

SEMANTIC 3D SCENE INTERPRETATION: A FRAMEWORK COMBINING OPTIMAL NEIGHBORHOOD SIZE SELECTION WITH RELEVANT FEATURES

Martin Weinmann^a, Boris Jutzi^a, Clément Mallet^b

^a Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Englerstr. 7, 76131 Karlsruhe, Germany - {martin.weinmann, boris.jutzi}@kit.edu

^b IGN, MATIS lab., Université Paris Est
73 avenue de Paris, 94160 Saint-Mandé, France - clement.mallet@ign.fr

Commission III, WG III/2

KEY WORDS: LIDAR, Point Cloud, Neighborhood Size Selection, Features, Feature Selection, Classification, Interpretation

ABSTRACT:

3D scene analysis by automatically assigning 3D points a semantic label has become an issue of major interest in recent years. Whereas the tasks of feature extraction and classification have been in the focus of research, the idea of using only relevant and more distinctive features extracted from optimal 3D neighborhoods has only rarely been addressed in 3D lidar data processing. In this paper, we focus on the interleaved issue of extracting relevant, but not redundant features and increasing their distinctiveness by considering the respective optimal 3D neighborhood of each individual 3D point. We present a new, fully automatic and versatile framework consisting of four successive steps: (i) optimal neighborhood size selection, (ii) feature extraction, (iii) feature selection, and (iv) classification. In a detailed evaluation which involves 5 different neighborhood definitions, 21 features, 6 approaches for feature subset selection and 2 different classifiers, we demonstrate that optimal neighborhoods for individual 3D points significantly improve the results of scene interpretation and that the selection of adequate feature subsets may even further increase the quality of the derived results.

1 INTRODUCTION

The automatic interpretation of 3D point clouds represents a fundamental issue in photogrammetry, remote sensing and computer vision. Nowadays, different subtopics are in the focus of research such as point cloud classification (Hu et al., 2013; Niemeyer et al., 2014; Xu et al., 2014), object recognition (Pu et al., 2011; Velizhev et al., 2012), creation of large-scale city models (Lafarge and Mallet, 2012) or urban accessibility analysis (Serna and Marcotegui, 2013). For all of them, it is important to cope with the complexity of 3D scenes caused by the irregular sampling and very different types of objects as well as the computational burden arising from both large 3D point clouds and a variety of available features.

For scene interpretation in terms of uniquely assigning each 3D point a semantic label (e.g. *ground*, *building* or *vegetation*), the straightforward approach is to extract respective geometric features from its local 3D structure. Thus, the features rely on a local 3D neighborhood which is typically chosen as spherical neighborhood with fixed radius (Lee and Schenk, 2002), cylindrical neighborhood with fixed radius (Filin and Pfeifer, 2005) or spherical neighborhood formed by a fixed number of the k closest 3D points (Linsen and Prutzsch, 2001). Once features have been calculated, the classification of each 3D point may be conducted via standard supervised learning approaches such as Gaussian Mixture Models (Lalonde et al., 2005), Support Vector Machines (Secord and Zakhor, 2007), AdaBoost (Lodha et al., 2007), a cascade of binary classifiers (Carlberg et al., 2009), Random Forests (Chehata et al., 2009) and Bayesian Discriminant Classifiers (Khoshelham and Oude Elberink, 2012). In contrast, contextual learning approaches also involve relationships among 3D points in a local neighborhood¹ which have to be inferred from the training data. Respective methods for classifying

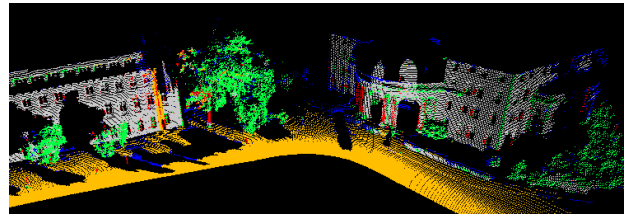


Figure 1: 3D point cloud with assigned labels (wire: blue, pole/trunk: red, façade: gray, ground: brown, vegetation: green).

point cloud data have been proposed with Associative and non-Associative Markov Networks (Munoz et al., 2009a; Shapovalov et al., 2010), Conditional Random Fields (Niemeyer et al., 2012), multi-stage inference procedures focusing on point cloud statistics and relational information over different scales (Xiong et al., 2011), and spatial inference machines modeling mid- and long-range dependencies inherent in the data (Shapovalov et al., 2013).

Since the semantic labels of nearby 3D points tend to be correlated (Figure 1), involving a smooth labeling is often desirable. However, exact inference is computationally intractable when applying contextual learning approaches. Instead, either approximate inference techniques or smoothing techniques are commonly applied. Approximate inference techniques remain challenging as there is no indication towards an optimal inference strategy, and they quickly reach their limitations if the considered neighborhood is becoming too large. In contrast, smoothing techniques may provide a significant improvement concerning classification accuracy (Schindler, 2012). All of these techniques exploit either the estimated probability of a 3D point belonging to each of the defined classes or the direct assignment of the respective label, and thus the results of a classification for individual 3D points. Consequently, it seems desirable to investigate sources for potential improvements with respect to classification accuracy.

¹This local neighborhood is typically different from the one used for feature extraction.

One potential improvement may address the design of utilized features as, despite the different neighborhood definitions, the pa-

parameterization of the neighborhood is still typically selected with respect to empirical a priori knowledge on the scene and identical for all 3D points. This raises the question about estimating the optimal neighborhood for each individual 3D point and thus increasing the distinctiveness of derived features. Respective approaches addressing this issue are based on local surface variation (Pauly et al., 2003; Belton and Lichti, 2006), iterative schemes relating neighborhood size to curvature, point density and noise of normal estimation (Mitra and Nguyen, 2003; Lalonde et al., 2005), or dimensionality-based scale selection (Demantké et al., 2011). Instead of mainly focusing on optimal neighborhoods, further approaches extract features based on different entities such as points and regions (Xiong et al., 2011; Xu et al., 2014). Alternatively, it would be possible to calculate features at different scales and later use a training procedure to define which combination of scales allows the best separation of different classes (Brodu and Lague, 2012).

Considering the variety of features which have been proposed for classifying 3D points, it may further be expected that there are more and less suitable features among them. For compensating lack of knowledge, however, often all extracted features are included in the classification process, and a respective feature selection has only rarely been applied in 3D point cloud processing. The main idea of such a feature selection is to improve the classification accuracy while simultaneously reducing both computational effort and memory consumption (Guyon and Elisseeff, 2003; Liu et al., 2010). Respective approaches allow to assess the relevance/importance of single features, rank them according to their relevance and select a subset of the best-ranked features (Chehata et al., 2009; Mallet et al., 2011; Khoshelham and Oude Elberink, 2012; Weinmann et al., 2013).

In this paper, we use state-of-the-art approaches for classifying 3D points and focus on the interleaved issue of deriving an optimal subset of relevant, but not redundant, features extracted from individual neighborhoods with optimal size. In comparison to seminal work addressing optimal neighborhood size selection (Pauly et al., 2003; Mitra and Nguyen, 2003; Demantké et al., 2011), we directly assess the order/disorder of 3D points in the local neighborhood from the eigenvalues of the 3D structure tensor. In comparison to recent work on feature selection for 3D lidar data processing (Mallet et al., 2011; Weinmann et al., 2013), we exploit entropy-based measures for (i) determining the optimal neighborhood size for each 3D point and (ii) removing irrelevant and redundant features in order to derive an adequate feature subset. Both of these issues are crucial for the whole processing chain, and it is therefore of great importance to avoid parameters or thresholds which are explicitly selected by human interaction based on empiric or heuristic knowledge.

In summary, the main contribution of our work is a fully automatic versatile framework which is based on

- determining the optimal neighborhood size for each individual 3D point by considering the order/disorder of 3D points within a covariance ellipsoid,
- extracting optimized 3D and 2D features from the derived optimal neighborhoods in order to optimally describe the local structure for each 3D point,
- selecting a compact and robust feature subset by addressing different intrinsic properties of the given training data via multivariate filter-based feature selection (based on both feature-class and feature-feature relations) in order to remove feature redundancy, and
- improving the classification accuracy by exploiting the derived feature subsets and state-of-the-art classifiers.

Our framework is generally applicable for interpreting 3D point cloud data acquired via airborne laser scanning (ALS), terrestrial laser scanning (TLS), mobile laser scanning (MLS), range imaging by 3D cameras or 3D reconstruction from images. While the selected feature subset may vary with respect to different datasets, the beneficial impact of both optimal neighborhood size selection and feature selection remains. Further extensions of the framework by involving additional features such as color/intensity or full-waveform features can easily be taken into account.

The paper is organized as follows. In Section 2, we explain the single components of our framework in detail. Subsequently, in Section 3, we evaluate the proposed methodology on MLS data acquired within an urban environment. The derived results are discussed in Section 4. Finally, in Section 5, concluding remarks are provided, and suggestions for future work are outlined.

2 METHODOLOGY

For semantically interpreting 3D point clouds, we propose a new methodology which involves neighborhood selection with optimal neighborhood size for each individual 3D point (Section 2.1), 3D and 2D feature extraction (Section 2.2), feature subset selection via feature-class and feature-feature correlation (Section 2.3), and supervised classification of 3D point cloud data (Section 2.4). A visual representation of the whole framework and its components is provided in Figure 2.

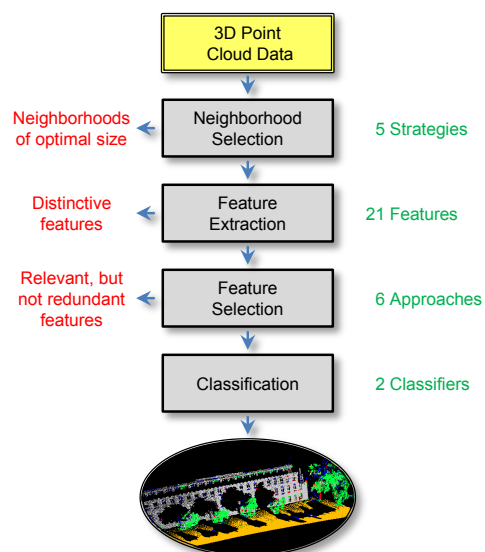


Figure 2: The proposed framework: the contributions are highlighted in red, and the quantity of attributes/approaches used for evaluation is indicated in green.

2.1 Neighborhood Selection

In general, we may face a varying point density in the captured 3D point cloud data. Since we do not want to assume a priori knowledge on the scene, we exploit the spherical neighborhood definition based on a 3D point and its k closest 3D points (Linsen and Prutzsch, 2001), which allows more flexibility with respect to the geometric size of the neighborhood. In order to avoid heuristically selecting a certain value for the parameter k , we focus on automatically estimating the optimal value for k .

Assuming a point cloud formed by a total number of N 3D points and a given value $k \in \mathbb{N}$, we may consider each individual 3D point $\mathbf{X} = (X, Y, Z)^T \in \mathbb{R}^3$ and the respective k neighbors

defining its scale. For describing the local 3D structure around \mathbf{X} , the respective 3D covariance matrix also known as 3D structure tensor $\mathbf{S} \in \mathbb{R}^{3 \times 3}$ is derived which is a symmetric positive-definite matrix. Thus, its three eigenvalues $\lambda_1, \lambda_2, \lambda_3 \in \mathbb{R}$ exist, are non-negative and correspond to an orthogonal system of eigenvectors. Since there may not necessarily be a preferred variation with respect to the eigenvectors, we consider the general case based on a structure tensor with rank 3. Hence, it follows that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$ holds for each 3D point \mathbf{X} .

From the eigenvalues of the 3D structure tensor, the surface variation C_λ (i.e. the change of curvature) with

$$C_\lambda = \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \quad (1)$$

can be estimated. For an increasing neighborhood size, the heuristic search for locations with significant increase of C_λ allows to find the critical neighborhood size and thus to select a respective value for k (Pauly et al., 2003). This procedure is motivated by the fact that occurring jumps indicate strong deviations in the normal direction. As alternative, it has been proposed to select the neighborhood size according to a consistent curvature level (Belton and Lichti, 2006).

Further investigations focus on extracting the dimensionality features of linearity L_λ , planarity P_λ and scattering S_λ according to

$$L_\lambda = \frac{\lambda_1 - \lambda_2}{\lambda_1} \quad P_\lambda = \frac{\lambda_2 - \lambda_3}{\lambda_1} \quad S_\lambda = \frac{\lambda_3}{\lambda_1} \quad (2)$$

which represent 1D, 2D and 3D features. As these features sum up to 1, they may be considered as the probabilities of a 3D point to be labeled as 1D, 2D or 3D structure (Demantké et al., 2011). Accordingly, a measure E_{dim} of unpredictability given by the Shannon entropy (Shannon, 1948) as

$$E_{\text{dim}} = -L_\lambda \ln(L_\lambda) - P_\lambda \ln(P_\lambda) - S_\lambda \ln(S_\lambda) \quad (3)$$

can be minimized across different scales k to find the optimal neighborhood size which favors one dimensionality the most. For this purpose, the radius has been taken into account, and the interval $[r_{\min}, r_{\max}]$ has been sampled in 16 scales, where the radii are not linearly increased since the radius of interest is usually closer to r_{\min} . The values r_{\min} and r_{\max} depend on various characteristics of the given data and are therefore specific for each dataset. However, the results are based on the assumption of particular shapes being present in the observed scene.

In order to avoid assumptions on the scene, we propose a more general solution to optimal neighborhood size selection. Since the eigenvalues correspond to the principal components, they span a 3D covariance ellipsoid. Consequently, we may normalize the three eigenvalues by their sum Σ_λ and consider the measure of eigenentropy E_λ given by the Shannon entropy according to

$$E_\lambda = -e_1 \ln(e_1) - e_2 \ln(e_2) - e_3 \ln(e_3) \quad (4)$$

where the e_i with $e_i = \lambda_i / \Sigma_\lambda$ for $i \in \{1, 2, 3\}$ represent the normalized eigenvalues summing up to 1. The eigenentropy thus provides a measure of the order/disorder of 3D points within the covariance ellipsoid². Hence, we propose to select the parameter k by minimizing the eigenentropy E_λ over varying values for k . For this purpose, we consider relevant statistics to start with $k_{\min} = 10$ samples which is in accordance to similar investigations (Demantké et al., 2011). As maximum, we select a

relatively high number of $k_{\max} = 100$ samples, and all integer values in $[k_{\min}, k_{\max}]$ are taken into consideration.

2.2 Feature Extraction

For feature extraction, we follow the strategy of deriving a variety of both 3D and 2D features (Weinmann et al., 2013), but we optimize their distinctiveness by taking into account the optimal neighborhood size of each individual 3D point. Based on the normalized eigenvalues e_1, e_2 and e_3 of the 3D structure tensor \mathbf{S} , we extract a feature set consisting of 8 eigenvalue-based features for each 3D point \mathbf{X} (Table 1). Additionally, we derive 6 further 3D features for characterizing the local neighborhood: absolute height Z , radius $r_{k\text{-NN}}$ of the spherical neighborhood, local point density D , verticality V which is derived from the vertical component of the normal vector, and maximum height difference $\Delta Z_{k\text{-NN}}$ as well as height variance $\sigma_{Z,k\text{-NN}}$ within the local neighborhood.

Linearity:	$L_\lambda = \frac{e_1 - e_2}{e_1}$
Planarity:	$P_\lambda = \frac{e_2 - e_3}{e_1}$
Scattering:	$S_\lambda = \frac{e_3}{e_1}$
Omnivariance:	$O_\lambda = \sqrt[3]{e_1 e_2 e_3}$
Anisotropy:	$A_\lambda = \frac{e_1 - e_3}{e_1}$
Eigenentropy:	$E_\lambda = -\sum_{i=1}^3 e_i \ln(e_i)$
Sum of eigenvalues:	$\Sigma_\lambda = e_1 + e_2 + e_3$
Change of curvature:	$C_\lambda = \frac{e_3}{e_1 + e_2 + e_3}$

Table 1: Eigenvalue-based 3D features.

Finally, we consider 7 features arising from the 2D projection of the 3D point cloud data onto a horizontally oriented plane. Four of them are directly derived: radius $r_{k\text{-NN},2D}$, local point density D_{2D} and sum $\Sigma_{\lambda,2D}$ as well as ratio $R_{\lambda,2D}$ of eigenvalues. The other three features are derived via the construction of a 2D accumulation map with discrete, quadratic bins of side length 0.25 m as number M of points, maximum height difference ΔZ and height variance σ_Z within the respective bin.

2.3 Feature Selection

The definition of adequate feature vectors remains a common and crucial issue for classification problems. Hence, the interest in feature selection techniques emerged for finding compact and robust subsets of relevant and informative features in order to gain predictive accuracy, improve computational efficiency with respect to both time and memory consumption, and retain meaningful features (Guyon and Elisseeff, 2003; Liu et al., 2010). By definition, a feature is statistically relevant if its removal from a feature set will reduce the prediction power. In general, feature selection methods can be categorized into filter-based methods, wrapper-based methods and embedded methods. As both wrapper-based and embedded feature selection methods involve a classifier, they generally yield a better performance than filter-based methods. In particular, embedded methods provide the capability of dealing with exhaustive feature sets as input and letting the classifier internally select a suitable feature subset during the training phase (Chehata et al., 2009; Tokarczyk et al., 2013).

²Note that the occurrence of eigenvalues identical to zero has to be avoided by adding an infinitesimal small value ε .

However, they face a relatively high computational effort and provide feature subsets which are only optimized with respect to the applied classifier. Hence, we focus on a filter-based method.

Due to their simplicity and efficiency, such filter-based methods are commonly applied. These methods are classifier-independent and only exploit a score function directly based on the training data. Univariate filter-based feature selection methods rely on a score function which evaluates feature-class relations and thus the relation between the values of each single feature across all observations and the respective label vector. In general, the score function may address different intrinsic properties of the given training data such as distance, information, dependency or consistency. Accordingly, a variety of possible score functions addressing a specific intrinsic property (Guyon and Elisseeff, 2003; Zhao et al., 2010) as well as a general relevance metric addressing different intrinsic properties (Weinmann et al., 2013) have been proposed. Multivariate filter-based feature selection methods rely on both feature-class and feature-feature relations in order to discriminate between relevant, irrelevant and redundant features.

Defining random variables X for the feature values and C for the classes, we can apply the general definition of the Shannon entropy $E(X)$ indicating the distribution of feature values x_a as

$$E(X) = - \sum_a P(x_a) \ln P(x_a) \quad (5)$$

and the Shannon entropy $E(C)$ indicating the distribution of (semantic) classes c_b as

$$E(C) = - \sum_b P(c_b) \ln P(c_b) \quad (6)$$

respectively. The joint Shannon entropy results in

$$E(X, C) = - \sum_{a,b} P(x_a, c_b) \ln P(x_a, c_b) \quad (7)$$

and can be used for deriving the mutual information

$$MI(X, C) = E(X) + E(C) - E(X, C) \quad (8)$$

$$= E(X) - E(X|C) \quad (9)$$

$$= E(C) - E(C|X) \quad (10)$$

$$= IG(X|C) \quad (11)$$

$$= IG(C|X) \quad (12)$$

which represents a symmetrical measure defined as *information gain* (Quinlan, 1986). Thus, the amount of information gained about C after observing X is equal to the amount of information gained about X after observing C . Following the definition, a feature X is regarded as more correlated to the classes C than a feature Y if $IG(C|X) > IG(C|Y)$. For feature selection, information gain is evaluated independently for each feature and features with a high information gain are considered as relevant. Consequently, those features with the highest values may be selected as relevant features. Information gain can also be derived via the conditional entropy, e.g. via $E(X|C)$ which quantifies the remaining uncertainty in X given that the value of the random variable C is known.

However, information gain is biased in favor of features with greater numbers of values since these appear to gain more information than others, even if they are not more informative (Hall, 1999). The bias can be compensated by considering the measure

$$SU(X, C) = 2 \frac{MI(X, C)}{E(X) + E(C)} \quad (13)$$

defined as *symmetrical uncertainty* (Press et al., 1988) with values in $[0, 1]$. Information gain and symmetrical uncertainty however are only measures for ranking features according to their relevance to the class and do not eliminate redundant features.

In order to remove redundancy, Correlation-based Feature Selection (CFS) has been proposed (Hall, 1999). Considering a subset of n features and taking the symmetrical uncertainty as correlation measure, we may define $\bar{\rho}_{XC}$ as average correlation between features and classes as well as $\bar{\rho}_{XX}$ as average correlation between different features. The relevance R of the feature subset results in

$$R(X_{1\dots n}, C) = \frac{n\bar{\rho}_{XC}}{\sqrt{n + n(n-1)\bar{\rho}_{XX}}} \quad (14)$$

which can be maximized by searching the feature subset space (Hall, 1999), i.e. by iteratively adding a feature to the feature subset (forward selection) or removing a feature from the feature subset (backward elimination) until R converges to a stable value.

For comparison only, we also consider feature selection exploiting a Fast Correlation-Based Filter (FCBF) (Yu and Liu, 2003) which involves heuristics and thus does not meet our intention of a fully generic methodology. For deciding whether features are relevant to the class or not, a typical feature ranking based on symmetrical uncertainty is conducted in order to determine the feature-class correlation. If the symmetrical uncertainty is above a certain threshold, the respective feature is considered to be relevant. For deciding whether a relevant feature is redundant or not, the symmetrical uncertainty among features is compared to the symmetrical uncertainty between features and classes in order to remove redundant features and only keep predominant features.

2.4 Classification

Based on given training data, a supervised classification of individual 3D points can be conducted by using the training data to train a classifier which afterwards should be able to generalize to new, unseen data. Introducing a formal description, the training set $\mathcal{X} = \{(\mathbf{x}_i, l_i)\}$ with $i = 1, \dots, N_{\mathcal{X}}$ consists of $N_{\mathcal{X}}$ training examples. Each training example encapsulates a feature vector $\mathbf{x}_i \in \mathbb{R}^d$ in a d -dimensional feature space and the respective class label $l_i \in \{1, \dots, N_C\}$, where N_C represents the number of classes. In contrast, the test set $\mathcal{Y} = \{\mathbf{x}_j\}$ with $j = 1, \dots, N_{\mathcal{Y}}$ only consists of $N_{\mathcal{Y}}$ feature vectors $\mathbf{x}_j \in \mathbb{R}^d$. If available, the respective class labels may be used for evaluation. For multi-class classification, we apply different classifiers. Following recent work on smooth image labeling (Schindler, 2012), we apply a classical (Gaussian) maximum-likelihood (ML) classifier as well as Random Forest (RF) classifier as representative of modern discriminative methods.

The classical ML classifier represents a simple generative model – the Gaussian Mixture Model (GMM) – which is based on the assumption that the classes can be represented by different Gaussian distributions. Hence, in the training phase, a multivariate Gaussian distribution is fitted to the given training data. For each new feature vector, the probability of belonging to the different classes is evaluated and the class with maximum probability is assigned. Since the decision boundary between any two classes in such a model represents a quadratic function (Schindler, 2012), the resulting classifier is also referred to as Quadratic Discriminant Analysis (QDA) classifier.

A Random Forest (Breiman, 2001) is an ensemble of randomly trained decision trees. In the training phase, individual trees are trained on randomly selected feature subsets of the given training

data. Thus, the trees are all randomly different from one another which results in a de-correlation between individual tree predictions and thus improved generalization and robustness (Criminisi and Shotton, 2013). For a new feature vector, each tree votes for a single class and a respective label is subsequently assigned according to the majority vote of all trees. We use a RF classifier with 100 trees and a tree depth of $\lfloor \sqrt{d} \rfloor$, where d is the dimension of the feature space.

Since we may often face an unbalanced distribution of training examples per class in the training set, which may have a detrimental effect on the training process (Criminisi and Shotton, 2013), we apply a class re-balancing which consists of resampling the training data in order to obtain a uniform distribution of randomly selected training examples per class. The alternative would be to exploit the known prior class distribution of the training set for weighting the contribution of each class.

3 EXPERIMENTAL RESULTS

We demonstrate the performance of the proposed methodology for two publicly available MLS benchmark datasets which are described in Section 3.1. The conducted experiments are outlined in Section 3.2. A detailed evaluation and a comparison of single approaches are presented in Section 3.3.

3.1 Datasets

For our experiments, we use the Oakland 3D Point Cloud Dataset³ (Munoz et al., 2009a) which is a labeled benchmark MLS dataset representing an urban environment. The dataset has been acquired with a mobile platform equipped with side looking SICK LMS laser scanners used in push-broom mode. A separation into training set \mathcal{X} , validation set \mathcal{V} and test set \mathcal{Y} is provided, and each 3D point is assigned one of the five semantic labels *wire*, *pole/trunk*, *façade*, *ground* and *vegetation*. After class re-balancing, the reduced training set encapsulates 1,000 training examples per class. The test set contains 1.3 million 3D points.

Additionally, we apply our framework on the Paris-rue-Madame database⁴ (Serna et al., 2014) acquired in the city of Paris, France. The point cloud data consists of 20 million 3D points and corresponds to a street section with a length of approximately 160 m. For data acquisition, the Mobile Laser Scanning (MLS) system L3D2 (Goulette et al., 2006) equipped with a Velodyne HDL32 was used, and annotation has been conducted in a manually assisted way. Since the annotation includes both point labels and segmented objects, the database contains 642 objects which are in turn categorized in 26 classes. We exploit the point labels of the six dominant semantic classes *façade*, *ground*, *cars*, *motorcycles*, *traffic signs* and *pedestrians*. All 3D points belonging to the remaining classes are removed since the number of samples per class is less than 0.05% of the complete dataset. For class re-balancing, we take into account that the smallest of the selected classes comprises little more than 10,000 points. In order to provide a higher ratio between training and testing samples across all

³The Oakland 3D Point Cloud Dataset is available online at http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/ (last access: 30 March 2014).

⁴Paris-rue-Madame database: MINES ParisTech 3D mobile laser scanner dataset from Madame street in Paris. ©2014 MINES Paris-Tech. MINES ParisTech created this special set of 3D MLS data for the purpose of detection-segmentation-classification research activities, but does not endorse the way they are used in this project or the conclusions put forward. The database is publicly available at <http://cmm.ensmp.fr/~serna/rueMadameDataset.html> (last access: 30 March 2014).

classes, we randomly select a training set \mathcal{X} with 1,000 training examples per class, and the remaining data is used as test set \mathcal{Y} .

3.2 Experiments

In the experiments, we first consider the impact of five different neighborhood definitions on the classification results:

- the neighborhood \mathcal{N}_{10} formed by the 10 nearest neighbors,
- the neighborhood \mathcal{N}_{50} formed by the 50 nearest neighbors,
- the neighborhood \mathcal{N}_{100} formed by the 100 nearest neighbors,
- the optimal neighborhood $\mathcal{N}_{\text{opt,dim}}$ for each individual 3D point when considering dimensionality features, and
- the optimal neighborhood $\mathcal{N}_{\text{opt},\lambda}$ for each individual 3D point when considering our proposed approach⁵.

The latter two definitions involving optimal neighborhoods are based on varying the scale parameter k between $k_{\text{min}} = 10$ and $k_{\text{max}} = 100$ with a step size of $\Delta k = 1$, and selecting the value with minimum Shannon entropy of the respective criterion. Subsequently, we focus on testing six different feature sets for each neighborhood definition:

- the whole feature set \mathcal{S}_{all} with all 21 features,
- the feature subset \mathcal{S}_{dim} covering the three dimensionality features L_λ , P_λ and S_λ ,
- the feature subset $\mathcal{S}_{\lambda,3D}$ covering the 8 eigenvalue-based 3D features,
- the feature subset \mathcal{S}_5 consisting of the five features $R_{\lambda,2D}$, V , C_λ , $\Delta Z_{k\text{-NN}}$ and $\sigma_{Z,k\text{-NN}}$ proposed in recent investigations (Weinmann et al., 2013),
- the feature subset $\mathcal{S}_{\text{CFES}}$ derived via Correlation-based Feature Selection, and
- the feature subset $\mathcal{S}_{\text{FCBF}}$ derived via the Fast Correlation-Based Filter.

The latter three feature subsets are based on either explicitly or implicitly assessing feature relevance. In case of combining feature subsets with RF-based classification, the tree depth of the Random Forest is determined as $\max\{\lfloor \sqrt{d} \rfloor, 3\}$, since at least 3 features are required for separating 5 or 6 classes. Note that the full feature set only has to be calculated and stored for the training data, whereas a smaller feature subset automatically selected during the training phase has to be calculated for the test data.

All implementation and processing was done in Matlab. In the following, the main focus is put on the impact of both optimal neighborhood size selection and feature selection on the classification results. We may expect that (i) optimal neighborhoods for individual 3D points significantly improve the classification results and (ii) feature subsets selected according to feature relevance measures provide an increase in classification accuracy.

3.3 Results and Evaluation

For evaluation, we consider five commonly used measures: (i) precision which represents a measure of exactness or quality, (ii) recall which represents a measure of completeness or quantity, (iii) F_1 -score which combines precision and recall with equal weights, (iv) overall accuracy (OA) which reflects the overall performance of the respective classifier on the test set, and (v) mean class recall (MCR) which reflects the capability of the respective classifier to detect instances of different classes. Since the results for classification may slightly vary for different runs, the mean

⁵The code is publicly available at <http://www.ipf.kit.edu/code.php>

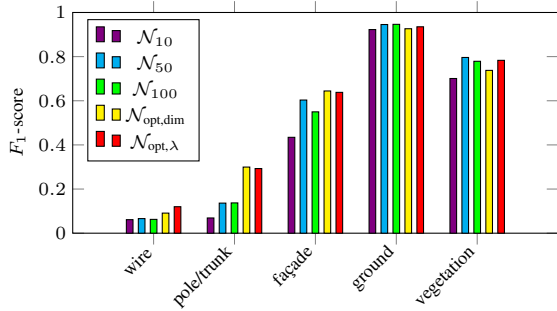


Figure 3: F_1 -scores for QDA-based classification.

values across 20 runs are used in the following in order to allow for more objective conclusions. Additionally, we consider that, for CFS and FCBF, the derived feature subsets may vary due to the random selection of training data, and hence determine them as the most often occurring feature subsets over 20 runs.

First, we test our framework on the Oakland 3D Point Cloud Dataset. Since the upper boundary $k = 100$ has been selected for reasons of computational costs, we have to take into account that it is likely to also represent 3D points which might favor a higher value. Accordingly, we consider the percentage of 3D points which are assigned neighborhoods with $k < 100$ neighbors which is 98.12% and 98.08% for $\mathcal{N}_{opt,dim}$ and $\mathcal{N}_{opt,\lambda}$. For QDA-based classification based on all 21 features, the derived recall and precision values for different neighborhood definitions are provided in Table 2 and Table 3, and the respective F_1 -scores are visualized in Figure 3. The recall and precision values when using a RF classifier are provided in Table 4 and Table 5, and the respective F_1 -scores are visualized in Figure 4. For both classifiers, it becomes visible that introducing an optimal neighborhood size for each individual 3D point has a beneficial impact on both recall and precision values, and consequently also on the F_1 -score. Exemplary results for RF-based classification using $\mathcal{N}_{opt,\lambda}$ and all 21 features are illustrated in Figure 1 and Figure 5.

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.662	0.522	0.434	0.882	0.616
\mathcal{N}_{50}	0.667	0.507	0.473	0.916	0.788
\mathcal{N}_{100}	0.606	0.417	0.472	0.916	0.767
$\mathcal{N}_{opt,dim}$	0.754	0.750	0.543	0.890	0.778
$\mathcal{N}_{opt,\lambda}$	0.791	0.765	0.519	0.906	0.829

Table 2: Recall values for QDA-based classification using all features and different neighborhood definitions.

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.032	0.037	0.614	0.967	0.812
\mathcal{N}_{50}	0.035	0.079	0.832	0.977	0.805
\mathcal{N}_{100}	0.033	0.082	0.659	0.979	0.791
$\mathcal{N}_{opt,dim}$	0.048	0.187	0.793	0.966	0.701
$\mathcal{N}_{opt,\lambda}$	0.065	0.181	0.829	0.966	0.742

Table 3: Precision values for QDA-based classification using all features and different neighborhood definitions.

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.705	0.684	0.503	0.981	0.668
\mathcal{N}_{50}	0.578	0.617	0.679	0.988	0.779
\mathcal{N}_{100}	0.513	0.579	0.631	0.987	0.724
$\mathcal{N}_{opt,dim}$	0.850	0.791	0.659	0.985	0.794
$\mathcal{N}_{opt,\lambda}$	0.862	0.798	0.672	0.985	0.809

Table 4: Recall values for RF-based classification using all features and different neighborhood definitions.

If, besides the neighborhood definitions, the different feature sets are also taken into account, we get a total number of 30 possible combinations. For each combination, the resulting overall accuracy and mean class recall value are provided in Table 6 and

Oakland	wire	pole/trunk	façade	ground	vegetation
\mathcal{N}_{10}	0.054	0.079	0.786	0.970	0.946
\mathcal{N}_{50}	0.048	0.196	0.845	0.979	0.942
\mathcal{N}_{100}	0.041	0.134	0.742	0.980	0.938
$\mathcal{N}_{opt,dim}$	0.080	0.219	0.832	0.976	0.950
$\mathcal{N}_{opt,\lambda}$	0.091	0.236	0.846	0.972	0.959

Table 5: Precision values for RF-based classification using all features and different neighborhood definitions.

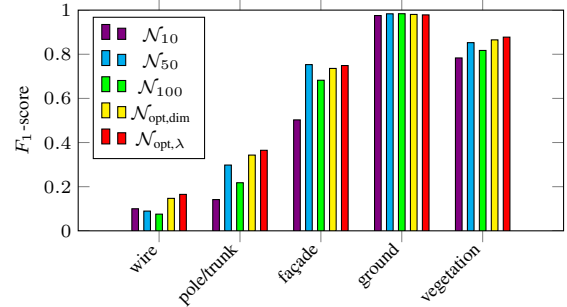


Figure 4: F_1 -scores for RF-based classification.

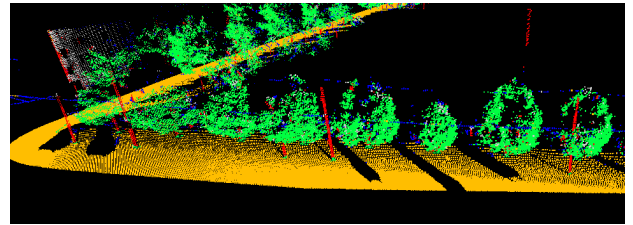


Figure 5: 3D point cloud with semantic labels assigned by the RF classifier (wire: blue, pole/trunk: red, façade: gray, ground: brown, vegetation: green).

Table 7 for QDA-based classification. The respective values for RF-based classification are provided in Table 8 and Table 9. Here, \mathcal{S}_{CFS} contains between 12 and 14 features, whereas \mathcal{S}_{FCBF} contains between 6 and 8 features. For both subsets, the respective features are distributed across all types of 3D and 2D features. The derived results clearly reveal that the feature subset \mathcal{S}_{dim} is not sufficient for obtaining adequate classification results. In contrast, using the feature subsets \mathcal{S}_5 , \mathcal{S}_{CFS} and \mathcal{S}_{FCBF} which are all based on feature relevance assessment yields classification results of better quality and, in particular when using a RF classifier, partially even a higher quality than the full feature set \mathcal{S}_{all} .

Oakland	\mathcal{S}_{all}	\mathcal{S}_{dim}	$\mathcal{S}_{\lambda,3D}$	\mathcal{S}_5	\mathcal{S}_{CFS}	\mathcal{S}_{FCBF}
\mathcal{N}_{10}	0.788	0.689	0.741	0.867	0.667	0.678
\mathcal{N}_{50}	0.850	0.771	0.822	0.927	0.725	0.762
\mathcal{N}_{100}	0.845	0.758	0.823	0.924	0.713	0.903
$\mathcal{N}_{opt,dim}$	0.837	0.371	0.798	0.910	0.715	0.687
$\mathcal{N}_{opt,\lambda}$	0.857	0.480	0.801	0.920	0.851	0.723

Table 6: Overall accuracy for QDA-based classification using different neighborhood definitions and different feature sets.

Oakland	\mathcal{S}_{all}	\mathcal{S}_{dim}	$\mathcal{S}_{\lambda,3D}$	\mathcal{S}_5	\mathcal{S}_{CFS}	\mathcal{S}_{FCBF}
\mathcal{N}_{10}	0.623	0.365	0.454	0.583	0.570	0.618
\mathcal{N}_{50}	0.670	0.509	0.588	0.673	0.633	0.699
\mathcal{N}_{100}	0.636	0.474	0.555	0.668	0.600	0.708
$\mathcal{N}_{opt,dim}$	0.743	0.440	0.561	0.666	0.694	0.703
$\mathcal{N}_{opt,\lambda}$	0.762	0.477	0.576	0.704	0.755	0.739

Table 7: Mean class recall values for QDA-based classification using different neighborhood definitions and different feature sets.

Since the RF classifier in combination with our approach for optimal neighborhood size selection ($\mathcal{N}_{opt,\lambda}$) yields high values for both overall accuracy and mean class recall, we select this combination for a test on the Paris-rue-Madame database. The obtained

Oakland	\mathcal{S}_{all}	\mathcal{S}_{dim}	$\mathcal{S}_{\lambda,3D}$	\mathcal{S}_5	\mathcal{S}_{CFs}	$\mathcal{S}_{\text{FCBF}}$
\mathcal{N}_{10}	0.875	0.579	0.742	0.887	0.873	0.857
\mathcal{N}_{50}	0.917	0.734	0.805	0.912	0.916	0.924
\mathcal{N}_{100}	0.901	0.728	0.814	0.901	0.902	0.920
$\mathcal{N}_{\text{opt,dim}}$	0.918	0.696	0.773	0.915	0.918	0.907
$\mathcal{N}_{\text{opt},\lambda}$	0.922	0.628	0.851	0.911	0.924	0.923

Table 8: Overall accuracy for RF-based classification using different neighborhood definitions and different feature sets.

Oakland	\mathcal{S}_{all}	\mathcal{S}_{dim}	$\mathcal{S}_{\lambda,3D}$	\mathcal{S}_5	\mathcal{S}_{CFs}	$\mathcal{S}_{\text{FCBF}}$
\mathcal{N}_{10}	0.708	0.483	0.598	0.686	0.699	0.702
\mathcal{N}_{50}	0.728	0.544	0.642	0.655	0.728	0.742
\mathcal{N}_{100}	0.687	0.504	0.612	0.638	0.693	0.697
$\mathcal{N}_{\text{opt,dim}}$	0.816	0.615	0.676	0.754	0.812	0.808
$\mathcal{N}_{\text{opt},\lambda}$	0.825	0.596	0.692	0.759	0.827	0.825

Table 9: Mean class recall values for RF-based classification using different neighborhood definitions and different feature sets.

recall and precision values using the feature sets \mathcal{S}_{all} and \mathcal{S}_{CFs} are provided in Table 10 as well as the resulting F_1 -scores. Based on the full feature set \mathcal{S}_{all} , the RF classifier provides an overall accuracy of 90.1% and a mean class recall of 77.6%, whereas based on the feature subset \mathcal{S}_{CFs} , a slight improvement to an overall accuracy of 90.5% and a mean class recall of 77.8% can be observed. A visualization for RF-based classification using $\mathcal{N}_{\text{opt},\lambda}$ and all 21 features is provided in Figure 6.

Paris	R	P	F_1	R	P	F_1
façade	0.957	0.962	0.960	0.958	0.964	0.961
ground	0.902	0.964	0.932	0.911	0.960	0.935
cars	0.606	0.755	0.672	0.603	0.768	0.676
motorcycles	0.639	0.123	0.206	0.657	0.136	0.225
traffic signs	0.974	0.055	0.105	0.978	0.058	0.109
pedestrians	0.575	0.019	0.036	0.559	0.020	0.038

Table 10: Recall (R), precision (P) and F_1 -score for RF-based classification involving all 21 features (left) and only the features in \mathcal{S}_{CFs} (right).

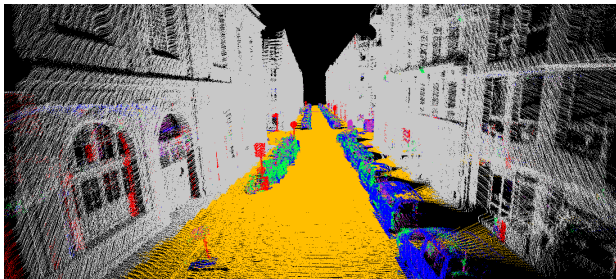


Figure 6: 3D point cloud with semantic labels assigned by the RF classifier (façade: gray, ground: brown, cars: blue, motorcycles: green, traffic signs: red, pedestrians: pink).

4 DISCUSSION

Certainly, a huge advantage of the proposed methodology is that it avoids the use of empiric or heuristic a priori knowledge on the scene with respect to neighborhood size. For the sake of generality, involving such data-dependent knowledge should not be an option and the optimal neighborhood of each individual 3D point should be considered instead. This is in accordance with the idea that the optimal neighborhood size may not be the same for different classes and furthermore depend on the respective point density. In the provided Tables 2-5, the class-specific classification results clearly reveal that the suitability of all three neighborhood definitions based on a fixed scale parameter may vary from one class to the other. Instead, the approaches based on optimal neighborhood size selection address this issue and hence provide a significant improvement in recall and precision, and thus also in the F_1 -score over all classes (Figure 3 and Figure 4).

In particular, the detailed evaluation provides a clear evidence that the proposed approach for optimal neighborhood size selection is beneficial in comparison to the other neighborhood definitions, since it often yields a significant improvement with respect to performance and behaves close to the best performance otherwise. A strong indicator for the quality of the derived results has been defined by the mean class recall, as only a high overall accuracy may not be sufficient for analyzing the derived results. For the Oakland 3D Point Cloud Dataset, for instance, we have an unbalanced test set and an overall accuracy of 70.5% can be obtained if only the instances of the class *ground* are correctly classified. This clear trend to overfitting becomes visible when considering the respective mean class recall of only 20.0%.

In comparison to other recent investigations based on a fixed scale parameter k (Weinmann et al., 2013), the recall values are significantly increased, and a slight improvement with respect to the precision values can be observed. Even in comparison to investigations involving approaches of contextual learning (Munoz et al., 2009b), our methodology yields higher precision values with approximately the same recall values over all classes.

Considering the different feature sets (Tables 6-9), it becomes visible that the feature subset \mathcal{S}_{dim} of the three dimensionality features L_λ , P_λ and S_λ is not sufficient for 3D scene interpretation. This might be due to ambiguities, since the classes *wire* and *pole/trunk* provide a linear behavior, whereas the classes *façade* and *ground* provide a planar behavior. This can only be adequately handled by considering additional features. Even when only using the feature subset $\mathcal{S}_{\lambda,3D}$ of the eigenvalue-based 3D features, the results are significantly worse than when using the full feature set \mathcal{S}_{all} . In contrast, the feature subsets derived via the three approaches for feature selection provide a performance close to the full feature set \mathcal{S}_{all} or even better. In particular, the feature subset \mathcal{S}_{CFs} derived via Correlation-based Feature Selection provides a good performance without being based on manually selected parameters such as the feature subset $\mathcal{S}_{\text{FCBF}}$ derived via the Fast Correlation-Based Filter.

5 CONCLUSIONS AND FUTURE WORK

In this paper, we have addressed the interleaved issue of optimally describing 3D structures by geometrical features and selecting the *best* features among them as input for classification. We have presented a new, fully automatic and versatile framework for semantic 3D scene interpretation. The framework involves optimal neighborhood size selection which is based on minimizing the measure of eigenentropy over varying scales in order to derive optimized features with higher distinctiveness in the subsequent step of feature extraction. Further applying the measure of entropy for feature selection, irrelevant and redundant features are recognized based on a relatively small training set and, consequently, these features do not have to be calculated and stored for the test set. In a detailed evaluation, we have demonstrated the significant and beneficial impact of optimal neighborhood size selection, and that the selection of adequate feature subsets may even further increase the quality of 3D scene interpretation.

For future work, we plan to address the step from individual 3D point classification to a spatially smooth labeling of nearby 3D points. This could be based on probabilistic relaxation or smooth labeling techniques adapted from image processing.

ACKNOWLEDGEMENTS

For collaboration, a stay abroad of the first author was supported by Karlsruhe House of Young Scientists (KHYS).

REFERENCES

- Belton, D. and Lichti, D. D., 2006. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, Part 5, pp. 44–49.
- Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5–32.
- Brodu, N. and 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, pp. 121–134.
- Carlberg, M., Gao, P., Chen, G. and Zakhor, A., 2009. Classifying urban landscape in aerial lidar using 3D shape analysis. *Proceedings of the IEEE International Conference on Image Processing*, pp. 1701–1704.
- Chehata, N., Guo, L. and Mallet, C., 2009. Airborne lidar feature selection for urban classification using random forests. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 3/W8, pp. 207–212.
- Criminisi, A. and Shotton, J., 2013. *Decision forests for computer vision and medical image analysis*. Advances in Computer Vision and Pattern Recognition, Springer, London, UK.
- Demantké, J., Mallet, C., David, N. and Vallet, B., 2011. Dimensionality based scale selection in 3D lidar point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 5/W12, pp. 97–102.
- Filin, S. and Pfeifer, N., 2005. Neighborhood systems for airborne laser data. *Photogrammetric Engineering & Remote Sensing*, 71(6), pp. 743–755.
- Goulette, F., Nashashibi, F., Abuhadrous, I., Ammoun, S. and Laurgeau, C., 2006. An integrated on-board laser range sensing system for on-the-way city and road modelling. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVI, Part 1.
- Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, pp. 1157–1182.
- Hall, M. A., 1999. Correlation-based feature subset selection for machine learning. PhD thesis, Department of Computer Science, University of Waikato, New Zealand.
- Hu, H., Munoz, D., Bagnell, J. A. and Hebert, M., 2013. Efficient 3-D scene analysis from streaming data. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2297–2304.
- Khoshelham, K. and Oude Elberink, S. J., 2012. Role of dimensionality reduction in segment-based classification of damaged building roofs in airborne laser scanning data. *Proceedings of the International Conference on Geographic Object Based Image Analysis*, pp. 372–377.
- Lafarge, F. and Mallet, C., 2012. Creating large-scale city models from 3D-point clouds: a robust approach with hybrid representation. *International Journal of Computer Vision*, 99(1), pp. 69–85.
- Lalonde, J.-F., Unnikrishnan, R., Vandapel, N. and Hebert, M., 2005. Scale selection for classification of point-sampled 3D surfaces. *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*, pp. 285–292.
- Lee, I. and Schenk, T., 2002. Perceptual organization of 3D surface points. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIV, Part 3A, pp. 193–198.
- Linsen, L. and Prautzsch, H., 2001. Natural terrain classification using three-dimensional lidar data for ground robot mobility. *Proceedings of Eurographics*, pp. 257–263.
- Liu, H., Motoda, H., Setiono, R. and Zhao, Z., 2010. Feature selection: an ever evolving frontier in data mining. *Proceedings of the Fourth International Workshop on Feature Selection in Data Mining*, pp. 4–13.
- Lodha, S. K., Fitzpatrick, D. M. and Helmbold, D. P., 2007. Aerial lidar data classification using AdaBoost. *Proceedings of the International Conference on 3-D Digital Imaging and Modeling*, pp. 435–442.
- Mallet, C., Bretar, F., Roux, M., Soergel, U. and Heipke, C., 2011. Relevance assessment of full-waveform lidar data for urban area classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), pp. S71–S84.
- Mitra, N. J. and Nguyen, A., 2003. Estimating surface normals in noisy point cloud data. *Proceedings of the Annual Symposium on Computational Geometry*, pp. 322–328.
- Munoz, D., Bagnell, J. A., Vandapel, N. and Hebert, M., 2009a. Contextual classification with functional max-margin Markov networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 975–982.
- Munoz, D., Vandapel, N. and Hebert, M., 2009b. Onboard contextual classification of 3-D point clouds with learned high-order Markov random fields. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2009–2016.
- Niemeyer, J., Rottensteiner, F. and Soergel, U., 2012. Conditional random fields for lidar point cloud classification in complex urban areas. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, pp. 263–268.
- Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp. 152–165.
- Pauly, M., Keiser, R. and Gross, M., 2003. Multi-scale feature extraction on point-sampled surfaces. *Computer Graphics Forum*, 22(3), pp. 81–89.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T., 1988. *Numerical recipes in C*. Cambridge University Press, Cambridge, UK.
- Pu, S., Rutzinger, M., Vosselman, G. and Oude Elberink, S., 2011. Recognizing basic structures from mobile laser scanning data for road inventory studies. *ISPRS Journal of Photogrammetry and Remote Sensing*, 66(6), pp. S28–S39.
- Quinlan, J. R., 1986. Induction of decision trees. *Machine Learning*, 1(1), pp. 81–106.
- Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing*, 50(11), pp. 4534–4545.
- Secord, J. and Zakhor, A., 2007. Tree detection in urban regions using aerial lidar and image data. *IEEE Geoscience and Remote Sensing Letters*, 4(2), pp. 196–200.
- Serna, A. and Marcotegui, B., 2013. Urban accessibility diagnosis from mobile laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 84, pp. 23–32.
- Serna, A., Marcotegui, B., Goulette, F. and Deschaud, J.-E., 2014. Paris-rue-Madame database: a 3D mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, pp. 819–824.
- Shannon, C. E., 1948. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3), pp. 379–423.
- Shapovalov, R., Velizhev, A. and Barinova, O., 2010. Non-associative Markov networks for 3D point cloud classification. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVIII, Part 3A, pp. 103–108.
- Shapovalov, R., Vetrov, D. and Kohli, P., 2013. Spatial inference machines. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2985–2992.
- Tokarczyk, P., Wegner, J. D., Walk, S. and Schindler, K., 2013. Beyond hand-crafted features in remote sensing. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3/W1, pp. 35–40.
- Velizhev, A., Shapovalov, R. and Schindler, K., 2012. Implicit shape models for object detection in 3D point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, pp. 179–184.
- Weinmann, M., Jutzi, B. and Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3D point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-5/W2, pp. 313–318.
- Xiong, X., Munoz, D., Bagnell, J. A. and Hebert, M., 2011. 3-D scene analysis via sequenced predictions over points and regions. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 2609–2616.
- Xu, S., Vosselman, G. and Oude Elberink, S., 2014. Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 88, pp. 1–15.
- Yu, L. and Liu, H., 2003. Feature selection for high-dimensional data: a fast correlation-based filter solution. *Proceedings of the International Conference on Machine Learning*, pp. 856–863.
- Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A. and Liu, H., 2010. Advancing feature selection research - ASU feature selection repository. Technical Report, School of Computing, Informatics, and Decision Systems Engineering, Arizona State University.