

AUTOMATIC DETECTION OF GREY INFRASTRUCTURE BASED ON VHR IMAGE

J. Pluto-Kossakowska¹

¹ Dept. of Geodesy and Cartography, Warsaw University of Technology, Poland – joanna.kossakowska@pw.edu.pl

Commission III, WG III/7

KEY WORDS: Machine learning, Maximum Likelihood, Random Forests, WorldView-2, Topographical Database update.

ABSTRACT:

Grey infrastructure is an integral part of the urban environment. Continuous modernization of architecture, construction, routes or services in that region leads to more and more new grey infrastructure appearing. The reason for this are constant migrations of people, dissemination of a healthy lifestyle or improvement of its level. Its growth is particularly noticeable in agglomerations where keeping the balance between sealed and vegetated area is very much concerned. Therefore, it is necessary to constantly monitor changes over time and thus update the databases containing information on land cover such as the Topographical Database. For this purpose VHR images were processed and analysed in terms of detection efficiency of topographical objects defined as grey infrastructure. This study presents the results of an analysis of the possibility of updating the land cover classes in the Topographical Database based on WorldView-2 satellite images. The methods used to detect grey infrastructure come from a machine learning approach such as Random Forests and parametric Maximum Likelihood classifier, resulting at a 90% level of accuracy. The other aim of the work was to analyse changes in the grey infrastructure on the basis of the Topographic Database at scale 1:10000 using a VHR satellite image. The analysis of its changes was carried out on the dynamically developing city of Warsaw.

1. INTRODUCTION

Grey infrastructure includes all topographic objects that create impervious surfaces. These are: roads, pavements, parking squares and buildings as well. Due to the variety of grey infrastructure objects, spatial data providing the accurate and detailed information is needed to perform such an analysis. VHR satellite images with GSD around 1 m are data that is sufficiently accurate to assess the current state of coverage by grey infrastructure objects. Figures 1, 2 and 3 show the changes in grey infrastructure as new buildings, squares or even roads still under construction (within yellow ellipses). Red lines or patches show grey infrastructure from the Topographical Database.



Figure 1. High buildings in dense urban area (test site “U”), the new ones indicated within yellow circle.

The goal of the research was to check the possibility of grey infrastructure detection in a VHR image using machine learning techniques and database updating.



Figure 2. New road under construction. The red lines define objects in the Topographical Database.



Figure 3. New buildings at residential area (test site “W”) in yellow lines, red patches – build-up area from Topographical Database.

1.1 Topographical Database (1:10000)

The Topographical Database (TD) at scale 1:10000 contains groups of classes like: water, transport, buildings, land cover, land use, protected areas and administrative borders. TD is very useful for many purposes in spacial management, especially for creating cartographical products such as maps, visualizations, spatial analysis or 3D models. It is a national "GIS" product in Poland updated periodically from different sources (i.e. another databases) by national and local authorities. To support any decision concerning the local or regional area it must be up-to-date. The updating process takes time and requires a lot of resources. Therefore any method to speed up the whole process is worth to consider and test. The existing land cover database update methods are mainly based on aerial images photointerpretation and manual editing of objects. Supporting this process, even partially by indicating that new objects or places need to be revised, can significantly accelerate it. The most effective method seems to be automatic classification of VHR satellite images, whose spatial resolution (GSD) corresponds to the detail of the database at a scale of 1:10000.

1.2 Classification algorithms

Automatic and specially machine-learning classification has become a major focus of the remote-sensing literature by many reviewers (e.g. Belgiu and Drăguț 2016; Maxwell et al., 2018). Machine-learning algorithms are generally able to model complex class signatures, can accept a variety of input predictor data, and do not make assumptions about the data distribution (i.e. are nonparametric). A wide range of studies have generally found that these methods tend to produce higher accuracy compared to traditional parametric classifiers, especially for complex data with a high-dimensional feature space (Maxwell et al., 2018).

Many different studies indicate different classifier accuracy and it is difficult to select an effective method a priori. However, most of them point to Support Vector Machine (SVM), Random Forests (RF), and boosted Decision Trees (DT) generally as to be reliable classification methods (Maxwell, 2018). RF requires fewer training samples and is faster to implement. The main advantages are: quick classification, understandable operating procedure, simple final form of the classification trees, allowing easy classification of new objects and resistance to outliers. The RF algorithm uses two basic parameters: k - the number of trees and m - the number of rules that can be created in each tree to make a decision. By reducing the number of rules, each tree is less strong and the correlation between trees is weaker, which increases the accuracy of the model. Therefore, it is important to optimize the k and m parameters to reduce the errors (Rodríguez-Galiano et al., 2012).

On the other hand parametric classifiers such as Maximum Likelihood (ML) are still very popular in many remote sensing software and it works for classes with normal distribution. The ML algorithm calculates the probability of belonging to a given pixel for each class (sample). The pixel is assigned to the class for which the probability is the highest. This method gives good results when the training samples have a normal distribution, so they must be selected with great care. If the class is highly variable (i.e. grey infrastructure), this algorithm can classify with a commission (Adamczyk, Będkowski, 2007). Therefore two classifiers were used to test the grey infrastructure recognition and its accuracy: Maximum Likelihood (ML) and Random Forests (RF). Both algorithms belong to a supervised classification approach, which requires training samples.

2. DATA AND METHODS

2.1 Test sites

The research area covered a part of Warsaw city (Poland) in two zones with varying degrees of urbanization. The first area is the highly urbanized Ursynów district with dense and compact buildings (named: "U"). There are a lot of tall multi-family buildings here (fig. 1). The second area is the Wilanów district (named: "W") of a more residential character with loose, low-rise buildings (Fig. 3). In both areas there are many new investments, such as family houses, new expressway (Fig. 2) and commercial or service buildings as well. Each site covered approximately 6 km².



Figure 4. Test areas in Warsaw city (Poland): Wilanów district "W" and Ursynów district "U"

2.2 Data sets

For grey infrastructure detection a WorldView-2 image was used (fig. 4). This image is a collection of 8 multispectral bands with GSD=1,80 m, acquired on the 2nd of September 2018. This image was pre-processed using the Gram-Schmidt pan-sharpening method (Craig, Brower, 2000) resulting in 8-bands with GSD=0,40 m.

In addition, a Topographical Database (TD) at scale 1:10000 was used to compare the results of the classification and to evaluate the accuracy. TD was older (2012) than the satellite image and the changes in infrastructure were noticeable (fig. 4). TD consists of different classes (features), but only part of them can determine the grey infrastructure. These classes are originally modelled as (fig. 5):

- polygons: buildings, places and squares, area under railroads or airports;
- lines: roads, pedestrian and bicycle paths, roundabouts.

Therefore the lines features need to be converted into polygons using buffer processing to compare with the classification results. There were also independent sets of random points (for each test site elaborated independently) used to assess image classification accuracy.

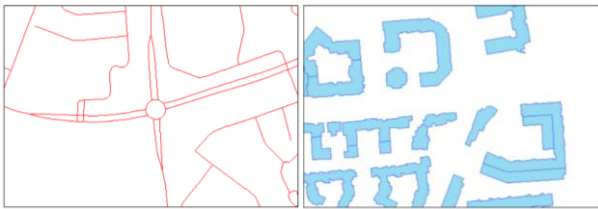


Figure 5. Models of grey infrastructure features in TD: lines (e.g. roads) and polygons (e.g. buildings).

2.3 Data processing and image classification

The image data was processed using a common approach to multispectral classification workflow in a few steps: 1. training, 2. calculation with classifiers and post-processing, 3. accuracy assessment, for both test sites separately.

1. The training step included elaboration of training samples and signature analysis. A total of 44 training fields were identified, with an area of 13 m², which is 0.2% of the total “W” area and 50 training fields for “U” test area. They included grey infrastructure features such as: buildings, roads, pavements, etc. but also natural features like: water, vegetation or bare soil (fig. 6).



Figure 6. Training samples for image classification: buildings, roads, bare soils, vegetated area and water.

The signature analysis was performed for both training sets, what led to merging or eliminating some of them. The signatures of grey infrastructure are quite similar to bare soil signatures (fig. 7), therefore, it might cause confusion between these two classes. The others were well separated and all of them have normal distribution. The average values of the standard deviation do not exceed 10 units, the minimum values fluctuate within 2-3 units, while the maximum values fluctuate to less than 20. The lowest values of the standard deviations can be observed in bands 2 and 5, and the largest in band 7.

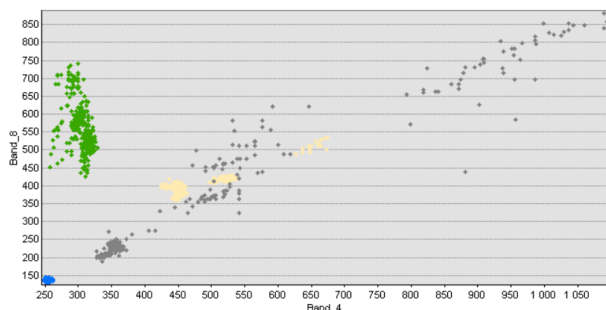


Figure 7. Example of scattergram for band 3 and 4 showing spectral distribution of the signatures: grey infrastructure (grey), bare soil (beige), vegetation (green), water (blue).

2. In the classification step, two different classifiers: Maximum Likelihood (parametric) and Random Forests (non-parametric), were used to test and check the efficiency in grey infrastructure detection.

The ML algorithm was performed in many variants using different values of reject fraction and a priori probability weighting. A reject fraction determines whether a cell will be classified based on its likelihood of being correctly assigned to one of the classes. A priori probability weighting specifies how a priori probabilities will be determined: “equal” — all classes will have the same a priori probability or “sample” — a priori probabilities will be proportional to the number of cells in each class relative to the total number of cells sampled in all classes. The best results were achieved for reject fraction equal to 0,1 and for probability weighting as “sample” for both test sites. Later, these results were used to compare with RF classification.

Random Forests classification was also performed in many variants, testing different values like: the number of trees (k), the number of rules for each tree (m) and the number of samples to use for defining each class (s). Increasing the number of trees (k) usually leads to higher accuracy rates, although in this test there was no improvement. The best results were achieved for k=100 and k=200. The maximum depth of each tree in the forest is another way of saying the number of rules (m) each tree is allowed to create to come to a decision. Trees with m=50 setting gave the optimum results. Increasing this number did not any better effects. The number of samples to use for defining each class (s) was tested as well, starting from 0 (all the samples from the training sites to train the classifier) to 1000 which is recommended for non-segmented images. Further increasing this number did not bring better results. After testing all RF variants, only two with the best results with k=100 and k=200 (m=50, s=1000) were taken to compare with ML.

In the post-processing stage a majority filter was used to eliminate random individual pixels that appear within other land cover classes. This helped to clean up and partially homogenize the features which appeared in the images. Only single pixels were dissolved to make objects more uniform (fig. 8).

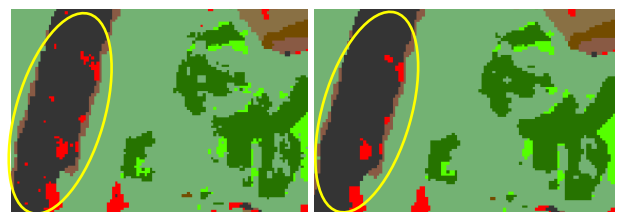


Figure 8. Example of majority filter work. Changes indicated within yellow lines.

After filtration, the aggregation of the classes into main groups was made. These classes represent the grey infrastructure like: buildings, roads and pavements, squares, but also natural green-blue infrastructure like: grasslands, bushes, forests, water.

3. RESULTS

3.1 Classification results

The final results allowed to compare both approaches ML and RF to grey infrastructure detection. Figure 9 shows the classified images for the “W” district using ML and RF (100 trees). The classification results are described and exemplified in detail in figures 10-11.



Figure 9. ML (left) and RF- 100 trees (right) results for “W” area.

In figure 10 there are visualized buildings classified using ML algorithm (upper image) and using RF algorithm (down image). In ML images buildings appeared more homogenized and smooth while in RF images some parts of the buildings were classified as water (fig. 10-a) what was in fact a shadow (blue colour). On the other hand there are some buildings in ML images not detected at all, while in RF images the same buildings are recognized at least partially (fig. 10-b). The buildings detection is very dependent on spectral properties and similarities to road cover. Yet both classes belong to one group of grey infrastructure.



Figure 10. Buildings obtained from ML (up) and RF (down) classification.

The transport paths are recognized by both ML and RF methods with similar effect. It seems that in RF images the roads and pavements are more consistent and reliable (fig. 11-a). On the other hand in RF images there are more commission errors for roads in the grassland class (compare fig. 11-b).

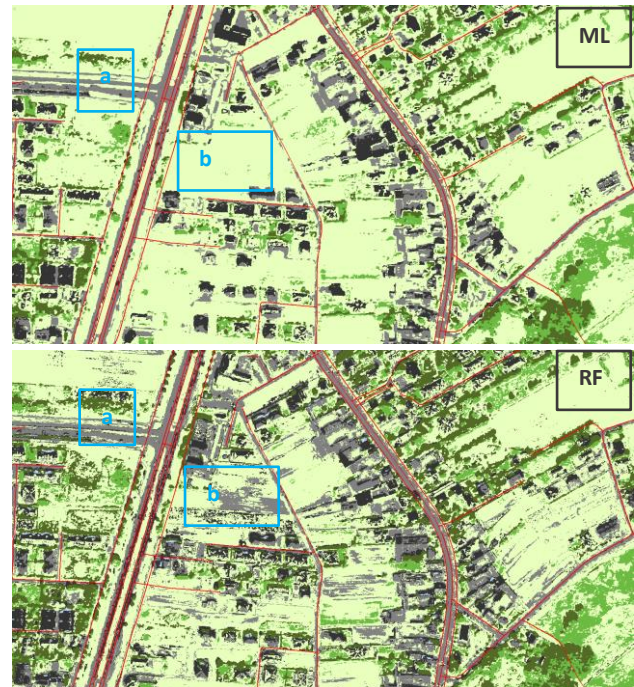


Figure 11. Roads obtained from ML (up) and RF (down) classification.

It seems that ML is more accurate for green cover while RF for grey infrastructure but the detailed accuracy assessment is required.

3.2 Accuracy assessment

For accuracy assessment a common practice is to randomly select hundreds of points and label their classification types by referencing reliable sources, such as field work and visual imagery interpretation. The reference points are then compared with the classification results at the same locations. In this research there were two independent sets of more than 1000 control points for both test areas. The points were generated by a random sampling, using equalized-stratified approach. This method creates points that are randomly distributed within each class, where each class has the same number of points.

The spatial distribution of classes was assessed using the methodology adopted from Congalton (Congalton, 2009). Based on observations an error matrix was generated and statistical parameters for particular classes were calculated: Producer's Accuracy (PA), User's Accuracy (UA) and the general parameters K - kappa and OA – overall accuracy (omega). The interpretation of these parameters and their usability is described in Congalton (Congalton, 2009) and Lillesand, Kiefer, Chipman (Lillesand et al., 2004). UA indicates the probability that a pixel classified into a given category actually represents that category on the ground. Hence it describes the possibility of correct class recognition in the field well. The Kappa value is a measure of agreement between classification and the reference data. Hence this measure is a good parameter to compare different classification results. Omega is a percentage of correctly

classified pixels. This value is the most commonly reported accuracy assessment statistic and well describes the classification result itself. Therefore the best indicator of the classification quality for a particular class is UA and the best parameter for comparison of different classifications is Kappa.

Tables 1-3 show the results for test area “W”, but similar results occurred at “U” test area. The ML classifier gave the best results in terms of overall accuracy with kappa=0,75, while in RF kappa varies between 0,68-0,69 for the variants with 100 trees and 200 trees respectively. The UA for roads is similar in each method and equals 0,69 for ML and 0,67 for RF (in both variants). The buildings are classified with the best accuracy and in ML images achieved UA=0,82, and in RF images: UA=0,71 (100) or UA=0,69 (200). The best classified class is water as a common and usual case, regardless of the method, with UA equal to or almost 1.

The low vegetation class reached a UA value of 0,67-0,68 (RF classifier, tab. 2-3) and 0,85 (ML classifier, tab. 1). This kind of vegetation is mostly grasslands and is sometimes confused with bare soil (mainly at cultivated areas) or with medium vegetation which usually consists of small bushes or low crops. There is a confusion between roads and buildings. Especially buildings with asbestos or concrete roofs are classified as a road class. Nevertheless both belong to the same group of grey infrastructure. The problematic class is a bare soil which usually appears at natural or agricultural areas, but sometimes also occurs as undeveloped land in build-up areas. It reached the poorest results in every method.

References/control areas									
Class	water	roads	bare soil	low veg.	medium veg.	buildings	high veg.	UA	
water	143	0	0	0	0	0	0	1,00	
roads	3	99	5	1	0	35	0	0,69	
bare soil	0	4	86	14	0	39	0	0,60	
low veg.	0	9	1	121	3	5	4	0,85	
medium veg.	0	1	0	32	93	1	16	0,65	
buildings	0	14	4	4	0	117	4	0,82	
high veg.	1	1	0	4	5	1	131	0,92	
PA	0,97	0,77	0,90	0,69	0,92	0,59	0,85		
Kappa	0,75							Omega	0,79

Table 1. Maximum Likelihood classification results for “W” area

References/control areas									
Class	water	roads	bare soil	low veg.	medium veg.	buildings	high veg.	UA	
water	145	2	0	0	1	3	1	0,95	
roads	0	81	5	4	0	31	0	0,67	
bare soil	0	3	57	4	0	33	0	0,59	
low veg.	0	10	30	138	8	17	4	0,67	
medium veg.	0	1	0	12	78	1	11	0,76	
buildings	0	29	4	5	0	96	1	0,71	
high veg.	2	2	0	13	14	17	138	0,74	
PA	0,99	0,63	0,59	0,78	0,77	0,48	0,89		
Kappa	0,69							Omega	0,73

Table 2. Random Forests (100) classification results for “W” area

References/control areas									
Class	water	roads	bare soil	low veg.	medium veg.	buildings	high veg.	UA	
water	145	2	0	0	1	5	1	0,94	
roads	0	82	5	4	0	31	0	0,67	
bare soil	0	2	68	10	0	34	0	0,60	
low veg.	0	12	19	127	9	14	6	0,68	
medium veg.	0	1	0	9	73	1	11	0,77	
buildings	0	27	4	12	0	99	1	0,69	
high veg.	2	2	0	14	18	14	136	0,73	
PA	0,99	0,64	0,71	0,72	0,72	0,50	0,88		
Kappa	0,68							Omega	0,73

Table 3. Random Forests (200) classification results for “W” area

The evaluation for grey infrastructure detection was performed for the aggregated classes as well. Overall accuracy for ML classification (table 4) was 87% for dense build-up area (“U”) and 92% for residential area (“W”). RF classification was also performed for “W” test site and resulted in an accuracy level of 89% for k=100 trees and 88% for k=200 trees (table 5). Other errors were calculated and are shown in tables 4 and 5. A commission error (CE) is the share of reference pixels in that class that have been “omitted” in the classification image. An omission error (OE) is the percentage of class pixels in the classification image which are falsely classified.

ML classification resulted slightly worst for the dense urban area “U” (omega=87%) than for the loose urban area “W” (omega=92%). This is because at densely built-up areas a lot of deep shadows from high buildings appear and confuse the classes. For the same reason the commission error is lower at “W” area (7%), however the omission error is higher due to many small family houses covered by the neighbouring trees.

Grey infrastructure	ML “U” test site	ML “W” test site
Omission Error	13%	19%
Commission Error	11%	7%
Overall accuracy	87%	92%
Kappa	0,85	0,86

Table 4. Maximum Likelihood classification of grey infrastructure accuracies.

Grey infrastructure	RF (100) “W” test site	RF (200) “W” test site
Omission Error	27%	27%
Commission Error	7%	10%
Overall accuracy	89%	88%
Kappa	0,81	0,80

Table 5. Random Forests classification of grey infrastructure accuracies for “W” test site.

There is no noticeable difference in accuracy levels for the RF classifier in two variants. RF with k=200 reached slightly lower results (omega=88%) compared to RF with k=100 (omega=89%). Both variants showed a quite high omission error (27%) while commission error was slightly better for RF (100) reaching 7% (table 5). It seems there is no need to use more trees-complex variant. Therefore RF with k=100 trees would be recommended as more efficient or ML otherwise.

Since built-up area is a very complex, mixed-class and detailed terrain, the achieved accuracy is satisfying. Both methods – ML and likewise RF, can be used to indicate the update needs in Topographical Database.

4. DISCUSSION AND CONCLUSIONS

The results of this research proved the high quality of grey infrastructure features detection. The other studies show similar effects in impervious surfaces detection. Results of Zhang’s work based on 1-m multispectral aerial images indicate a separation of impervious surfaces from barren land, vegetation, and water, with user’s accuracies ranging from 69% to 90% and producer’s accuracies from 88% to 95% (Zhang et al., 2020). The other study on land cover classification was performed for the West Virginia using GEOBIA and RF machine learning on orthophotography (Maxwell et al, 2019). The best classification accuracy obtained was 96.7% (Kappa = 0.886). Forests, low vegetation, and water were mapped with user’s and producer’s accuracies above 85%. In contrast, the impervious and mixed developed classes were more difficult to map reaching 80% of PA (Maxwell et al, 2019). The fast expansion of satellite and aerial imagery resources is outpacing the capacity of conventional effort of collecting ground truth through field surveys or on-screen digitalization (Zhang et al., 2020). This automatic manner significantly reduces the effort of visual interpretation especially where the databases cross-check is not available.

4.1 Database updating in grey infrastructure

To check the possibility of updating, the Topographical Database at scale 1:10000 with status for 2012 was used. Comparison with classification results (fig. 13 and fig. 14) showed the needs in editing and insertion of the new features. Figure 12 presents residential area “W” where 112 new buildings were detected out of around 1300 total which gives 9% objects to update. There are also 10 new roads compared to 380 roads in total which gives 5% in total length.

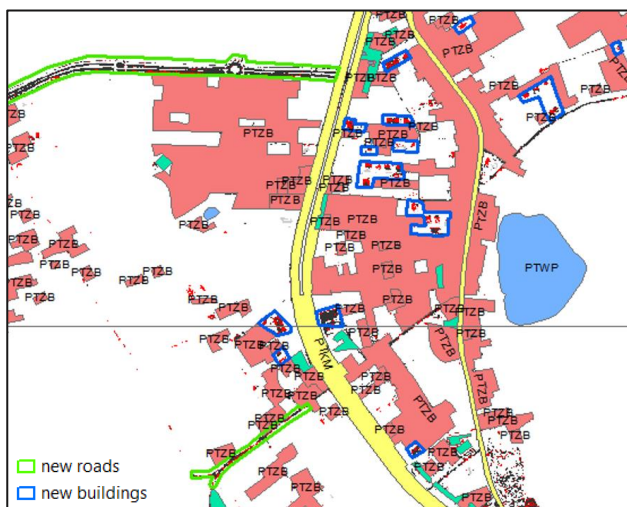


Figure 12. Classes presenting grey infrastructure according to TD class names and the indication of places to update (blue and green lines). Land cover classes are: PTZB – build-up area, PTKM – transport area, PTWP – water area.

The figure 13 presents dense build-up area “U” with 10 new features. Since this area is already very much developed there are rather new constructions as roads or squares rather than buildings or houses. Nevertheless a new grey infrastructure area of almost

0,5km² arises in a total of 5km² in test “U” site. The main reason is the new long expressway investment.

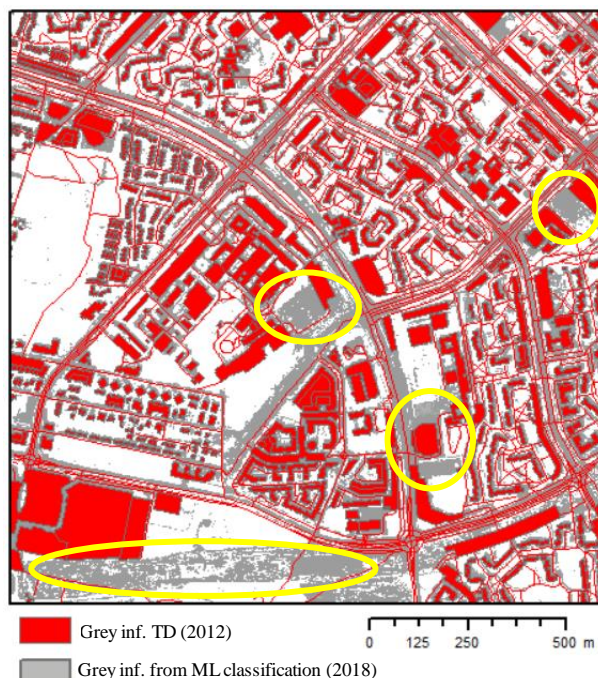


Figure 13. Comparison of grey infrastructure from classification and from Topographical Database. Yellow lines indicate the new constructions.

The precision and detail of this Database at scale 1:10000 requires the exact vertices of feature polygons (i.e. buildings) or feature lines (i.e. roads). Therefore the detected new objects can be pointed out and flagged to input as a new feature by manual editing. The other way to use the output from the classification is to compare and check with land cover classes existing within this Database as shown in figure 12. Land cover classes like: PTZB – build-up area, PTKM – transport area, PTWP – water area, represent the surface with dominant features, but not exact object borders. Still the automatic process of updating requires more investigation.

4.2 Conclusions

This research showed the potential of grey infrastructure detection using VHR imagery. Based on multispectral properties the supervised classification reached a level of 92% of overall accuracy with kappa=0,86. The best producer’s accuracy for building was 0,82 and for roads – 0,69. In dense built-up site the results were slightly worse due to the shadow and close constructions, than in residential areas with spread out buildings.

The introduced and tested fast method of grey infrastructure detection based on ML or RF can be an alternative to very much elaborative Artificial Neural Networks or other classifiers demanding a lot of training samples or object-based more complex approaches. However, the contextual information could support better distinguishing and separating buildings and roads from each other. This would effectively support the process of entering class features into the database.

There are also some additional questions that should be explored in order to further use VHR data to map large extents. There is a need to explore the automatic obtaining of the training data and models produced in one location to map other areas. The other

issue is to elaborate the automatic process of database updating without manual intervention. For now it seems that flagging places is very limited and will not guarantee the exact outline of the features. Yet, both processes could speed up the obtaining of the information about grey infrastructure.

REFERENCES

Adamczyk J., Będkowski K., 2007: *Metody cyfrowe w teledetekcji. Digital Methods in Remote Sensing* [2nd edition]. Publisher SGGW Warsaw, Poland. ISBN: 978-83-7244-898-9.

Belgiu, M., and L. Drăguț. 2016: Random Forest in Remote Sensing: A Review of Applications and Future Directions. *ISPRS Journal of Photogrammetry and Remote Sensing* 114: 24–31. doi:10.1016/j.isprsjprs.2016.01.011.

Congalton R. G., Green K. 2009: *Assessing the Accuracy of Remotely Sensed Data. Principles and Practices*. Second Edition. CRC Press Taylor & Francis Group

Craig A. L., and Brower B. V. 2000: Process for Enhancing the Spatial Resolution of Multispectral Imagery using Pan-Sharpener. *US Patent 6,011,875*, issued January 4, 2000.

Lillesand T. M. Kiefer R. W. Chipman J. W. 2004: *Remote Sensing and Image Interpretation*, 5th edition, Wiley

Maxwell A. E., Warner T. A. Fang F., 2018: Implementation of machine-learning classification in remote sensing: an applied review, *International Journal of Remote Sensing*, 39:9, 2784-2817, doi: 10.1080/01431161.2018.1433343

Maxwell A. E., Strager M. P., Warner T. A., Ramezan C. A., Morgan A. N. and Pauley C. E., 2019: Large-Area, High Spatial Resolution Land Cover Mapping Using Random Forests, GEOBIA, and NAIP Orthophotography: Findings and Recommendations *Remote Sensing*, vol. 11, 1409; doi:10.3390/rs11121409

Rodriguez-Galiano V.F., Ghimire B., J. Rogan, M. Chica-Olmo, J.P. Rigol-Sanchez 2012: An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, Vol. 67, 93-104. doi.org/10.1016/j.isprsjprs.2011.11.002.

Zhang H., Gorelick S. M. and Zimba P. V. 2020: Extracting Impervious Surface from Aerial Imagery Using Semi-Automatic Sampling and Spectral Stability *Remote Sensing*, Vol. 12, 506; doi:10.3390/rs12030506

ABBREVIATIONS

ANN – Artificial Neural Networks
CE – Commission Error
GI – Grey Infrastructure
GSD – Ground Sampling Distance
ML – Maximum Likelihood
OA – Overall Accuracy
OE – Omission Error
RF – Random Forests
SVM – Support Vector Machine
TD – Topographical Database
VHR – Very High Resolution

ACKNOWLEDGEMENTS

I would like to thank Joanna Giczan and Weronika Wiśniewska, students from Warsaw University of Technology (Poland) who assisted with necessary data processing and analysis.

Revised March 2020