# CONTEXTUAL CLASSIFICATION OF POINT CLOUD DATA BY EXPLOITING INDIVIDUAL 3D NEIGBOURHOODS

M. Weinmann[a], A. Schmidt[b], C. Mallet[c], S. Hinz[a], F. Rottensteiner[b], B. Jutzi[a]

[a] Institute of Photogrammetry and Remote Sensing, Karlsruhe Institute of Technology (KIT)
Englerstr. 7, 76131 Karlsruhe, Germany - {martin.weinmann, stefan.hinz, boris.jutzi}@kit.edu
[b] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover
Nienburger Str. 1, 30167 Hannover, Germany - {alena.schmidt, rottensteiner}@ipi.uni-hannover.de
[c] Université Paris-Est, IGN, SRIG, MATIS
73 avenue de Paris, 94160 Saint-Mandé, France - clement.mallet@ign.fr

**Commission III, WG III/4**

**KEY WORDS:** Lidar, Laser Scanning, Point Cloud, Features, Classification, Contextual Learning, 3D Scene Analysis, Urban

**ABSTRACT:**

The fully automated analysis of 3D point clouds is of great importance in photogrammetry, remote sensing and computer vision. For reliably extracting objects such as buildings, road inventory or vegetation, many approaches rely on the results of a point cloud classification, where each 3D point is assigned a respective semantic class label. Such an assignment, in turn, typically involves statistical methods for feature extraction and machine learning. Whereas the different components in the processing workflow have extensively, but separately been investigated in recent years, the respective connection by sharing the results of crucial tasks across all components has not yet been addressed. This connection not only encapsulates the interrelated issues of neighborhood selection and feature extraction, but also the issue of how to involve spatial context in the classification step. In this paper, we present a novel and generic approach for 3D scene analysis which relies on (*i*) individually optimized 3D neighborhoods for (*ii*) the extraction of distinctive geometric features and (*iii*) the contextual classification of point cloud data. For a labeled benchmark dataset, we demonstrate the beneficial impact of involving contextual information in the classification process and that using individual 3D neighborhoods of optimal size significantly increases the quality of the results for both pointwise and contextual classification.

## 1 INTRODUCTION

The fully automated analysis of 3D point clouds has become a topic of major interest in photogrammetry, remote sensing and computer vision. Recent research addresses a variety of topics such as object detection (Pu et al., 2011; Velizhev et al., 2012; Bremer et al., 2013; Serna and Marcotegui, 2014), extraction of curbstones and road markings (Zhou and Vosselman, 2012; Guan et al., 2014), urban accessibility analysis (Serna and Marcotegui, 2013), or the creation of large-scale city models (Lafarge and Mallet, 2012). A crucial task for many of these applications is point cloud classification, which aims at assigning a semantic class label to each 3D point of a given point cloud. Due to the complexity of 3D scenes caused by the irregular sampling of 3D points, varying point density and very different types of objects, point cloud classification has also become an active field of research, e.g. (Guo et al., 2014; Niemeyer et al., 2014; Schmidt et al., 2014; Weinmann et al., 2014; Xu et al., 2014).

Most of the approaches for point cloud classification consider the different components of the classification process (i.e. neighborhood selection, feature extraction and classification) independently from each other. However, it would seem desirable to connect these components by sharing the results of crucial tasks across all of them. Such a connection would not only be relevant for the interrelated problems of neighborhood selection and feature extraction, but also for the question of how to involve spatial context in the classification task.

In this paper, we focus on the combination of (*i*) feature extraction from individual 3D neighborhoods and (*ii*) contextual classification of point cloud data. This is motivated by the fact that such a combination provides further important insights into the interrelated issues of neighborhood selection, feature extraction and contextual classification. Using features extracted from individual neighborhoods has a significantly beneficial impact on the individual classification of points (Weinmann et al., 2014). On the other hand, using contextual information might even have more influence on the classification accuracy, because it takes into account that class labels of neighboring 3D points tend to be correlated. Consequently, this paper addresses the question whether the use of features extracted from neighborhoods of individual size still improves the classification accuracy when contextual classification is applied, and whether it is beneficial to use the same neighborhood definition for contextual classification. We propose a novel and generic approach for 3D scene analysis which relies on individually optimized 3D neighborhoods for both feature extraction and contextual classification. Considering different neighborhood definitions as the basis for feature extraction, we use a Conditional Random Field (CRF) (Lafferty et al., 2001) for contextual classification and compare the respective classification results with those obtained when using a Random Forest classifier (Breiman, 2001). As the unary terms of the CRF are also based on a Random Forest classifier, we can quantify the influence of the context model on the classification results.

After reflecting related work in Section 2, we explain the different components of our methodology in Section 3. Subsequently, in Section 4, we evaluate the proposed methodology on a labeled point cloud dataset representing an urban environment and discuss the derived results. Finally, in Section 5, concluding remarks and suggestions for future work are provided.

## 2 RELATED WORK

When focusing on point cloud classification, different strategies may be involved for each component of the processing workflow.

### 2.1 Fixed vs. Individual 3D Neighborhoods

In order to describe the local 3D structure at a given 3D point, the spatial arrangement of 3D points within the local neighborhood is typically taken into consideration. The respective local neighborhood may be defined as a spherical (Lee and Schenk, 2002) or cylindrical (Filin and Pfeifer, 2005) neighborhood with fixed radius. Alternatively, the local neighborhood can be defined to consist of the $k \in \mathbb{N}$ nearest neighbors either on the basis of 3D distances (Linsen and Prautzsch, 2001) or 2D distances (Niemeyer et al., 2014). The latter definition based on the $k$ nearest neighbors offers more flexibility with respect to the absolute neighborhood size and is more adaptive to varying point density. All these neighborhood definitions, however, rely on a scale parameter (i.e. either a radius or $k$), which is commonly selected to be identical for all 3D points and determined via heuristic or empiric knowledge on the scene. As a result, the derived scale parameter is specific for each dataset.

In order to obtain a solution taking into account that the selection of a scale parameter depends on the local 3D structure as well as the local point density, an individual neighborhood size can be determined for each 3D point. In this context, most approaches rely on a neighborhood consisting of the $k$ nearest neighbors and thus focus on optimizing $k$ for each individual 3D point. This optimization may for instance be based on the 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), dimensionality-based scale selection (Demantké et al., 2011) or eigenentropy-based scale selection (Weinmann et al., 2014). In particular, the latter two approaches have proven to be suitable for point cloud data acquired via mobile laser scanning, and a significant improvement of classification results can be observed in comparison to the use of fixed 3D neighborhoods with identical scale parameter (Weinmann et al., 2014).

### 2.2 Single-Scale vs. Multi-Scale Features

Given a 3D point and its local neighborhood, geometric features may be derived from the spatial arrangement of all 3D points within the neighborhood. For this purpose, it has been proposed to sample geometric relations such as distances, angles and angular variations between 3D points within the local neighborhood (Osada et al., 2002; Rusu et al., 2008; Blomley et al., 2014). However, the individual entries of the resulting feature vectors are hardly interpretable, and consequently, other investigations focus on deriving interpretable features. Such features may for instance be obtained by calculating the 3D structure tensor from the 3D coordinates of all points within the local neighborhood (Pauly et al., 2003). The eigenvalues of the 3D structure tensor may directly be applied for characterizing specific shape primitives (Jutzi and Gross, 2009). In order to obtain more intuitive features which also indicate linear, planar or volumetric structures, a set of features derived from these eigenvalues has been presented (West et al., 2004) which is nowadays commonly applied in lidar data processing. This standard feature set may be complemented by further geometric features derived from angular statistics (Munoz et al., 2009), height and local plane characteristics (Mallet et al., 2011), height characteristics and curvature properties (Schmidt et al., 2012; Schmidt et al., 2013), or basic properties of the neighborhood and characteristics of a 2D projection (Weinmann et al., 2013; Weinmann et al., 2014). Furthermore, the combination with full-waveform and echo-based features has been proposed (Chehata et al., 2009; Mallet et al., 2011; Niemeyer et al., 2011).

When deriving features at a single scale, one has to consider that a suitable scale (in the form of either fixed or individual 3D neighborhoods) is required in order to obtain an appropriate description of the local 3D structure. As an alternative to selecting such an appropriate scale, we may also derive features at multiple scales and subsequently involve a classifier in order to define which combination of scales allows the best separation of different classes (Brodu and Lague, 2012). In this context, features may even be extracted by considering different entities such as points and regions (Xiong et al., 2011; Xu et al., 2014) or by involving a hierarchical segmentation based on voxels, blocks and pillars (Hu et al., 2013). However, multi-scale approaches result in feature spaces of higher dimension, so that it may be advisable to use appropriate feature selection schemes in order to gain predictive accuracy while at the same time reducing the extra computational burden in terms of both time and memory consumption (Guyon and Elisseeff, 2003).

### 2.3 Individual vs. Contextual Classification

Based on the derived feature vectors, classification is typically conducted in a supervised way, where the straightforward solution consists of an independent classification of each 3D point relying only on its individual feature vector. The list of respective classification methods that have been used for lidar data processing includes classical Maximum Likelihood classifiers based on 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). Such an individual point classification may be carried out very efficiently, but there is a severe drawback, namely the noisy appearance of the classification results.

In order to account for the fact that the class labels of neighboring 3D points tend to be correlated, contextual classification approaches may be applied which also involve a model of the relations between 3D points in a local neighborhood. For that purpose, statistical models of context have been increasingly used for point cloud classification, e.g. Associative and non-Associative Markov Networks (Munoz et al., 2009; Shapovalov et al., 2010), Conditional Random Fields (Lim and Suter, 2009; Schmidt et al., 2012; Niemeyer et al., 2014), Simplified Markov Random Fields (Lu and Rasmussen, 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). Some methods are based on point cloud segments, e.g. (Shapovalov et al., 2010), whereas others directly classify points, e.g. (Niemeyer et al., 2014). As segment-based methods heavily depend on the quality of the results of the segmentation algorithm, we prefer point-based techniques. Typically, statistical models for context, e.g. in a Conditional Random Field (CRF), are based on interactions between neighboring point pairs, and the considerations made about the size of a local neighborhood (Section 2.1) also apply to the selection of the set of points interacting with a given point. However, existing investigations are usually based on a radius search or on the $k$ nearest neighbors either in 2D or in 3D, involving either a fixed radius or a fixed value for $k$. In (Niemeyer et al., 2011), the impact of varying the radius of a cylindrical neighborhood for defining the set of neighbors is investigated. The results indicate a saturation effect when increasing that radius, so that the average number of involved neighbors is 7, but in each experiment the radius is fixed. In this paper, we want to investigate the effect of using individual 3D neighborhoods of optimal size for defining the edges of a CRF.

## 3 METHODOLOGY

The proposed methodology for point cloud classification consists of (*i*) neighborhood selection, (*ii*) feature extraction and (*iii*) contextual classification. Instead of treating these components separately, we focus on sharing the result of the crucial task of neighborhood selection across all components. Details are explained in the subsequent sections.

### 3.1 Estimation of Optimal Neighborhoods

We start from a point cloud consisting of $N_P$ points $\mathbf{X}_i \in \mathbb{R}^3$ with $i \in \{1, \ldots, N_P\}$. In order to obtain flexibility with respect to the absolute neighborhood size, we employ neighborhoods consisting of the $k \in \mathbb{N}$ nearest neighbors. As we intend to avoid an empirical selection of an appropriate fixed scale parameter $k$ which is identical for all points, we focus on the generic selection of individual neighborhoods described by an optimized scale parameter $k$ for each 3D point $\mathbf{X}_i$, where the optimization relies on a specific energy function. This strategy is motivated by the fact that the distinctiveness of geometric features calculated from the neighboring points is increased when involving individually optimized neighborhoods (Weinmann et al., 2014).

The energy functions used to define the optimal neighborhood size are based on the covariance matrix calculated from the 3D coordinates of a given 3D point $\mathbf{X}_i$ and its $k$ nearest neighbors. This covariance matrix is also referred to as the 3D structure tensor. Denoting the eigenvalues of the 3D structure tensor by $\lambda_{1,i}, \lambda_{2,i}, \lambda_{3,i} \in \mathbb{R}$, where $\lambda_{1,i} \geq \lambda_{2,i} \geq \lambda_{3,i} \geq 0$, two recent approaches for selecting individual neighborhoods can be applied. On the one hand, the *dimensionality features* of linearity $L_{\lambda,i}$, planarity $P_{\lambda,i}$ and scattering $S_{\lambda,i}$ with

$$L_{\lambda,i} = \frac{\lambda_{1,i} - \lambda_{2,i}}{\lambda_{1,i}} \quad P_{\lambda,i} = \frac{\lambda_{2,i} - \lambda_{3,i}}{\lambda_{1,i}} \quad S_{\lambda,i} = \frac{\lambda_{3,i}}{\lambda_{1,i}} \quad (1)$$

sum up to 1 and may be used in order to derive the Shannon entropy (Shannon, 1948) representing the energy function $E_{\text{dim},i}$ for *dimensionality-based scale selection* (Demantké et al., 2011):

$$E_{\text{dim},i} = -L_{\lambda,i}\ln(L_{\lambda,i}) - P_{\lambda,i}\ln(P_{\lambda,i}) - S_{\lambda,i}\ln(S_{\lambda,i}). \quad (2)$$

Alternatively, we may normalize the three eigenvalues by their sum $\sum_j \lambda_{j,i}$ in order to obtain the normalized eigenvalues $\epsilon_{j,i}$ with $\epsilon_{j,i} = \lambda_{j,i} / \sum_j \lambda_{j,i}$ for $j \in \{1, 2, 3\}$, summing up to 1, and we can use the Shannon entropy of these normalized eigenvalues as the basis of the energy function $E_{\lambda,i}$ for *eigenentropy-based scale selection* (Weinmann et al., 2014):

$$E_{\lambda,i} = -\epsilon_{1,i}\ln(\epsilon_{1,i}) - \epsilon_{2,i}\ln(\epsilon_{2,i}) - \epsilon_{3,i}\ln(\epsilon_{3,i}). \quad (3)$$

For each 3D point $\mathbf{X}_i$, the energy functions $E_{\text{dim},i}$ and $E_{\lambda,i}$ are calculated for varying values of $k$, and the value yielding the minimum entropy is selected to define the *optimal* neighborhood size. Note that minimizing $E_{\text{dim},i}$ corresponds to favoring dimensionality features which are as dissimilar as possible from each other, whereas minimizing $E_{\lambda,i}$ corresponds to minimizing the disorder of points within the neighborhood. Similarly to (Weinmann et al., 2014), we vary the scale parameter $k$ between $k_{\text{min}} = 10$ and $k_{\text{max}} = 100$ with $\Delta k = 1$.

### 3.2 Feature Extraction

We involve the same feature set as (Weinmann et al., 2014), which has been shown to give good results in point cloud classification. This feature set consists of both *3D features* and *2D features*.

A group of 3D features represents basic properties of the neighborhood such as absolute height $H_i$ of the center point $\mathbf{X}_i$, radius $r_{k\text{-NN},i}$ of the neighborhood, maximum difference $\Delta H_{k\text{-NN},i}$ and standard deviation $\sigma_{H,k\text{-NN},i}$ of height values within the neighborhood, local point density $D_i$, and verticality $V_i$. Further 3D features are based on the normalized eigenvalues of the 3D structure tensor and consist of linearity $L_{\lambda,i}$, planarity $P_{\lambda,i}$, scattering $S_{\lambda,i}$, omnivariance $O_{\lambda,i}$, anisotropy $A_{\lambda,i}$, eigenentropy $E_{\lambda,i}$, the sum $\Sigma_{\lambda,i}$ of eigenvalues and the change of curvature $C_{\lambda,i}$.

Particularly in urban environments, we may face a variety of man-made objects which, in turn, are characterized by almost perfectly vertical structures (e.g. building façades, walls, poles, traffic signs or curbstone edges). For this reason, we also involve features based on a 2D projection of a given 3D point $\mathbf{X}_i$ and its $k$ nearest neighbors onto a horizontal plane $\mathcal{P}$. Exploiting the projected 3D points, we may easily obtain the respective radius $r_{k\text{-NN},2\text{D},i}$ and point density $D_{2\text{D},i}$ in 2D. Furthermore, we derive the covariance matrix of the 2D coordinates of these points in the projection plane, i.e. the 2D structure tensor, whose eigenvalues provide additional features, namely their sum $\Sigma_{\lambda,2\text{D},i}$ and their ratio $R_{\lambda,2\text{D},i}$. Finally, we derive features resulting from a 2D projection of all 3D points onto $\mathcal{P}$ and a subsequent spatial binning. For that purpose, we discretize the projection plane and define a 2D accumulation map with discrete, quadratic bins with a side length of 0.25 m as proposed in (Weinmann et al., 2013). The additional features for describing a given 3D point $\mathbf{X}_i$ are represented by the number $N_{B,i}$ of points as well as the maximum difference $\Delta H_{B,i}$ and standard deviation $\sigma_{H,B,i}$ of height values within the respective bin.

All the extracted features are concatenated to a feature vector and, since the geometric features describe different quantities, a normalization $[\cdot]_n$ across all feature vectors is involved which normalizes the values of each dimension to the interval $[0, 1]$. Thus, the 3D point $\mathbf{X}_i$ is characterized by a 21-dimensional feature vector $\mathbf{f}_i$ with

$$\begin{aligned} \mathbf{f}_i \quad = \quad & [H_i, r_{k\text{-NN},i}, \Delta H_{k\text{-NN},i}, \sigma_{H,k\text{-NN},i}, D_i, V_i, L_{\lambda,i}, \\ & P_{\lambda,i}, S_{\lambda,i}, O_{\lambda,i}, A_{\lambda,i}, E_{\lambda,i}, \Sigma_{\lambda,i}, C_{\lambda,i}, r_{k\text{-NN},2\text{D},i}, \\ & D_{2\text{D},i}, \Sigma_{\lambda,2\text{D},i}, R_{\lambda,2\text{D},i}, N_{B,i}, \Delta H_{B,i}, \sigma_{H,B,i}]_n^T \quad (4) \end{aligned}$$

which is used as input for the classification of that point.

### 3.3 Classification Based on Conditional Random Fields

We use a *Conditional Random Field* (*CRF*) (Lafferty et al., 2001; Kumar and Hebert, 2006) for classification. CRFs are undirected graphical models that allow to model interactions between neighboring objects to be classified, and, thus, to model local context. The underlying graph $G(n, e)$ consists of a set of nodes $n$ and a set of edges $e$, the latter being responsible for the context model. In our case, similarly to (Niemeyer et al., 2014), the nodes $n_i \in n$ correspond to the 3D points $\mathbf{X}_i$ of the point cloud, whereas the edges $e_{ij} \in e$ connect neighboring pairs of nodes $(n_i, n_j)$. Consequently, the number of nodes in the graph is identical to the number $N_P$ of points to be classified. It is the goal of classification to assign a class label $c_i \in \{c^1, \ldots, c^L\}$ to each 3D point $\mathbf{X}_i$ (and thus to each node $n_i$ of the graph), where $L$ is the number of classes, superscripts indicate specific class labels corresponding to an object type, and subscripts indicate the class label of a given point. Due to the mutual dependencies between the class labels at neighboring points induced by the edges of the graph, the class labels of all points have to be determined simultaneously. We collect the class labels of all points in a vector $\mathbf{C} = [c_1, \ldots, c_i, \ldots, c_{N_P}]^T$. Denoting the combination of all input data by $\mathbf{x}$, we want to determine the configuration of class

labels that maximizes the posterior probability $p(\mathbf{C}|\mathbf{x})$ (Kumar and Hebert, 2006):

$$p(\mathbf{C}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \left( \prod_{i \in n} \phi(\mathbf{x}, c_i) \prod_{i \in n} \prod_{j \in N_i} \psi(\mathbf{x}, c_i, c_j) \right). \quad (5)$$

Here, $Z(\mathbf{x})$ is a normalization constant called the partition function. As it does not depend on the class labels, it can be neglected in classification. The functions $\phi(\mathbf{x}, c_i)$ are called *association potentials*; they provide local links between the data $\mathbf{x}$ and the local class labels $c_i$. The functions $\psi(\mathbf{x}, c_i, c_j)$, referred to as *interaction potentials*, are responsible for the local context model, providing the links between the class labels $(c_i, c_j)$ of the pair of nodes connected by the edge $e_{ij}$ and the data $\mathbf{x}$. $N_i$ denotes the set of neighbors of node $n_i$ that are linked to $n_i$ by an edge. Details about our definitions of the individual terms and the local neighborhood are given in the subsequent subsections.

**3.3.1 Association Potentials:** Any local discriminative classifier whose output can be interpreted in a probabilistic way can be used to define the association potentials $\phi(\mathbf{x}, c_i)$ in Equation 5. Note that the data $\mathbf{x}$ appear without an index in the argument list, which means that the association potential for node $n_i$ may depend on all the data (Kumar and Hebert, 2006). This is usually considered by defining site-wise feature vectors $\mathbf{f}_i(\mathbf{x})$, in our case one such vector per 3D point $\mathbf{X}_i$ to be classified. We use the feature vectors $\mathbf{f}_i$ defined according to Equation 4 as site-wise vectors $\mathbf{f}_i(\mathbf{x})$, whose components are functions of the data within a neighborhood of point $\mathbf{X}_i$. In our experiments, we will compare different variants of these feature vectors based on different definitions of the local neighborhood used for computing the features as defined in Section 3.1. The association potential can be defined as the posterior probability of a local discriminative classifier based on $\mathbf{f}_i(\mathbf{x})$ (Kumar and Hebert, 2006):

$$\phi(\mathbf{x}, c_i) = p\left(c_i | \mathbf{f}_i(\mathbf{x})\right). \quad (6)$$

For individual point classification, a good trade-off between classification accuracy and computational effort can be achieved by using a Random Forest classifier (Breiman, 2001). Such a Random Forest consists of a pre-defined number $N_T$ of random decision trees which are trained independently on different subsets of the given training data, where the subsets are randomly drawn with replacement. The random sampling results in randomly different decision trees and thus in diversity in terms of de-correlated hypotheses across the individual trees. In the classification, the site-wise feature vectors $\mathbf{f}_i(\mathbf{x})$ are classified by each tree. Each tree casts a vote for one of the class labels $c^l$. Usually, the majority vote over all class labels is used as the classification output, because it can be expected to result in improved generalization and robustness. In order to use the output of a Random Forest for the association potential, we define the posterior of each class label $c^l$ to be the ratio of the number $N_l$ of votes cast for that class and the number $N_T$ of involved decision trees:

$$p\left(c_i = c^l | \mathbf{f}_i(\mathbf{x})\right) = \frac{N_l}{N_T}. \quad (7)$$

The most important parameters of a Random Forest are the number $N_T$ of trees to be used for classification, the minimum allowable number $n_{\text{min}}$ of training points for a tree node to be split, the number of active variables $n_a$ to be used for the test in each tree node, and the maximum depth $d_{\text{max}}$ of each tree. For our experiments, we use the Random Forest implementation of openCV[1].

[1]The openCV documentation for Random Forests is available at http://docs.opencv.org/modules/ml/doc/random_trees.html (accessed 5 February 2015).

**3.3.2 Interaction Potentials:** Just as the association potentials, the interaction potentials can be based on the output of a discriminative classifier (Kumar and Hebert, 2006). In (Niemeyer et al., 2014), a Random Forest is used as discriminative classifier delivering a posterior $p\left(c_i, c_j | \mu_{ij}(\mathbf{x})\right)$ for the occurrence of the class labels $(c_i, c_j)$ at two neighboring points given an observed interaction feature vector $\mu_{ij}$ (the concatenated node feature vectors). Thus $\psi(\mathbf{x}, c_i, c_j) = p\left(c_i, c_j | \mu_{ij}(\mathbf{x})\right)$ is used to define the interaction potential. The derived results show that such a model delivers a better classification performance for classes having a relatively small number of instances in a point cloud. However, in order to apply such an approach, it is a prerequisite to have a sufficient number of training samples for each type of class transition; if the original number of classes is $N_l$, one would need enough training samples for $N_l \times N_l$ such transitions, which may be prohibitive. Consequently, we use a simpler model, namely a variant of the *contrast-sensitive Potts model* (Boykov and Jolly, 2001) for the interaction potentials:

$$\log\left(\psi(\mathbf{x}, c_i, c_j)\right) = \\ = \delta_{c_i c_j} \cdot w_1 \cdot \frac{N_a}{N_{k_i}} \cdot \left[ w_2 + (1 - w_2) \cdot e^{-\frac{d_{ij}(\mathbf{x})^2}{2 \cdot \sigma^2}} \right]. \quad (8)$$

In this equation, $d_{ij}\left(\mathbf{x}\right)^2 = \|\mathbf{f}_i\left(\mathbf{x}\right) - \mathbf{f}_j\left(\mathbf{x}\right)\|^2$ is the square of the Euclidean distance between the node feature vectors $\mathbf{f}_i\left(\mathbf{x}\right)$ and $\mathbf{f}_j\left(\mathbf{x}\right)$ of the two nodes connected by the edge $e_{ij}$. Furthermore, $\delta_{c_i c_j}$ represents the Kronecker delta returning 1 if the class labels $c_i$ and $c_j$ are identical and 0 otherwise. The parameter $\sigma$ is the average square distance between the feature vectors at neighboring training points, $N_a$ is the average number of edges connected to a node in the CRF and $N_{k_i}$ is the number of neighbors of node $n_i$. The weight parameter $w_1$ influences the impact of the interaction potential on the classification results. The normalization of the interaction potential by the ratio $N_a/N_{k_i}$ is required for the interaction potentials to have an equal total impact on the classification of all nodes (Wegner et al., 2011). The model in Equation 8 will result in a data-dependent smoothing of the classification results. The second weight parameter $w_2 \in [0, 1]$ describes the degree to which smoothing will depend on the data.

**3.3.3 Definition of the Neighborhood:** An important question in the application of a CRF is the definition of an appropriate neighborhood $\mathcal{N}_i$ for each node $n_i$. For images, one can for instance use the four neighbors defined on the image grid (Kumar and Hebert, 2006). For point clouds, such a simple definition is impossible. Typically, the definition of the local neighborhood is based on the $k$ nearest neighbors or on all neighbors within a fixed radius of the node $n_i$. In both cases, a cylindrical or a spherical neighborhood can be used, i.e. the search for neighbors can be carried out using a 2D or a 3D neighborhood. In case of airborne laser scanning data, it has been shown that a 2D neighborhood is to be preferred, because in an urban area building façades will only receive a relatively small number of laser points, and the height differences between neighboring points (in 2D) carry a lot of information (Niemeyer et al., 2014). The method described in this paper is designed for data acquired by laser scanners on mobile mapping devices, where one has to deal with many points on building façades, in which case a cylindrical neighborhood does not make much sense. Consequently, we use the $k$ nearest neighbors in 3D of each point to define the edges of the graph. However, selecting a single value for $k$ may not be appropriate in case of varying point density. Hence, we use the neighborhood size as defined in Section 3.1 for *spatially varying definitions* of the local neighborhood. For performance reasons, we have to apply stricter limits to the size of the local neighborhood than for the size of the local neighborhood used to extract the features.

Thus, if the neighborhood size determined according to one of the methods defined in Section 3.1 is larger than a threshold $k_{\text{max,CRF}}$, it will be set to $k_{\text{max,CRF}}$. In our experiments, we will compare several such definitions of the neighborhood size, some of them using a neighborhood with fixed scale parameter $k$. For variants with variable $k$, the average number $N_a$ of neighbors in Equation 8 will only be based on the actual number of neighbors per node (that is, after enforcing the threshold $k_{\text{max,CRF}}$).

**3.3.4 Training and Inference:** In order to determine the parameters of our classifier, we need training data, i.e. a set of 3D points with known class labels. The parameters of the two types of potentials are trained independently from each other. In case of the association potentials, this involves the training of a Random Forest classifier, where we randomly select an identical number $N_S$ of training samples per class. This is required because otherwise a class with many samples might lead to a bias towards that class in training (Chen et al., 2004). Note that for classes with a small number of training samples, this might result in a duplication of training samples. For the interaction potentials, the parameter $\sigma$ is determined as the average square distance between neighboring points in the training data based the same local neighborhood that is used for the definition of the graph in classification. The weight parameters $w_1$ and $w_2$ could be set based on a technique such as cross validation (Shotton et al., 2009). Here, they are set to values that were found empirically.

For inference, i.e. for the determination of the label configuration $\mathbf{C}$ maximizing the posterior in Equation 5 once the parameters of the potentials are known, we use Loopy Belief Propagation (Frey and MacKay, 1998), a standard optimization technique for graphs with cycles.

## 4 EXPERIMENTAL RESULTS

In the following, we present the involved dataset, describe the conducted experiments and discuss the derived results.

### 4.1 Dataset

A benchmark point cloud dataset representing an urban environment has been released with the Oakland 3D Point Cloud Dataset[2] (Munoz et al., 2009). The data have been collected in the vicinity of the CMU campus in Oakland, USA, with a mobile laser scanning system. This system captures the local 3D geometry with side looking SICK LMS laser scanners used in push-broom mode. After acquisition, the dataset has been split into a training set consisting of approximately 37k points and a test set with about 1.3M points. The reference class labels were assigned to the points in a semi-automatic annotation process. Thus, the classification task consists of assigning each 3D point a semantic label from the set {*wire* (*w*), *pole/trunk* (*p/t*), *façade* (*f*), *ground* (*g*), *vegetation* (*v*)}. The distribution of the classes in the test set is very inhomogeneous, with 70.5% and 20.2% of the data belonging to classes *g* and *v*, respectively. Class *f* constitutes 8.4% of the points, whereas the two remaining classes (*w* and *p/t*) only consist of 0.3% and 0.6% of the points, respectively.

### 4.2 Experiments

For our experiments, we use five different variants of the definition of the neighborhood for computing the features described in Section 3.2. Three variants (denoted by $\mathcal{N}_{10}$, $\mathcal{N}_{50}$ and $\mathcal{N}_{100}$) are based on fixed scale parameters (thus a fixed neighborhood) of $k = 10$, 50 and 100, respectively, for all points of the point cloud. For variant $\mathcal{N}_{\text{opt,dim}}$ the optimal neighborhood derived via dimensionality-based scale selection is used, whereas for variant $\mathcal{N}_{\text{opt},\lambda}$ the optimal neighborhood is derived via eigenentropy-based scale selection (cf. Section 3.1). For each variant of the feature vectors, two variants of the Random Forest classifier based on different settings are compared. In variant $\text{RF}_{100}$ the Random Forest consists of 100 trees with a maximum tree depth of $d_{\text{max}} = 4$ which are trained on 1,000 training samples per class ($N_S$ = 1,000), whereas in variant $\text{RF}_{200}$ we train 200 trees with a maximum tree depth of $d_{\text{max}} = 15$ on 10,000 training samples per class. In both variants, a node is only split if it is reached by at least $n_{\text{min}} = 20$ training samples, and the number of features for each test ($n_a$) is set to the square root of the number of features, following the recommendations of the openCV implementation. The first setting is a standard one, whereas the second one is expected to lead to a slightly improved performance due to the larger number of training samples and to the larger number of trees, though at the cost of a higher computational effort.

First, we apply a classification solely based on the association potentials to the dataset, i.e. on the results of the two variants of the Random Forest classifier; the respective classification variants are denoted by $\text{RF}_{100}$ and $\text{RF}_{200}$, respectively. After that, we apply the contrast-sensitive Potts model in a CRF-based classification. We use $w_2 = 0.5$, a value found empirically; in a set of experiments not reported here for lack of space, we found that changes of that parameter had very little influence on the results. The chosen value gives equal influence of the data-dependent and the data-independent terms of the interaction potential. We compare three different values of the weight $w_1$ ($w_1 = 1.0$, $w_1 = 5.0$ and $w_1 = 10.0$) to show its impact on the classification results; the respective classification variants are referred to as $\text{CRF}^1_{N_T}$, $\text{CRF}^5_{N_T}$ and $\text{CRF}^{10}_{N_T}$, respectively, where $N_T$ is either 100 or 200, depending on whether the association potential was based on $\text{RF}_{100}$ or on $\text{RF}_{200}$. The size of the neighborhood for each node of the graph is based on the one for the definition of the features, but thresholded by a parameter $k_{\text{max,CRF}}$. For variant $\mathcal{N}_{10}$, we connect each point to its 10 nearest neighbors, whereas for $\mathcal{N}_{50}$ and $\mathcal{N}_{100}$ the number of neighbors is set to $k_{\text{max,CRF}} = 15$. For the other variants, we use $k_{\text{max,CRF}} = 25$, but vary the size of the neighborhood according to the one used for the definition of the features. This results in an average number of $N_a = 21$ neighbors for $\mathcal{N}_{\text{opt,dim}}$ and $N_a = 15$ neighbors for $\mathcal{N}_{\text{opt},\lambda}$.

As a consequence of these definitions, we carry out 40 experiments. In each case, the test set is classified, and the resulting labels are compared to the reference labels on a per-point basis. We determine the confusion matrices and derive the overall accuracy (OA), completeness (cmp), correctness (cor) and quality ($q$) of the results. For most experiments, we only report OA and $q$, the latter being a compound metric indicating a good trade-off between omission and commission errors (Heipke et al., 1997).

### 4.3 Results and Discussion

The overall accuracy achieved in all experiments is summarized in Table 1, whereas the quality $q$ for the five classes is shown in Tables 2-6. Some results are visualized in Figure 1. Looking at the numbers in Table 1, one can get the impression that the classification performs reasonably well in all cases, the lowest value of overall accuracy being 85.3% ($\text{RF}_{100}$, $\mathcal{N}_{10}$). The best overall accuracy is better than that by about 10% (95.5% for $\text{CRF}^5_{200}$, $\mathcal{N}_{\text{opt},\lambda}$). However, these results are dominated by the excellent discrimination of class *g* from the others, which is expressed by a quality of 92.3% - 98.4% for that class (cf. Table 5), which,

[2]The Oakland 3D Point Cloud Dataset is publicly available at http://www.cs.cmu.edu/∼vmr/datasets/oakland_3d/cvpr09/doc/ (accessed 5 February 2015).

as mentioned above, accounts for 70.5% of all points in the test set. The quality is still reasonable for class *v*, which contains the second largest part of the data (20.2%), though the variation is much larger (61.8% - 88.7%; cf. Table 6). For the other classes, in particular for *w*, it is very low, and whereas for *p/t* and *f* it can be improved considerably by varying the neighborhood definitions and the classifier, for class *w* the best result is $q = 11.7\%$, with a variation of about 8% between variants (cf. Table 2). The main reason for the poor quality numbers of classes *w* and *p/t* is a low correctness for these classes, i.e. there are many false positives (for an example, cf. Table 7). In both cases, this is due to a relatively large number of misclassified points that actually correspond to class *f*. In case of poles/trunks, structures appearing like semicolumns in the façades are frequently misclassified as *p/t*. Misclassifications between *f* and *w* frequently occur at façades that are orthogonal to the road so that they show a more sparse point distribution than those facing the roads. In any case, we have to put the relatively high values for overall accuracy into perspective: some classes can be differentiated well, independently from the classification setup, whereas wires of power lines (*w*) cannot be differentiated using any of the methods compared here, and the main difference between the individual experiments is in the quality of the differentiation of the classes *p/t* and *f*.

Comparing the results based on a Random Forest classifier consisting of 100 trees ($\mathrm{RF}_{100}$, $\mathrm{CRF}^1_{100}$, $\mathrm{CRF}^5_{100}$, $\mathrm{CRF}^{10}_{100}$) to those based on 200 trees, it is obvious that using more trees and more training data leads to a slightly better classification performance. The increase in OA by using 200 trees is in the order of 0.2% - 3.6% for all variants (cf. Table 1). The difference in *q* is largest for the variants based on a fixed neighborhood. This is particularly the case for the class *f* for variants $\mathcal{N}_{10}$ and $\mathcal{N}_{50}$. Here, the ordering is reversed, and the variants based on $\mathrm{RF}_{100}$ achieve a considerably better performance (cf. Table 4), though at the price of other misclassifications. However, these versions are not the best-performing ones for that class, and for the variants based on a variable neighborhood the differences in *q* in Table 4 are smaller, in particular for the versions based on a CRF.

Of the variants using a fixed neighborhood, $\mathcal{N}_{50}$ performs best in nearly all indices. $\mathcal{N}_{10}$ performs considerably worse in OA and particularly in the quality of classes *p/t* and *f*. This also holds for the largest constant neighborhood, $\mathcal{N}_{100}$, though to a lesser degree. A neighborhood size of 50 points seems to give a relatively good trade-off between smoothing and allowing changes at class boundaries. If no interactions are considered ($\mathrm{RF}_{100}$ and $\mathrm{RF}_{200}$), the variants based on a variable neighborhood perform slightly worse than $\mathcal{N}_{50}$ in overall accuracy, with $\mathcal{N}_{\mathrm{opt},\lambda}$ performing slightly better than $\mathcal{N}_{\mathrm{opt,dim}}$ in quality for the "small" classes (*w*, *p/t*, *f*) if $\mathrm{RF}_{200}$ is used as the base classifier.

Involving contextual information in the classification process improves nearly all classification indices. The improvement in overall accuracy varies between about 1% and 5% (cf. Table 1). It is most pronounced for the variant having the poorest OA if no interactions are considered ($\mathrm{RF}_{100}$, $\mathcal{N}_{10}$). Apart from this single example, it is in general better for the variants having an adaptive

| | $\mathcal{N}_{10}$ | $\mathcal{N}_{50}$ | $\mathcal{N}_{100}$ | $\mathcal{N}_{\mathrm{opt,dim}}$ | $\mathcal{N}_{\mathrm{opt},\lambda}$ |
|---|---|---|---|---|---|
| $\mathrm{RF}_{100}$ | 85.3 | 91.2 | 90.0 | 90.8 | 91.0 |
| $\mathrm{CRF}^1_{100}$ | 90.0 | 92.8 | 91.2 | 94.5 | 94.2 |
| $\mathrm{CRF}^5_{100}$ | 89.2 | 93.5 | 90.8 | 94.3 | 94.5 |
| $\mathrm{CRF}^{10}_{100}$ | 90.3 | 93.2 | 90.7 | 95.1 | 94.5 |
| $\mathrm{RF}_{200}$ | 88.6 | 93.9 | 92.5 | 92.4 | 93.5 |
| $\mathrm{CRF}^1_{200}$ | 90.7 | 94.7 | 93.3 | 95.1 | 95.4 |
| $\mathrm{CRF}^5_{200}$ | 91.5 | 94.8 | 94.4 | 94.7 | 95.5 |
| $\mathrm{CRF}^{10}_{200}$ | 91.4 | 94.6 | 93.2 | 95.3 | 94.9 |

Table 1. Overall accuracy OA [%] achieved in all experiments.

| | $\mathcal{N}_{10}$ | $\mathcal{N}_{50}$ | $\mathcal{N}_{100}$ | $\mathcal{N}_{\mathrm{opt,dim}}$ | $\mathcal{N}_{\mathrm{opt},\lambda}$ |
|---|---|---|---|---|---|
| $\mathrm{RF}_{100}$ | 4.3 | 4.7 | 3.6 | 5.1 | 7.4 |
| $\mathrm{CRF}^1_{100}$ | 6.1 | 4.5 | 3.8 | 6.7 | 9.6 |
| $\mathrm{CRF}^5_{100}$ | 5.8 | 5.1 | 4.0 | 5.8 | 11.6 |
| $\mathrm{CRF}^{10}_{100}$ | 6.3 | 4.9 | 4.2 | 11.7 | 10.3 |
| $\mathrm{RF}_{200}$ | 6.7 | 8.4 | 5.5 | 7.4 | 8.1 |
| $\mathrm{CRF}^1_{200}$ | 9.2 | 10.0 | 6.3 | 10.0 | 9.9 |
| $\mathrm{CRF}^5_{200}$ | 10.2 | 9.7 | 6.8 | 8.6 | 10.5 |
| $\mathrm{CRF}^{10}_{200}$ | 9.7 | 9.7 | 6.3 | 10.2 | 10.0 |

Table 2. Quality *q* [%] for class *w* achieved in all experiments.

| | $\mathcal{N}_{10}$ | $\mathcal{N}_{50}$ | $\mathcal{N}_{100}$ | $\mathcal{N}_{\mathrm{opt,dim}}$ | $\mathcal{N}_{\mathrm{opt},\lambda}$ |
|---|---|---|---|---|---|
| $\mathrm{RF}_{100}$ | 7.6 | 24.0 | 19.2 | 31.5 | 30.0 |
| $\mathrm{CRF}^1_{100}$ | 9.5 | 33.5 | 26.4 | 55.4 | 46.6 |
| $\mathrm{CRF}^5_{100}$ | 8.0 | 30.8 | 22.0 | 42.1 | 32.3 |
| $\mathrm{CRF}^{10}_{100}$ | 9.3 | 24.9 | 19.2 | 36.1 | 40.6 |
| $\mathrm{RF}_{200}$ | 11.5 | 30.4 | 11.3 | 28.3 | 32.7 |
| $\mathrm{CRF}^1_{200}$ | 13.3 | 38.4 | 12.9 | 41.0 | 53.6 |
| $\mathrm{CRF}^5_{200}$ | 12.8 | 41.9 | 25.6 | 43.8 | 51.4 |
| $\mathrm{CRF}^{10}_{200}$ | 13.9 | 39.7 | 15.2 | 43.2 | 55.4 |

Table 3. Quality *q* [%] for class *p/t* achieved in all experiments.

| | $\mathcal{N}_{10}$ | $\mathcal{N}_{50}$ | $\mathcal{N}_{100}$ | $\mathcal{N}_{\mathrm{opt,dim}}$ | $\mathcal{N}_{\mathrm{opt},\lambda}$ |
|---|---|---|---|---|---|
| $\mathrm{RF}_{100}$ | 38.6 | 60.3 | 53.8 | 52.2 | 53.8 |
| $\mathrm{CRF}^1_{100}$ | 51.5 | 68.0 | 56.4 | 66.8 | 65.5 |
| $\mathrm{CRF}^5_{100}$ | 51.3 | 71.4 | 52.8 | 70.5 | 68.0 |
| $\mathrm{CRF}^{10}_{100}$ | 52.7 | 69.6 | 52.2 | 70.8 | 69.7 |
| $\mathrm{RF}_{200}$ | 34.1 | 63.6 | 53.6 | 56.1 | 59.7 |
| $\mathrm{CRF}^1_{200}$ | 39.7 | 65.6 | 53.3 | 67.5 | 69.4 |
| $\mathrm{CRF}^5_{200}$ | 39.7 | 67.2 | 65.9 | 67.8 | 69.5 |
| $\mathrm{CRF}^{10}_{200}$ | 41.0 | 64.7 | 53.6 | 68.3 | 69.3 |

Table 4. Quality *q* [%] for class *f* achieved in all experiments.

| | $\mathcal{N}_{10}$ | $\mathcal{N}_{50}$ | $\mathcal{N}_{100}$ | $\mathcal{N}_{\mathrm{opt,dim}}$ | $\mathcal{N}_{\mathrm{opt},\lambda}$ |
|---|---|---|---|---|---|
| $\mathrm{RF}_{100}$ | 92.3 | 97.2 | 97.1 | 97.1 | 95.6 |
| $\mathrm{CRF}^1_{100}$ | 94.1 | 97.3 | 96.9 | 98.4 | 96.5 |
| $\mathrm{CRF}^5_{100}$ | 94.2 | 98.0 | 96.8 | 97.7 | 96.6 |
| $\mathrm{CRF}^{10}_{100}$ | 94.5 | 98.1 | 96.6 | 97.2 | 96.3 |
| $\mathrm{RF}_{200}$ | 94.0 | 97.3 | 97.5 | 96.8 | 97.5 |
| $\mathrm{CRF}^1_{200}$ | 94.5 | 97.2 | 96.8 | 98.0 | 98.0 |
| $\mathrm{CRF}^5_{200}$ | 95.3 | 97.4 | 97.8 | 97.5 | 98.1 |
| $\mathrm{CRF}^{10}_{200}$ | 95.7 | 97.2 | 98.3 | 98.0 | 97.3 |

Table 5. Quality *q* [%] for class *g* achieved in all experiments.

| | $\mathcal{N}_{10}$ | $\mathcal{N}_{50}$ | $\mathcal{N}_{100}$ | $\mathcal{N}_{\mathrm{opt,dim}}$ | $\mathcal{N}_{\mathrm{opt},\lambda}$ |
|---|---|---|---|---|---|
| $\mathrm{RF}_{100}$ | 61.8 | 69.9 | 66.7 | 73.4 | 73.6 |
| $\mathrm{CRF}^1_{100}$ | 77.1 | 77.3 | 71.5 | 86.2 | 86.0 |
| $\mathrm{CRF}^5_{100}$ | 73.9 | 78.3 | 70.7 | 87.7 | 87.5 |
| $\mathrm{CRF}^{10}_{100}$ | 78.0 | 77.2 | 71.7 | 87.9 | 87.1 |
| $\mathrm{RF}_{200}$ | 72.0 | 79.1 | 77.5 | 79.1 | 82.4 |
| $\mathrm{CRF}^1_{200}$ | 79.5 | 83.0 | 82.3 | 87.8 | 88.6 |
| $\mathrm{CRF}^5_{200}$ | 82.8 | 82.7 | 80.7 | 87.7 | 88.7 |
| $\mathrm{CRF}^{10}_{200}$ | 80.1 | 82.2 | 77.7 | 88.5 | 86.7 |

Table 6. Quality *q* [%] for class *v* achieved in all experiments.

| $R\backslash C$ | *w* | *p/t* | *f* | *g* | *v* | cmp |
|---|---|---|---|---|---|---|
| *w* | 0.25 | 0.01 | 0.00 | 0.00 | 0.03 | 88.8 |
| *p/t* | 0.01 | 0.47 | 0.01 | 0.01 | 0.09 | 78.8 |
| *f* | 1.32 | 0.21 | 6.07 | 0.00 | 0.78 | 72.4 |
| *g* | 0.45 | 0.02 | 0.03 | 69.26 | 0.78 | 98.2 |
| *v* | 0.35 | 0.08 | 0.31 | 0.05 | 19.39 | 96.1 |
| cor | 10.7 | 59.7 | 94.6 | 99.9 | 92.0 | |

Table 7. Confusion matrix, completeness (cmp) and correctness (cor) for the variant $\mathrm{CRF}^5_{200}$ using neighborhood $\mathcal{N}_{\mathrm{opt},\lambda}$. The numbers in the confusion matrix are the respective percentage of the whole test set. $R\backslash C$: Reference (rows) vs. classification results (columns).

neighborhood than for $\mathcal{N}_{50}$, in the order of 2% for the first and of 1% for the latter if 200 trees are used for the association potential. Consequently, the variant $\mathcal{N}_{\mathrm{opt},\lambda}$ performs better than $\mathcal{N}_{50}$ in all cases, the margin being in the order of 1%. If $\mathrm{RF}_{100}$ is used for the association potential, this also holds for $\mathcal{N}_{\mathrm{opt,dim}}$, whereas in case 200 trees are used $\mathcal{N}_{\mathrm{opt,dim}}$ performs similar to $\mathcal{N}_{50}$. Again,

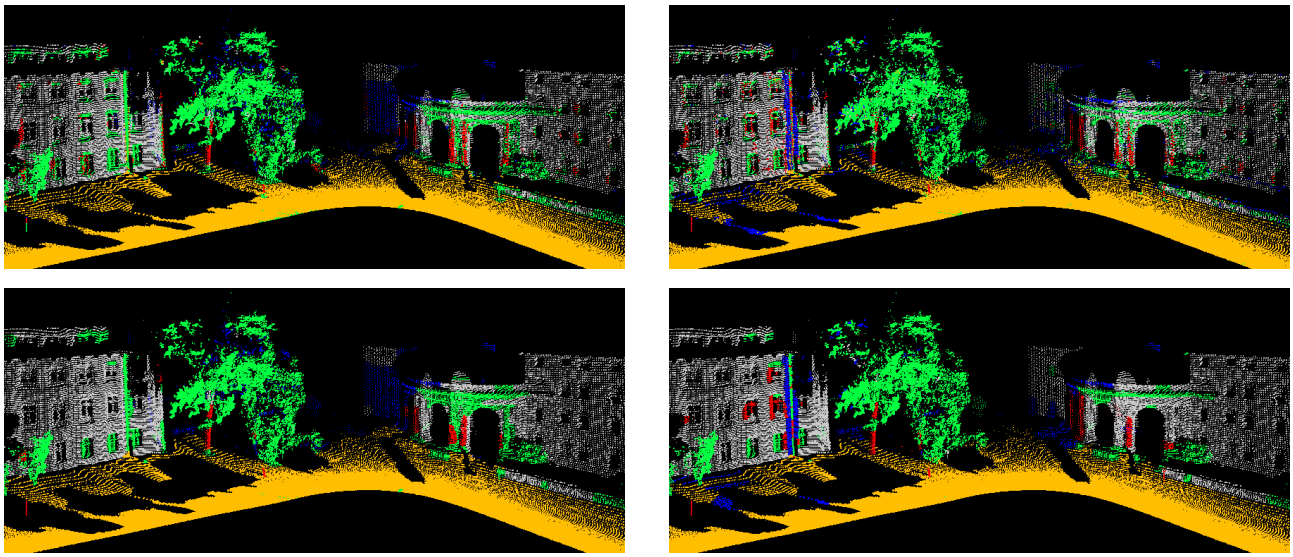

Figure 1. Classified 3D point clouds for the neighborhoods $\{\mathcal{N}_{50}, \mathcal{N}_{\text{opt},\lambda}\}$ (left and right column) and the classifiers $\{\text{RF}_{200}, \text{CRF}^5_{200}\}$ (top and bottom row) when using a standard color encoding (*wire*: blue; *pole/trunk*: red; *façade*: gray; *ground*: brown; *vegetation*: green). Note the noisy appearance of the results for individual point classification (top row).

the differences in quality for the classes *w*, *p/t* and *f* show higher variations. It becomes obvious that if the better base classifier ($\text{RF}_{200}$) is used, these classes are differentiated best by using an adaptive neighborhood as in variant $\mathcal{N}_{\text{opt},\lambda}$, in case of class *p/t* by a large margin. The weight of the interaction potential does have an impact on the results, but at least in those cases where 200 trees are used for the association potentials, the effect of changing the weight in the range tested here is relatively low compared to the impact of using the interactions in the first place. The value $w_1 = 5.0$ seems to be a good trade-off in this application.

One can see from our results that the main impact of using interactions in classification consists of a considerable improvement in the classification performance of classes that are not dominant in the data, which is consistent with the findings in (Niemeyer et al., 2014) for airborne laser scanning data. In the case of mobile laser scanning data, it might in fact be those classes one is mainly interested in. The most dominant class *g* can easily be distinguished from the remaining data by simply considering height, and the respective completeness and correctness numbers do not vary much. In contrast, *p/t* might for instance be a class of major interest for mapping urban infrastructure. When using a fixed neighborhood $\mathcal{N}_{50}$ and a Random Forest without interactions (variant $\text{RF}_{200}$), the completeness and the correctness of the results are $52.5\%$ and $42.0\%$, respectively, resulting in a quality of $30.4\%$ (Table 3). Nearly half of the points on poles or trunks are not correctly detected, and more than half of the points classified as *p/t* are in fact not situated on poles or trunks. Using the neighborhood $\mathcal{N}_{\text{opt},\lambda}$ and a CRF ($\text{CRF}^5_{200}$), these numbers are increased to a completeness of $78.8\%$ and a correctness of $59.7\%$ (cf. Table 7), which results in a quality of $51.4\%$ and certainly provides a better starting point for subsequent processes.

## 5 CONCLUSIONS

In this paper, we have presented a generic approach for automated 3D scene analysis. The novelty of this approach addresses the interrelated issues of (*i*) neighborhood selection, (*ii*) feature extraction and (*iii*) contextual classification, and it consists of using individual 3D neighborhoods of optimal size for the subsequent steps of feature extraction and contextual classification. The results derived on a standard benchmark dataset clearly indicate the beneficial impact of involving contextual information in the classification process and that using individual 3D neighborhoods of optimal size significantly increases the quality of the results for both pointwise and contextual classification.

For future work, we want to carry out deeper investigations concerning the influence of the amount of training data as well as the influence of the number of different classes on the classification results for different datasets. Moreover, we intend to exploit the results of contextual point cloud classification for extracting single objects in a 3D scene such as trees, cars or traffic signs.

## 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-5, pp. 44–49.

Blomley, R., Weinmann, M., Leitloff, J. and 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*, Vol. II-3, pp. 9–16.

Boykov, Y. Y. and Jolly, M.-P., 2001. Interactive graph cuts for optimal boundary and region segmentation of objects in N-D images. *Proceedings of the IEEE International Conference on Computer Vision*, Vol. 1, pp. 105–112.

Breiman, L., 2001. Random forests. *Machine Learning*, 45(1), pp. 5–32.

Bremer, M., Wichmann, V. and Rutzinger, M., 2013. Eigenvalue and graph-based object extraction from mobile laser scanning point clouds. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-5/W2, pp. 55–60.

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-3/W8, pp. 207–212.

Chen, C., Liaw, A. and Breiman, L., 2004. *Using random forest to learn imbalanced data*. Technical Report, University of California, Berkeley, USA.

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-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.

Frey, B. J. and MacKay, D. J. C., 1998. A revolution: belief propagation in graphs with cycles. *Proceedings of the Neural Information Processing Systems Conference*, pp. 479–485.

Guan, H., Li, J., Yu, Y., Wang, C., Chapman, M. and Yang, B., 2014. Using mobile laser scanning data for automated extraction of road markings. *ISPRS Journal of Photogrammetry and Remote Sensing*, 87, pp. 93–107.

Guo, B., Huang, X., Zhang, F. and Sohn, G., 2014. Classification of airborne laser scanning data using JointBoost. *ISPRS Journal of Photogrammetry and Remote Sensing*, 92, pp. 124–136.

Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, pp. 1157–1182.

Heipke, C., Mayer, H., Wiedemann, C. and Jamet, O., 1997. Evaluation of automatic road extraction. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXII/3-4W2, pp. 151–160.

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.

Jutzi, B. and 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*, Vol. XXXVIII-1-4-7/W5.

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.

Kumar, S. and Hebert, M., 2006. Discriminative random fields. *International Journal of Computer Vision*, 68(2), pp. 179–201.

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.

Lafferty, J. D., McCallum, A. and Pereira, F. C. N., 2001. Conditional random fields: probabilistic models for segmenting and labeling sequence data. *Proceedings of the International Conference on Machine Learning*, pp. 282–289.

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-3A, pp. 193–198.

Lim, E. H. and Suter, D., 2009. 3D terrestrial lidar classifications with super-voxels and multi-scale conditional random fields. *Computer-Aided Design*, 41(10), pp. 701–710.

Linsen, L. and Prautzsch, H., 2001. Local versus global triangulations. *Proceedings of Eurographics*, pp. 257–263.

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.

Lu, Y. and Rasmussen, C., 2012. Simplified Markov random fields for efficient semantic labeling of 3D point clouds. *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2690–2697.

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., 2009. Contextual classification with functional max-margin Markov networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 975–982.

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.

Niemeyer, J., Wegner, J. D., Mallet, C., Rottensteiner, F. and Soergel, U., 2011. Conditional random fields for urban scene classification with full waveform lidar data. *Photogrammetric Image Analysis*, LNCS 6952, Springer, Heidelberg, Germany, pp. 233–244.

Osada, R., Funkhouser, T., Chazelle, B. and Dobkin, D., 2002. Shape distributions. *ACM Transactions on Graphics*, 21(4), pp. 807–832.

Pauly, M., Keiser, R. and Gross, M., 2003. Multi-scale feature extraction on point-sampled surfaces. *Computer Graphics Forum*, 22(3), pp. 81–89.

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.

Rusu, R. B., Marton, Z. C., Blodow, N. and Beetz, M., 2008. Persistent point feature histograms for 3D point clouds. *Proceedings of the International Conference on Intelligent Autonomous Systems*, pp. 119–128.

Schmidt, A., Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of full waveform lidar data in the Wadden Sea. *IEEE Geoscience and Remote Sensing Letters*, 11(9), pp. 1614–1618.

Schmidt, A., Rottensteiner, F. and Soergel, U., 2012. Classification of airborne laser scanning data in Wadden Sea areas using conditional random fields. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX-B3, pp. 161–166.

Schmidt, A., Rottensteiner, F. and Soergel, U., 2013. Monitoring concepts for coastal areas using lidar data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XL-1/W1, pp. 311–316.

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. and Marcotegui, B., 2014. Detection, segmentation and classification of 3D urban objects using mathematical morphology and supervised learning. *ISPRS Journal of Photogrammetry and Remote Sensing*, 93, pp. 243–255.

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-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.

Shotton, J., Winn, J., Rother, C. and Criminisi, A., 2009. TextonBoost for image understanding: multi-class object recognition and segmentation by jointly modeling texture, layout, and context. *International Journal of Computer Vision*, 81(1), pp. 2–23.

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.

Wegner, J. D., Soergel, U. and Rosenhahn, B., 2011. Segment-based building detection with conditional random fields. *Proceedings of the Joint Urban Remote Sensing Event*, pp. 205–208.

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.

Weinmann, M., Jutzi, B. and Mallet, C., 2014. Semantic 3D scene interpretation: a framework combining optimal neighborhood size selection with relevant features. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. II-3, pp. 181–188.

West, K. F., Webb, B. N., Lersch, J. R., Pothier, S., Triscari, J. M. and Iverson, A. E., 2004. Context-driven automated target detection in 3-D data. *Proceedings of SPIE*, Vol. 5426, pp. 133–143.

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.

Zhou, L. and Vosselman, G., 2012. Mapping curbstones in airborne and mobile laser scanning data. *International Journal of Applied Earth Observation and Geoinformation*, 18(1), pp. 293–304.