ELEKTRONIKAS UN DATORZINĀTŅU INSTITŪTS

INSTITUTE OF ELECTRONICS AND COMPUTER SCIENCE

Monitoring Earth Resources: Nonparametric Classification of Remote Sensing Data

Institute of Electronics and Computer Science
14 Dzerbenes St., LV-1006, Riga, Latvia

Abstract

The task of classifying multispectral image pixels into selected land cover categories is considered. A robust classification approach is described for designing a practicable classifier using completely nonparametric unsupervised clustering with consequent association of the clusters with target categories using multiple sources of testing and training data. The clustering used is primarily based on ranking and grouping. Completely nonparametric cluster union and cleaning procedures are presented. The software implementation and complexity of the methodology are discussed. The approach aims at getting the highest possible classification accuracy under real conditions for images with more than 100 million pixels.

Introduction

The necessity to classify objects into predefined categories in the area of satellite imagery is widely acknowledged [1-3]. For several reasons a number of classification methods have been designed and tested using limited, arguably insufficient data. In addition, most of them use strict assumptions about the distribution and quality of the data, testing and training data [1-3]. Global coverage (high resolution regularly renewed) satellite data from a multitude of sensors recently became publicly available in downloadable format [4], however reliable testing and training data are still scarce. Under more adequate assumptions we propose foundations for a new approach to the classification of satellite image objects. The robustness of the approach and absence of assumptions about the distribution of the data ensure applicability of the methodology to a wide variety of satellite images. The quality of the testing and training is also taken into account.

1. Categories of interest

The categories of interest are considered to be wide classes of objects. Examples are coniferous forest, deciduous forest, agricultural field, bog, water. Those generally consist of a large variety of subclasses. For example, agricultural fields consist of various kinds of cultivated and uncultivated fields, and forests consist of different species and varieties of trees. Some categories, such as forests, have additional spatial criteria describing them. These criteria vary in different countries and may change with time. All subclasses must be taken care of in order to accurately classify images into such categories. It is assumed that the categories don’t overlap by definition; otherwise the intersection of the overlapping categories can be processed as separate category.

2. Classification approach

The major principle was that all assumptions accepted both explicit and implicit must be as realistic as possible.

Classification approach was elaborated on the basis of the following considerations. The data from satellite image, related to a certain category, have some distribution in the multidimensional feature space. The distribution is unknown and usually multimodal. The dense areas of distribution for the particular category are not localized but scattered in the whole feature space. It reminds us of a set of "isles" scattered in space and having various shapes (see Figure 1). It is assumed that subclasses of different categories – "isles" – are separable from each other along pixel density gradient lines. The distributions change from image to image. It might occur that some pixels of the image are mixed i.e. they correspond to several categories or subclasses. Sentinel-2 multispectral data [5] are used in this article as exemplary.

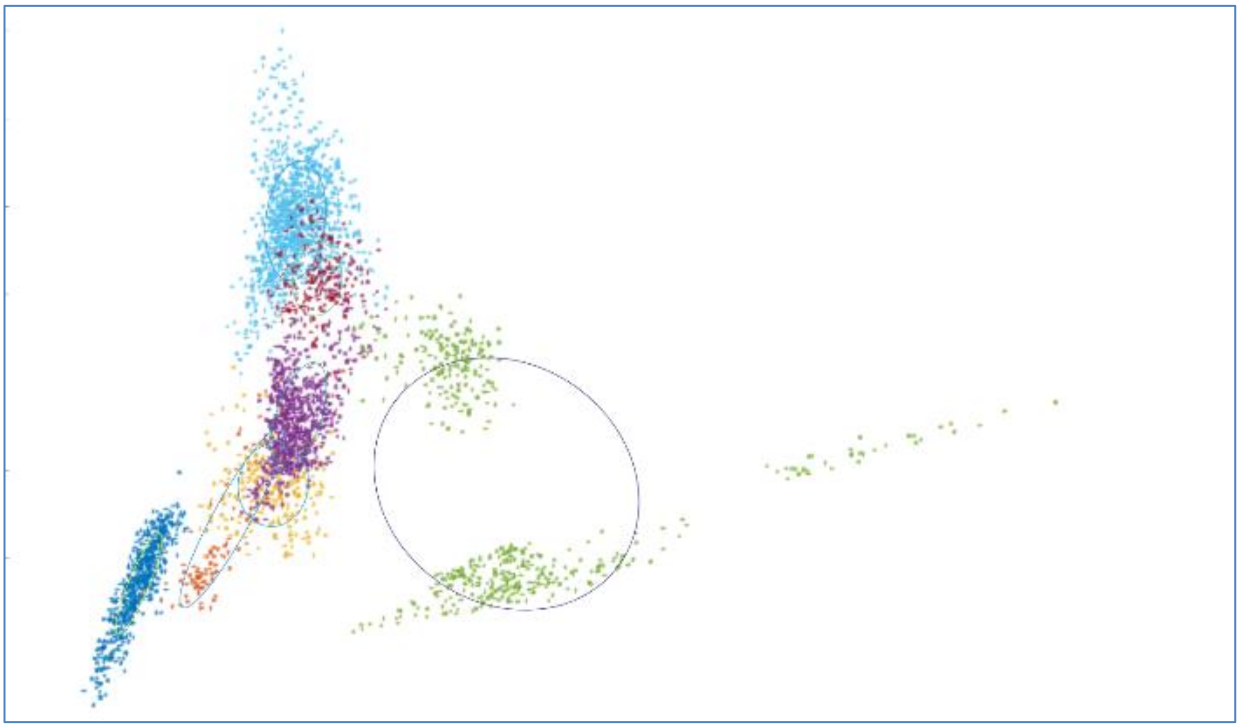

Figure 1. Example of distributions in the multispectral feature space. Legend: dark blue – coniferous trees, red – deciduous trees, yellow – apple trees, purple – fields (type I), green – fields (type II), light blue – fields (type III), dark red – bushes. Ellipses are fixed Mahalanobis distances obtained by assuming Gaussian distribution.

The testing and training data from single source is considered to contain significant imprecisions. In an extreme case, up to 20% of testing and training data are considered corrupt. The sources of potential problems are:

– testing and training data are old; there is too small number of samples (if any) available for some subclasses;
– the present pixel size does not provide for the necessary spatial resolution;
– forgery;

– preprocessing and sensor issues;
– random errors;
– high cost of the testing and training data acquisition and verification.

For large areas, for example a country, currently it is seldom (if at all) possible to regularly get verified test and training data from single source, and many data sources currently are not public.

3. Robustness of classifier

Robustness, in general, is synonymous to sustainability, stability, being strong, performing well in the large variety of situations, overcoming non-standard conditions. Hereby it is assumed that any practicable classifier should meet the following requirements:

1) Stability. Stability to the small random changes in the whole data set. In case of small deviations in the data set, (including testing and training data) the accuracy degradation of the classifier should also remain small, even if all pixels of the set suffer deviation simultaneously.

2) Robustness to outliers. Individual pixels, including pixels from the testing and training data, very unusual or extremely corrupted does not produce perceptible degradation of classification accuracy.

3) Invariance. Classification accuracy should be invariant to data deviations due to variations of lighting and other circumstances allowed at the data acquisition process.

4) Computation time limitations. Classifier software should provide for completion of all the computation tasks in predefined time limits for any set of valid input data.

For research and development some of above requirements are not of crucial importance. They are important when the classification tools are part of toolbox used in the field of current economic (macroeconomic) problems.

4. Basics of methodology

To classify the image data with properties mentioned in section 3 and meet the robustness requirements described in section 4, the nonparametric statistics [6,7] is adequate underlying theory. Nonparametric statistics do not make any assumptions concerning distribution laws of the data in question. The ranking and the grouping are the main principles used in the methodology we propose. The following approach is proposed to address the problems of testing and training data described in the section 3: the unsupervised clustering is performed first, and then the obtained clusters are associated with the categories for which the testing and training data are available. Due to the shortage of testing and training data, association is partly done manually by means of utility software.

Pixels which cause discrepancies between the testing and training data and the unsupervised clustering results are extracted and processed separately. The rest of the data, which are not involved in the mentioned discrepancies are used as the additional information to resolve such discrepancies. Manual use of relevant utility software is part of the process.

It is usual that for the particular category or subclass some methods provide more precise results than others. We recognize that it is feasible to involve several methods. That is, we accept compounding results of different methods.

The pixel set on which the results do not contradict each other is considered reliably classified (although it does not exclude cases where some methods are simultaneously mistaken – the “blind spot”).

Let us call the set on which results of methods contradict to each other the “contradictory set”. It is proposed to reclassify the contradictory set using the reliably classified set as additional information about the categories, and remove the similarity from the work of the methods (i.e. remove covariance) in a nonparametric way. The idea is akin to how artificial neural networks (ANN) [2] use outputs of the multiple agents and re-learn using the results. The difference is that the compounding of results aims to work in a robust way and benefits from prior knowledge about performance of different methods; reliably classified set is used as additional testing and training data. Even if robustness is absent in some regions of the feature space, the compounding of results allows to find these regions relatively easy and deal with them separately, when for ANN, especially the ones that are large and changing from image to image, this could be difficult [8].

It is proposed to employ a similar approach when using multiple sources of the testing and training data and the data of different nature (e.g. multispectral and SAR data).

5. Nonparametric similarity between pixel and other pixels

To compute nonparametric similarity between a particular pixel and other pixels in a meaningful way, it is proposed to rank the other pixels with respect to that pixel. The ranking is performed as follows: for the value of the current pixel the 11 closest values of pixels (including the pixel itself) have rank 1 – they are the most similar, next closest 10 values of pixels have rank 2, and so on. Such ranking is done for each pixel (see section 10 on complexity). Such ranking is easy in the one-dimensional space. In the multidimensional feature space, we must account for the covariance (i.e. for the common information provided by different features) in the neighborhood of the pixel in the space. It should be accounted also in the nonparametric way, which significantly complicates the computation. As an intermediate substitute, the locally parametric similarity based on the Mahalanobis distance using the covariance matrix in the neighborhood of the pixel was used. It is not possible to obtain the practical robust classifier without application of completely nonparametric similarity. The similarity here is not necessarily a distance in mathematical sense, but the properties of reflexivity and non-negativity hold.

6. Growth and union of clusters

Initial cluster is created starting from some pixel by adding the neighbors with the rank 1 (i.e. pixels with largest similarity between the initial and other pixels). At the start there are as many clusters as pixels, see section 10 for adjustments. Clusters are grown in an iterative way. The cluster is united with other clusters when the cardinality of their intersection is greater than 80% of the minimum cardinality of the pair. In simple words, clusters are united when they overlap too much. If for all clusters all intersections are with smaller cardinality, then the cluster growth is performed as follows. For each pixel in the cluster, pixels with rank 1, i.e. the most similar ones, are added to the cluster. After that the union is continued. In practice, it is impossible to grow or unite large clusters if the clusters count is also large, hence only part of clusters is grown or united, see section 10. During this procedure, one pixel can be included in many clusters; the pixel is assigned to the group of clusters to which it is close. The principle is “do not rely on the single value”, that is central in statistics. The idea of the cluster growth and union is similar to maximum likelihood and Bayesian classifiers as well as some unsupervised learning algorithms, where the cluster is described by a distribution function (mostly by a center and covariance in the area around the center), and the distances between clusters and pixels are computed based on the assumption of the distribution in the area; then the closest (most likely) cluster is assigned to the pixel. The difference is that here the cluster is grown in the direction of where the nearest neighbors of its pixels are, while the neighbors considered to be computed in a nonparametric way. The criteria for the pixel inclusion in the cluster and the cluster union doesn’t rely on the assumptions about the distribution, also multiple close clusters can be assigned to the pixel. The growth and union are similar to DBSCAN [9]. The difference is that multiple clusters can be assigned to the pixel and clusters that overlap extensively are united.

When some cluster is grown, this growth happens in the direction of increasing density of pixels, because, for each pixel in the cluster, the neighbour pixels are more likely to be from the region with highest density in the closest regions in the feature space. If the clusters are united on the base of their intersection, as was described above, then the result of the union will likely be the shape containing the closest dense regions.

Still, the use of the nonparametric similarity is mandatory for the operation.

7. Cluster cleaning

Usually the unsupervised classification methods don't allow cleaning of the resulting clusters. The clusters can contain false positives (pixels erroneously included in the cluster) and false negatives (pixels erroneously not included in the cluster). The assignment of the multiple likely clusters to the pixel allows performing of the cleaning. The cleaning is done after the previous iteration of growth or union. If the cardinality of the intersection of the multiple clusters exceeds 95%, then the clusters are considered to be multiple instances of the single cluster. From these multiple instances of the single cluster only such pixels are taken, whose relative frequency among instances is more than 95%. I.e. pixels that occur frequently enough among instances are taken, and pixels that occur very rarely are thrown away from the cluster. The taken pixels create the new, cleaned, cluster. The empty and equal clusters are eliminated. Note that only clusters with similar sizes are united. The parameters provided here are not necessarily optimized to equally account for both types of errors. The union and growth continues after the further cleaning doesn't change any cluster. The union and growth stops when only a single cluster remains or there are no changes in clusters.

8. Redistribution of pixels in clusters

During the earlier described process of cluster formation it can easily occur, that some pixels appear included simultaneously in several clusters. Most of the tasks require that in the end each pixel is associated with one and only one cluster. To achieve status where such requirement is satisfied, the following iterative process is used.

For each pixel, which belongs to more than one cluster, is calculated the number of neighbors in each cluster. If the number of neighbors in some cluster is smaller than 85% of the total number of neighbors, the pixel is excluded from that cluster. Thus we obtain a new set of clusters and we can perform the next iterative step in the same manner. In such iterative process the number N of pixels associated with more than one cluster decreases. If after some iteration step N=0 the task is completed. If the N=0 cannot be reached, then additional procedures are used, not discussed here.

Our method is similar to the support vector machine (SVM), where pixels on the borders between clusters are separated in the optimal way given some assumptions about the shapes of the borders. The difference is that our method does not require such assumptions. The mixed pixels are not accounted for in this procedure. Further the union and growth of clusters is made on the set of clusters where some number of pixels may be associated with more than one cluster.

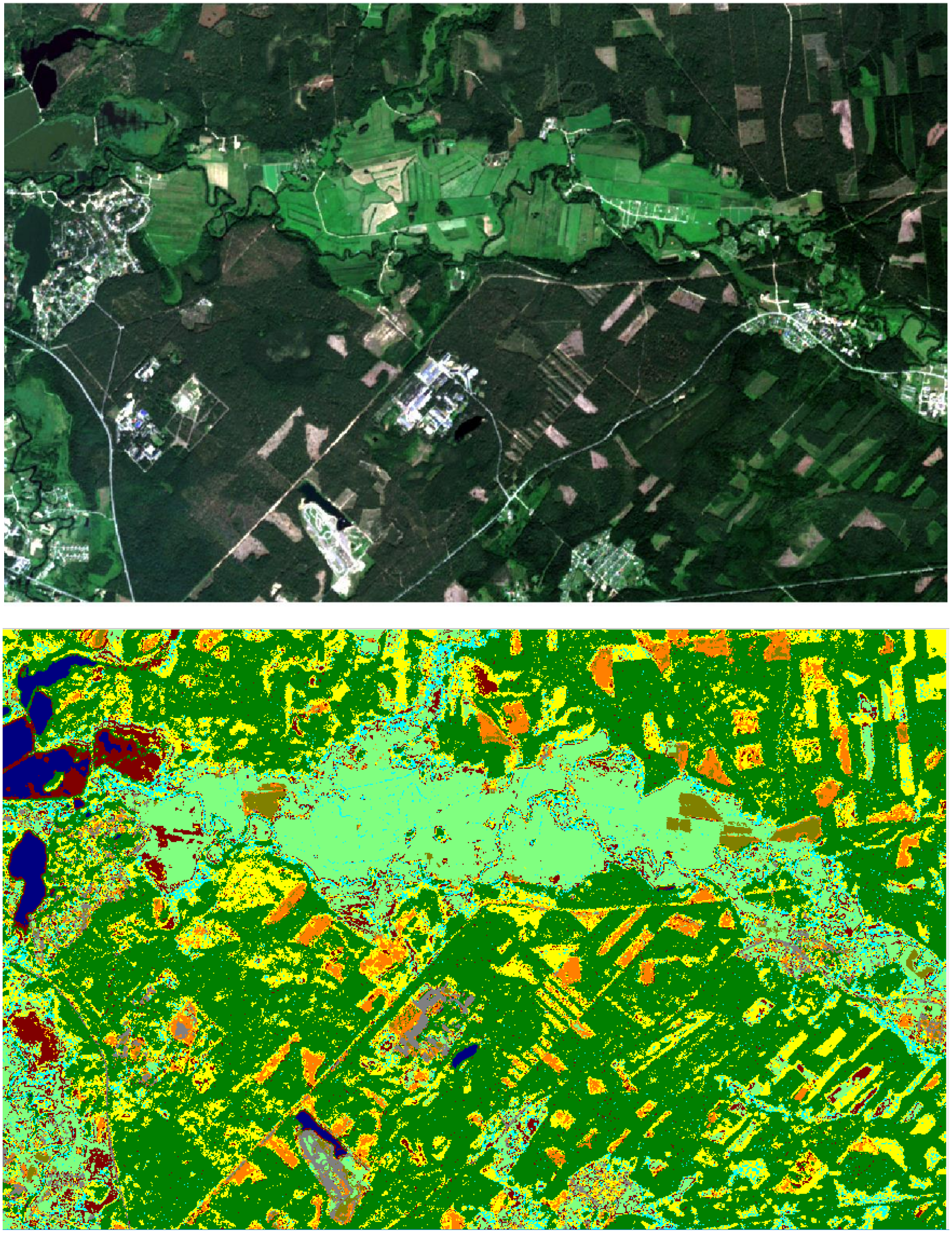

Figure 2. Initial image (upper) and the result of the per-pixel classification (lower) using the methodology proposed in this paper. Bands 2 – 8, 8A, 11 – 13 of the Sentinel-2 image taken on 10th September, 2016 and containing 480 000 pixels were processed. Center coordinates: 56,97°N, 24.43°E. Colors: dark green – coniferous forest, yellow – young forest, orange – cut forest, light blue - deciduous forest, light green – meadow green, dark red – wetland, dark blue – water, brown – arable land, gray – other (roads, buildings, etc.)

9. Preprocessing, computational complexity and adjustments

The distance computation time grows at least quadratically with the number of pixels. Fortunately the task is convenient for parallel computing; at cluster union and growth parallel computing is possible, but it is more sophisticated. The speed of the cluster union is significantly increased by processing only those clusters which have been previously changed, but again computation time grows at least quadratically with the number of pixels. In similar manner grows the necessary memory size. Note that insufficient RAM will terminate or significantly slow down the computation. However, most operations consist of ranking, sorting and arithmetic operations using only integer numbers or are binary, which allow for significant optimizations. Binary sparse matrices are used to store the data of pixels associations with clusters, and for other purposes. Use of sparse matrices allows for significant increase of computation speed and decrease memory requirements. MATLAB was used but Julia looks more promising. The threshold is put on the matrix size and, to decrease the necessary memory size and increase computation speed, the union or growth of clusters is not performed if the threshold was exceeded. The threshold increases when cluster count decreases, which allows for relatively homogenous growth of clusters. The neighborhood calculation for each pixel can be speeded up by pre-calculating large enough neighborhood (e.g. 15000 pixels) using the Euclidean distance.

10. Discussion and further work

The approach doesn't account for missing values, those considered to be relatively simple to deal with by replacement or exclusion. To account for the mixed pixels, the fuzzy classification is the general approach [2]. It is considered here that the mixed pixels can be separated into a separate set and processed additionally. It might be reasonable to quantify the levels of belonging to the categories when extending the nonparametric approach to include mixed pixels. The hyperspectral imaging here is not accounted for. The proposed classification is considered here as per-pixel classification. Additional spatial criteria are considered easy to implement in comparison to design of the per-pixel classifier. The approach is designed to work with images containing up to 100 million pixels and more, meanwhile it is tested on images containing up to 480 000 pixels with future expansion and reliable accuracy assessment in progress. For the time-sensitive applications of image processing, not necessarily for images originated from satellite, such as images processed by self-driving cars, the methodology might need significant change to increase speed.

References

[1] Landgrebe, D.A. 2003. *Signal Theory Methods in Multispectral Remote Sensing*, John Wiley & Sons, Inc., Hoboken, New Jersey, 2003. DOI= https://doi.org/10.1002/0471723800 .

[2] Lillesand, T., Kiefer, R.W., Chipman, J. 2008. *Remote Sensing and Image Interpretation 6th Edition*, John Wiley & Sons, Inc., USA.

[3] Grahn, H.F., Geladi P. 2007. *Techniques and Applications of Hyperspectral Image Analysis*, John Wiley & Sons, Ltd., England. DOI= https://doi.org/10.1002/9780470010884 .

[4] Copernicus Open Access Hub, https://scihub.copernicus.eu/, last visited: 8 Jan. 2018.

[5] ESA Sentinel Online, User Guides, Sentinel-2 MSI, https://earth.esa.int/web/sentinel/user-guides/sentinel-2-msi/resolutions/spatial, last visited: 8 Jan. 2018.

[6] Hollander, M., Wolfe, D.A., Chicken, E. 2014. *Nonparametric Statistical Methods 3rd Edition*, John Wiley & Sons, Inc., Hoboken, New Jersey.

[7] Wasserman, L. 2006. *All of Nonparametric Statistics*, Springer science + Business Media, LLC, USA.

[8] Zheng, S., Song, Y., Leung, T., Goodfellow, I. 2016. Improving the Robustness of Deep Neural Networks via Stability Training, *CVPR 2016, 2016 IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, pp. 4480-4488, 2016. DOI= https://doi.org/10.1109/CVPR.2016.485 .

[9] Ester, M., Kriegel, H.-P., Sander, J., Xu, X. 1996. Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise, *KDD-96 Proceedings*, AAAI Press. pp. 226–231.