

## FEATURE RELEVANCE ASSESSMENT FOR THE SEMANTIC INTERPRETATION OF 3D POINT CLOUD DATA

Martin Weinmann<sup>a</sup>, Boris Jutzi<sup>a</sup>, Clément Mallet<sup>b</sup>

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

<sup>b</sup> IGN/SR MATIS, 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, Features, Feature Relevance, Classification, Interpretation, Semantization

### ABSTRACT:

The automatic analysis of large 3D point clouds represents a crucial task in photogrammetry, remote sensing and computer vision. In this paper, we propose a new methodology for the semantic interpretation of such point clouds which involves feature relevance assessment in order to reduce both processing time and memory consumption. Given a standard benchmark dataset with 1.3 million 3D points, we first extract a set of 21 geometric 3D and 2D features. Subsequently, we apply a classifier-independent ranking procedure which involves a general relevance metric in order to derive compact and robust subsets of versatile features which are generally applicable for a large variety of subsequent tasks. This metric is based on 7 different feature selection strategies and thus addresses different intrinsic properties of the given data. For the example of semantically interpreting 3D point cloud data, we demonstrate the great potential of smaller subsets consisting of only the most relevant features with 4 different state-of-the-art classifiers. The results reveal that, instead of including as many features as possible in order to compensate for lack of knowledge, a crucial task such as scene interpretation can be carried out with only few versatile features and even improved accuracy.

### 1 INTRODUCTION

The automatic interpretation of large point clouds with irregularly distributed 3D points is a task of major interest in photogrammetry, remote sensing and computer vision. Aiming at uniquely labeling each 3D scene point with a semantic label (e.g. *ground*, *building* or *vegetation*), this task should be suited for very general scene structures, different acquisition techniques, and different types of 3D point cloud data. Airborne Laser Scanning (ALS) for instance is used for capturing large-scale 3D environments with almost homogeneous point density. However, the local point density still remains relatively low and reaches up to only about 50 pts/m<sup>2</sup>. In contrast, Terrestrial Laser Scanning (TLS) and Mobile Laser Scanning (MLS) are applied for capturing dense and accurate 3D information representing local object surfaces, but the density of the measured 3D points depends on their distance to the scanning unit. As a consequence, an appropriate interpretation of the captured data has to face certain challenges arising from either low or varying point density.

We address the interpretation of large and densely sampled environments, i.e. the analysis of TLS and MLS point clouds, and focus on a detailed description of 3D structures. Hence, the main goal is to assign a semantic label to each 3D point in the scene. For this purpose, it has to be taken into account that 3D point clouds contain valuable geometric information which is implicitly represented by the spatial arrangement of single 3D points. In order to adequately describe this implicit information, the local neighborhood around each 3D point is typically recovered. Subsequently, geometric 3D features are extracted from the spatial arrangement of all 3D points within the neighborhood. Depending on the device used for data acquisition, additional features may be available such as intensity when using terrestrial laser scanners or range cameras. In case of full-waveform lidar data, further features can be derived from the returned echoes and the shape of the waveform.

In addition to different types of features which may be correlated, redundant or even irrelevant, the scheme used for classification is of great importance. Standard approaches for 3D point classification focus on classifying each 3D point individually. For this purpose, both unsupervised and supervised classification schemes have been proposed. Unsupervised approaches have for instance been proposed with a method based on hierarchical *K*-means clustering of the 3D point cloud (Chehata et al., 2008) or a method exploiting the analytical consideration of eigenvalues in order to describe object structures (Jutzi and Gross, 2009). For supervised classification, schemes based on Gaussian Mixture Models (Lalonde et al., 2006), Support Vector Machines (Lodha et al., 2006; Mallet et al., 2011), AdaBoost (Lodha et al., 2007) or Random Forests (Chehata et al., 2009) have been proposed. In contrast to these standard approaches, other techniques incorporate contextual information into the classification process, e.g. by using graphical models such as Associative Markov Networks (Munoz et al., 2008; Munoz et al., 2009a; Munoz et al., 2009b), non-Associative Markov Networks (Shapovalov et al., 2010; Shapovalov and Velizhev, 2011), Conditional Random Fields (Niemeyer et al., 2012) or Simplified Markov Random Fields (Lu and Rasmussen, 2012). Further approaches have been proposed with multi-stage inference procedures exploiting point cloud statistics and learning relational information over fine and coarse scales (Xiong et al., 2011) as well as classification trees resulting in sequences of rule-based classifications (Xu et al., 2012). Very recently, spatial inference machines have been presented which are based on modeling mid-range and long-range dependencies inherent in the data and thus account for spatial semantic context (Shapovalov et al., 2013).

For modeling context with a graphical model, the contextual relationships among 3D points in a local neighborhood have to be learned from a training set. Hence, exact inference is intractable and therefore only approximate inference techniques can be applied. Investigations show that the performance of recent approaches (Munoz et al., 2008; Munoz et al., 2009b) can signif-

icantly be increased by only introducing improved learning techniques for modeling interactions between single elements (Munoz et al., 2009a).

Instead of using complex models and finding the optimal training technique for learning contextual relationships, we focus on the use of good and versatile features which provide a higher interpretability compared to numerous interactions between a large number of elements. Since the consideration of individual points does not include context, these features should be computed with a certain spatial support, i.e. within a local 3D neighborhood. Consequently, in this paper, we address two important issues: 1) which features can be calculated for each 3D point of a point cloud and 2) how meaningful are these features for further tasks such as 3D scene analysis? Aiming to reach general applicability, we only exploit geometric features for 3D scene analysis and neglect echo-based features as well as full-waveform features which are also included for similar approaches (Chehata et al., 2009; Mallet et al., 2011), but not always available. Furthermore, we explicitly want to avoid a classifier-dependent feature selection scheme which has for instance been proposed for segment-based classification (Khoshelham and Oude Elberink, 2012). The main contribution of this paper is a new methodology which

- exploits a variety of geometric 3D and 2D features for characterizing the local neighborhood of 3D points,
- analyzes the relevance of these features according to a general relevance metric addressing different intrinsic properties of the given data,
- provides compact and robust subsets of versatile features with general applicability, and
- allows for reliably performing subsequent tasks of a standard point cloud processing pipeline such as the semantic interpretation of 3D point cloud data.

The paper is organized as follows. We first introduce the proposed methodology in Section 2 and provide a detailed description of each step in the processing chain. Subsequently, in Section 3, we evaluate this methodology and demonstrate its performance on a standard benchmark dataset for 3D point cloud classification. In Section 4, we discuss the results of our approach and compare them to state-of-the-art methods. Finally, conclusions and suggestions for future work are provided in Section 5.

## 2 METHODOLOGY

The proposed methodology exploits spatial 3D information represented as 3D point clouds. Hence, special geometric properties are of importance for adequately analyzing the 3D scene structure. These are derived by considering the local neighborhood of each 3D point (Section 2.1) and extracting 3D features encapsulating the geometric relations between the respective 3D points (Section 2.2). As façades and poles are usually oriented exactly in vertical direction, all 3D points can be projected onto a horizontally oriented plane, and further properties can be described with 2D features (Section 2.3). All the information derived from the local neighborhood and thus the full set of features is typically used for classifying a large number of unorganized 3D points with respect to predefined semantic labels (Section 2.4). This however remains challenging with respect to both processing time and memory consumption. Hence, the selection of a compact and robust subset of versatile features is desirable (Section 2.5). The selected feature subset should be generally applicable for a variety of subsequent tasks such as 3D scene analysis, point cloud registration, and autonomous navigation. For this reason, it should not depend on the use of a learning machine and not only address a single characteristic, but rather involve various criteria.

### 2.1 Neighborhood Selection

Once spatial 3D information about the local environment has been collected in form of 3D point clouds, the next step towards scene analysis typically consists of recovering the local 3D neighborhood  $\mathcal{V}$  of each 3D point  $\mathbf{X} = (X, Y, Z)$ . This can for instance be done by applying a spherical neighborhood definition where all points within a sphere of fixed radius  $r_s$  form the local neighborhood (Lee and Schenk, 2002). Alternatively, a cylindrical neighborhood definition can be applied which takes into account all the 3D points whose 2D projection onto the ground plane is within a circular area of fixed size  $r_c$  (Filin and Pfeifer, 2005). A further definition is based on selecting the  $k$  closest 3D points which does not include a fixed spatial neighborhood size (Linsen and Prutzsch, 2001). Furthermore, an approach for selecting the optimal neighborhood size based on dimensionality features has been proposed (Demantké et al., 2011).

### 2.2 3D Feature Extraction

Based on the spatial information of all 3D points within the local neighborhood  $\mathcal{V}$ , invariant moments representing geometric properties can be calculated for each 3D point (Maas and Vosselman, 1999) as well as the respective 3D covariance matrix known as 3D structure tensor  $\mathbf{S}$  (Jutzi and Gross, 2009). The eigenvalues  $\lambda_1, \lambda_2$  and  $\lambda_3$  of the structure tensor  $\mathbf{S}$  with  $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq 0$  can directly be used to describe the local 3D structure or, alternatively, further measures based on these eigenvalues can be derived which encapsulate special geometric properties (West et al., 2004; Toshev et al., 2010; Mallet et al., 2011). For describing the local dimensionality, the measures of linearity  $L_\lambda$ , planarity  $P_\lambda$  and scatter (i.e. sphericity)  $S_\lambda$  provide information about the presence of a linear 1D structure, a planar 2D structure or a volumetric 3D structure. Further measures are provided by omnivariance  $O_\lambda$ , anisotropy  $A_\lambda$ , eigenentropy  $E_\lambda$  and the sum of eigenvalues denoted as  $\Sigma_\lambda$ . Exploiting the change of curvature denoted as  $C_\lambda$  has also been proposed (Rusu, 2009). The formal definitions of these measures are provided with

$$\begin{aligned} L_\lambda &= \frac{\lambda_1 - \lambda_2}{\lambda_1} & P_\lambda &= \frac{\lambda_2 - \lambda_3}{\lambda_1} \\ S_\lambda &= \frac{\lambda_3}{\lambda_1} & O_\lambda &= \sqrt[3]{\lambda_1 \lambda_2 \lambda_3} \\ A_\lambda &= \frac{\lambda_1 - \lambda_3}{\lambda_1} & E_\lambda &= -\sum_{i=1}^3 \lambda_i \ln(\lambda_i) \\ \Sigma_\lambda &= \lambda_1 + \lambda_2 + \lambda_3 & C_\lambda &= \frac{\lambda_3}{\lambda_1 + \lambda_2 + \lambda_3} \end{aligned} \quad (1)$$

and show a characteristic behavior depending on the local 3D structure. In addition to features derived from the 3D structure tensor  $\mathbf{S}$ , there are further measures which are important, namely the verticality  $V$  (Demantké et al., 2012) and the local point density  $D$  which are defined as

$$V = 1 - n_z \quad D = \frac{k+1}{\frac{4}{3} \pi r_{k\text{-NN}}^3} \quad (2)$$

where  $n_z$  is the third component of the normal vector  $\mathbf{n}$ , and  $r_{k\text{-NN}}$  represents the radius of the spherical neighborhood defined by a 3D point and its  $k$  closest neighbors. Here, the definition of the local point density has been adapted from the respective definition in 2D (Lari and Habib, 2012).

### 2.3 2D Feature Extraction

As façades and poles are usually oriented in vertical direction, further features can be derived by projecting all 3D points onto a

horizontally oriented plane  $\mathcal{P}$ . In particular, the local point density  $D_{2D}$  in the new 2D representation can be derived based on the radius  $r_{k\text{-NN},2D}$  of the circular neighborhood defined by a 2D point and its  $k$  neighbors (Lari and Habib, 2012). Additionally, the eigenvalues  $\lambda_{1,2D}$  and  $\lambda_{2,2D}$  of the structure tensor  $\mathbf{S}_{2D}$  in 2D as well as their ratio

$$R_{\lambda,2D} = \frac{\lambda_{2,2D}}{\lambda_{1,2D}} \quad (3)$$

can be calculated which indicate meaningful 2D structures in a plane. The sum  $\sum \lambda_{\lambda,2D}$  of these eigenvalues is also used as feature. Further introducing a binning based on a discrete rectangular raster (e.g. with quadratic bins of side length 0.25 m), a quantization of occurrences at certain locations  $(X, Y)$  on the plane  $\mathcal{P}$  yields an accumulation map  $\mathcal{M}$  (Monnier et al., 2012). Since each 3D point is mapped into a bin, the entry  $M(X, Y)$  of the accumulation map  $\mathcal{M}$  reveals how many 3D points voted for the same bin. High values  $M(X, Y)$  thus indicate the presence of a local vertical structure at  $(X, Y)$ . Such a binning allows for deriving the maximum height difference  $\Delta Z$  of the 3D points mapped into each bin and the standard deviation  $\sigma_Z$  of all height values  $Z$  in a similar way to approaches based on height features derived from cylindrical neighborhoods (Mallet et al., 2011).

#### 2.4 Classification of 3D Points

Once features have been extracted, these can be used for classifying 3D points with respect to certain data characteristics. For this purpose, the whole point set is divided into a training set and a test set. The training set  $\mathcal{Y} = \{s_i\}$  consists of  $n$  single training samples  $s_i$  with  $i = 1, \dots, n$ . Each training sample  $s_i = (\mathbf{x}_i, l_i)$  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. The test set  $\mathcal{Y}^* = \{s_i^*\}$  consists of  $m$  single training samples  $s_i^*$  which encapsulate new feature vectors  $\mathbf{x}_i^* \in \mathbb{R}^d$  in the  $d$ -dimensional feature space and, optionally, for later being able to evaluate the classification results, the respective class labels  $l_i^* \in \{1, \dots, n_C\}$ .

A very simple method for classification is offered by a Nearest Neighbor (NN) classifier which assigns each feature vector the class label of the closest training example. For this purpose, a distance function has to be defined which might be based on the Euclidean distance, a general Minkowski metric or other distance metrics. Additional information is involved when applying a  $k$  Nearest Neighbor ( $k$ -NN) classifier (Cover and Hart, 1967) which, for each feature vector, selects the  $k$  nearest samples in the training data and classifies according to the majority vote of their class labels. Hence, a  $k$ -NN classifier simply performs kernel density estimation with a uniform kernel and variable bandwidth. Both methods do neither require parameter estimation nor the assumption of a certain model.

In contrast, a Naive Bayesian (NB) classifier (John and Langley, 1995) is a probabilistic classifier which is based on Bayes' theorem and the naive assumption of all features being conditionally independent. For this reason, independent random variables  $X_j$  with  $j = 1, \dots, d$  and  $L$  for the  $d$  corresponding attributes  $x_j$  of a feature vector  $\mathbf{x}$  and the class label  $l$  are introduced. In the classification process, a new feature vector of a test set  $\mathcal{Y}^*$  is assigned the most likely class label according to

$$\hat{l} = \arg \max_{l \in \{1, \dots, n_C\}} P(L = l) \prod_{j=1}^d P(X_j = x_j | L = l) \quad (4)$$

where the probabilities  $P(L = l)$  and  $P(X_j = x_j | L = l)$  are determined based on the training set  $\mathcal{Y}$ . However, since con-

ditional independence is assumed, correlated features cannot be modeled appropriately.

A further supervised classifier is represented by the Support Vector Machine (SVM) classifier (Cortes and Vapnik, 1995). Being able to generalize in case of small training sets and robust to noise, SVMs have become very popular in recent years. In general, a SVM is a binary classifier trained to linearly separate two classes by constructing a hyperplane or a set of hyperplanes in a high-dimensional feature space. Since the linear separation in the feature space often is not possible, a kernel function is introduced which implicitly maps the training data into a new feature space of higher dimensionality where the data is linear separable. For solving the problem of multi-class classification, a SVM classifier composed of several binary SVMs and provided in the LIBSVM package (Chang and Lin, 2011) has been applied which is based on a one-against-one approach. Thus, for each pair of classes, a SVM is trained to distinguish samples of one class from samples of the other class. Since with TLS and MLS, many objects of similar shapes or at least similar geometrical behaviors are typically acquired (e.g. poles, wires, trunks or traffic lights), such a strategy may allow a better training step and subsequent discrimination of classes closely located in the feature space. In order to optimally select the parameter  $\gamma$  representing the width of a (Gaussian) radial basis function (RBF) kernel and the parameter  $C$  penalizing classification errors, a cross-validation based on a grid-search in a suitable subspace  $(\gamma, C)$  is applied. For this purpose, the training set  $\mathcal{Y}$  is divided into two disjoint sets  $\mathcal{Y}_1$  and  $\mathcal{Y}_2$ . Subsequently, at each point on the discrete grid, the performance of the whole classifier is evaluated by training on the new training set  $\mathcal{Y}_1$  with the current parameter choices and testing on the validation set  $\mathcal{Y}_2$ . Resulting from this, those values  $(\gamma, C)$  are selected which yield the best performance.

#### 2.5 Feature Relevance Assessment and Feature Selection

Once a variety of features has been extracted, it has to be considered that these may contain redundant or irrelevant information with respect to the classification task. Although in theory many classifiers are considered to be insensitive to the given dimensionality, such redundant or irrelevant information has been proved to influence their performance in practice. Hence, it is often desirable to select a compact subset of the most relevant features which allows for classification without significant loss of predictive information. Such a feature selection is important to defy the curse of dimensionality, to improve class separability and to facilitate interpretability. Furthermore, a significant increase of efficiency with respect to both processing time and memory consumption can be expected due to much less involved data.

Recent approaches to feature selection mainly apply filter-based methods, wrapper-based methods or embedded methods (Guyon and Elisseeff, 2003). The filter-based selection requires the definition of a performance evaluation function providing a score which can directly be calculated from the training data. Such a function typically exploits a common measure for distance, information, dependency or consistency. In contrast, the wrapper-based selection directly involves the performance of the learning machine in order to evaluate each subset and thus conduct feature selection. The embedded methods carry out feature selection during the training process, i.e. parameter space of the learning machine and feature selection space are searched simultaneously. In general, wrapper-based methods yield a better performance than filter-based methods. However, they have a risk of overfitting, require a huge computational effort and provide feature subsets optimized for the specified classifier only. Hence, the selected subset may be less suited in a different context which might for

instance be important for subsequent tasks such as object detection. The embedded methods also provide a classifier-dependent selection which might be less suited for other tasks. Since the filter-based selection evaluates intrinsic properties of the given data, the respective results exhibit more generality.

In order to avoid a classifier-dependent solution for deriving compact and robust feature subsets, we apply a filter-based feature selection where relevance is directly calculated from training data without involving learning techniques. Thus, feature selection is a small independent component in the processing chain which yields reproducible results. The order of the derived scores is used for ranking the features. As different criteria of the data may be relevant for scene analysis (Zhao et al., 2010), we not only use a single measure, but several measures<sup>1</sup>:

- The measure  $c_\chi$  results from a  $\chi^2$ -test which is used as a test of independence in order to assess whether a class label is independent of a particular feature. Following the provided implementation, a higher value indicates more relevance.
- The F-score or Fisher score  $c_{\text{Fisher}}$  represents the ratio between interclass and intraclass variance. A larger value indicates that a feature is more likely to be discriminative.
- The Gini Index  $c_{\text{Gini}}$  provides a statistical measure of dispersion and thus an inequality measure which quantifies a feature's ability to distinguish between classes. Smaller values of the Gini Index indicate more relevant features.
- The Information Gain  $c_{\text{IG}}$  is a measure revealing the dependence between a feature and a class label. Higher values represent more relevant features.
- The Pearson correlation coefficient  $c_{\text{Pearson}}$  indicates to which degree a feature is correlated with a class label. A higher correlation results in higher values and thus higher relevance.
- The ReliefF measure  $c_{\text{ReliefF}}$  indicates the contribution of a feature to the separation of samples from different classes. If samples with the same label have close values and samples with different labels are well discriminated, this measure provides a higher value.
- The measure  $c_t$  results from applying a  $t$ -test on each feature and checking how effective it is for separating classes. Following the provided implementation, a higher value indicates more relevance.

These measures encapsulate very different properties according to which feature subsets can be selected. The Gini Index and the Information Gain are commonly used for node splitting in decision trees. Typically, a single relevance metric is utilized for ranking the importance of the features according to the respective score. In order to realize a general ranking scheme taking into account very different aspects which might represent relevant properties of the given data, we apply a combined measure where the global rank of each feature is derived according to the ascending order of the mean rank across all measures. Feature selection can thus be realized by only keeping a feature subset consisting of the best-ranked features.

### 3 EXPERIMENTAL RESULTS

We demonstrate the performance of the proposed methodology for a standard MLS benchmark dataset which is described in Section 3.1. The parameter settings for the experiments as well as a detailed evaluation is presented in Section 3.2.

<sup>1</sup>Some of these measures are also included in the ASU Feature Selection Repository and available at <http://featureselection.asu.edu> (last access: 30 May 2013).

### 3.1 Dataset

For evaluation, we use the Oakland 3-D Point Cloud Dataset<sup>2</sup> (Munoz et al., 2008; 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  $\mathcal{Y}_1$ , validation set  $\mathcal{Y}_2$  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*.

### 3.2 Experiments

Assuming that a single 3D point can be characterized by the spatial relationships to its 50 closest neighbors, the whole set of 3D and 2D features is calculated. These features exhibit different scalings and therefore each feature is normalized to the interval  $[0, 1]$ . Subsequently, the combined relevance metric is calculated for sorting the features according to their relevance (Figure 1).

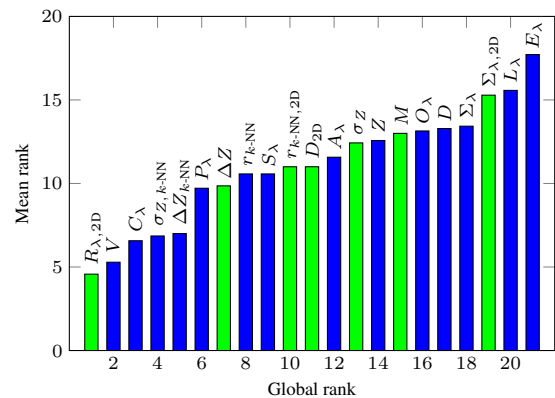


Figure 1: Mean rank of the derived 3D features (blue) and those features derived from considerations in 2D (green).

Following the principle of forward selection, we begin with only the most relevant feature. Subsequently, the derived order of the features is used to successively train and test the classifiers with one additional feature per iteration. Between the features with global rank 5 and global rank 6, the relevance metric shows a significant change (Figure 1). For this reason, we assume the five best-ranked features to be a meaningful subset in the experiments. As the number of samples per class varies significantly for both training set  $\mathcal{Y}_1$  and validation set  $\mathcal{Y}_2$ , we reduce these sets to equal class size by selecting all samples of the smallest class and randomly selecting the same number of samples for all other classes. We apply a multi-class SVM classifier exploiting a Gaussian kernel. The respective parameters ( $\gamma, C$ ) are derived via cross-validation where training is performed on the reduced training set  $\mathcal{Y}_{1,r}$  and validation is carried out with the reduced validation set  $\mathcal{Y}_{2,r}$ . The performance of the SVM classifier is evaluated on the test set  $\mathcal{Y}^*$  and compared with the performance of a Nearest Neighbor (NN) classifier, a  $k$  Nearest Neighbor ( $k$ -NN) classifier with  $k = 50$  and a Naive Bayesian (NB) classifier. The resulting overall accuracies are depicted in Figure 2. For being able to compare the results to other recent approaches, the confusion matrices for SVM-based classification exploiting all features and only the five best-ranked features are depicted in Table 1 and Table 2. The precision values can be seen as a measure of exactness or quality, whereas recall represents a measure of completeness or quantity. Finally, the overall classification accuracies for different feature subsets are provided in Table 3.

<sup>2</sup>The Oakland 3-D Point Cloud Dataset is available online at [http://www.cs.cmu.edu/~vmr/datasets/oakland\\_3d/cvpr09/doc/](http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/) (last access: 30 May 2013).

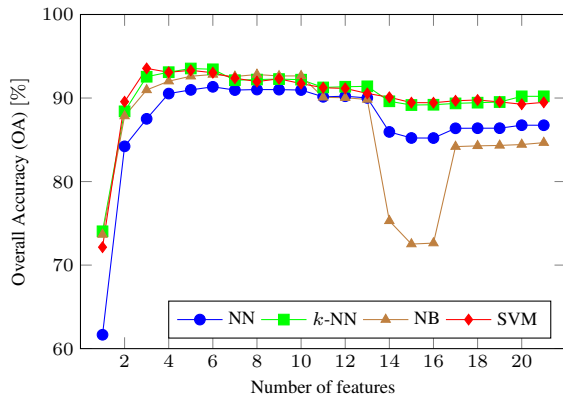


Figure 2: Overall classification accuracies for different classifiers: The size of the feature subset is iteratively increased according to the ascending global rank depicted in Figure 1.

	W	P/T	F	G	V	recall
W	2414	538	300	41	501	0.636
P/T	1552	4691	219	41	1430	0.591
F	10480	33298	63335	62	3937	0.570
G	12477	2785	223	914461	4200	0.979
V	24027	24349	13600	5229	200120	0.749
precision	0.047	0.071	0.815	0.994	0.952	

Table 1: Confusion matrix resulting from SVM-based classification with all 3D and 2D features (W: wire, P/T: pole/trunk, F: façade, G: ground, V: vegetation). The reference labels are provided in the left column, the estimated labels are represented in the upper row. The overall classification accuracy is 89.48%.

	W	P/T	F	G	V	recall
W	1831	125	724	208	906	0.483
P/T	1378	4729	386	8	1432	0.596
F	7506	5577	88664	237	9128	0.798
G	4680	501	191	917418	11356	0.982
V	19042	10308	11277	3550	223148	0.835
precision	0.053	0.223	0.876	0.996	0.907	

Table 2: Confusion matrix resulting from SVM-based classification with only the five best-ranked features. The overall classification accuracy is 93.32%.

	NN	k-NN	NB	SVM
all features	86.74	90.21	84.66	89.48
only 3D features	85.70	88.64	85.80	88.34
only 2D features	82.66	83.80	77.09	80.06
eigenvalue-based 3D features	80.36	82.42	83.92	84.81
$\{L_\lambda, P_\lambda, S_\lambda\}$	62.91	75.72	76.90	75.26
five best-ranked features	90.97	93.53	92.61	93.32

Table 3: Overall classification accuracies in % for different feature subsets and different classifiers.

#### 4 DISCUSSION

Considering feature ranking, a clear trend is visible. The proposed scheme reveals that the ratio  $R_{\lambda,2D}$  of the eigenvalues in 2D, the verticality  $V$  and the change of curvature  $C_\lambda$  are highly relevant, even when considering different relevance measures. The feature  $R_{\lambda,2D}$  which we propose in this paper is a strong feature for detecting façades. The verticality  $V$  is well-suited to distinguish between façades and ground, and the change of curvature  $C_\lambda$  provides a strong hint for distinguishing between planar structures such as façades or ground, and non-planar structures representing vegetation. The height difference  $\Delta Z_{k-NN}$  and the height variance  $\sigma_{Z,k-NN}$  within the local neighborhood are also assigned lower and thus better global ranks. In contrast, the features  $\Sigma_\lambda$ ,  $\Sigma_{\lambda,2D}$  and  $E_\lambda$  are always among the least relevant features which

is consistent to similar investigations involving echo-based and full-waveform features (Mallet et al., 2011). Interestingly and against a priori knowledge, also the linearity  $L_\lambda$  does not seem to be a relevant feature here (Figure 1). Indeed, when considering the behavior of this feature, it turns out that high values are mainly caused by edges due to occlusions or borders of the field-of-view which are not supposed to be linear objects. Furthermore, it becomes visible that the global ranks of features derived from considerations in 2D are distributed across the whole scale, i.e. there are more and less relevant features among them.

For the semantic 3D point cloud classification, the results indicate that the more features are used, the more complicated is the border separating the classes. A simpler border however results in better generalization and less overfitting. The performance of all classifiers is even significantly improved by selecting only a small subset of versatile features (Figure 2, Tables 1-3). Including the height  $Z$  with global rank 14 has a negative effect on the overall classification accuracy, in particular for the NB classifier. This might be due to the fact that elements of the different classes appear at different heights. The reduction to a small subset of five features also saves both processing time and memory consumption with respect to feature extraction as well as classification.

Moreover, the classification based the five best-ranked geometric features significantly outperforms classification using the set of all eigenvalue-based 3D features which are commonly used in literature as well as the subset  $\{L_\lambda, P_\lambda, S_\lambda\}$  often used for describing 1D, 2D and 3D structures (Table 3). The derived small feature subset even yields higher precision/recall values in comparison to recent investigations where Associative Markov Networks and thus graphical models involving contextual information have been used (Munoz et al., 2009b). A visual representation of the classification results for the subset consisting of the five best-ranked features is depicted in Figure 3.

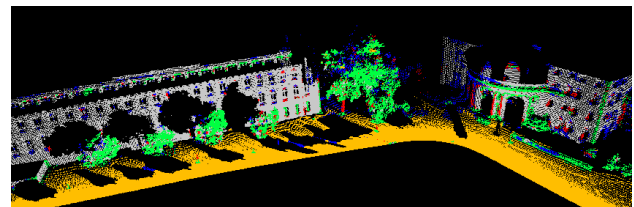


Figure 3: Visual representation of the classification results for the subset consisting of the five best-ranked features (W: blue, P/T: red, F: gray, G: brown, V: green).

#### 5 CONCLUSIONS AND FUTURE WORK

In this paper, a new methodology for feature relevance assessment and its beneficial impact on the semantic interpretation of 3D point cloud data are presented. Involving a variety of geometric 3D and 2D features, a general relevance metric addressing different intrinsic properties of the given data is exploited in order to derive compact and robust subsets of versatile features. The results reveal that the use of only a small feature subset is profitable for crucial tasks such as the semantic interpretation of 3D point clouds. The improvements not only address processing time and memory consumption, but also classification accuracy. For future research, it would be desirable to apply compact and robust subsets of versatile features in order to support point cloud registration, to facilitate navigation of autonomous vehicles equipped with range cameras or to alleviate an even more detailed scene analysis up to object level. For these tasks, it would also be helpful to automatically find the optimal size of the feature subset. Additionally, contextual information should be taken into account. Promising results can be expected.

## REFERENCES

- Chang, C.-C. and Lin, C.-J., 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2(3), pp. 27:1–27:27.
- Chehata, N., David, N. and Bretar, F., 2008. Lidar data classification using hierarchical K-means clustering. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXVII, Part B3b, pp. 325–330.
- 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.
- Cortes, C. and Vapnik, V., 1995. Support-vector networks. *Machine Learning*, 20(3), pp. 273–297.
- Cover, T. and Hart, P., 1967. Nearest neighbor pattern classification. *IEEE Transactions on Information Theory*, 13(1), pp. 21–27.
- 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.
- Demantké, J., Vallet, B. and Paparoditis, N., 2012. Streamed vertical rectangle detection in terrestrial laser scans for facade database production. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 1-3, pp. 99–104.
- Filin, S. and Pfeifer, N., 2005. Neighborhood systems for airborne laser data. *Photogrammetric Engineering & Remote Sensing*, 71(6), pp. 743–755.
- Guyon, I. and Elisseeff, A., 2003. An introduction to variable and feature selection. *Journal of Machine Learning Research*, 3, pp. 1157–1182.
- John, G. H. and Langley, P., 1995. Estimating continuous distributions in Bayesian classifiers. *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*, pp. 338–345.
- 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, Part 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.
- Lalonde, J.-F., Vandapel, N., Huber, D. and Hebert, M., 2006. Natural terrain classification using three-dimensional lidar data for ground robot mobility. *Journal of Field Robotics*, 23(10), pp. 839–861.
- Lari, Z. and Habib, A., 2012. Alternative methodologies for estimation of local point density index: Moving towards adaptive lidar data processing. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. XXXIX, Part B3, pp. 127–132.
- 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 Prantusch, H., 2001. Natural terrain classification using three-dimensional lidar data for ground robot mobility. *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.
- Lodha, S. K., Kreps, E. J., Helmbold, D. P. and Fitzpatrick, D., 2006. Aerial lidar data classification using support vector machines. *Proceedings of the International Symposium on 3D Data Processing, Visualization, and Transmission*, pp. 567–574.
- Lu, Y. and Rasmussen, C., 2012. Simplified Markov random fields for efficient semantic labeling of 3D point clouds. *Proceedings of the IEEE/RISJ International Conference on Intelligent Robots and Systems*, pp. 2690–2697.
- Maas, H.-G. and Vosselman, G., 1999. Two algorithms for extracting building models from raw laser altimetry data. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54(2-3), pp. 153–163.
- 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.
- Monnier, F., Vallet, B. and Soheilian, B., 2012. Trees detection from laser point clouds acquired in dense urban areas by a mobile mapping system. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-3, pp. 245–250.
- 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., 2008. Directional associative Markov network for 3-D point cloud classification. *Proceedings of the International Symposium on 3D Data Processing, Visualization and Transmission*, pp. 63–70.
- 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.
- Rusu, R. B., 2009. Semantic 3D object maps for everyday manipulation in human living environments. PhD thesis, Computer Science department, Technische Universität München, Germany.
- Shapovalov, R. and Velizhev, A., 2011. Cutting-plane training of non-associative Markov network for 3D point cloud segmentation. *Proceedings of the IEEE International Conference on 3D Digital Imaging, Modeling, Processing, Visualization and Transmission*, pp. 1–8.
- 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*.
- Toshev, A., Mordohai, P. and Taskar, B., 2010. Detecting and parsing architecture at city scale from range data. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 398–405.
- 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., Oude Elberink, S. J. and Vosselman, G., 2012. Entities and features for classification of airborne laser scanning data in urban area. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. I-4, pp. 257–262.
- 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.