

CROPS IDENTIFICATION BY USING SATELLITE IMAGES AND ALGORITHM FOR CALCULATING ESTIMATES

N.S. Vorobiova

Samara National Research University, Samara, Russia
Image Processing Systems Institute - Branch of the Federal Scientific Research Centre "Crystallography and Photonics" of Russian Academy of Sciences, Samara, Russia

Abstract. The paper proposes a crop identification method based on the algorithm for calculating estimates using satellites images. The classification features are the set of time series values, built by the sequence of satellite images, and geographic coordinate of field – latitude. Method was tested by using Terra/MODIS images and ground-based information. The comparison of the classification quality of proposed method with classification quality of the classifier based on Mahalanobis Distance is given.

Keywords: time series, vegetation index, NDVI, satellite images, crops identification, crops recognition, algorithm for calculating estimates.

Citation: Vorobiova NS. Crops identification by using satellite images and algorithm for calculating estimates. CEUR Workshop Proceedings, 2016; 1638: 419-427. DOI: 10.18287/1613-0073-2016-1638-419-427

1 Introduction

Currently, management in agricultural sector is an area of active development and implementation of methods that apply remotely sensed data (hereinafter - RSD) to solve production tasks [1]. One of such tasks is the crop identification by using satellite images. The solution of this task is very important for such events:

- real-time monitoring;
- control of crop lands usage,
- verification of information provided by farmers about crops seeded on fields;
- mapping land usage in areas with no information from the farmers about crops seeded on fields.

The literature suggests a large number of methods to solve crop recognition task by using satellite images [2-6]. The most important methods are those methods which are suitable for usage over large areas (regional scale and more). Territory of large area,

in contrast to the farm or group of farms, is difficult to explore by land, and remotely sensed data provide this opportunity [9, 10].

In this paper crop recognition method based on algorithm for calculating estimates is proposed. This method is suitable for usage in areas of regional scale. The method allows to take into account the geographical position of the field and to use the time series with gaps in observation days. Gaps occur due to the presence of clouds on satellite images at a certain day of observation. The ability to use data with gaps does not require an additional procedure of time series interpolation. The proposed detection method is based on an evaluation of the proximity between classified and learning objects. The quality of the proposed method is compared with the quality of the classifier based on the Mahalanobis distance.

2 Input data and preprocessing

All data for research were obtained for the Samara Region. The paper uses satellite images and ground data about fields for the year 2014. The description of input data and preprocessing methods is given below.

2.1 Satellite images

Data from satellite Terra/MODIS are used to construct time series. The raw data recorded on the satellite are passed through ground processing and are saved in the form of so-called products [7]. The product is the result of some processing run on the raw data or on data already processed to a certain level. In this paper red and near-infrared channels of daily MOD09GQ product are used to construct time series. The product MOD09GQ has passed through radiometric and atmospheric correction, is geo-referenced, and is a minimum level of combination of daily 250 m data.

Preprocessing of MOD09GQ products (for each day) includes the following steps:

1. Mosaic composing from two tiles (h20v03, h21v03) which cover the territory of the Samara region. Tile is the cell of MODIS sinusoidal cell.
2. Pixel-by-pixel synchronization with the images for previous days: setting of uniform map projection and resolution; setting of uniform horizontal and vertical size; setting a uniform binding of the corner points.
3. Formation of cloud mask.

2.2 Ground data

Ground data about fields is provided by farmers and are necessary for training the classifier and evaluation of its quality. The following ground data is provided for each field:

- the boundary and the square of the field;
- crops seeded on the field or type of field usage such as fallows or unused lands.

The boundaries of the fields were superimposed on a cover composed by satellite images of medium resolution (20-30 m) in order to verify the homogeneity of fields. The heterogeneity is caused by the following factors: the presence of several crops or types of usage on the field, the use of only part of the field for sowing, sprouts heterogeneity. All identified inhomogeneous fields were divided into the corresponding number of homogeneous regions. The study of homogeneity was conducted using a segmentation method [9]. As a result, new boundaries of plots were obtained. New plots boundaries are used for the time series calculation and crops identification.

In this study, all crops and types of field usage have been divided into the following groups: perennial grasses, unused land, winter crops, fallows, early spring crops, late spring crops. Such division into classes will be used further in the classification algorithm based on calculating estimates.

2.3 Time series

Normalized difference vegetation index NDVI was selected for the construction of time series. Value of index is calculated according to the formula:

$$NDVI = \frac{\rho_{NIR} - \rho_{RED}}{\rho_{NIR} + \rho_{RED}}, \quad (1)$$

where ρ_{NIR}, ρ_{RED} – values of reflected radiation in the near infrared and the red regions of the spectrum, respectively.

Terra/MODIS images received during the period from March 1 to September 30 for year 2014 were used to construct the NDVI time series. An index value for each plot is calculated as follows:

- calculation of the values averaged by all pixels of the plot for red and near-infrared channels ρ_{NIR}, ρ_{RED} ;
- calculation of the NDVI value on the plot by averaged values ρ_{NIR}, ρ_{RED} according to the formula (1).

NDVI values are sorted by the date when the image was taken. Sorted NDVI values form NDVI time series of the plot.

3 Crop identification method

To detect crops it is offered to use the classification method based on the algorithm for calculating estimates (hereinafter – ACE). Class of recognition algorithms based on the calculating estimates was proposed by Zhuravlev YI [8]. ACE is not a predetermined algorithm but a model of recognition algorithm. And in each case the model of recognition algorithm must be specified. A description of the proposed method for crop identification, which is a specification of ACE model, is given below.

3.1 Model of ACE for crop identification

The input data for the algorithm is a set of reference objects and recognizable objects. All objects are characterized by the following features: time series values and geographical coordinate – latitude. The task is to classify a set of recognizable objects in the predefined classes. A priori information is given in a table 1.

Table 1. A priori information

Class	Objects	Features					
		x_0	x_1	x_2	...	x_{N-1}	δ
Ω_0	ω_0	$\omega_{0,0}$	$\omega_{0,1}$	$\omega_{0,2}$		$\omega_{0,N-1}$	δ_0
	ω_1	$\omega_{1,0}$	-	$\omega_{1,2}$		$\omega_{1,N-1}$	δ_1
		
Ω_1	ω_p	-	$\omega_{p,1}$	$\omega_{p,2}$		-	δ_p
		
					...		

The following conventions are used in Table 1: x_i, δ – features: time-series value of the day i and latitude of the object, respectively; Ω_m – class $m, m = \overline{0, M-1}, M$ – number of classes; ω_p - a reference object of a some class Ω_m . The object ω_p has the following features: latitude δ_p and time series values $\omega_{p,n}, n = \overline{0, N-1}$. Here $\omega_{p,n}$ – the value of the time series of the object ω_p in a day n ; N – the total number of observation days. Both reference objects and recognizable objects can have gaps in the values of the time series. Specifying ACE model assumes setting the following sub-paragraphs.

3.2 System of support feature sets

In this method *system of support feature sets* consists of a single set comprising all the features.

3.3 Proximity function

The proximity ρ of the recognized object a and the reference object ω_p is calculated as follows:

$$\rho = e^{-(k \cdot \rho_1 + (1-k) \rho_2)}, \quad (2)$$

where ρ_1 – the value that characterizes proximity of two objects by the time series values, ρ_2 – the value that characterizes proximity of two objects by latitude, k – parameter determining the weight ρ_1 and ρ_2 in the final value of proximity. Values ρ_1 and ρ_2 are calculated by the formulas (3) and (4), respectively.

$$\rho_1 = \sum_{n=0}^{N-1} (\omega_{p,n} - a_n)^2 \quad (3)$$

$$\rho_2 = |\delta_p - \delta_a| \quad (4)$$

The following conventions are used in the formulas (3) and (4): $a_n, n = \overline{0, N-1}$ and δ_a – the features of object a : set of time series values and latitude, respectively. Value ρ_1 is calculated only for the days n on which both the object ω_p and the object a have the values of time series $\omega_{p,n}$ and a_n , respectively.

The value of proximity function $f(\omega_p, a)$ between reference object ω_p and recognized object a is calculated as follows:

$$f(\omega_p, a) = \begin{cases} 1, & \rho > T \\ 0, & \rho \leq T \end{cases}, \quad (5)$$

where T – proximity threshold.

3.4 Estimate of proximity for a class.

Estimate of proximity $\Gamma(\Omega_j)$ of object a to a class Ω_j is calculated so:

$$\Gamma(\Omega_j) = \sum_{\omega_p \in \Omega_j} f(\omega_p, a), \quad (6)$$

3.5 Decision rule

Classification of recognized object a will be done in class Ω_c according to the decision rule. We define two variants of decision rules:

1) $c = \arg \max_{m=0, M-1} \Gamma(\Omega_m)$; 2) $c = \arg \max_{m=0, M-1} \frac{\Gamma(\Omega_m)}{q_m}$, where q_m – the number of elements in a class Ω_m .

4 Experiments

4.1 Set of objects for testing

Testing sample of 6424 plots has been formed to assess the quality of the proposed classifier. Quality evaluation has been conducted using cross-validation. Testing sample was divided five times into training and control sample in the ratio of 2: 1.

4.2 Classification by using algorithm for calculating estimates

Procedure of searching the parameters k and T has been executed for each option of the decision rule. Searching procedure allows finding such parameters k and T on which the best classification results are achieved. Search has been made by a brute force method with a step of 0.01 for parameter k and 0.001 for parameter T . The best results of classification for the decision rule 1 are achieved when $k=0.98$ and $T=0.9945$. For decision rule 2 best results are achieved when $k=0.97$ and $T=0.9945$. The results of crop identification using the ACE are shown in Tables 2 and 3.

Table 2. The results of classification from class to class using the ACE. Decision rule 1

Class	Decision rule 1, $k=0.98$, $T=0.9945$, $Q=0.72$					
	perennial grasses	unused lands	winter crops	fallows	early spring crops	late spring crops
perennial grasses	0.18	0.54	0.13	0.06	0.02	0.03
unused lands	0.02	0.87	0.02	0.03	0.04	0.01
winter crops	0.00	0.04	0.89	0.02	0.02	0.02
fallows	0.01	0.19	0.04	0.58	0.08	0.08
early spring crops	0.00	0.08	0.01	0.04	0.69	0.17
late spring crops	0.00	0.07	0.01	0.07	0.06	0.78

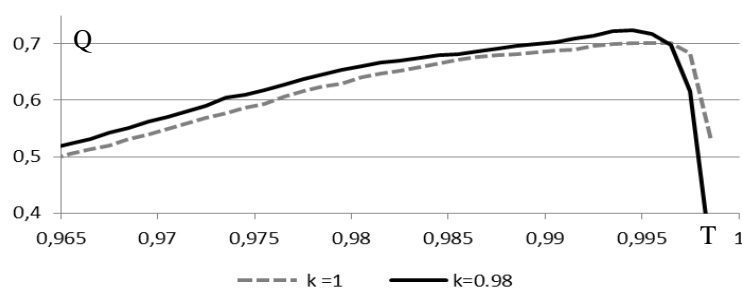
Figure 1 shows the dependencies of the total probability of correct classification Q of the threshold T . Latitude is not considered when $k=1$. The graphs show that Q value is higher the in the case when the latitude is taken into account.

4.3 The classifier on Mahalanobis Distance

To assess the quality of the proposed classifier based on ACE the classification of the test sample according to Mahalanobis distance [9] has been made using cross-validation scheme mentioned above. The results of classification from class to class are given in Table 4. The total probability of correct classification is 0.64.

Table 3. The results of classification from class to class using the ACE. Decision rule 2

Class	Decision rule 2, $k = 0.98$, $T = 0.9945$, $Q = 0.72$					
	perennial grasses	unused lands	winter crops	fallows	early spring crops	late spring crops
perennial grasses	0.40	0.42	0.08	0.03	0.01	0.03
unused lands	0.17	0.77	0.01	0.02	0.02	0.01
winter crops	0.07	0.03	0.86	0.01	0.01	0.02
fallows	0.12	0.19	0.03	0.53	0.06	0.07
early spring crops	0.04	0.09	0.01	0.04	0.68	0.14
late spring crops	0.03	0.07	0.01	0.07	0.07	0.75

**Fig. 1.** Dependency of total probability Q of the threshold T **Table 4.** The results of classification from class to class according to Mahalanobis distance

Class	perennial grasses	unused lands	winter crops	fallows	early spring crops	late spring crops
perennial grasses	0.51	0.22	0.11	0.09	0.06	0.01
unused lands	0.10	0.70	0.02	0.07	0.07	0.04
winter crops	0.08	0.05	0.75	0.05	0.04	0.04
fallows	0.06	0.11	0.04	0.61	0.09	0.09
early spring crops	0.06	0.06	0.03	0.11	0.55	0.19
late spring crops	0.03	0.04	0.02	0.10	0.12	0.68

5 Conclusion

The paper proposes a method of crop identification based on the algorithm for calculating estimates. The advantages of the proposed algorithm: the usage of time series with gaps, accounting of the geographical position of the field. As seen from the results of classification, ACE method gives the total probability of correct classification ($Q = 0.72$) higher than the classifier by Mahalanobis Distance (0.64). However, the value 0.72 is not satisfactory, and improvement of the quality of the classification is required. A detailed study of the testing sample showed that some of the data are unreliable: incorrect crops for some fields are given as well as the division of crops into classes was carried out nonoptimally. Therefore, the direction of future research is to study the issue of division crops into groups, so that the best classification quality is achieved at the highest possible splitting crops into classes.

6 Acknowledgements

This work was financially supported by RFBR, project № 16-37-00043_mol_a «Development of methods of using data from geoinformation systems in remote sensing data processing» and project № 16-29-09494_ofi_m «Methods of computer processing of multispectral remote sensing data for vegetation areas detection in special forensics».

References

1. Agriculture. URL: <http://sovzond.ru/industry-solutions/agriculture/> [in Russian]
2. Wardlow BD, Egbert SL. Large-area crop mapping using time-series MODIS 250 m NDVI data: An assessment for the U.S. Central Great Plains. *Remote sensing of environment*, 2008; 112(3): 1096-1116.
3. Wardlow BD. An Evaluation of Time-Series MODIS 250-meter Vegetation Index Data for Crop Mapping in the U.S. Central Great Plains. Michigan: ProQuest Information and Learning Company, 2005.
4. Brown JC, Kastens JH, Coutinho AC, Victoria DC, Bishop CR. Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data. *Remote Sensing of Environment*, 2013; 130: 39–50.
5. Plotnikov DE, Bartalev SA, Zharko VO, Mihailov VV, Prosyannikova OI. An experimental assessment of crop types recognisability using time-series of intra-seasonal spectral reflectance measurements by satellite sensor. *Current problems in remote sensing of the Earth from space*, 2011; 8(1): 199-208. [in Russian]
6. Plotnikov DE, Bartalev SA, Loupian EA. The recognition features to map arable lands based on multi-annual MODIS Earth observation data. *Current problems in remote sensing of the Earth from space*, 2010; 7(1): 330-341. [in Russian]
7. MODIS Website. URL: <http://modis.gsfc.nasa.gov/>
8. Zhuravlev JI, Kamilov MM Tulyaganov ShE. Algorithms for calculating estimates. Tashkent: “FAN” Publisher, 1974; 119 p. [in Russian]

9. Vorobiova NS, Denisova AYu, Kuznetsov AV, Belov AM, Chernov AV, Myasnikov VV. How to Use Geoinformation Technologies and Space Monitoring for Controlling the Agricultural Sector in Samara Region. *Pattern Recognition and Image Analysis. Advances in Mathematical Theory and Applications*, 2015; 25(2): 347–353.
10. Vorobiova NS, Timbay EI. Geoinformation system of agricultural lands inventory and control development. *Computer Optics*, 2009; 33(3): 340-344.