Hybrid Approach to Classification of Remote Sensing Data

Artem A. Andreiev

***Abstract* —** Remote sensing engages two well-known approaches for data classifying: supervised and unsupervised ones. Supervised classification requires training samples of classes selected by an expert. However, those classes, as a rule, are subjective as well as selected training samples are not accurate. On the other hand, unsupervised classification needs no training samples and forms objective clusters. The disadvantage of the unsupervised approach is that formed clusters, unlike the training samples selected by an expert, could not be interpreted reasonably. A hybrid approach to classification is developed to solve the problem described above. The proposed approach to classification consists of three steps: 1) subdividing training samples of expert-selected classes into objective subclusters by unsupervised classification; 2) performing supervised classification of remote sensing data using subdivided training samples; 3) merging obtained subclasses into the expert-selected classes. This approach reduces the inaccuracy and subjectiveness of the selected samples. Therefore, it increases the final classification accuracy in comparison with both of the mentioned types of classification.

Keywords—remote sensing; unsupervised classification; supervised classification; training samples subdividing; subclasses.

I. Introduction

One of the main tasks of remote sensing is satellite image classification. This classification is used for such tasks as land-cover change detection [1], forecasting gas and oil potential of subsoil plots [2] and detection of any other objects [3]. Remote sensing engages two well-known approaches for data classifying: supervised and unsupervised ones.

Supervised classification requires the use of training samples for each of the classes. These samples should be representative. The aim of this approach is to assign each of pixels to that class to which sample the pixel is similar in its characteristics. Some of the most common methods of supervised classification are Maximum Likelihood, Mahalanobis Distance, Support Vector Machine [4,5].

Unlike supervised classification, unsupervised classification methods do not require a training sample. These methods provide for the automatic distribution of pixels to a certain number of clusters basing on the analysis of the statistical distribution of pixel characteristics. Among the methods of unsupervised classification К-means [6], K-medians [7], K-medoids [8] and Cure [9] can be highlighted.

However, disadvantage of the unsupervised approach is that formed clusters, unlike the training samples selected by an expert, could not be interpreted reasonably. On the other hand, classes, selected by an expert for supervised classification, as a rule, are subjective as well as selected training samples are not accurate. To solve this problem a hybrid approach to classification is developed.

II. Method

As any method of supervised classification, the proposed approach to classification requires an input image and training samples of each of the classes. A series of requirements should be satisfied in input training samples [10]. The key requirements here are completeness, sufficiency, and purity.

A.A. Andreiev Scientific Centre for Aerospace Research of the Earth, National Academy of Sciences of Ukraine, Kyiv
(e-mail: a.a.andreev@casre.kiev.ua).

The scheme of the proposed approach to classification of remote sensing data is shown in Fig. 1.

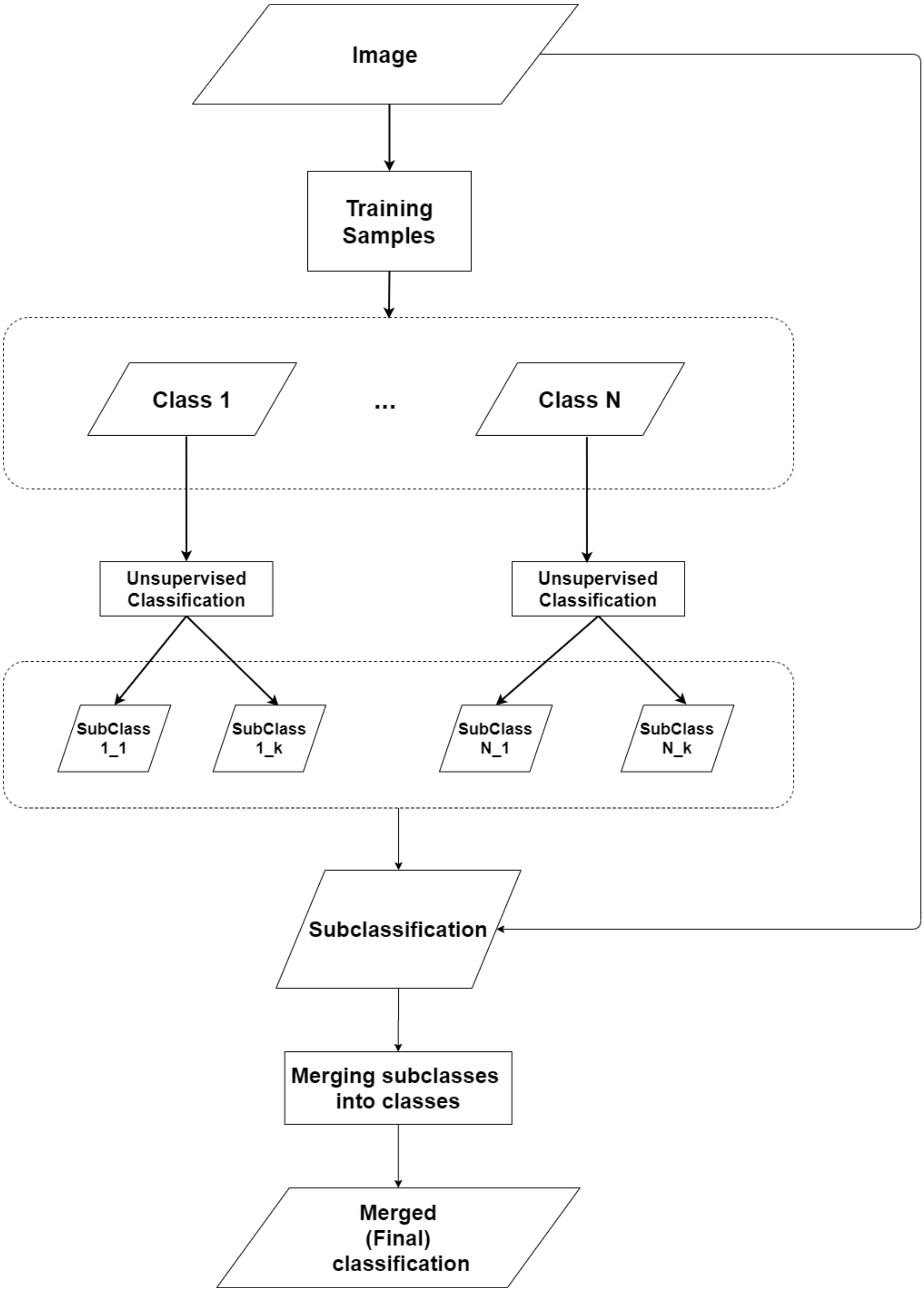

Fig 1. The scheme of hybrid approach to classification of remote sensing data

The first step is clustering of training samples. Due to subjectiveness of expert-selected classes, they should be subdivided into objective subclusters by unsupervised classification. This procedure provides a certain number of subclusters for each class of training samples. The number of subclusters could be determined taking into account the density of each class or its intersections with other classes.

In the second step, the classification of the image into subclasses is performed using subdivided training samples. This procedure is named subclassification.

The final step is to merge the subclasses into classes selected by an expert. To create merged classification, the initial class of each pixel's subclass should be determined.

III. RESULTS AND DISCUSSIONS

Proposed hybrid approach to classification of remote sensing data was applied to Landsat-OLI8 multispectral satellite image of the test site in Ivano-Frankivsk region, Ukraine (Fig. 2). Square of the study area is 247.91 km^2.

Fig. 2. Landsat-OLI8 multispectral satellite image, Ivano-Frankivsk region (Ukraine), August 9, 2018, 30 m input spatial resolution

The following six classes were selected for classification: artificial surfaces, croplands, grasslands, tree-covered areas, water bodies and bare rocks. The training samples were selected for each of the mentioned classes.

TABLE I
Description of the land cover classes

Land Cover Class	Description
Artificial surfaces	Urban public and industrial built-up areas, transport units and construction sites
Croplands	Arable land, permanent crops, fallow lands, heterogeneous agricultural areas, open soils
Grasslands	Natural herbaceous vegetation, permanent grasslands of natural origin, pastures
Tree-covered areas	Broadleaved and coniferous forests, roadside tree lines, areas with tree cover more than 30%
Water bodies	Rivers, reservoirs, streams
Bare rocks	Sand and gravel sediments of the river valleys

Initial classification was obtained by using maximum likelihood method of supervised classification. Result of classification is shown in Fig. 3.

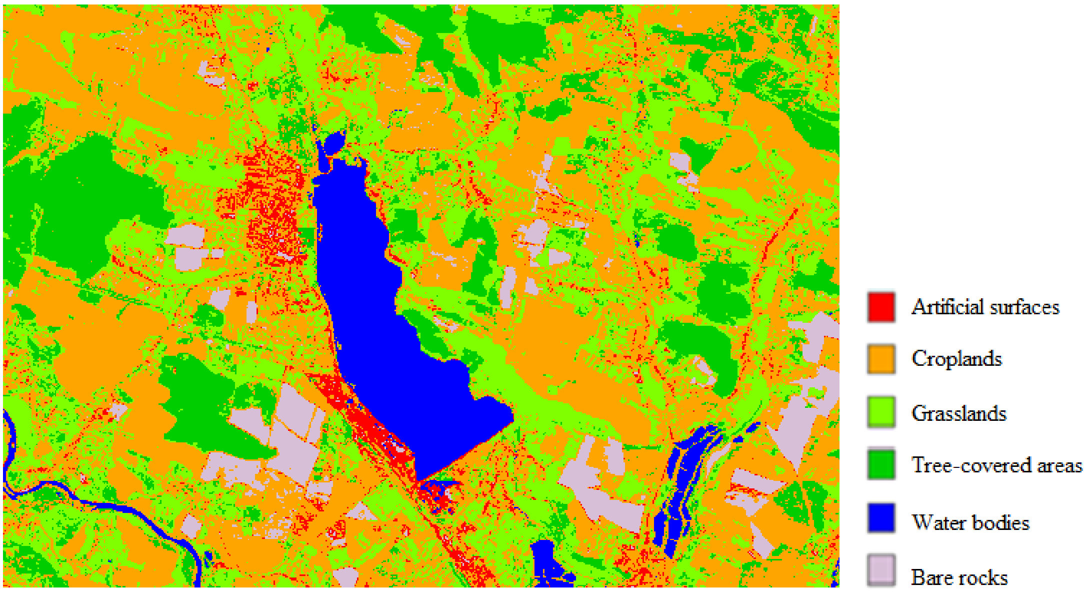

Fig. 3 Initial classification of study area

The proposed hybrid approach to classification involves unsupervised classification for clustering the training samples. For this example, K-Means was chosen as a method of unsupervised classification. This method requires a number of clusters for each of the classes. The chosen numbers are shown in table 2.

TABLE II
Number Of Subclasses For Each Of The Classes

Class name	Number of subclasses
Artificial surfaces	2
Croplands	5
Grasslands	4
Tree-covered areas	2
Water bodies	1
Bare rocks	2

After subdividing training samples into clusters, the procedure of subclassification was performed. The method of supervised classification used for this procedure was maximum likelihood one. The result is shown in fig. 4.

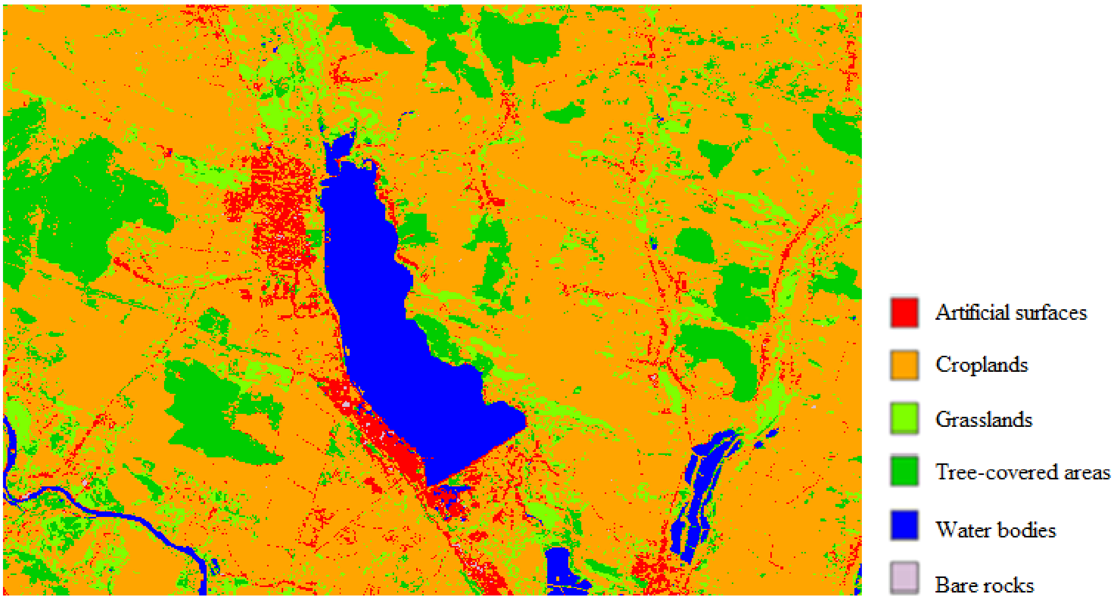

Fig. 4. Classification of study area using proposed approach

According to [11], the procedure of accuracy assessment was performed. The unproportionate stratified random sampling was used to select 10 pixels for verification from each of the classes. The total amount of the verification sample was 60 pixels. As the result of this procedure, producer's accuracy (PA), user's accuracy (UA) and overall accuracy (OA) were obtained for each classification.

OA of initial classification of study area is 63%. Values of UA and PA are shown in table 1.

TABLE III
Accuracy Assessment Of Initial Classification

		Truth data							
		1	2	3	4	5	6	Total (pixels)	UA (%)
Classification result	1. Artificial surfaces	6	0	0	0	0	0	6	100.00
	2. Croplands	2	7	2	0	0	10	21	33.33
	3. Grasslands	1	2	7	2	0	0	12	58.33
	4. Tree-covered areas	0	1	1	8	0	0	10	80.00
	5. Water bodies	0	0	0	0	10	0	10	100.00
	6. Bare rocks	1	0	0	0	0	0	1	0.00
	Total (pixels)	10	10	10	10	10	10		
	PA (%)	60.00	70.00	70.00	80.00	95.83	0.00		

OA of classification of study area by proposed approach is 83%. Values of UA and PA are shown in table 3.

TABLE IV
Accuracy Assessment Of Classification Using Proposed Approach

		Truth data							
		1	2	3	4	5	6	Total	UA (%)
Classification result	1. Artificial surfaces	7	0	0	0	0	1	8	87.50
	2. Croplands	2	7	2	0	0	10	10	80.00
	3. Grasslands	1	2	7	2	0	0	13	69.23
	4. Tree-covered areas	0	1	1	8	0	0	9	88.89
	5. Water bodies	0	0	0	0	10	0	10	100
	6. Bare rocks	1	0	0	0	0	0	10	80.00
	Total	10	10	10	10	10	10		
	PA (%)	70.00	80.00	90.0	80.00	100.00	62.50		

The proposed approach provides classification, which OA is 20% higher than initial one. It is worth noting that UA and PA of "Bare rocks" class have increased from 0% to 62.50% and 80% compared to initial classification, respectively.

IV. Conclusions

The presented hybrid approach to classification of remote sensing data is developed to solve the problem of subjectiveness of training samples selected by an expert. This approach allows to combine both supervised and unsupervised types of classification.

It was shown in the example, that subdividing initial classes into subclasses could significantly increase an accuracy of classification if the appropriate number of subclasses for each of the classes is determined.

Further research should be aimed at developing an algorithm for determination of the most appropriate number of subclasses for each of the classes.

REFERENCES

[1] S. A. Stankevich and A. A. Kozlova, “Long-term land cover change computer-aided mapping by remote sensed imagery”, *2015 International Conference on Information and Digital Technologies*, Zilina, 2015, pp. 339-341.

[2] M. O. Popov, M. V. Topolnytskyi, O. V. Titarenko, S. A. Stankevich, and A. A. Andreiev, “Forecasting Gas and Oil Potential of Subsoil Plots via Co-analysis of Satellite, Geological, Geophysical and Geochemical Information by Means of Subjective Logic”, *WSEAS Transactions on Computer Research*, 2415-1521, Vol. 8, 2020, pp. 90-101.

[3] S.A. Stankevich and M.I. Gerda, “Small-Size Target’s Automatic Detection in Multispectral Image using Equivalence Principle”, CERes Journal, Volume 6, Issue 1, Pages 1-9, 2020

[4] B. Tso and P. M. Mather, *Classification Methods for Remotely Sensed Data*, 2nd ed.; CRC Press: Boca Raton, USA, 2009.

[5] L. Bruzzone and B. A. Demir, “A review of modern approaches to classification of remote sensing data”, in *Land use and land cover mapping in Europe*, I. Manakos, M. Braun, Eds. Springer: Dordrecht, Netherlands, 2014, pp 127–143.

[6] J. Macqueen, “Some methods for classification and analysis of multivariate observations”, in *5-th Berkeley Symposium on Mathematical Statistics and Probability*, pages 281–297, 1967.

[7] A.K. Jain and R. C. Dubes, *Algorithms for Clustering Data*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1988.

[8] L. Kaufman and P.J. Rousseeuw, *Finding Groups in Data: an introduction to cluster analysis*. Wiley, 1990.

[9] S. Guha, R. Rastogi, and K. Shim, “Cure: An efficient clustering algorithm for large databases”, *SIGMOD Rec.*, 27(2):73–84, June 1998.

[10] W.G. Cochran, *Sampling Techniques*. New York: John Wiley & Sons, 1977.

[11] M.O. Popov, “Methodology of accuracy assessment of classification of objects on space images”, *J. Autom. Inf. Sci.* 2007, *39*, pp. 1-10 (In Russian).