

VHR Semantic Labeling by Random Forest Classification and Fusion of Spectral and Spatial Features on Google Earth Engine

M. Kakooei and Y. Baleghi*

Electrical & Computer Engineering Department, Babol Noshirvani University of Technology, Babol, Iran

Received 08 April 2019; Revised 03 February 2019; Accepted 31 March 2020

*Corresponding author: y.baleghi@nit.ac.ir (Y. Baleghi).

Abstract

Semantic labeling is an active field in remote sensing applications. Although handling highly detailed objects in a Very High Resolution (VHR) optical image, and the VHR Digital Surface Model (DSM) is a challenging task, it can improve the accuracy of the semantic labeling methods. In this paper, a semantic labeling method is proposed by fusion of optical and normalized DSM data. The spectral and spatial features are fused into a heterogeneous feature map to train the classifier. The evaluation database classes are impervious surface, building, low vegetation, tree, car, and background. The proposed method is implemented on the Google Earth Engine. The method consists of several levels. First, the Principal Component Analysis (PCA) is applied to the vegetation indices to find the maximum separable color space between the vegetation and non-vegetation areas. The Gray Level Co-occurrence Matrix (GLCM) is computed to provide the texture information as the spatial features. Several Random Forests (RFs) are trained with an automatically selected train dataset. Several spatial operators follow the classification to refine the result. The LeafLess Tree (LLT) feature is used to solve the underestimation problem in the tree detection. The area, and major and minor axes of the connected components are used to refine building and car detection. The evaluation shows a significant improvement in the tree, building, and car accuracy. The overall accuracy and Kappa coefficient are appropriate.

Keywords: *Very High Resolution Semantic Labeling, Spatial Feature, Google Earth Engine, Grey Level Co-Occurrence Matrix, Random Forest, Leafless Tree.*

1. Introduction

Semantic labeling is an active applicant research topic in the Remote Sensing (RS) image processing field, especially in urban areas [1-3]. Semantic labeling is one of the most important components of RS applications such as urban planning, vegetation investigation, and soil management [4]. In the past, spatial resolution of RS imagery was coarse and the Ground Sampling Distance (GSD) of images was greater than a meter. Since a coarse resolution image does not contain detailed information, the High Resolution (HR) and Very High Resolution (VHR) images have been considered recently [5]. On the other hand, small objects like cars are visible in VHR aerial imagery with a meter to sub-meter resolution [1, 6]. Therefore, VHR imagery has created new interests in extracting the man-made structures in urban areas [7].

Increasing the spatial resolution (moving from low resolution to VHR) causes more uncertainty in decision-making, which leads to more complex algorithms [8]. Therefore, scientists are facing new challenges to provide a high accuracy in VHR images [9].

The spectral information of different classes are more similar in the HR and VHR images than in the low spatial resolution scenes [10]. Therefore, accuracy of the typical and traditional classification methods decreases if they are applied to VHR images. In other words, the traditional algorithms can doubtfully support VHR images [11].

Feature extraction plays an important role in the preprocessing step [12], and the solution key is mixing the spectral information with the spatial features [13]. The spatial information like texture and structure (e.g. edges) contains more detailed information in the VHR images than in the coarse

resolution images [14]. This leads to an increasing demand for using the geometrical (spatial) features. Morphological profile (MP) is an example of popular spatial operators [7]. Gray-Level Co-Occurrence Matrix (GLCM) is another popular feature to describe an image texture [15]. The GLCM operations are applied to neighborhood pixels (definite windows size) [16].

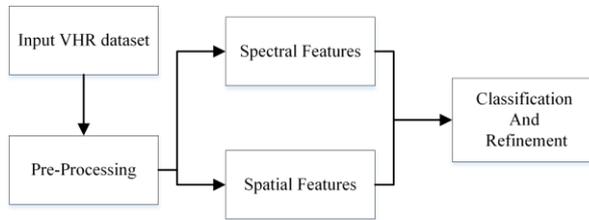


Figure 1. Typical framework of the semantic labeling methods.

Figure 1 shows a typical framework of the VHR image semantic labeling methods in a high level of abstraction. The pre-processing step may include some noise-removal and/or color adjustment algorithms. Then the spectral and/or spatial features are used to classify the scene into pre-defined classes. An initial semantic labeling map may be refined by connected component specifications. The performance of each box in figure 1 could be improved by the researchers. Each imagery platform has its unique specification. Thus fusion of multi-sensor imagery may increase the accuracy of the VHR image analysis [6]. Fusion can occur in different levels in which data-level refers to mixing the information of one sensor with another [5]. Fusion of the satellite, aerial, and UAV images has been utilized by Kakooei and Baleghi [17] for building a damage assessment. Fusion of a Digital Surface Model (DSM) and optical images can facilitate building detection [18] and small object extraction [10]. DSM generation is based on the radar imagery that follows geometric distortion removal (i.e. layover, foreshortening, and shadows) [19]. Hussain and Shan [20] have fused optical VHR aerial images with elevation information and city zoning maps for an urban land cover classification. Combination of the optical and long-wave infrared images has been utilized by Guan et al. [21] to improve land classification. Although VHR images contain fine detailed objects, they have a large amount of noise that can potentially mislead the classification. An extension of mean filter can smooth the noise of VHR images as the pre-processing step [22]. In [23], spectral features are used to detect some special non-urban areas like vegetation and water. Then they proposed a structural feature in the refinement step. An automatic method has been

proposed by Ozdarici Ok [24] to detect individual trees. It is based on vegetation extraction, fast radial symmetry (FRS) transform, and simple object-based hierarchical operations. Kakooei and Baleghi [17] have mentioned that a leafless tree does not contain any leaf to be classified as tree class by the spectral features. Therefore, a spatial feature named LeafLess Tree (LLT) has been proposed to improve the tree detection accuracy. The proposed LLT feature is also used in this work to improve tree detection accuracy.

The spatial-based method is a part of many previous VHR image analysis algorithms. MPs require Structural Elements (SEs) to apply different operations to a definite neighborhood. SEs are typically disk-shaped. Bellens et al. [7] have proposed new linear SEs and utilized partial morphological reconstruction to preserve the shape of objects. Qin [25] has proposed a new special feature that is based on mean shift. Li et al. [26] have arranged the spatial features of land cover to mix with common land cover indices to improve the classification accuracy.

Classification can be improved by following a feature extraction step. Regniers et al. [27] have proposed a supervised classification method, in which the features are based on the wavelet multivariate models. Huang et al. [10] have extracted a wavelet-based feature to fuse the spectral and spatial information in a multi-scale procedure. Xu and Li [14] have extracted the shape features (Hu's moments, Zernike moments, and wavelet moments) from segmented objects. These features were classified by an object-based method. Chaib et al. [28] have proposed a method containing scale-invariant feature transform, and sparse principal component analysis (sPCA).

Convolutional Neural Network (CNN) creates a hierarchy of decision nodes to classify objects. Volpi and Tuia [2] have presented a CNN-based method for semantic labeling. The designed method contains down-sampling and up-sampling sections. Deep features play an important role in VHR detailed images [29]. Sherrah [1] has applied Deep Convolutional Neural Networks (DCNNs) to VHR remote sensing images for semantic labeling. A DCNN-based decision-level fusion approach has been proposed by Liu et al. [6] for semantic labeling. Fusion was based on a probabilistic graphical model.

Random Forest (RF) is a popular classification algorithm. The RF classifier creates multiple decision trees, in which a train dataset is selected randomly. RF is fast and it is an appropriate classifier for high-dimensional data [30]. Du et al. [31] have trained many high-dimensional features

by a random forest classifier to classify buildings into seven categories. Abe et al. [32] have compared the RF and Support Vector Machine (SVM) performance in hyperspectral land classification. RF can handle vegetation mapping [33], analysis of gross primary production upscaling [34], and building irregularity detection [35].

As many satellites are orbiting the earth to acquire images, satellite data collection is becoming a big data. Managing this large amount of data is tough for the developers, and requires time and effort. Besides, processing requires a high computational capability. The developers desire to focus on algorithms instead of data and hardware [36]. Google has provided a developing environment for remote sensing developers named Google Earth Engine (GEE). It is a parallel cloud-based platform that provides free access to many satellites [37]. Recently, several developers have used GEE to implement their methods. Miettinen et al. [38] have proposed an automatic decision tree to classify land cover into 11 classes on GEE. GEE can be used for mapping an urban area, rice paddy, and flood. This is also useful to estimate forest change, water change, and crop yield [39].

In this work, the spectral and spatial features are fused into a heterogeneous feature map to propose a new semantic labeling method. The VHR optical and VHR DSM data are fused in the first step to produce the spectral features. Some GLCM features are extracted from VHR optical data, as the spatial features. These spectral and spatial features are used to train the classifier. Random forest classifier is trained by an automatically selected train dataset. The final refinement is based on the LeafLess Tree (LLT) feature and some popular connected component specifications such as area, and major and minor axes. The process is totally automatic and implemented on GEE.

The evaluation dataset that contains the VHR optical image and DSM data is described in the next section. Then our proposed method is presented step by step in detail. It is implemented on GEE, and the evaluation result is presented on the VHR dataset.

2. Data source

The utilized data sources were from two different sensors that were provided by ISPRS Commission II/4 in the 2D semantic labeling contest in 2017 ("ISPRS 2D semantic labeling contest-Potsdam")

[40]. This dataset covers part of the Potsdam city in Germany. It comprises 38 patches of optical and DSM images with the same size. Only the ground truth of 24 patches is provided.

It contains aerial VHR optical images in four bands including Red, Green, Blue, and Near Infra-Red (RGB and NIR).

It also contains VHR DSM and normalized DSM (nDSM) data in the area of interest. GSD of optical and DSM images is 5 cm and the size of each image tile is 6000×6000 .

All image patches are uploaded to the GEE servers. An overview of 24 patches are shown in figure 2. Figures 2a and 2b illustrate the VHR optical images. Figure 2a shows an RGB image that is presented in the R-G-B format. The visualization of figure 2b is NIR-R-G. Figure 2c displays the VHR nDSM of the dataset. Figure 2d shows the ground truth comprising six classes. These classes are:

- Impervious surfaces (RGB: 255, 255, 255)
- Building (RGB: 0, 0, 255)
- Low vegetation (RGB: 0, 255, 255)
- Tree (RGB: 0, 255, 0)
- Car (RGB: 255, 255, 0)
- Clutter/background (RGB: 255, 0, 0)



Figure 2. An overview of the dataset on the GEE. All 24 patches are shown. (a) RGB image. (b) NIR-R-G image. (c) nDSM image. (d) Ground truth.

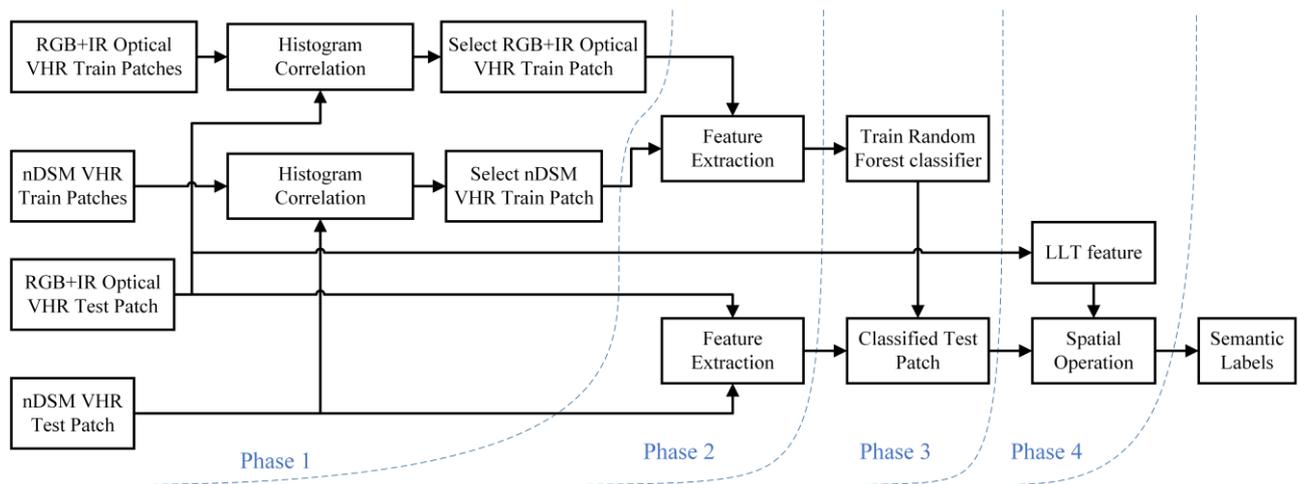


Figure 3. The framework of the proposed method.

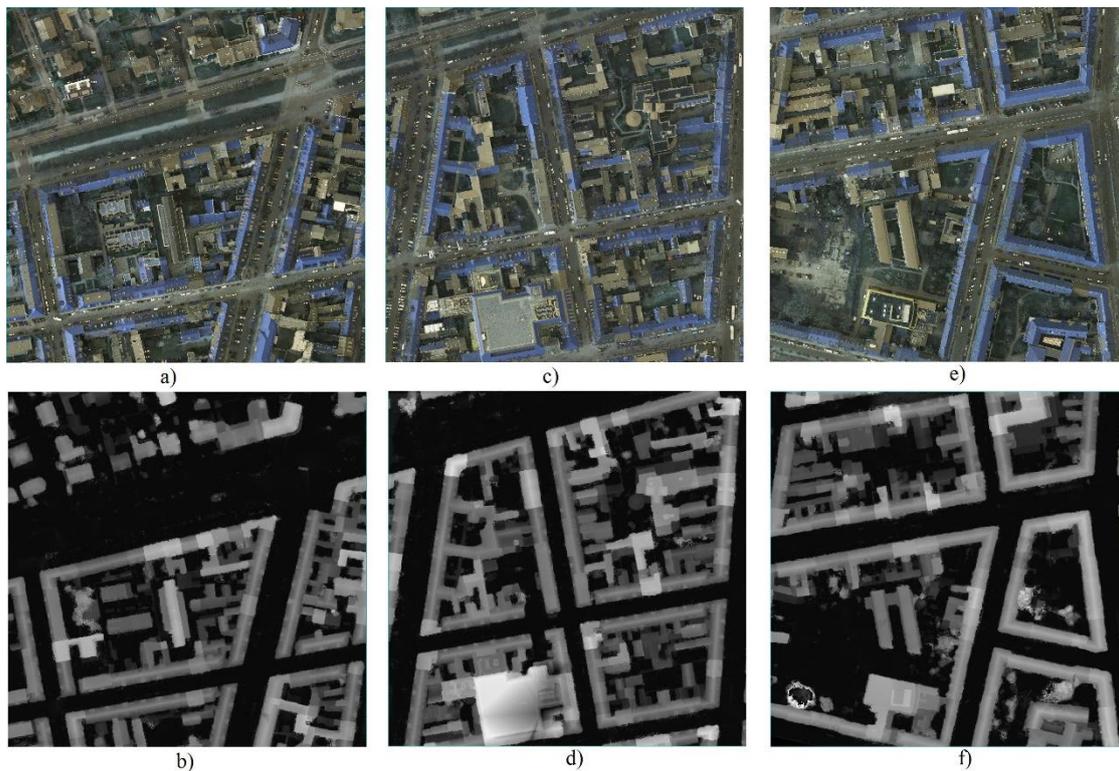


Figure 4. Similarity investigation based on histogram correlation (a) Patch 5-11 RGB (b) Patch 5-11 nDSM (c) Patch 5-12 RGB (d) Patch 5-12 nDSM (e) Patch 6-12 RGB (f) Patch 6-12 nDSM.

3. Proposed method

This section contains five sub-sections. The method framework is shown in figure 3 and briefly described in the following sub-section. Furthermore, the purpose of all phases is explained in this sub-section. Then in the next four sub-sections, the phases are detailed.

3.1. Method framework

An overview of the multi-level structure of the proposed method is shown in figure 3. At the left side of the flowchart, the VHR RGB+NIR optical image and VHR nDSM data are divided into the train and test dataset patches. First, the histogram

correlation between each test patch and all the available training patches are calculated.

For each test patch, two patches are selected from the dataset to train the classifier. These patches are the most similar ones to test them in the RGB+NIR and nDSM histograms, and they are selected automatically. Thus it is possible to use just one training patch if both the RGB+NIR and nDSM test histograms are similar to one training patch.

In the second phase, the features are extracted from the train and test patches. Then the RF classifier is trained in the third phase. This classifier is used to classify the test patch into the desired classes.

Phase 4 is dedicated to spatial operations to refine and improve the classification result. The LeafLess

Tree (LLT) feature is extracted from the optical image to improve tree detection. The area, major axis, and minor axis properties are utilized to improve the car and building detection. Besides, some typical morphology operations like opening and closing are applied to create the final semantic labels.

3.2. Phase 1: Selecting train patches

Cross-correlation between the histogram of each test patch and the histogram of all training patches can find the most similar training patches to the test patch. For instance, the experimental investigation indicates that patch No. 5-11 (in Figures 4a and 4b) is similar to the patch No. 5-12 in the nDSM histogram (Figure 4d), and it is similar to patch No. 6-12 in the RGB+NIR histogram (Figure 4e). These RGB and nDSM patches are shown in figure 4. It means that these two patches (No. 5-12 and 6-12) are sampled to train the RF classifier to classify patch No. 5-11.

3.3. Phase 2: Feature extraction

The feature extraction step is divided into three phases in figure 5. A heterogeneous feature map is the outcome of this section in phase 2.3. The first phase (phase 2.1) is based on three indices including Normalized Difference Vegetation Index (NDVI), Enhanced Vegetation Index (EVI), and

Green Ratio Vegetation Index (GRVI). These indices are calculated in (1) to (3).

$$NDVI = \frac{NIR - R}{NIR + R} \tag{1}$$

$$EVI = 2.5 \frac{NIR - R}{NIR + 6 \times R - 7.5 \times B + 1} \tag{2}$$

$$GRVI = \frac{NIR}{G} \tag{3}$$

The PCA algorithm is applied to this 3-band image, and the band with the largest Eigen value is selected to be one band of the heterogeneous feature map. This band improves the separability between the vegetation area (low vegetation and tree) and the other classes. Figure 6a shows this band for patch No. 5-11.

The gray level image is constructed by RGB bands in phase 2.3. Some GLCM features are calculated based on the Gray and NIR images as the classification features. These features are Angular Second Moment (ASM), Inverse difference Moment (IDM), Image correlation 1d (IMCORR1), Image correlation 2d (IMCORR2), Sum of Entropy (SENT), and Difference of Entropy (DENT). These features have been described in [41, 42]. For instance, ASM, IDM, IMCORR1 and DENT of Gray image are shown in figures 6b to 6e.

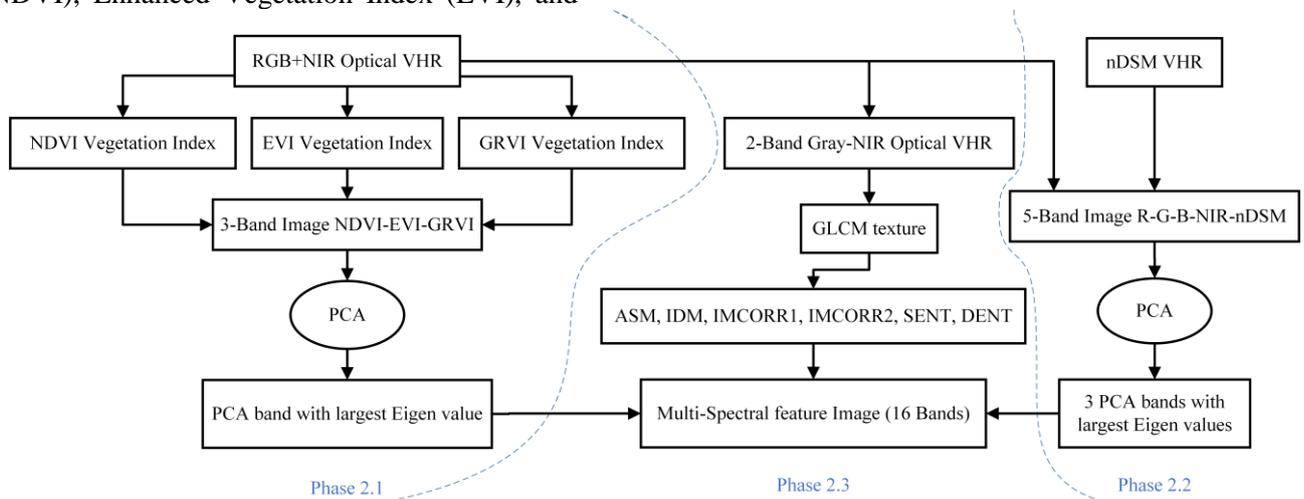


Figure 5. Heterogeneous Feature Map Flowchart.

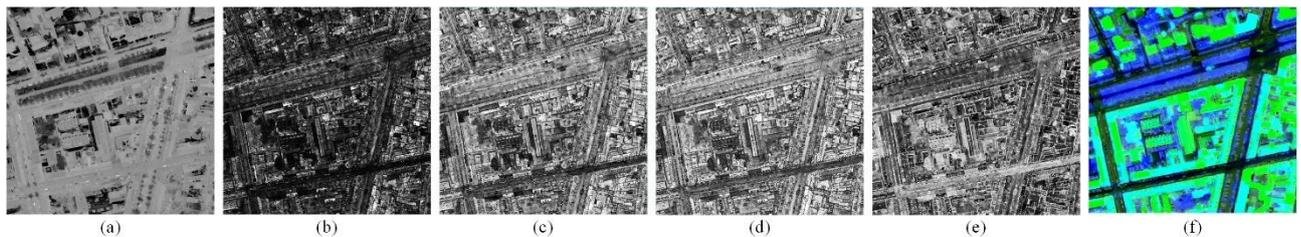


Figure 6. Bands of heterogeneous feature map (a) PC1 of vegetation PCA (b) Gray ASM (c) Gray IDM (d) Gray IMCORR1 (e) Gray DENT (f) Three PCA bands of 5-band.

PCA is applied to a constructed 5-band image in phase 2.2. These bands are Red, Green, Blue, NIR, and nDSM. Three PCA bands with the largest Eigen values are selected as the classification features, which are shown in RGB false color in figure 6f for patch No. 5-11. The final heterogeneous feature map contains sixteen bands:

- One band is selected from phase 2.1 (PCA of vegetation indices)
- Six bands are from GLCM of Gray image
- Six bands are from GLCM of NIR image
- Three bands are selected from phase 2.2 (PCA of VHR data)

3.4. Phase 3: Random forest classification

An overview of Random Forest classification is shown in figure 7. Three separate classifiers are trained to improve the classification result. The classifier that is indicated with number 1 is a 4-class classifier, in which low vegetation and tree impervious surface and background are merged, compared to the original one. There are two 2-class classifiers that are indicated by numbers 2 and 3 to generate the final 6-class map.

There are different systems available to combine the classifiers such as serial, parallel, and embedded modes [43]. The proposed method utilizes a serial combination of classifiers.

Assume that there are n numbers of samples with p features, and RF uses n_T number of trees. The training complexity of the RF classification in big O notation is $O(n^2 \sqrt{pn_T})$, and the prediction complexity is $O(pn_T)$.

The serial classifiers work on sub-samples of the original samples n . Assume that there are n_s number of serial classifiers, in which each classifier works on s_i samples ($s_i \leq n$). In the worst case, the training and prediction complexities are $O(n^2 \sqrt{pn_T n_s})$ and $O(pn_T n_s)$, respectively. Thus, while the number of classes does not directly play a role in the computation complexity, different numbers of classes can result in different numbers of serial classifiers that affect the complexity.

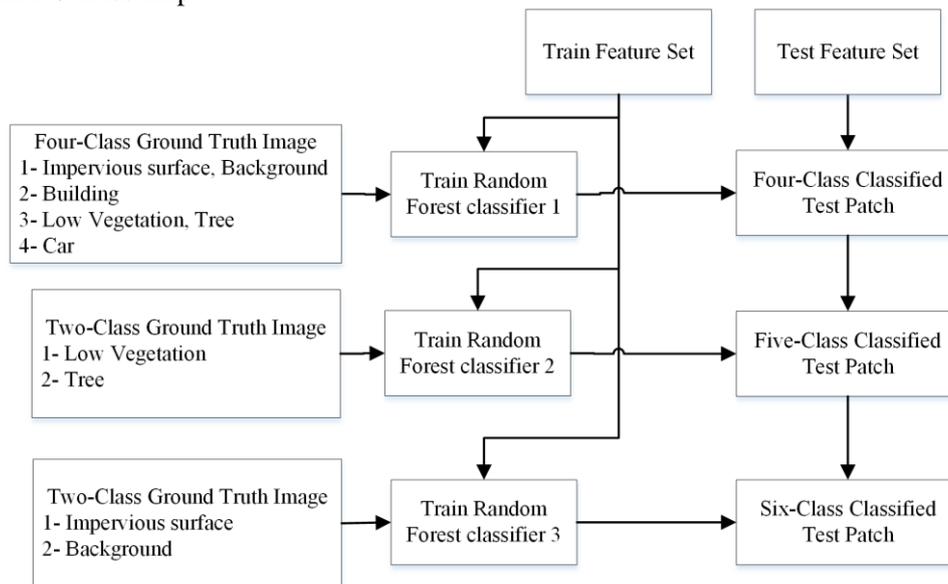


Figure 7. Multi-level Random Forest Classification.

3.5. Phase 4: Spatial operations

Spectral features are insufficient to handle the VHR semantic labelling methods. Thus the proposed method uses the GLCM features to contribute some texture information in the classification. It is followed by some spatial operations in this phase, which are defined according to the structure of different objects. A feature named LeafLess Tree (LLT) that has been proposed in [44] is utilized to improve tree detection. The structure of car and building

connected components are considered to define some spatial operators to refine the results.

The spatial operations in this step include morphological closing, filtering by size, and thresholding on major and minor axes. A flowchart of the spatial operations is shown in figure 8, which includes three phases that are applied hierarchically. In order to simplify the explanation, each phase is explained in a challenge and solution manner. In figure 9a, ground truth of patch No. 5-11 is shown. Figures 9b and 9c show the results of classification and spatial refinement, respectively.

Phase 4.1: Refining car detection

Challenge 1: Windshield is not classified as car. Windshield is probably detected and classified as shadow since it appears dark. Abundance of

shadow in the impervious surface and low vegetation classes is more than car class, which will force the classifier to classify windshield as a non-car class.

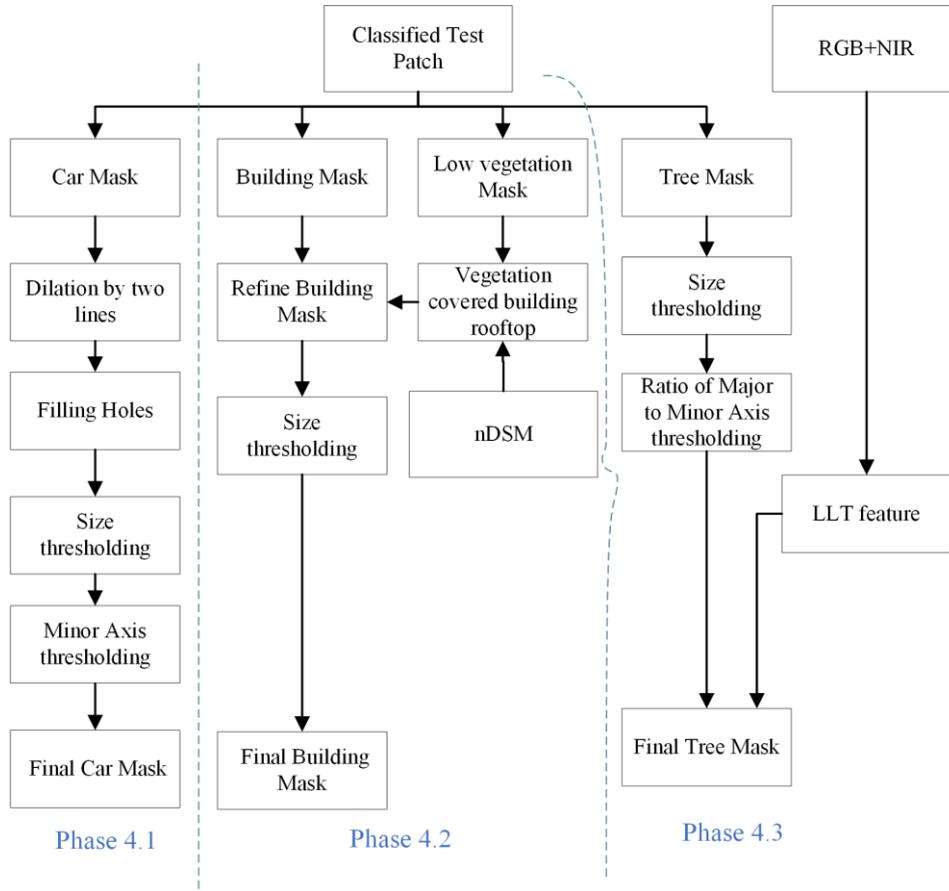


Figure 8. Spatial Operations to Refine Classification Result.

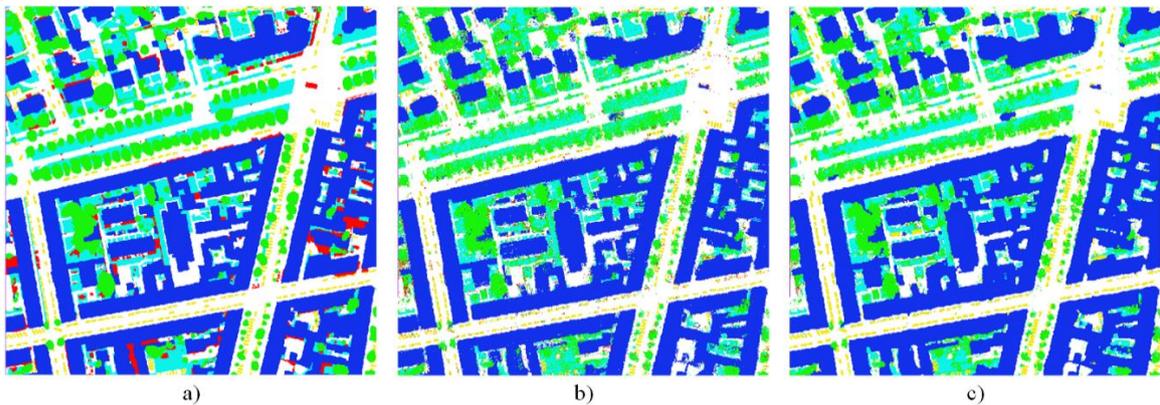


Figure 9. Spatial Operation on patch No. 5-11 a) Ground truth b) Classified image c) Result of refinement by spatial operators.

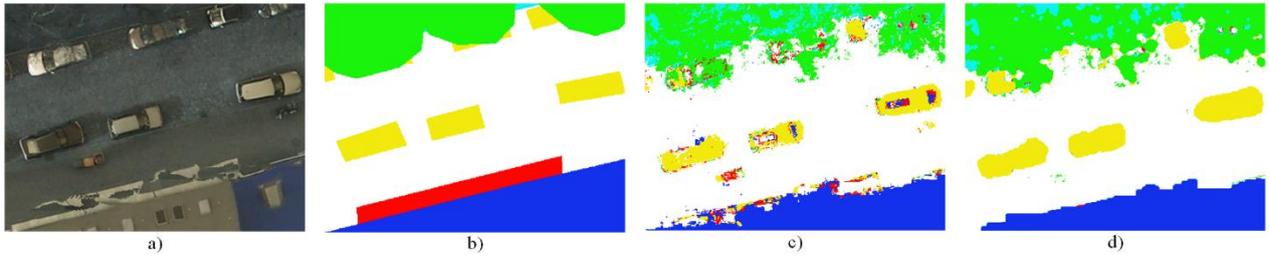


Figure 10. Car refinement on a part of patch No. 5-11 a) RGB view b) Ground truth c) Classified image d) Result of refinement by spatial operators.

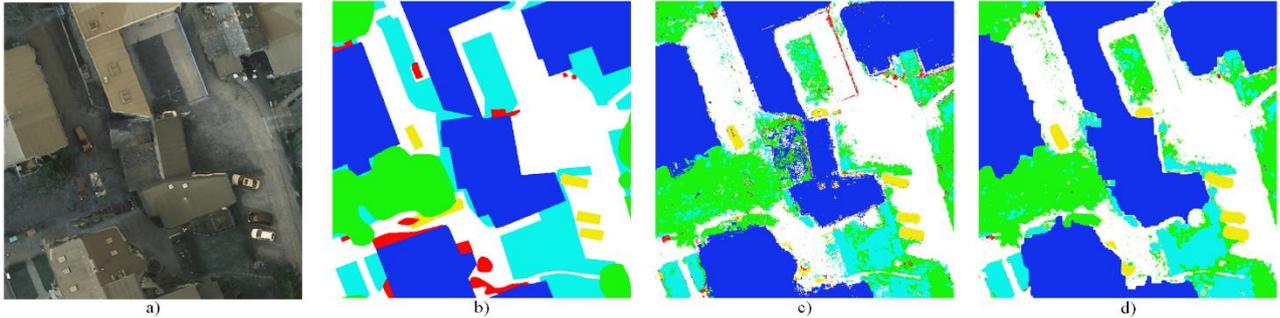


Figure 11. Building refinement on a part of patch No. 5-11 a) RGB view b) Ground truth c) Classified image with false negative rooftop as vegetation and small points d) Refining building detection on rooftop.

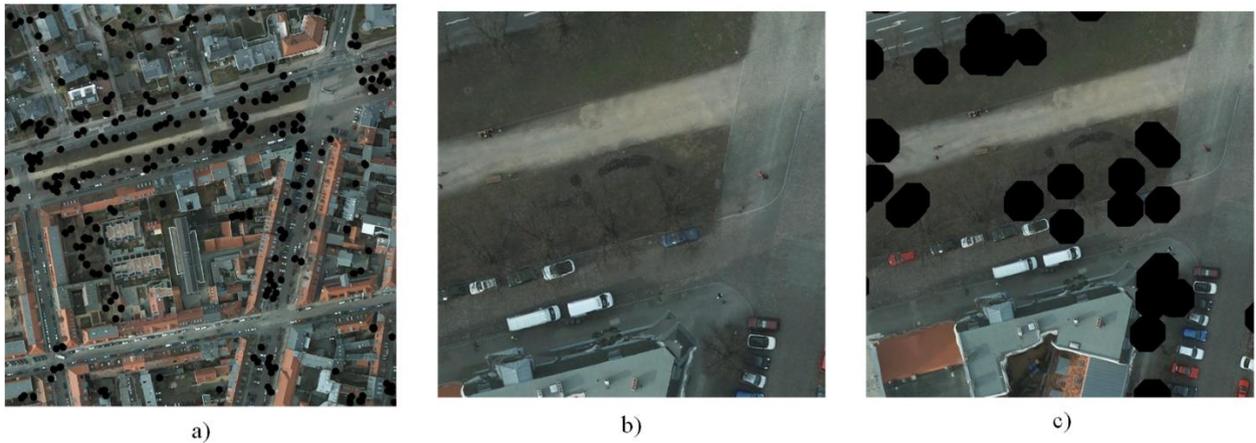


Figure 12. Tree refinement on a part of patch No. 5-11 a) Leafless tree detection b) Part of original image c) Leafless tree detection in a part of image.

Solution 1: Morphological dilation will be applied to car connected components. Two vertical and horizontal lines are used as the Structural Element (SE) to connect the car's hood and trunk to the roof. This operation will put the windshield in the hole. Filling the holes is utilized to put the windshield in the car class.

Challenge 2: Small points and large areas are detected as car.

Solution 2: Thresholding on size of connected component refines the result.

Challenge 3: The connected component size is correct but the structure seems wrong. For example, sometimes street lines are detected as car.

Solution 3: Thresholding on the minor axis of the connected component removes this false detection. The minor axis of car should be at least greater than 1 m.

Figure 10 shows a part of patch No. 5-11, in which three cars are in the middle of the street. The car windshields are detected after refinement in figure 10d. The lines are removed.

Phase 4.2: Refining building detection

Challenge 1: Vegetation on rooftop of building is detected as low vegetation.

Solution 1: Using the nDSM data, if the nDSM value of low vegetation classes are similar to building classes, they will be corrected and the

class will be changed from low vegetation to building.

Challenge 2: Small areas are detected as building.

Solution 2: The connected components with small areas are removed by thresholding.

In the center of figure 11c, there is a mixture detection of building and vegetation, which is refined in 11d. There are some small points that are removed from image compared to ground truth in figure 11b.

Phase 4.3: Refining tree detection

Challenge 1: The distributed small points are classified as trees.

Solution 1: Thresholding on size of tree connected components to remove the small sparse points. Figure 13a contains several small points, which are removed in figure 13b after refinement.

Challenge 2: Dense fence shrub and similar structures may be detected as tree.

Solution 2: Eccentricity of a region is defined as the ratio of the major axis to its minor axis. Eccentricity should be less than a specified threshold.

Challenge 3: Leafless trees are not detected.

Solution 3: The LLT feature is utilized. The LLT feature is based on the structure of leafless tree branches in the VHR optical images. A special gradient operator was proposed to find gradient in four directions. Mixing with directional filters and maximum operator finds the point of the center of the tree in the orthophoto image [44]. Figure 12a shows leafless tree detection by LLT in patch No. 5-11. A part of patch is provided in figures 12b and 12c for a better visualization.

3.6. Ground truth overestimation

The provided ground truth seems to be overestimated, especially in tree detection. As our proposed method is based on fusion of the spectral and spatial features, it can detect tree branches in the image, while the ground truth is overestimated in the tree class. Figure 13a shows the initial classified image, which is refined by spatial operators to create the final sematic labelling map in figure 13b. The ground truth is shown in figure 13c. It seems that our proposed method is accurate in all classes except tree. Considering the RGB image in figure 13d, our proposed method preserves the structure of tree, which is not defined well in ground truth. This issue affected the evaluation results of the next section, especially in the tree class.

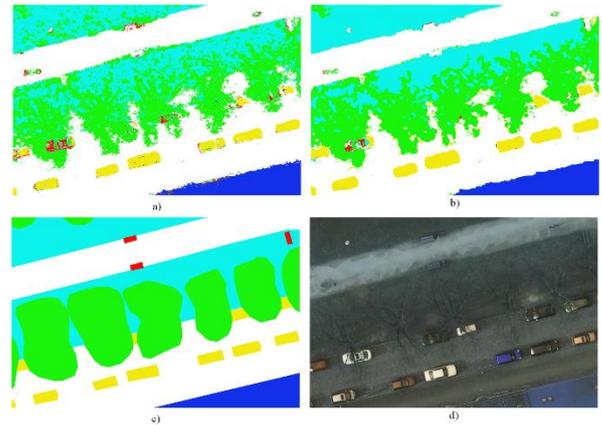


Figure 13. Ground truth overestimation a) Result of classification b) Result of spatial refinement c) Ground truth d) Original RGB image.

4. Evaluation

The utilized dataset comprises 24 images with ground truth. Three recent published papers that have used the same dataset are selected for comparison. The test dataset of [1, 2, 6] methods are summarized in table 1. Our test dataset patches contain a mixture of [1, 2, 6] dataset patches.

Table 1. Test Dataset in Our Evaluation and Previous Works.

Method	Test dataset
[1]	2-11, 2-12, 4-10, 5-11, 6-7, 7-8, 7-10
[2]	2-12, 3-12, 4-12, 5-12, 6-12, 7-12
[6]	4-10, 6-8, 6-11
Proposed	2-11, 2-12, 3-12, 4-10, 4-12, 5-11, 6-7, 6-8, 6-11, 7-8

Performance of the serial classification and spatial operations is analyzed in table 2. In this table, the Overall Accuracy (OA) and Kappa (K) coefficient values are shown for a typical classification, multi-level classification, and spatial refinement. OA is calculated by summing the number of correctly classified pixels and dividing by the total number of pixels. Calculation of the K coefficient in (4) shows the degree of accuracy and reliability of the proposed method.

$$k = \frac{N \sum_{i=1}^n m_{i,i} - \sum_{i=1}^n (F_i D_i)}{N^2 - \sum_{i=1}^n (F_i D_i)} \quad (4)$$

where i is the class number, N is the total number of classified pixels, $m_{i,i}$ is the number of pixels belonging to the class i that have been classified as class i , D_i is the total number of predicted pixels belonging to class i , and F_i is the total number of pixels belonging to class i .

Comparing the results of the typical classification and multiple classification shows the performance of the utilized multi-level classification method. Furthermore, the effect of spatial refinement is evaluated and illustrated in table 2.

The table verifies a significant improvement in using a serial combination of classifiers. On the other hand, although there is a little improvement in utilizing the spatial refinement, the most important achievement is about achieving more accurate results in human-made classes. The producer accuracy of all classes is plotted in Figures 14 to 18 to demonstrate the effects of the

designed steps. Considering the accuracy of each class, there is a little decrease in the impervious surface and low vegetation classes but show significant improvements in the building, tree, and car classes, which are man-made objects in urban areas.

Table 2. Overall accuracy and Kappa coefficient of test patches in the proposed method.

Patch number	Typical classification result		Multi-level classification result		Spatial refinement	
	Overall accuracy	Kappa coefficient	Overall Accuracy	Kappa Coefficient	Overall Accuracy	Kappa Coefficient
2-11	71.5558	0.5951	79.1282	0.7072	81.8967	0.7478
2-12	69.7295	0.5527	83.3584	0.7368	83.1100	0.7415
3-12	76.389	0.6884	83.7557	0.7836	84.8757	0.7987
4-10	73.0919	0.647	83.74758	0.7816	85.4160	0.8043
4-12	84.5974	0.7862	90.9518	0.8718	91.7267	0.8828
5-11	82.3244	0.7478	87.3597	0.8188	87.9796	0.8273
6-7	60.2285	0.4622	85.2256	0.7467	85.4232	0.7533
6-8	77.0864	0.7	85.2216	0.8029	86.5333	0.8205
6-11	85.651	0.777	90.3974	0.8487	90.4832	0.8485
7-8	78.0161	0.7026	84.5269	0.7867	85.3667	0.7970

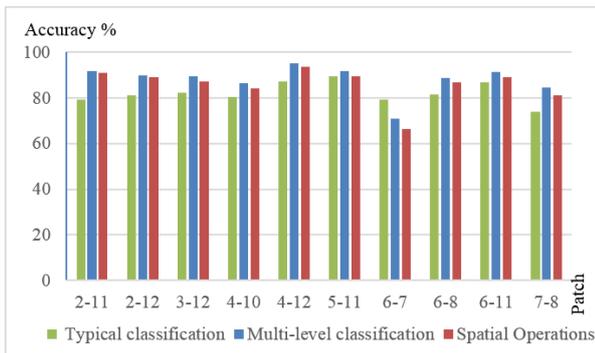


Figure 14. Impervious surface accuracy of the proposed method in test patches.

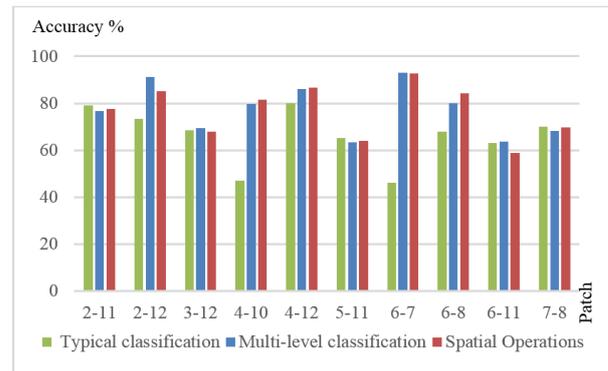


Figure 16. Low vegetation accuracy of the proposed method in test patches.

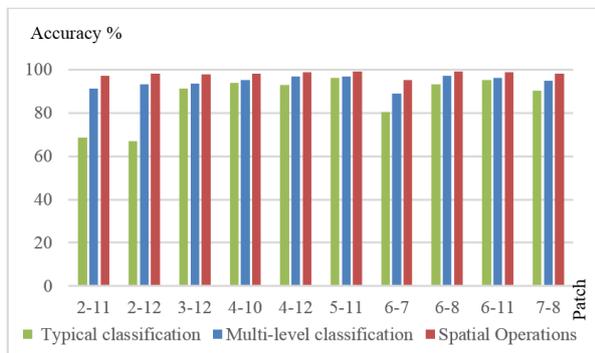


Figure 15. Building accuracy of the proposed method in test patches.

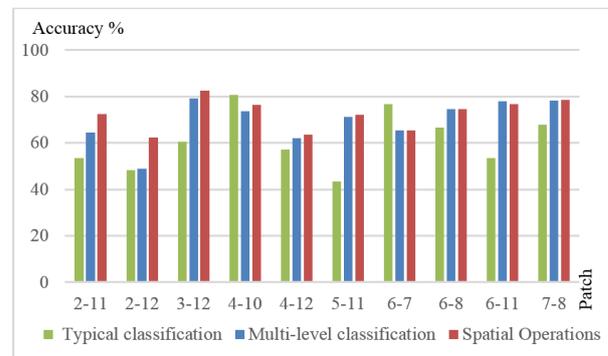


Figure 17. Tree accuracy of the proposed method in test patches.

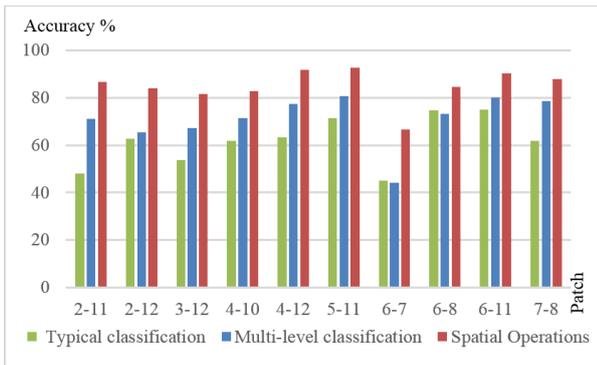


Figure 18. Car accuracy of the proposed method in test patches.

The visual comparison is based on part of the image patches that is available in [1, 2, 6]. These images are shown in Figure 19. The visual comparison includes RGB, ground truth, reference result, and the result of our proposed method. Some purple eclipses are added to the results to emphasize on

the differences. Figures 19a to 19d compare the result of our proposed method versus the method proposed by Sherrah [1]. Our proposed method detects cars in the main street, which are ignored in the reference result.

Volpi and Tuia [2] have proposed a method that cannot identify the background properly. This is shown by comparing the ground truth and their results in Figures 19f and 19g, respectively. Our proposed method detects the background in Figure 19h.

The visual-based comparison of our proposed method against reference [6] is shown in the last row of Figure 19. As it shows, the quality of building detection is better in our proposed method. The bottom eclipse shows that our proposed method properly detects trees, which are falsely detected as low vegetation in [6].

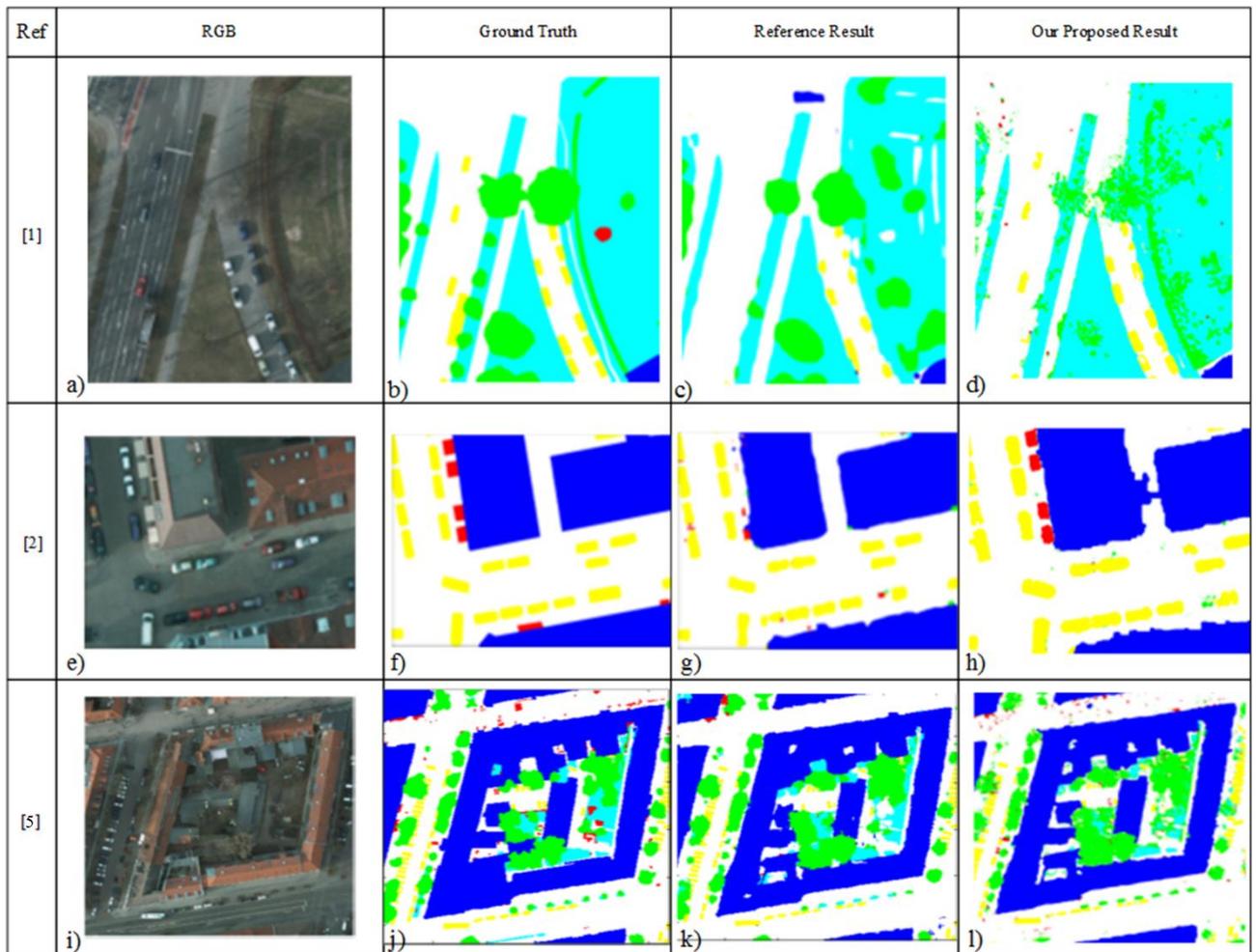


Figure 19. Visual result comparison between our proposed method and the previous works. The second column includes the original RGB images, where the corresponding ground truth is shown in the third column. The fourth column shows the reported results in the previous works, noted in the first column. The last column is the result of our proposed method.

5. Conclusion

A multi-level method was proposed in this work for semantic labelling, which includes fusion of spectral and spatial features. The VHR optical RGB+NIR image and nDSM data are fused for object classification. The impervious surface, building, low vegetation, tree, car, and background were detected in different levels. Although the procedure of the proposed method is not deep, the results are comparable to deep models.

Our proposed method comprises four phases. First, the optical and nDSM images of the area of interest are compared to dynamically find the best training set, which is found in a histogram-based manner. Feature set is extracted in the following phase, which includes the spectral and spatial features. Then three Random Forest (RF) classifiers are trained to classify the selected image into the desired classes.

The spatial operations refine the classification result. It contains a new feature named LLT to find the leafless tree structures, which are typically hard to find in a spectral analysis. It considers the structure of car and building to improve the detection accuracy of these man-made structures. Although our evaluation does not show a high improvement in the overall accuracy and Kappa coefficient, our non-deep proposed method with spatial refinements gets similar results compared to deep methods. It is also implemented in GEE that is fast and of interest by many developers recently.

Acknowledgments

The authors would like to acknowledge the provision of the datasets by ISPRS and BSF Swissphoto, released in conjunction with the ISPRS. Thanks to Markus Gerke for providing the normalised DSMs. The authors also acknowledge the funding support of Babol Noshirvani University of Technology through Grant program No. BNUT/370123/98.

References

[1] Sherrah, J. (2016). Fully Convolutional Networks for Dense Semantic Labelling of High-Resolution Aerial Imagery. arXiv:1606.02585.

[2] Volpi, M., & Tuia, D. (2017). Dense semantic labeling of subdecimeter resolution images with convolutional neural networks. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 55, no. 2, pp. 881-893.

[3] kashaf, S., & Nezamabadi-pour, H. (2019). MLIFT: Enhancing Multi-label Classifier with Ensemble Feature Selection. *Journal of AI and Data Mining*, vol. 7, no. 3, pp. 355-365.

[4] Tehrany, M. S., Pradhan, B., & Jebuv, M. N. (2014). A comparative assessment between object and pixel-based classification approaches for land use/land cover mapping using SPOT 5 imagery. *Geocarto International*, vol. 29, no. 4, pp. 351-369.

[5] Carleer, A., & Wolff, E. (2006). Urban land cover multi-level region-based classification of VHR data by selecting relevant features. *International Journal of Remote Sensing*, vol. 27, no. 6, pp. 1035-1051.

[6] Liu, Y., Piramanayagam, S., Monteiro, S. T., & Saber, E. (2017). Dense semantic labeling of very-high-resolution aerial imagery and LiDAR with fully convolutional neural networks and higher-order crfs. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Honolulu, USA, 2017.

[7] Bellens, R., Gautama, S., Martinez-Fonte, L., Philips, W., Chan, J. C.-W., & Canters, F. (2008). Improved classification of VHR images of urban areas using directional morphological profiles. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 46, no. 10, pp. 2803-2813.

[8] Li, G., & Wan, Y. (2015). A new combination classification of pixel-and object-based methods. *International Journal of Remote Sensing*, vol. 36, no. 23, pp. 5842-5868.

[9] López, J. A., Verdiguier, E. I., Chova, L. G., Marí, J. M., Barreiro, J. R., Valls, G. C., & Maravilla, J. C. (2011). Land cover classification of VHR airborne images for citrus grove identification. *ISPRS journal of photogrammetry and remote sensing*, vol. 66, no. 1, pp. 115-123.

[10] Huang, X., Zhang, L., & Li, P. (2008). A multiscale feature fusion approach for classification of very high resolution satellite imagery based on wavelet transform. *International Journal of Remote Sensing*, vol. 29, no. 20, pp. 5923-5941.

[11] Kakooei, M., & Baleghi, Y. (2020a). Shadow detection in very high resolution RGB images using a special thresholding on a new spectral-spatial index. *Journal of Applied Remote Sensing*, vol. 14, no. 1, 016503.

[12] Imani, M., & Ghassemian, H. (2017). Feature extraction of hyperspectral images using boundary semi-labeled samples and hybrid criterion. *Journal of AI and Data Mining*, vol. 5, no. 1, pp. 39-53.

[13] Kakooei, M., & Baleghi, Y. (2019). Spectral Unmixing of Time Series Data to Provide Initial Object Seeds for Change Detection on Google Earth Engine. *27th Iranian Conference on Electrical Engineering (ICEE)*, 2019.

[14] Xu, H., & Li, P. (2010). Urban land cover classification from very high resolution imagery using spectral and invariant moment shape information. *Canadian Journal of Remote Sensing*, vol. 36, no. 3, pp. 248-260.

- [15] Huang, X., & Zhang, L. (2013). An SVM ensemble approach combining spectral, structural, and semantic features for the classification of high-resolution remotely sensed imagery. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 51, no. 1, pp. 257-272.
- [16] Hall-Beyer, M. (2004). GLCM Texture: A Tutorial v. 2.7. 1, on-line document. In.
- [17] Kakooei, M., & Baleghi, Y. (2017). Fusion of satellite, aircraft, and UAV data for automatic disaster damage assessment. *International Journal of Remote Sensing*, vol. 38, no. 8-10, pp. 2511-2534.
- [18] Suliman, A., & Zhang, Y. (2015). Development of line-of-sight digital surface model for co-registering off-nadir VHR satellite imagery with elevation data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 5, pp. 1913-1923.
- [19] Kakooei, M., Nascetti, A., & Ban, Y. (2018). Sentinel-1 Global Coverage Foreshortening Mask Extraction: An Open Source Implementation Based on Google Earth Engine. *IEEE International Geoscience and Remote Sensing Symposium, IGARSS 2018*.
- [20] Hussain, E., & Shan, J. (2016). Urban building extraction through object-based image classification assisted by digital surface model and zoning map. *International Journal of Image and Data Fusion*, vol. 7, no. 1, pp. 63-82.
- [21] Guan, X., Liao, S., Bai, J., Wang, F., Li, Z., Wen, Q., Chen, T. (2017). Urban land-use classification by combining high-resolution optical and long-wave infrared images. *Geo-spatial Information Science*, vol. 20, no. 4, pp. 299-308.
- [22] ZhiYong, L., Shi, W., Benediktsson, J. A., & Gao, L. (2018). A modified mean filter for improving the classification performance of very high-resolution remote-sensing imagery. *International Journal of Remote Sensing*, vol. 39, no. 3, pp. 770-785.
- [23] Doustfatemeh, I., & Baleghi, Y. (2016). Comprehensive urban area extraction from multispectral medium spatial resolution remote-sensing imagery based on a novel structural feature. *International Journal of Remote Sensing*, vol. 37, no. 18, pp. 4225-4242.
- [24] Ozdarici-Ok, A. (2015). Automatic detection and delineation of citrus trees from VHR satellite imagery. *International Journal of Remote Sensing*, vol. 36, no. 17, pp. 4275-4296.
- [25] R. Qin, "A mean shift vector-based shape feature for classification of high spatial resolution remotely sensed imagery," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, Vol. 8, No. 5, pp. 1974-1985, 2015.
- [26] Li, M., Stein, A., Bijker, W., & Zhan, Q. (2016). Urban land use extraction from Very High Resolution remote sensing imagery using a Bayesian network. *ISPRS journal of photogrammetry and remote sensing*, vol. 122, pp. 192-205.
- [27] Regniers, O., Bombrun, L., Lafon, V., & Germain, C. (2016). Supervised classification of very high resolution optical images using wavelet-based textural features. *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 6, pp. 3722-3735.
- [28] Chaib, S., Gu, Y., & Yao, H. (2016). An informative feature selection method based on sparse PCA for VHR scene classification. *IEEE Geoscience and Remote Sensing Letters*, vol. 13, no. 2, pp. 147-151.
- [29] Zhao, W., Du, S., Wang, Q., & Emery, W. J. (2017). Contextually guided very-high-resolution imagery classification with semantic segments. *ISPRS journal of photogrammetry and remote sensing*, vol. 132, pp. 48-60.
- [30] Belgiu, M., & Drăguț, L. (2016). Random forest in remote sensing: A review of applications and future directions. *ISPRS journal of photogrammetry and remote sensing*, vol. 114, pp. 24-31.
- [31] Du, S., Zhang, F., & Zhang, X. (2015). Semantic classification of urban buildings combining VHR image and GIS data: An improved random forest approach. *ISPRS journal of photogrammetry and remote sensing*, vol. 105, pp. 107-119.
- [32] Abe, B., Olugbara, O., & Marwala, T. (2014). Experimental comparison of support vector machines with random forests for hyperspectral image land cover classification. *Journal of Earth System Science*, vol. 123, no. 4, pp. 779-790.
- [33] Feng, Q., Liu, J., & Gong, J. (2015). UAV remote sensing for urban vegetation mapping using random forest and texture analysis. *Remote Sensing*, vol. 7, no. 1, pp. 1074-1094.
- [34] Tramontana, G., Ichii, K., Camps-Valls, G., Tomelleri, E., & Papale, D. (2015). Uncertainty analysis of gross primary production upscaling using Random Forests, remote sensing and eddy covariance data. *Remote Sensing of Environment*, vol. 168, pp. 360-373.
- [35] Kakooei, M., & Baleghi, Y. (2020c). A two-level fusion for building irregularity detection in post-disaster VHR oblique images. *Earth Science Informatics*, vol. 13, no. 2, pp. 459-477.
- [36] Agapiou, A. (2017). Remote sensing heritage in a petabyte-scale: satellite data and heritage Earth Engine® applications. *International Journal of Digital Earth*, vol. 10, no. 1, pp. 85-102.
- [37] Trianni, G., Lisini, G., Angiuli, E., Moreno, E., Dondi, P., Gaggia, A., & Gamba, P. (2015). Scaling up to national/regional urban extent mapping using landsat data. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 7, pp. 3710-3719.
- [38] Miettinen, J., Shi, C., & Liew, S. C. (2017). Towards automated 10–30 m resolution land cover

mapping in insular South-East Asia. *Geocarto International*, pp. 1-15.

[39] Gorelick, N., Hancher, M., Dixon, M., Ilyushchenko, S., Thau, D., & Moore, R. (2017). Google Earth Engine: Planetary-scale geospatial analysis for everyone. *Remote Sensing of Environment*, vol. 202, pp. 18-27.

[40] ISPRS 2D semantic labeling contest Potsdam website (2018), Available: <http://www2.isprs.org/commissions/comm3/wg4/2d-sem-label-potsdam.html>.

[41] Haralick, R. M., & Shanmugam, K. (1973). Textural features for image classification. *IEEE Transactions on systems, man, and cybernetics*, vol. 6, pp. 610-621.

[42] Connors, R. W., Trivedi, M. M., & Harlow, C. A. (1984). Segmentation of a high-resolution urban scene using texture operators. *Computer vision, graphics, and image processing*, vol. 25, no. 3, pp. 273-310.

[43] Jiang, A., Xiao, S., Wei, L., & Zhu, Y. (2017). Research on Multiple Classifiers Combination Method for Remote Sensing Images. *International Conference on Smart Vehicular Technology, Transportation, Communication and Applications*, 2017.

[44] Kakooei, M., & Baleghi, Y. (2018). Leaf-Less-Tree feature for semantic labeling applications on Google Earth Engine. *9th International Symposium on the Telecommunications (IST)*, 2018.

برچسب زنی مفهومی تصاویر VHR با طبقه بندی جنگل تصادفی و ادغام ویژگی‌های طیفی و مکانی بر روی موتور Google Earth

محمد کاکوئی و یاسر بالغی*

دانشگاه صنعتی نوشیروانی بابل، دانشکده مهندسی برق و کامپیوتر، بابل، ایران.

ارسال ۲۰۱۹/۰۴/۰۸؛ بازنگری ۲۰۲۰/۰۲/۰۳؛ پذیرش ۲۰۲۰/۰۳/۳۱

چکیده:

یکی از زمینه‌های فعال در کاربردهای سنجش از دور، برچسب‌زنی مفهومی است. اشیاء در تصاویر نوری و مدل رقومی سطح (DSM) با رزولوشن خیلی بالا (VHR) دارای جزئیات زیادی هستند و می‌توانند دقت برچسب‌زنی مفهومی را افزایش دهد، اما مدیریت آنها یک کار چالش برانگیز است. در این مقاله، یک روش برچسب‌زنی مفهومی با ادغام تصاویر نوری و DSM ارائه شده است. ویژگی‌های طیفی و مکانی اشیاء ادغام شده‌اند تا برای آموزش طبقه‌بند استفاده شوند. برای ارزیابی از پایگاه داده‌ای استفاده شده است که کلاس‌های آن عبارت‌اند از: سطوح غیر قابل نفوذ، ساختمان، سبزه‌زار، درخت، ماشین و پس‌زمینه. روش پیشنهادی بر روی موتور Google Earth پیاده‌سازی شده است. این روش از چندین سطح تشکیل شده است. ابتدا، تحلیل مؤلفه‌های اصلی (PCA) به شاخص‌های سبزی‌نگی اعمال می‌شود تا باندی با حداکثر فاصله بین مناطق سبز و غیر سبز ایجاد شود. ماتریس ماتریس هم‌رخداد سطح خاکستری (GLCM) برای محاسبه ویژگی‌های مکانی مورد استفاده قرار می‌گیرد. چندین طبقه‌بند جنگل تصادفی (RF) به صورت خودکار آموزش داده می‌شود. پس از طبقه‌بندی، چندین عملگر مکانی به تصحیح نتایج طبقه‌بندی می‌پردازند. ویژگی درخت بدون برگ LeafLess (LLT) Tree برای بهبود شناخت درخت و افزایش دقت آن استفاده می‌شود. اندازه مساحت، محور اصلی و محور فرعی برای تصحیح شناختی ساختمان و ماشین استفاده می‌شود. ارزیابی نشان می‌دهد که ارتقای قابل توجهی در شناسایی درخت، ساختمان و ماشین ایجاد شده است. دقت کل و ضریب کاپا نیز مقدار مناسبی دارند.

کلمات کلیدی: برچسب زنی مفهومی تصاویر با رزولوشن خیلی بالا، ویژگی مکانی، موتور Google Earth، GLCM، ماتریس هم‌رخداد سطح خاکستری، جنگل تصادفی، درخت بدون برگ.