the values in different channels. Data from channels is requested in the form of standard PNG images. Each image includes 3 channels of data displayed as red, green and blue.

Downloading the data required for exploring large areas also requires a large number of downloads. Therefore, a specially developed script is used to load data. This script generates a request for satellite data for a specified location within a specified date period. When requesting data from the service, the access key is transmitted which has been obtained earlier. This key is updated hourly and can be obtained as a result of another request to the service. Therefore, when exchanging with the hub, the size of uploaded images is checked, and if it is too small, this indicates that the service did not provide data, therefore, you need to request a new key is to be requested, and the script takes care of that. For interaction with the site, the standard cURL utility [3] is used, which is part of almost all operating systems. After the request is generated, the script is passed to this utility as a parameter.

3. Fertile soil detection

The algorithm for detecting fertile soils is based on a fairly simple idea of comparing images of a region in the spring-summer and late autumn periods. The first period was chosen because it is in this period that the vegetation cover is the strongest and is well determined from satellite images using standard methods. The second period is characterized by almost no vegetation, and in these images, the surface that was covered with vegetation looks like soil. Thus, comparing these images, one can draw conclusions about the location and presence of productive soils in the observed area, since the soil on which nothing grows is either degraded or unsuitable for agricultural purposes.

The input data contains, in addition to spectral images, the results of preliminary classification of images of the region, obtained by processing the data of the Copernicus Open Access Hub [2]. This classification results are tied to each requested area of the earth's surface and are represented as a pixel mask M_{scl} of the image of this area. The values of each pixel in this mask correspond to specific types of ground surface or types of clouds. This classification mask is used to segment areas containing agricultural soils. These areas are covered with vegetation in spring and summer. Therefore, segmenting such areas in the images of the late autumn period, we will obtain data on the distribution of fertile soils. Thus, it is necessary to form two new masks from the original summer mask. The M_{summer_veg} mask represents the vegetation cover only, while the M_{summer_soil} mask represents the soil. Two similar masks are also formed out of the autumn classification mask. The first M_{autumn_soil} mask represents the soil cover, and the second M_{autumn_veg} mask represents the vegetation cover. The $M_{goodsoil}$ fertile soil mask is calculated by applying a bitwise AND to the summer vegetation mask and the autumn soil distribution mask.

$$M_{goodsoil} = M_{summer\_veg} \& M_{autumn\_soil} \qquad (1)$$

The second mask contains additional information about the location of the fertile soil hidden under the vegetation. This mask can be used to obtain a complete mask of fertile soil $M_{f\_goodsoil}$.

$$M_{f\_goodsoil} = M_{goodsoil} \mid (M_{summer\_soil} \& M_{autumn\ veg}) \qquad (2)$$

Figure 1 shows satellite images of the same region in the visible range in the spring-summer and autumn periods, the corresponding classification masks and the resulting masks of fertile soils. The last mask shows the distribution of fertile soils in the area not covered by clouds, and it is used later for marking when training the recognition system.
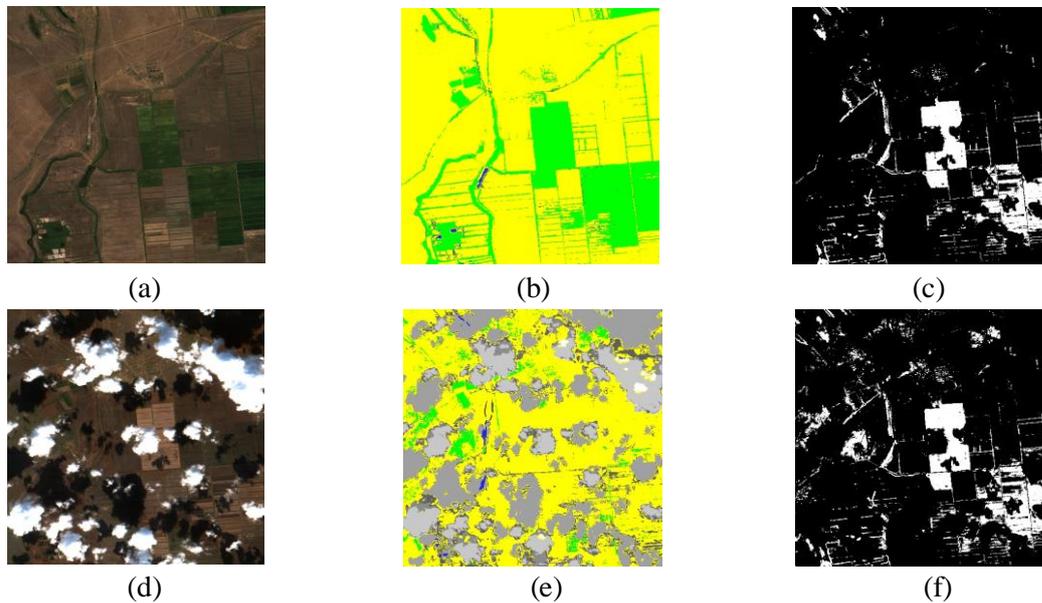
Figure 1. Images of the region and corresponding masks: (a) image of the region – August; (b) classification mask – August; (c) image of the region - October; (d) classification mask – October; (e) the fertile soil mask without autumn vegetation; (f) the complete fertile soil mask; ■ – vegetation; ■ – bare soil; ■ – clouds

4. Experimental results

For training and testing the classifiers for recognition fertile soils, a set of multispectral images and calculated masks of fertile soils were used. This dataset was formed from 379 pairs of sets of combined spectral images of subregions. Each pair of sets for the same subregion consists of four combined images taken in August and four combined ones taken at the end of October. For these pairs of images, masks of fertile soils were calculated, which were also included in the dataset.

We investigated the following classifiers – Bayesian classifier, Random forest, and SVM classifier [4-6]. All these classifiers are based on heuristic features. We used a vector of normalized differential indices as a feature vector. Each index is calculated based on the values of two different spectral channels. The calculation of the corresponding index V_{ij} was carried out similarly to the procedure for calculating the known vegetation index NDVI.

$$V_{ij} = (B_i - B_j) / (B_i + B_j) \tag{3}$$

As a result of the experiments, the best accuracy was achieved with the dimension of the feature vector equal to 8. In this case, the following indices were used: V_{47}; V_{46}; V_{89}; V_{38}; V_{92}; V_{01}, V_{48}; V_{57}. These indices are calculated using the following channel pairs: (B06 B8A), (B06 B08), (B09 B11), (B05 B09), (B8A B04), (B02 B03), (B06 B09), (B07 B8A). The results of soil type recognition for the studied classifiers are shown in Table 2.

Table 2. The results of soil type recognition

Classifier	Precision of determining fertile soil	Precision of determination of sand
Bayesian	0.834	0.933
Random forest	0.945	0.958
SVM	0.929	0.963

4. Conclusion

We propose a method for detecting fertile soils in specified regions based on processing multispectral satellite images. The data required for detection is requested during interaction with the hub service for accessing remote database. In addition to spectral images, we also use the results of their preprocessing, which are also requested from this database. The feature vector used for classification is a set of indices similar to the well-known NDVI index. The paper presents experimental estimates of the recognition quality for the classifiers under study, which are used to detect fertile soils. Based on the analysis of the estimates obtained, the choice of the type of classifier for the detection of fertile soils is made.

References

[1] ESA Earth Observation Portal. Available at: https://directory.eoportal.org/web/eoportal/ satellite-missions/c-missions/copernicus-sentinel-2/ (accessed 30. 06. 2021)

[2] Copernicus Open Access Hub. Available at: https://scihub.copernicus.eu/dhus/ (accessed 02. 07. 2021)

[3] Command line tool and library for transferring data with URLs. Available at: https://curl.se/ (accessed 14.05.2021)

[4] Cutler A., Cutler D.R. and Stevens J.R. Random Forests // Ensemble Machine Learning: Methods and Applications, pp. 157-176, Chapter5, Springer, 2011

[5] Cristianini N, Shawe T. J. An Introduction to Support Vector Machines and other kernel-based learning methods. Cambridge University Press, UK: Cambridge, 2000

[6] Domingos P., Pazzani M. On the optimality of the simple Bayesian classifier under zero-one loss // Machine Learning, V. 29 (2/3), pp. 103–137, 1997. - DOI:10.1023/A:1007413511361.