

# Investigations on the Potential of Binary and Multi-class Classification for Object Extraction from Airborne Laser Scanning Point Clouds

MARTIN WEINMANN<sup>1</sup>, ROSMARIE BLOMLEY<sup>1</sup>, MICHAEL WEINMANN<sup>2</sup> & BORIS JUTZI<sup>1</sup>

*Abstract: In this paper, we address the strategies of a binary classification and a multi-class classification for the pointwise semantic labeling of airborne laser scanning data. For both strategies, we use a collection of spherical and cylindrical neighborhoods as the basis for extracting geometric multi-scale features for each point of a considered point cloud. The extracted features, in turn, are provided as input to a standard Random Forest classifier. The results achieved for multi-class classification indicate a better classification across different classes, which is important for a subsequent spatial regularization. The results achieved for binary classification addressing the detection of cars, buildings and trees, respectively, show the potential for a subsequent extraction of individual objects.*

## 1 Introduction

Nowadays, the semantic interpretation of airborne laser scanning (ALS) point cloud data still commonly relies on a supervised classification. However, the small amount of training data which is typically available for this task not only hinders the use of modern deep learning techniques, but also the use of standard contextual classification techniques such as Conditional Random Fields (CRFs). The latter can cope with the irregular point sampling, while providing a theoretically well-founded approach for point cloud classification. However, CRFs require a sufficient number of training examples to 1) relate the data representations derived for individual 3D points and the defined classes to obtain an initial labeling and then 2) robustly infer relations among neighboring 3D points to improve the initial labeling by imposing spatial regularity. Thereby, the initial labeling is often achieved by using a classifier with probabilistic output, and this output, in turn, is used to define the unary potential also known as the association potential of a CRF. In most cases, the unary potentials are obtained via classic supervised classification techniques like Support Vector Machines or Random Forest classifiers.

When analyzing the performance of classic supervised classification techniques for point cloud classification, different influencing factors have to be taken into account. The most prominent ones are represented by the number and similarity of the defined classes, the set of involved features and the classification strategy itself. The latter becomes particularly important for applications which focus on object extraction in terms of detecting only objects corresponding to a specific class (e.g. “Car”, “Building” or “Tree”), while all remaining classes could also be merged to a common class “Background”.

---

<sup>1</sup> Karlsruhe Institute of Technology, Institute of Photogrammetry and Remote Sensing, Englerstraße 7, D-76131 Karlsruhe, E-Mail: [Martin.Weinmann, Rosmarie.Bломley, Boris.Jutzi]@kit.edu

<sup>2</sup> University of Bonn, Institute of Computer Science II, Friedrich-Ebert-Allee 144, D-53113 Bonn, E-Mail: mw@cs.uni-bonn.de

In this paper, we investigate the potential of different classification strategies for object extraction from ALS data. As baseline, we take into account the standard strategy of a multi-class classification which delivers a labeling with respect to multiple labels as indicated in Fig. 1. In addition, we focus on the strategy of a binary classification with respect to different objects of interest as e.g. represented by the semantic classes “Car”, “Building” and “Tree”. For each classification strategy, the classification itself is performed using a Random Forest classifier and based on a set of geometric features that has recently been proven to be appropriate for classifying ALS data (BLOMLEY & WEINMANN 2017). To demonstrate the performance of our classification approach, we perform tests on a publicly available ALS benchmark dataset for which a corresponding semantic labeling is provided.

After briefly summarizing related work (Section 2), we explain the proposed methodology for the semantic classification of ALS data in detail (Section 3). Subsequently, we present and discuss the derived experimental results with a specific focus on binary and multi-class classification of ALS data given user-defined class labels such as “Car”, “Building” and “Tree” (Section 4 and Section 5). Finally, we provide concluding remarks as well as suggestions for future work (Section 6).

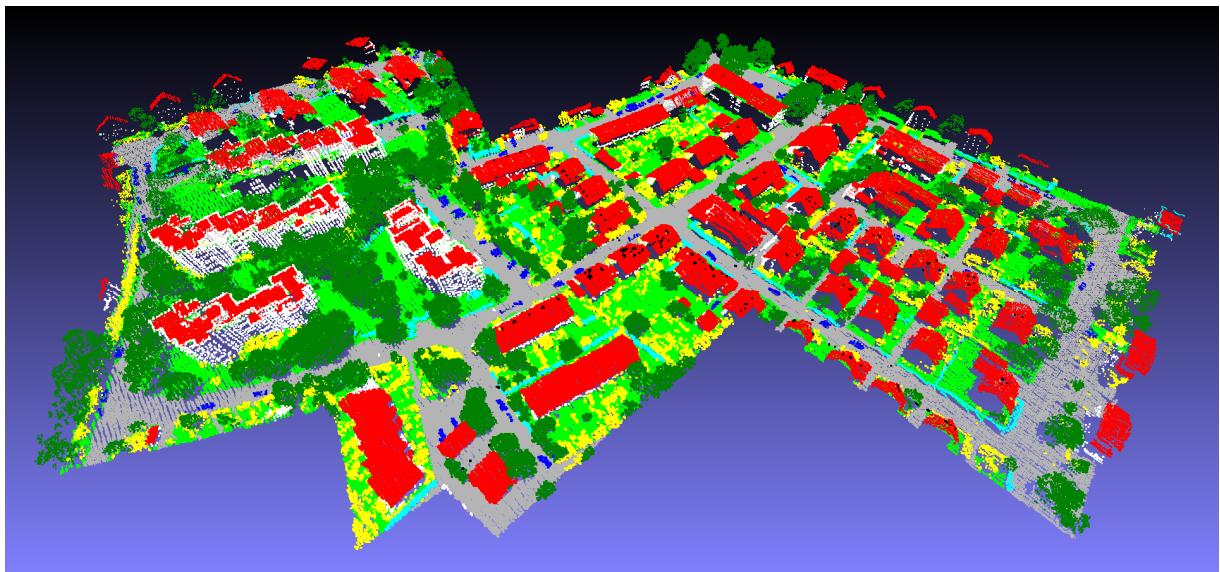


Fig. 1: ALS point cloud colored with respect to the classes “Roof” (red), “Façade” (white), “Impervious Surfaces” (gray), “Car” (blue), “Tree” (dark green), “Low Vegetation” (bright green), “Shrub” (yellow), “Fence/Hedge” (cyan) and “Powerline” (black). The point cloud comprises about 754k 3D points

## 2 Related Work

In this section, we summarize related work and thereby focus on the definition of appropriate local neighborhoods (Section 2.1) as the basis for extracting geometric features (Section 2.2) which, in turn, are provided as input to a classification framework (Section 2.3).

## 2.1 Neighborhood Definition

Commonly applied neighborhood definitions are represented by a cylindrical neighborhood (FILIN & PFEIFER 2005) or a spherical neighborhood (LEE & SCHENK 2002; LINSEN & PRAUTZSCH 2001). Such neighborhoods are parameterized by a scale parameter represented by either a radius (FILIN & PFEIFER 2005; LEE & SCHENK 2002) or the number of nearest neighbors (LINSEN & PRAUTZSCH 2001). This scale parameter is typically selected by involving knowledge about the scene and data, and an identical value is chosen for all points of the considered point cloud. Intuitively, however, a suitable value for the scale parameter might also depend on the local 3D structure and, consequently, a locally adaptive neighborhood definition seems to be more appropriate. This can for instance be achieved via dimensionality-based scale selection (DEMANTKÉ et al. 2011) favoring a highly dominant behavior of one of the dimensionality features (i.e. linearity, planarity and sphericity) or via eigenentropy-based scale selection (WEINMANN et al. 2015a) favoring the minimal disorder of 3D points. Indeed, the consideration of locally adaptive neighborhoods has been proven to be advantageous in comparison to the consideration of identically parameterized neighborhoods for all 3D points.

Instead of considering local point cloud characteristics at a single scale, it has also been proposed to consider a collection of such neighborhoods to derive a multi-scale representation for the local 3D structure. This allows a description of geometric properties at different scales and thereby implicitly accounts for the way in which these properties change across scales. Straightforward approaches for defining a multi-scale neighborhood only use differing values of the involved scale parameter. In this regard, a collection of cylindrical neighborhoods with infinite extent in the vertical direction and radii of 1m, 2m, 3m and 5m has been proposed (NIEMEYER et al. 2014) as well as a collection of spherical neighborhoods with different radii (BRODU & LAGUE 2012). Further multi-scale neighborhoods have for instance been proposed with a multi-scale voxel neighborhood (HACKEL et al. 2016) and with a collection of neighborhoods of different scale and type, e.g. in the form of voxels, blocks and pillars (HU et al. 2013), in the form of spatial bins, planar segments and local neighborhoods (GEVAERT et al. 2016) or in the form of cylindrical and spherical neighborhoods (BLOMLEY & WEINMANN 2017).

In this paper, we use a multi-scale, multi-type neighborhood composed of several cylindrical and spherical neighborhoods as this has been proven to be favorable for the classification of airborne laser scanning data compared to the use of a collection of spherical neighborhoods or a collection of cylindrical neighborhoods (BLOMLEY & WEINMANN 2017).

## 2.2 Feature Extraction

The spatial arrangement of 3D points within a defined local neighborhood can be encoded with a variety of geometric features. In this regard, metrical features describing local point cloud characteristics by evaluating certain geometric measures within the neighborhood are often used such as shape measures which are rather intuitive and represent one single property of the local neighborhood by a single value (WEST et al. 2004; JUTZI & GROSS 2009; MALLET et al. 2011; WEINMANN et al. 2015a). Furthermore, sampled features are often used which focus on a sampling of specific properties within the neighborhood e.g. in the form of histograms (OSADA et al. 2002; BLOMLEY et al. 2014). All extracted features are typically concatenated to a feature vector which is provided as input for classification.

In this paper, we focus on the consideration of both metrical features and sampled features to account for different geometric properties of the 3D points within the local neighborhood. In addition, we take into account that the scene is not necessarily flat and therefore use the spatial coordinates to approximate the local topography of the scene and thus derive the height above ground as additional feature for each 3D point. Instead of an accurate ground filtering of LiDAR data for automatically generating a digital terrain model (MONGUS & ZALIK 2012; SITHOLE & VOSSELMAN 2004; KRAUS & PFEIFER 1998), we assume that a coarse approximation of the local topography is already sufficient to derive normalized heights.

### 2.3 Classification

To classify the derived feature vectors, the concept of a supervised classification is typically applied. Given representative training data, the internal parameters of the classifier are tuned so that the classifier can afterwards well generalize to new data and assign appropriate class labels. Among the standard approaches for supervised classification, a Support Vector Machine (MALLET et al. 2011) or a Random Forest classifier (HACKEL et al. 2016; BLOMLEY & WEINMANN 2017) are often used. As such classifiers treat each 3D point individually by only considering the respective feature vector, a visualization of the derived labeling might reveal a “noisy” behavior.

To take into account that the labels of neighboring 3D points tend to be correlated and a spatial regularity of the derived labeling should be given, contextual information preserved in the feature vectors and labels of neighboring points can be involved. Such a contextual classification is often realized by using a CRF (NIEMEYER et al. 2014; WEINMANN et al. 2015b; STEINSIEK et al. 2017) or Associative and non-Associative Markov Networks (MUNOZ et al. 2009; SHAPOVALOV et al. 2010), but in general a diversity of structured regularization approaches can be used (LANDRIEU et al. 2017).

In this paper, we investigate the potential of different classification strategies for object extraction from ALS data given a small amount of training data. For this purpose, we follow the standard strategy of a multi-class classification which delivers a labeling with respect to multiple labels, and we additionally involve the strategy of a binary classification with respect to different objects of interest. For both cases, we make use of a Random Forest classifier which is a representative of modern discriminative methods.

## 3 Methodology

To investigate the potential of binary and multi-class classification for object extraction from ALS point clouds, we propose a framework consisting of three components. Assuming input data in the form of only spatial coordinates of 3D points without additional information, the first component exploits the given spatial information to derive suitable multi-scale, multi-type neighborhoods (Section 3.1). These multi-scale, multi-type neighborhoods represent the basis for describing the local 3D structure at different scales by extracting a set of low-level geometric features (Section 3.2). The extracted geometric features, in turn, are provided as input to a standard supervised classification approach (Section 3.3), whereby the strategies of a binary classification or a multi-class classification may be followed.

### 3.1 Neighborhood Definition

As the selected neighborhood definition should allow appropriately describing local point cloud characteristics, we focus on the combination of (i) a cylindrical multi-scale neighborhood, (ii) a spherical multi-scale neighborhood, (iii) a locally adaptive spherical neighborhood and (iv) spatial bins as the basis for feature extraction:

- The cylindrical multi-scale neighborhood consists of four cylindrical neighborhoods which are aligned along the vertical direction, have infinite extent in the vertical direction and are parameterized by radii of 1m, 2m, 3m and 5m, respectively (NIEMEYER et al. 2014).
- The spherical multi-scale neighborhood consists of four spherical neighborhoods which are parameterized by radii of 1m, 2m, 3m and 5m, respectively.
- The locally adaptive spherical neighborhood is derived via eigenentropy-based scale selection (WEINMANN et al. 2015a) which delivers an individual spherical neighborhood comprising the optimal number of nearest neighbors with respect to the Euclidean distance in 3D space.
- The spatial bins are derived by partitioning the scene with respect to a horizontally oriented plane into quadratic bins with a side length of 20m, and the bins in turn are only used as the basis for approximating the topography of the considered scene (BLOMLEY & WEINMANN 2017).

Thus, a collection of 10 neighborhoods is used to achieve an advanced multi-scale, multi-type neighborhood providing the basis for feature extraction.

### 3.2 Feature Extraction

On the basis of the defined local neighborhoods, we derive a set of low-level geometric features comprising (i) covariance features, (ii) geometric 3D properties, (iii) shape distributions and (iv) the normalized height feature:

- The covariance features are extracted from the 3D structure tensor, a 3D covariance matrix derived from the spatial coordinates of all points within the considered local neighborhood. The eigenvalues of the 3D structure tensor are normalized by their sum and then used to define the features of linearity, planarity, sphericity, omnivariance, anisotropy, eigenentropy, sum of eigenvalues and change of curvature (WEST et al. 2004; PAULY et al. 2003).
- The geometric 3D properties are derived by evaluating intuitive geometric point cloud statistics within the considered local neighborhood. In this regard, we use the features represented by the local point density, the verticality, and the maximum difference as well as the standard deviation of the height values corresponding to those points within the local neighborhood (WEINMANN et al. 2015a). For the locally adaptive spherical neighborhood derived via eigenentropy-based scale selection, we additionally consider the radius of the local neighborhood as feature.

- The shape distributions (OSADA et al. 2002) have originally been introduced to describe the shape of complete objects. More specifically, shape distributions are histograms of shape values typically derived from random point samples by applying distance metrics or angular metrics. In this regard, the angle between any three random points (A3), the distance of one random point from the centroid of all points within the neighborhood (D1), the distance between two random points (D2), the square root of the area spanned by a triangle between three random points (D3) and the cubic root of the volume spanned by a tetrahedron between four random points (D4) represent commonly used metrics. In our work, we use the adaptation of shape distributions to describe local point cloud characteristics within a selected neighborhood (BLOMLEY et al. 2014). For each of the metrics A3, D1, D2, D3 and D4, we randomly select 255 minimal point samples from the considered neighborhood, evaluate the metric for each point sample, put the resulting values in a histogram and finally consider the distribution of histogram counts. Thereby, we use histograms with 10 bins and we estimate the binning thresholds in an adaptive procedure based on 500 exemplary local neighborhoods (BLOMLEY et al. 2014).
- The normalized height feature is derived from an approximation of the scene topography as shown in Fig. 2. First, absolute height minima are determined on a large grid for which we define a sampling distance of 20m. Subsequently, a linear interpolation is performed among those coarsely gridded minimum values and evaluated on a fine grid of 0.5m sampling distance. Finally, the difference between the height value of a 3D point and the topographic height value of the corresponding grid cell is considered to derive a normalized height value for that 3D point (BLOMLEY & WEINMANN 2017).

Thus, we derive 62 features per neighborhood parameterized by a fixed radius, 63 features for a neighborhood determined via eigenentropy-based scale selection, and the normalized height feature that is additionally used.

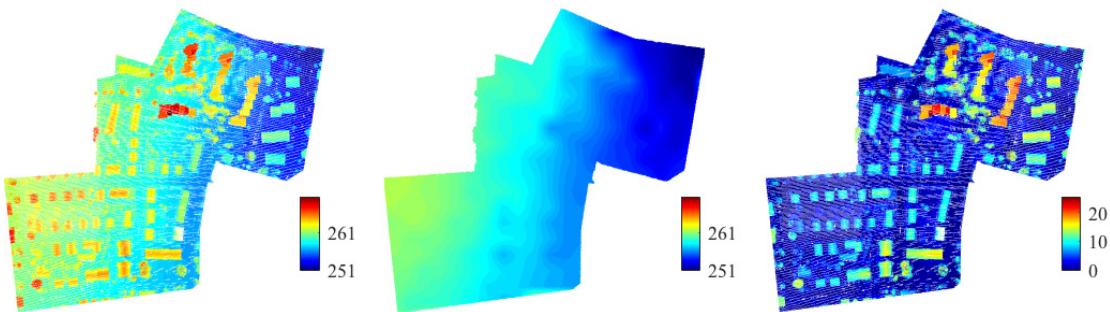


Fig. 2: Height minima on a 0.5m grid (left), approximation of the scene topography (center) and the normalized heights (right). The considered scene is the same as in Fig. 1

As the considered features address different quantities and may therefore be associated with different units as well as a different range of values which, in turn, might have a negative impact on the classification results, we introduce a normalization of the derived feature vectors. For the covariance features, the geometric 3D properties and the normalized height feature, we use a linear mapping to the interval [0,1]. Thereby, we reduce the effect of outliers by determining the

new range of the data via the 1<sup>st</sup> percentile and the 99<sup>th</sup> percentile of the training data (BLOMLEY & WEINMANN 2017). For the shape distributions, the normalization is achieved by dividing each histogram count by the total number of pulls from the local neighborhood (i.e. by 255).

### 3.3 Classification

For classification, we focus on a supervised classification and use a Random Forest classifier (BREIMAN 2001) which relies on the principle of ensemble learning, where the main idea consists in strategically generating a set of weak learners and combining them to achieve a strong learner. In the training stage, the weak learners are generated via bootstrap aggregating also known as “bagging” (BREIMAN 1996). More specifically, different subsets of the training data are randomly drawn with replacement so that an individual decision tree can be trained for each subset. Thereby, each decision tree is trained via the successive splitting of the considered training data to smaller subsets whose instances belong to the same class, respectively. In the prediction stage, the generated ensemble of randomly trained decision trees is used to assign a label to new feature vectors. Each of the decision trees casts a vote for one of the defined classes, and the majority vote across all decision trees is used to estimate the respective class label.

## 4 Experimental Results

To demonstrate the performance of our framework, we perform tests on a benchmark dataset (Section 4.1). Involving a variety of commonly used evaluation metrics (Section 4.2), we provide a detailed quantitative assessment of the quality of the derived classification results (Section 4.3).

### 4.1 Dataset

For our experiments, we use the Vaihingen Dataset (CRAMER 2010; ROTTENSTEINER et al. 2012) which is provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) and available upon request. This dataset has been acquired with a Leica ALS50 system over a small German village with many detached buildings surrounded by trees and small multi-story buildings. In total, the dataset consists of about 1.166M 3D points and is split into a training scene with about 754k 3D points and a test scene with about 412k 3D points. While a reference labeling with respect to the classes “Powerline”, “Low Vegetation”, “Impervious Surfaces”, “Car”, “Fence/Hedge”, “Roof”, “Façade”, “Shrub” and “Tree” is provided for the training scene (see Fig. 1), it is missing for the test scene so that derived classification results have to be submitted to the organizers of the ISPRS Benchmark on 3D Semantic Labeling who perform the evaluation externally.

### 4.2 Evaluation Metrics

For performance evaluation, we consider commonly used evaluation metrics that allow quantifying the quality of derived classification results on a per-point basis. We consider global evaluation metrics represented by the overall accuracy (OA), the kappa-index ( $\kappa$ ) and the mean  $F_1$ -score across all classes ( $mF_1$ ). Furthermore, we take into account that an imbalanced distribution of the occurrence of single classes might introduce a bias in the global evaluation

metrics and we therefore additionally consider the classwise evaluation metrics represented by recall (R), precision (P) and F<sub>1</sub>-score (F<sub>1</sub>), where the latter is a compound metric combining recall and precision with equal weights.

### 4.3 Results

First, we focus on the strategy of a multi-class classification. We train the involved Random Forest classifier and thereby take into account that an imbalanced distribution of training examples across all classes might have a detrimental effect on the training process (CRIMINISI & SHOTTON 2013). Accordingly, we introduce a class re-balancing by randomly sampling an identical number of training examples per class for the training phase. For those classes for which less training examples are available, this results in a duplication of training examples. Using 10,000 training examples per class to train the involved classifier, the classification results achieved for the test scene correspond to an overall accuracy of 68.1%, a kappa index of 60.5% and a mean F<sub>1</sub>-score of 52.6%. To better account for the variability given in some of the classes, we also use 100,000 training examples per class to train the classifier. The respective classification results achieved for the test scene correspond to an overall accuracy of 71.5%, a kappa index of 64.3% and a mean F<sub>1</sub>-score of 58.3%, i.e. a significant improvement of the classification results is given ( $\Delta OA = 3.4\%$ ,  $\Delta \kappa = 3.8\%$ ,  $\Delta mF_1 = 5.7\%$ ). For this case, the classification results are visualized in Fig. 3 and the corresponding confusion matrix as well as more details about classwise evaluation metrics are provided in Tab. 1.

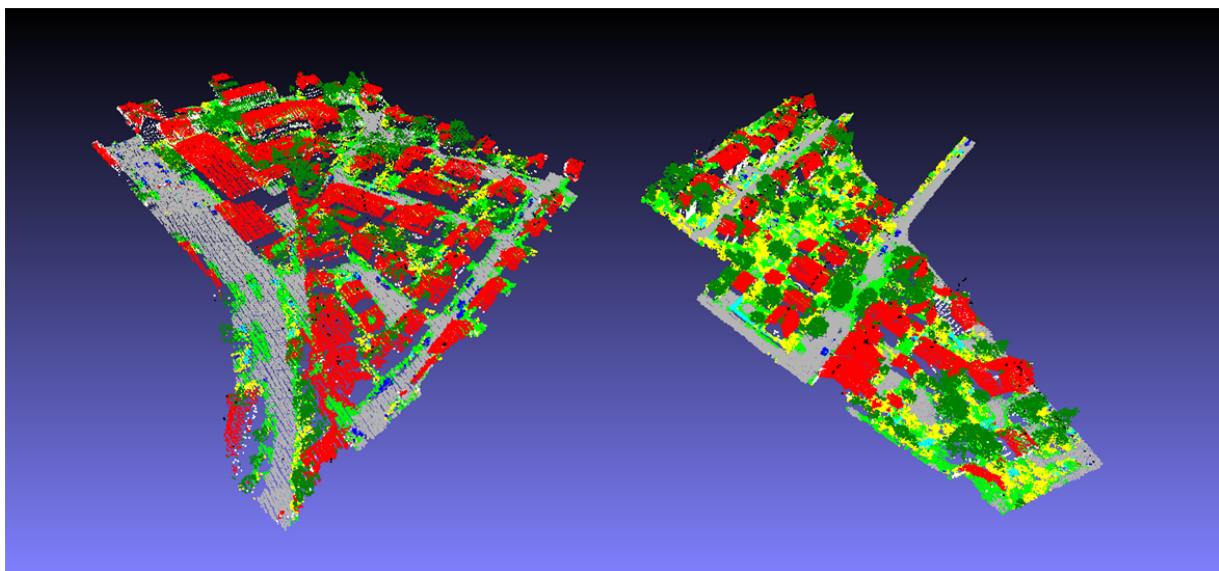


Fig. 3: Classified point cloud colored with respect to the classes “Roof” (red), “Façade” (white), “Impervious Surfaces” (gray), “Car” (blue), “Tree” (dark green), “Low Vegetation” (bright green), “Shrub” (yellow), “Fence/Hedge” (cyan) and “Powerline” (black). The point cloud comprises about 412k 3D points

Tab. 1: Confusion matrix and the classwise evaluation metrics of recall, precision and  $F_1$ -score (in %) corresponding to the result of multi-class classification ( $OA=71.5\%$ ,  $\kappa=64.3\%$ ,  $mF_1=58.3\%$ )

| Class                         |                     | Estimated Label |          |            |      |             |       |        |       |       |
|-------------------------------|---------------------|-----------------|----------|------------|------|-------------|-------|--------|-------|-------|
|                               |                     | Power-line      | Low Veg. | Imp. Surf. | Car  | Fence/Hedge | Roof  | Façade | Shrub | Tree  |
| Reference Label               | Powerline           | 432             | 1        | 0          | 0    | 0           | 44    | 20     | 2     | 101   |
|                               | Low Vegetation      | 0               | 56068    | 15233      | 125  | 1692        | 8859  | 595    | 14128 | 1990  |
|                               | Impervious Surfaces | 1               | 19068    | 80162      | 146  | 118         | 1502  | 38     | 850   | 101   |
|                               | Car                 | 0               | 644      | 112        | 1082 | 428         | 121   | 7      | 1280  | 34    |
|                               | Fence/Hedge         | 0               | 934      | 48         | 53   | 1212        | 256   | 87     | 4059  | 773   |
|                               | Roof                | 181             | 1534     | 183        | 7    | 503         | 90920 | 2164   | 2494  | 11062 |
|                               | Façade              | 26              | 365      | 13         | 33   | 67          | 808   | 5668   | 1468  | 2776  |
|                               | Shrub               | 7               | 3177     | 239        | 152  | 1182        | 549   | 546    | 15351 | 3615  |
|                               | Tree                | 28              | 546      | 25         | 7    | 277         | 1977  | 1266   | 6575  | 43525 |
| <b>Recall</b>                 |                     | 72.0            | 56.8     | 78.6       | 29.2 | 16.3        | 83.4  | 50.5   | 61.9  | 80.3  |
| <b>Precision</b>              |                     | 64.0            | 68.1     | 83.5       | 67.4 | 22.1        | 86.6  | 54.5   | 33.2  | 68.0  |
| <b><math>F_1</math>-Score</b> |                     | 67.8            | 61.9     | 81.0       | 40.7 | 18.8        | 84.9  | 52.4   | 43.2  | 73.6  |

Besides the strategy of a multi-class classification, we also focus on the strategy of a binary classification with respect to the classes “Car”, “Building” and “Tree”, respectively. While the classes “Car” and “Tree” are coincident with their counterparts in the provided reference labeling, the class “Building” is composed of the classes “Roof” and “Façade”. Again, we randomly sample an identical number of 100,000 training examples per class to train the classifier. For the binary classification addressing “Car” and “Background”, we achieve an overall accuracy of 99.0%, a kappa index of 43.7% and a mean  $F_1$ -score of 71.9%. For the binary classification addressing “Building” and “Background”, we achieve an overall accuracy of 91.4%, a kappa index of 79.5% and a mean  $F_1$ -score of 89.8%. For the binary classification addressing “Tree” and “Background”, we achieve an overall accuracy of 89.0%, a kappa index of 62.6% and a mean  $F_1$ -score of 81.1%. The classwise evaluation metrics for all three cases are provided in Tab. 2. A comparison to Tab. 1 reveals that the binary classification leads to a better detection of cars and buildings, while the detection of trees is worse in comparison to the results achieved with the strategy of a multi-class classification.

Tab. 2: Classwise evaluation metrics of recall, precision and  $F_1$ -score (in %) corresponding to the results of a binary classification with respect to “Car” (left), “Building” (center) and “Tree” (right).

|                               | Car  | Background |                               | Building | Background |                               | Tree | Background |
|-------------------------------|------|------------|-------------------------------|----------|------------|-------------------------------|------|------------|
| <b>Recall</b>                 | 44.1 | 99.5       | <b>Recall</b>                 | 87.7     | 92.9       | <b>Recall</b>                 | 91.6 | 88.7       |
| <b>Precision</b>              | 44.3 | 99.5       | <b>Precision</b>              | 83.7     | 94.8       | <b>Precision</b>              | 55.1 | 98.6       |
| <b><math>F_1</math>-Score</b> | 44.2 | 99.5       | <b><math>F_1</math>-Score</b> | 85.7     | 93.9       | <b><math>F_1</math>-Score</b> | 68.8 | 93.9       |

The derived results for multi-class classification and binary classification are visualized in Fig. 4 for the test scene with about 412k 3D points.

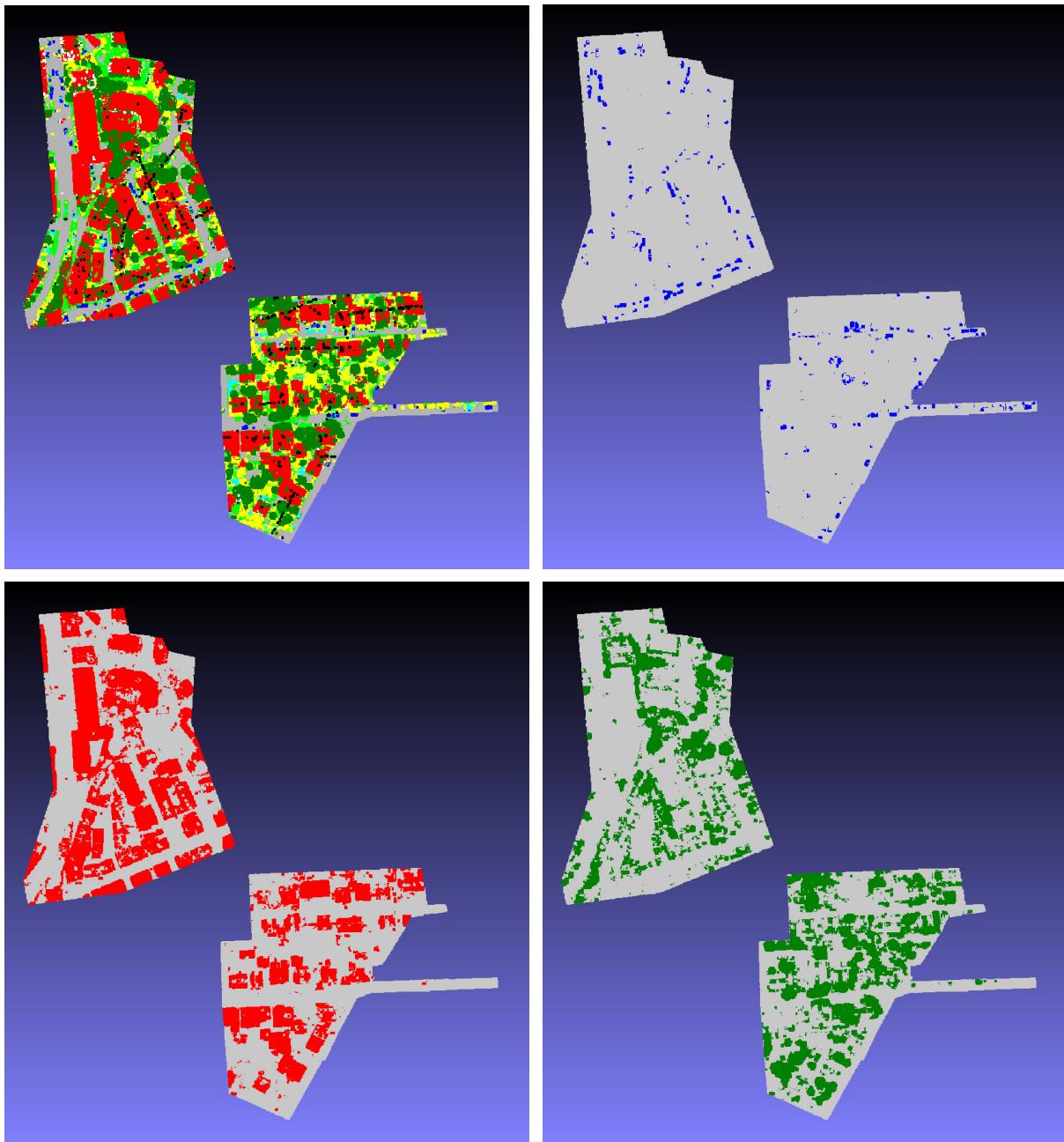


Fig. 4: Classified point clouds: the top left part shows the result of a multi-class classification with respect to "Roof" (red), "Façade" (white), "Impervious Surfaces" (gray), "Car" (blue), "Tree" (dark green), "Low Vegetation" (bright green), "Shrub" (yellow), "Fence/Hedge" (cyan) and "Powerline" (black); the top right part shows the result of a binary classification with respect to "Car" (blue) and "Background" (gray); the bottom left part shows the result of a binary classification with respect to "Building" (red) and "Background" (gray); the bottom right part shows the result of a binary classification with respect to "Tree" (green) and "Background" (gray).

## 5 Discussion

The derived results for multi-class classification indicate that the Vaihingen Dataset with the nine defined classes is rather challenging for the proposed methodology, as the achieved overall accuracy is only about 71.5%. The reason for this is that only geometric features are provided as input to the classifier and that several of the defined classes are characterized by a high geometric similarity. Particularly the classes “Low Vegetation” and “Impervious Surfaces” as well as the classes “Shrub” and “Tree” exhibit a similar geometric behavior and therefore misclassifications among these classes occur quite often (Tab. 1). This is in accordance with other recent investigations involving the same dataset (STEINSIEK et al. 2017; BLOMLEY & WEINMANN 2017). A more detailed comparison to the results achieved with other classification approaches is provided in Tab. 3 and reveals that the proposed methodology involving a Random Forest for classification outperforms other approaches using the same classifier. The significant improvement across several classes thereby comes at the cost of a loss with respect to the predominant classes represented by “Low Vegetation” and “Impervious Surfaces”. The mean  $F_1$ -score is however 8.3% higher than in recent investigations (STEINSIEK et al. 2017), which indicates a better classification of the different classes. Consequently, interpreting the votes of the involved Random Forest classifier in a probabilistic way would provide a better association potential for a CRF imposing spatial regularity on the derived classification results which, in turn, is likely to allow the CRF to further increase the quality of the classification results.

Tab. 3: Classwise  $F_1$ -scores as well as the overall accuracy (OA) and the mean  $F_1$ -score across all classes ( $mF_1$ ) (in %) for different approaches (RF<sup>1</sup>: classification based on a Random Forest (STEINSIEK et al. 2017); CRF<sup>1</sup>: classification based on a CRF approach (STEINSIEK et al. 2017); RF<sup>2</sup>: classification based on a Random Forest (BLOMLEY & WEINMANN 2017); RF<sup>3</sup>: classification based on the proposed approach).

|                  | $F_1$      |          |            |      |             |      |        |       |      | OA   | $mF_1$ |
|------------------|------------|----------|------------|------|-------------|------|--------|-------|------|------|--------|
|                  | Power-line | Low Veg. | Imp. Surf. | Car  | Fence/Hedge | Roof | Façade | Shrub | Tree |      |        |
| RF <sup>1</sup>  | 14.3       | 65.8     | 86.1       | 24.9 | 19.8        | 84.8 | 43.9   | 40.8  | 69.5 | 71.0 | 50.0   |
| CRF <sup>1</sup> | 69.8       | 73.8     | 91.5       | 58.2 | 29.9        | 91.6 | 54.7   | 47.8  | 80.2 | 80.5 | 66.4   |
| RF <sup>2</sup>  | 32.1       | 57.9     | 80.0       | 44.1 | 17.5        | 81.8 | 47.5   | 41.5  | 70.9 | 68.1 | 52.6   |
| RF <sup>3</sup>  | 67.8       | 61.9     | 81.0       | 40.7 | 18.8        | 84.9 | 52.4   | 43.2  | 73.6 | 71.5 | 58.3   |

Furthermore, the derived results reveal an improvement for the classes “Car” and “Building” if a binary classification is considered instead of a multi-class classification (Tab. 1 and Tab. 2). However, the binary classification delivers worse results for the class “Tree” which might be due to shrubs revealing geometrically similar characteristics. Indeed, there are many misclassifications among the classes “Shrub” and “Tree” in the case of a multi-class classification (Tab. 1). As the elements of the class “Shrub” are added to the class “Background” for binary classification, the relevant classes “Tree” and “Background” contain more similar instances which, in turn, results in lower detection rates. A consideration of the results for binary classification also reveals that the class “Building” can easily be identified, while the class “Car” seems to be more challenging (Tab. 2). The latter is due to the similar geometric behavior of the

classes “Car”, “Fence/Hedge” and “Shrub” for the given point density, which results in many misclassifications between these classes in the case of a multi-class classification (Tab. 1). Note that, for a binary classification, the overall accuracy does not necessarily provide an appropriate conclusion about the objects of interest. For the classification with respect to the classes “Car” and “Background”, the overall accuracy is 99.0%, although the detection rates are rather low for the class “Car”. This is due to the small amount of 3D points that should be labeled as “Car” in comparison to a huge amount of 3D points that should be labeled as “Background”. Hence, even classifying all 3D points as “Background” would still result in a high overall accuracy, while the kappa-index allows reasoning about the class separability and hence would indicate a poor classification result. For this example, the mean  $F_1$ -score across all classes would also indicate a poor classification result as it accounts for the appropriate detection of both defined classes.

## 6 Conclusions

In this paper, we have investigated the potential of binary and multi-class classification as the basis for object extraction from ALS data. We have focused on the use of a multi-scale, multi-type neighborhood composed of several cylindrical and spherical neighborhoods as the basis for extracting a variety of geometric features. The extracted features, in turn, have been provided as input for classification, whereby the focus has been put on the strategies of binary and multi-class classification based on a Random Forest classifier. The classification results derived for a benchmark dataset have clearly revealed the potential of the proposed methodology for pointwise classification. Regarding the strategy of a multi-class classification, the achieved results indicate a significantly better classification across all classes in comparison to other similar approaches which tend to better classify the dominant classes. Regarding the strategy of a binary classification, we have addressed the detection of cars, buildings and trees, respectively. The achieved classification results provide a good initial labeling for a subsequent extraction and counting of single objects in the scene, yet additional effort is required to account for misclassifications and only retain segments which correspond to real objects in the scene.

In future work, we intend to improve the classification results by first using the proposed methodology to achieve an initial labeling and then imposing spatial regularity by using spatial regularization techniques (NIEMEYER et al. 2014; STEINSIEK et al. 2017; LANDRIEU et al. 2017). Thereby, the proposed methodology allows for a better classification of different classes and thus a better initial labeling which serves as input to the CRF via the association potentials and this, in turn, is likely to allow the CRF to further increase the quality of the classification results compared to other approaches. Furthermore, we plan to address the step towards the extraction of individual objects in the scene.

## 7 Acknowledgements

The Vaihingen Dataset was kindly provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (CRAMER 2010): <http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html>.

## 8 References

- BLOMLEY, R., WEINMANN, M., LEITLOFF, J. & JUTZI, B., 2014: Shape distribution features for point cloud analysis – A geometric histogram approach on multiple scales. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **II(3)**, 9-16.
- BLOMLEY, R. & WEINMANN, M., 2017: Using multi-scale features for the 3D semantic labeling of airborne laser scanning data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **IV(2/W4)**, 43-50.
- BREIMAN, L., 1996: Bagging predictors. *Machine Learning*, **24**(2), 123-140.
- BREIMAN, L., 2001: Random forests. *Machine Learning*, **45**(1), 5-32.
- BRODU, N. & LAGUE, D., 2012: 3D terrestrial LiDAR data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology. *ISPRS Journal of Photogrammetry and Remote Sensing*, **68**, 121-134.
- CRAMER, M., 2010: The DGPF-test on digital airborne camera evaluation – Overview and test design. *PFG Photogrammetrie – Fernerkundung – Geoinformation*, **2/2010**, 73-82.
- CRIMINISI, A. & SHOTTON, J., 2013: Decision forests for computer vision and medical image analysis. Springer, London, UK.
- DEMANTKÉ, J., MALLET, C., DAVID, N. & VALLET, B., 2011: Dimensionality based scale selection in 3D LiDAR point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **38**(5/W12), 97-102.
- FILIN, S. & PFEIFER, N., 2005: Neighborhood systems for airborne laser data. *Photogrammetric Engineering & Remote Sensing*, **71**(6), 743-755.
- GEVAERT, C.M., PERSELLO, C. & VOSSELMAN, G., 2016: Optimizing multiple kernel learning for the classification of UAV data. *Remote Sensing*, **8**(12), 1-22.
- HACKEL, T., WEGNER, J.D. & SCHINDLER, K., 2016: Fast semantic segmentation of 3D point clouds with strongly varying density. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **III(3)**, 177-184.
- HU, H., MUÑOZ, D., BAGNELL, J. A. & HEBERT, M., 2013: Efficient 3-D scene analysis from streaming data. *Proceedings of the IEEE International Conference on Robotics and Automation*, 2297-2304.
- JUTZI, B. & GROSS, H., 2009: Nearest neighbour classification on laser point clouds to gain object structures from buildings. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **38**(1-4-7/W5), 1-6.
- KRAUS, K. & PFEIFER, N., 1998: Determination of terrain models in wooded areas with airborne laser scanner data. *ISPRS Journal of Photogrammetry and Remote Sensing*, **53**(4), 193-203.
- LANDRIEU, L., RAGUET, H., VALLET, B., MALLET, C. & WEINMANN, M., 2017: A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, **132**, 102-118.
- LEE, I. & SCHENK, T., 2002: Perceptual organization of 3D surface points. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, **34**(3A), 193-198.

- LINSEN, L. & PRAUTZSCH, H., 2005: Local versus global triangulations. Proceedings of Eurographics, 257-263.
- MALLET, C., BRETAR, F., ROUX, M., SOERGEL, U. & HEIPKE, C., 2011: Relevance assessment of full-waveform LiDAR data for urban area classification. ISPRS Journal of Photogrammetry and Remote Sensing, **66**(6), S71-S84.
- MONGUS, D. & ZALIK, B., 2012: Parameter-free ground filtering of LiDAR data for automatic DTM generation. ISPRS Journal of Photogrammetry and Remote Sensing, **67**, 1-12.
- MUNOZ, D., BAGNELL, J. A., VANDAPEL, N. & HEBERT, M., 2009: Contextual classification with functional max-margin Markov networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 975-982.
- NIEMEYER, J., ROTTENSTEINER, F. & SOERGEL, U., 2014: Contextual classification of LiDAR data and building object detection in urban areas. ISPRS Journal of Photogrammetry and Remote Sensing, **87**, 152-165.
- OSADA, R., FUNKHOUSER, T., CHAZELLE, B. & DOBKIN, D., 2002: Shape distributions. ACM Transactions on Graphics, **21**(4), 807-832.
- PAULY, M., KEISER, R. & GROSS, M., 2003: Multi-scale feature extraction on point-sampled surfaces. Computer Graphics Forum, **22**(3), 81-89.
- ROTTENSTEINER, F., SOHN, G., JUNG, J., GERKE, M., BAILLARD, C., BENITEZ, S. & BREITKOPF, U., 2012: The ISPRS benchmark on urban object classification and 3D building reconstruction. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **I**(3), 293-298.
- SHAPOVALOV, R., VELIZHEV, A. & BARINOVA, O., 2010: Non-associative Markov networks for 3D point cloud classification. The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **38**(3A), 103-108.
- SITHOLE, G. & VOSSELMAN, G., 2004: Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. ISPRS Journal of Photogrammetry and Remote Sensing, **59**(1-2), 85-101.
- STEINSIEK, M., POLEWSKI, P., YAO, W. & KRZYSTEK, P., 2017: Semantische Analyse von ALS- und MLS-Daten in urbanen Gebieten mittels Conditional Random Fields. Tagungsband der 37. Wissenschaftlich-Technischen Jahrestagung der DGPF, 521-531.
- WEINMANN, M., JUTZI, B., HINZ, S. & MALLET, C., 2015a: Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. ISPRS Journal of Photogrammetry and Remote Sensing, **105**, 286-304.
- WEINMANN, M., SCHMIDT, A., MALLET, C., HINZ, S., ROTTENSTEINER, F. & JUTZI, B., 2015b: Contextual classification of point cloud data by exploiting individual 3D neighborhoods. ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, **II** (3/W4), 271-278.
- WEST, K.F., WEBB, B.N., LERSCH, J.R., POTIER, S., TRISCARI, J.M. & IVERSON, A.E., 2004: Context-driven automated target detection in 3-D data. Proceedings of SPIE, **5426**, 133-143.