CONTEXTUAL CLASSIFICATION OF POINT CLOUDS USING A TWO-STAGE CRF

J. Niemeyer [a, *], F. Rottensteiner [a], U. Soergel [b], C. Heipke [a]

[a] Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover, Germany -
(niemeyer, rottensteiner, heipke)@ipi.uni-hannover.de
[b] Institute of Geodesy, Technische Universität Darmstadt, Germany -
soergel@geod.tu-darmstadt.de

Commission III, WG III/4

KEY WORDS: Classification, Point Cloud, Conditional Random Fields, Contextual, LiDAR, Urban

ABSTRACT:

In this investigation, we address the task of airborne LiDAR point cloud labelling for urban areas by presenting a contextual classification methodology based on a Conditional Random Field (CRF). A two-stage CRF is set up: in a first step, a point-based CRF is applied. The resulting labellings are then used to generate a segmentation of the classified points using a Conditional Euclidean Clustering algorithm. This algorithm combines neighbouring points with the same object label into one segment. The second step comprises the classification of these segments, again with a CRF. As the number of the segments is much smaller than the number of points, it is computationally feasible to integrate long range interactions into this framework. Additionally, two different types of interactions are introduced: one for the local neighbourhood and another one operating on a coarser scale.
This paper presents the entire processing chain. We show preliminary results achieved using the Vaihingen LiDAR dataset from the ISPRS Benchmark on Urban Classification and 3D Reconstruction, which consists of three test areas characterised by different and challenging conditions. The utilised classification features are described, and the advantages and remaining problems of our approach are discussed. We also compare our results to those generated by a point-based classification and show that a slight improvement is obtained with this first implementation.

1. INTRODUCTION

The classification of airborne LiDAR point clouds is challenging for urban areas due to the large amount of different objects located close to each other. However, particularly in these regions an accurate classification result is desirable since it is often an important step for object detection and reconstruction, for example for the generation of a three-dimensional city model.

In the last years, classification techniques incorporating contextual knowledge such as Markov Random Fields (MRF) and Conditional Random Fields (CRF) have become more and more popular for point cloud labelling, (*e.g.* Anguelov et al., 2005; Shapovalov et al., 2010). In these kinds of probabilistic approaches the random variables to be labelled are represented as nodes of a graph, which are connected by edges modelling the interactions. Spatial dependencies between the object classes can be trained to improve the results because some object classes are more likely to occur next to each other than others; for instance, it is more probable that cars are situated on a street than on grassland. However, most applications in the related work do not exploit the full potential of graphical models: up to now, they mainly make use of relatively simple models for the interactions such as the Potts model and the contrast-sensitive Potts model (Boykov and Jolly, 2001). Both models favour neighbouring points to have the same object class by penalising label changes. Relations between different types of objects are not trained in these cases, which tends to lead to an over-smoothing effect (Schindler, 2012). Small objects, such as cars, might be eliminated for this reason. Niemeyer et al. (2011) showed that the use of a more complex multi-class model for the joint probability of all class labels at neighbouring sites, rather than a binary model for the probability of the two labels being equal, leads to better results in terms of completeness and correctness values. This gain in accuracy however comes at an expense of higher computational effort and a larger amount of fully labelled reference data being required during training.

The work of Shapovalov et al. (2010) has its focus on the classification of airborne LiDAR points discerning five object classes, namely *ground*, *building*, *tree*, *low vegetation*, and *car*. The authors applied a non-associative Markov Network, which is able to model all class relations instead of only preferring the same labels for both nodes linked by an edge. First, the data are over-segmented, then a segment-based CRF classification is performed. Whereas this aspect helps to cope with noise and computational complexity, the result heavily depends on the segmentation. Small objects of sub-segment size cannot be detected, and important object details might be lost, which is, of course, a drawback of all segment-based algorithms. The authors show that using a segmented point cloud leads to a loss of 1 %-3 % in overall accuracy in their experiments due to segmentation errors and due to the fact that classes having few samples such as cars might be merged with the background. Whereas this does not seem to be much, it may become relevant if the classes of interest are the ones most affected by these problems. Both, the point-based as well as the segment-based classification, have advantages and disadvantages. While using segments leads to less computational costs, a point-wise classification avoids the segmentation errors.

A particular limitation of graphical models, not only for point cloud classification, is the scale of the context. In most state-of-the-art work, CRF only consider local context between directly neighbouring points or segments, respectively. Considering long range context in a pairwise CRF operating on points would usually correspond to many more edges in the graphical model making inference intractable. There is again a difference between the point and segment-based classifications since the latter is able to model context in a larger scale whereas the former is usually lim-

[*] Corresponding author

This contribution has been peer-reviewed.
doi:10.5194/isprsarchives-XL-3-W2-141-2015

ited to local context which might lead to some remaining errors due to ambiguities of local point cloud information.

Shapovalov et al. (2013) classified point clouds of indoor scenes, building a graphical model on point cloud segments. They consider long range dependencies by so-called structural links, also based on spatial directions such as the vertical, the direction to the sensor or the direction to the nearest wall. In an indoor scenario, walls can be detected using heuristics (Shapovalov et al., 2013). However, such approaches do not carry over to airborne data easily. CRF were also used by Lim and Suter (2007) for the point-wise classification of terrestrial LiDAR data. The authors coped with the computational complexity by adaptive point reduction. In further work they first segmented the points and then classified the resulting superpixels. The authors also considered both a local and a regional neighbourhood. Introducing multiple scales into a CRF represented by long range links between superpixels improved the classification accuracy by 5 % to 10 % (Lim and Suter, 2009). This result shows the importance of considering larger regions instead of only a very local neighbourhood of each 3D point for a correct classification. An alternative to long range edges, which might lead to a huge computational burden if points are to be classified individually, is the computation of multi-scale features, enabling a better classification of points with locally similar features. Although belonging to different objects, the variation of the regional neighbourhood can support the discrimination between the object types, and hence lead to a correct labelling.

A further option to incorporate more regional context into the classification process are CRF with higher order potentials. Najafi et al. (2014) set up a non-associative variant of this approach for point clouds. They first performed a segmentation and then applied the CRF to the segments. Overlapping segments in 2D were considered by a higher order potential. The authors additionally modelled the object class relation in the vertical direction with a pattern based potential. This is useful for terrestrial scans, but in the airborne case the derived features for the vertical are not very expressive due to missing point data for example on façades. Although higher order potentials are becoming more and more important, it is still difficult to apply them to point clouds (in particular for point-based classification) due to the extensive computational costs. Inference for such models is a challenging task and up to now only very few nodes can be combined to form a higher order clique for non-associative interaction models which currently restricts the expressive power of this framework. In the case of Najafi et al. (2014) only up to six segments were combined to one higher order clique to deal with this problem. Xiong et al. (2011) showed how point-based and region-based classification of LiDAR data can interact in a pairwise CRF. They proposed a hierarchical sequence of relatively simple classifiers applied to segments and points. Starting either with an independent classification of points or segments, in subsequent steps the output of the previous step is used to define context features that help to improve the classification results. In each classification stage, the results of the previous stage are taken as input, and, unlike with a single CRF, it is not guaranteed that a global optimum is reached (Boykov and Jolly, 2001; Kumar and Hebert, 2006). Luo and Sohn (2014) applied two asymmetric (pairwise) CRF for short range and long range interactions separately on terrestrial scan line profiles and evaluated the experiment on a terrestrial laser point cloud. The final label for each laser point is determined by finding the maximum product of both CRF posteriors. While the short range CRF has a smoothing effect on the point labels, a CRF for long range interaction models the structure of the scene. It was found that the introduction of the long ranges context was able to eliminate some misclassification errors such as trees on buildings or buildings on the top of trees. In our application of classifying airborne LiDAR data (Niemeyer et al., 2014), we observed similar confusion problems in the results of a CRF operating on points only locally.

The aim of this paper is to present and investigate a new two-stage CRF framework, which was inspired by the approaches of Luo and Sohn (2014), Xiong et al. (2011), and Albert et al. (2014). The latter work described a two-step method for the land use and land cover classification based on another kind of data, namely aerial images. Both of our CRFs for the point cloud classification make use of the complex, asymmetric interaction model, and hence learn all joint probabilities of the classes. The first classification step operates on individual points, referred to as CRF_P, in order to avoid the smoothing effects of a segmentation. Based on these results a segmentation is performed taking into account the actual class labels. The results represent the input for the second CRF-based classification (CRF_S), in which the nodes in the graph correspond to segments. The intention of this framework is to incorporate local as well as long range context. Wrongly classified points of CRF_P can then be correctly labelled by using this additional information. Therefore, two interaction potentials are trained in CRF_S: one for the local neighbourhood and another one for more regional context. We distinguish between the eight classes *natural soil*, *road*, *gable roof*, *flat roof*, *cars*, *low vegetation*, *trees*, and *façades&fences* and present preliminary results.

2. CONDITIONAL RANDOM FIELDS

We start with a brief overview of the CRF, which belong to the family of undirected graphical models. The underlying graph $G(n, e)$ consists of nodes n and edges e. We assign class labels y_i to each node $n_i \in n$ based on observed data $\mathbf{x}$. The vector $\mathbf{y} \in \Omega$ contains the labels y_i for all nodes, and hence has the same number of elements as n. The amount of object classes to distinguish is indicated by c. The graph edges e_{ij} are used to model the relations between pairs of adjacent nodes n_i and n_j, and thus enable representing contextual relations. For that purpose, edges link each node n_i to adjacent nodes ($n_j \in N_i$). The actual definitions of the graphs we use in our method is explained later. They are different for CRF_P and CRF_S.

CRF are discriminative classifiers that model the posterior distribution $p(\mathbf{y}|\mathbf{x})$ directly (Kumar and Hebert, 2006):

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})} \left(\prod_{i \in n} \phi_i(\mathbf{x}, y_i) \cdot \prod_{i,j \in e} \psi_{ij}(\mathbf{x}, y_i, y_j) \right). \quad (1)$$

The two terms $\phi_i(\mathbf{x}, y_i)$ and $\psi_{ij}(\mathbf{x}, y_i, y_j)$ in Eq. 1 are called the unary and pairwise potentials, respectively. e is the set of edges for the adjacent nodes n_i and n_j. The partition function Z($\mathbf{x}$) acts as normalisation constant, turning potentials into probabilities.

In the unary potentials, the data are represented by node feature vectors $\mathbf{h}_i(\mathbf{x})$. For each node n_i such a vector is determined taking into account not only the data $\mathbf{x}_i$ observed at that point, but also at the points in a certain neighbourhood. The particular definition of the node features used in our experiments are described in Sections 3.1 and 3.3. Using these node feature vectors $\mathbf{h}_i(\mathbf{x})$, the unary potential $\phi_i(\mathbf{x}, y_i)$, linking the data to the class labels, determines the most probable label of the c classes for a single node given its feature vector. ϕ is modelled to be proportional to the probability for y_i given the data:

$$\phi_i(\mathbf{x}, y_i) \propto p(y_i|\mathbf{h}_i(\mathbf{x})). \quad (2)$$

This is a very general formulation which allows to use any discriminative classifier with a probabilistic output for the unary potential (Kumar and Hebert, 2006). In this work, we chose a Random Forest (RF) classifier (Breiman, 2001) for the computation of the unary potential.

An RF is a bootstrap ensemble classifier based on decision-trees. It consists of a number T of trees grown in a training step. For the generation, each internal node of any tree corresponds to a test to find the best feature as well as the corresponding threshold to split the data into two parts. In order to classify an unknown sample from the dataset, each tree casts a vote for the most likely class based on its features which are presented to the trees. Further details can be found in Hänsch (2014). Dividing the sum of all votes for a class by the total number of trees defines a probability measure which is used to model the potential. The maximum number of samples used for training, the maximum depth, the minimum number of samples for a split, and the number of trees in the forest are the main parameters that have to be adapted[1].

The second term $\psi_{ij}(\boldsymbol{x}, y_i, y_j)$ in Eq. 1 represents the pairwise potential and incorporates the contextual relations explicitly in the classification process. It models the dependencies of a node n_i from its adjacent node n_j by comparing both node labels and considering the observed data **x**. In the pairwise potentials, the data are represented by interaction feature vectors $\boldsymbol{\mu}_{ij}(\mathbf{x})$ which are computed for each edge e_{ij}. In contrast to simple models such as the Potts model and the contrast-sensitive Potts model, a training of all local relations between the object classes does not only lead to a smoothing effect but is also able to learn that certain class relations may be more likely than others given the data (Niemeyer et al., 2014). In this case the potential is modelled to be proportional to the joint posterior probability of two node labels y_i and y_j given $\boldsymbol{\mu}_{ij}(\mathbf{x})$:

$$\psi_{ij}(\boldsymbol{x}, y_i, y_j) \propto p(y_i, y_j | \boldsymbol{\mu}_{ij}(\mathbf{x})). \tag{3}$$

This information is used to improve the quality of classification, with the drawback of having to determine more parameters. We apply an RF classifier to obtain the probabilities for the interactions in a similar way as to the unary potentials. The only difference is that now c^2 classes are trained and distinguished for the pairwise potential because each object class relation is considered to be a single class.

In the context of graphical models, inference is the task of determining the optimal label configuration based on maximising $p(\mathbf{y}|\mathbf{x})$ for given parameters. For large graphs with cycles exact inference is computationally intractable and approximate methods have to be applied. We use the max-sum version of the standard message passing algorithm Loopy Belief Propagation (LBP) (Frey and MacKay, 1998). Independent RF classifiers have to be trained for the unary and pairwise potentials. In order to learn the interactions of object classes, a fully labelled reference point cloud is needed.

3. FRAMEWORK

In computer vision and remote sensing, it was shown that typical structures of man-made objects can be used as contextual knowledge to improve classification results (Lim and Suter, 2009; Luo and Sohn, 2014). For instance, objects are often regularly distributed and have a certain relative arrangement. This information can be used to further increase the detection rate of these objects. In our previous approach (Niemeyer et al., 2014), these structures could not be modelled due to the computational complexity, because too many interactions would have been necessary. In order to avoid this computational intractability, we first apply the point-based classification (CRF_P) and introduce a second CRF based on segments (CRF_S), which is applied after the first stage. For the second CRF a segmentation is necessary. The final label for the segment is then assigned to each point belonging to this segment. The basic idea and main motivation of this framework is that some remaining classification errors of CRF_P might be eliminated by utilising more regional context information between segments instead of points. For example, the typical surrounding of trees might be modelled in this case in order to detect and correct *roof* points which were wrongly classified as *tree* canopy. This kind of error appeared a few times in Niemeyer et al. (2014) due to the features of those points. Locally the planarity of the points is high, because the LiDAR data was obtained in summertime under leaf-on conditions. Nearly all of the laser pulses are reflected from the canopy and do not penetrate the tree far enough to recognise points within the trees. This appearance is similar to that of building roofs and the points are often classified as *building*. This problem is difficult to solve incorporating only a local, point-wise neighbourhood. The second stage CRF_S operating on segments, and, hence, on a larger region, might learn that a small building segment (obtained from CRF_P) surrounded only by tree segments is likely to be labelled as *tree* instead. Of course this assumption requires a good segmentation, because the segments are the entities for CRF_S and the basis for new features.

By applying this sequential approach, context can potentially be introduced in two ways. On the one hand, contextual features may represent the local and the regional neighbourhood. For instance, a histogram of the number of segments per object class in a predefined neighbourhood might support a correct classification of a segment in CRF_S. The distance to other segments of a certain object type, such as the distance of a building to the closest street, can also be taken into account as a contextual feature. Up to now, these kinds of features have not been implemented. The features we use are explained in Sections 3.1 and 3.3. On the other hand, larger scale context is additionally introduced by the interactions of the two graphical models. While local interactions are mainly modelled by the point-based CRF_P, the segment-based CRF_S is able to represent long range interactions of regional level.

In the following subsections the three components CRF_P, segmentation, and CRF_S are described in more detail.

3.1 Point-Based Classification CRF_P

In the case of the point-based classification, each point represents a node of the graph and is linked by edges to its k nearest neighbours in 2D. This corresponds to a cylindrical neighbourhood which was identified to be more expressive than a spherical neighbourhood (Niemeyer et al., 2011).

After constructing the graph, a node feature vector $\mathbf{h}_i(\mathbf{x})$ consisting of 36 elements is extracted for each node n_i. The features, which have been shown to lead to good results (Chehata et al., 2009; Niemeyer et al., 2014), are:

1. intensity;
2. ratio of echo number per point and number of echoes in the waveform (a point cloud with multiple echoes is used, see Section 4.1);
3. height above DTM;

[1] OpenCV Reference - Random Trees, http://docs.opencv.org/modules/ml/doc/random_trees.html (accessed 15/01/2015).

4. approximated plane (points in a spherical neighbourhood of 1 m radius are considered): standard deviation and sum of the absolute residuals, direction and variance of normal vector;
5. variance of point elevations in a cylinder and in a sphere of 1 m radius;
6. ratio of point density in a cylinder and a sphere of 1 m radius;
7. eigenvalue-based features in a sphere of 1 m radius: three eigenvalues (λ_1, λ_2, λ_3), omnivariance, planarity, anisotropy, sphericity, eigenentropy, scatter (Chehata et al., 2009);
8. point density in a sphere of 1 m radius;
9. principal curvatures $k1$ and $k2$, mean and Gaussian curvature in a sphere of 1 m radius;
10. variation of intensity, omnivariance, planarity, anisotropy, sphericity, point density, number of returns, $k1$, $k2$, mean curvature, and Gaussian curvature in a sphere of 1 m radius.

The DTM for the feature *height above DTM* is generated using robust filtering (Kraus and Pfeifer, 1998) as implemented in the commercial software package *SCOP++*[2]. To derive the interaction feature vector $\boldsymbol{\mu}_{ij}(\mathbf{x})$, we concatenate the original feature vectors $\mathbf{h}_i(\mathbf{x})$ and $\mathbf{h}_j(\mathbf{x})$ of both nodes to one vector described by 70 elements. All features of nodes and interactions are scaled to the range [0,1].

3.2 Segmentation

Based on the results of the point-based classification, segments are extracted in the next step. For this task, a Conditional Euclidean Clustering (Rusu, 2009) is applied as implemented in the Point Cloud Library (Rusu and Cousins, 2011). It is a variant of a region growing algorithm connecting points which are close to each other and meet additional conditions. In our case the points are allowed to have a distance of d_{max} and must have the same label from the point-based classification to be assigned to the same segment. There is no minimum size for the segments in order to consider each point. This leads to a segmented point cloud with the advantage of having a prior for the semantic meaning for each entity, potentially enabling the extraction of more meaningful features for a following segment-based classification.

The prior information about the class labels makes it possible to process the classes differently and introduce more model-based knowledge. On the one hand, for example, classes consisting of planes can be segmented by a RANSAC plane detection model. Large segments detected by the Conditional Euclidean Clustering step can be split up then, enabling a more accurate extraction of the features such as the normal vectors and planarity. In our study, this is performed for *gable roofs* which are comprised of two roof planes in most cases. On the other hand, it is also possible to merge small segments which might represent one larger plane. We combine the segments of *façades* which are not detected as a connected segment in the first clustering due to a lower point density in these areas.

3.3 Segment-Based Classification CRF_S

The segments generated in the way described in Section 3.2 are the main entities for CRF_S and represent the nodes in this second graphical model. It is advantageous not to apply only one single model for the interactions but to introduce a distance-dependent interaction model as reported by Luo and Sohn (2014). We introduce two types of edges, namely edges for close range (e_{cr}) and edges for long range interactions (e_{lr}). The indices i and j in the interaction potential of Eq. 1 indicate that the type of interaction potential may vary with the relation to certain edges. Correspondingly the interaction potential of CRF_S consists of the two terms ψ_{cr} for close range and ψ_{lr} for long range interactions, respectively. The model is described by

$$p(\mathbf{y}|\mathbf{x}) = \frac{1}{Z(\mathbf{x})}\left(\prod_{i\in n}\phi\cdot\prod_{i,j\in e_{cr}}\psi_{cl}\cdot\prod_{i,j\in e_{lr}}\psi_{lr}\right). \quad (4)$$

The parameters of the potentials are the same as in Eq. 1. Both interaction potentials are designed by RF classifiers in this study and have to be trained separately. The motivation for the separation is that two different interaction types should be considered in CRF_S. On the one hand, we want to model expressive local edges in order to obtain a good result. They have a high influence on the output. For this classifier, two segments are linked by an edge if points of one segment lie within a sphere of a radius r_{cr}. On the other hand also the geometric arrangements in a coarser scale should be considered as it was reported to introduce helpful information, (*e.g.* He et al., 2004; Lim and Suter, 2009; Gould et al., 2008). The aim of the second interaction potential ψ_{lr} is to model these relations. In this case, a larger neighbourhood is obviously considered. Long range interactions for the graph are introduced by linking segments that are outside of r_{cr} but having points within a radius of r_{lr}. Potentially they can be extended to cover the entire scene if good contextual long range features are available.

The availability of segments enables us to extract a new set of features for nodes and interactions. Computing the mean values and standard deviations of all point features within a certain segment is one option. In particular the heights and their variations are important. Taking into account the eigenvalues as well as the normal vectors is useful to separate *tree* segments from *roofs*, for example. Furthermore, more expressive segment-based features such as the number of points or the maximum difference in point elevation within one segment can be used. We construct the following feature vector with 20 elements for nodes:

1. means of point-based eigenvalues ($\lambda_{1,mean} - \lambda_{3,mean}$);
2. mean and standard deviation of point-based normal directions;
3. mean of point-based residuals of an approximated plane;
4. number of points in segment;
5. mean and standard deviation of point-based intensity values;
6. mean and standard deviation of point-based height above DTM;
7. maximum difference in elevation in one segment;
8. means of the resulting point-based CRF_P class beliefs per segment.

The interaction feature vector is again based on the concatenation of the two adjacent node feature vectors. However, the beliefs are not taken into account for the interactions, which results in an interaction feature vector consisting of 24 elements.

4. EXPERIMENTS

4.1 Test site

The performance of our method is evaluated on the LiDAR benchmark data set of Vaihingen, Germany (Cramer, 2010) from the 'ISPRS Test Project on Urban Classification and 3D Building

[2] http://www.trimble.com/imaging/inpho.aspx?tab=Geo-Modeling_Module

Reconstruction' (Rottensteiner et al., 2014). The data set was acquired in August 2008 by a Leica ALS50 system with a mean flying height of 500 m above ground and a 45° field of view. The average strip overlap is 30 % and the point density in the test areas is approximately 8 points/m^2. Multiple echoes and intensities were recorded. However, only very few points (2.3 %) are multiple returns, as the acquisition was in summertime under leaf-on conditions. Hence, the vertical point distribution within trees is such that most points describe only the canopy.

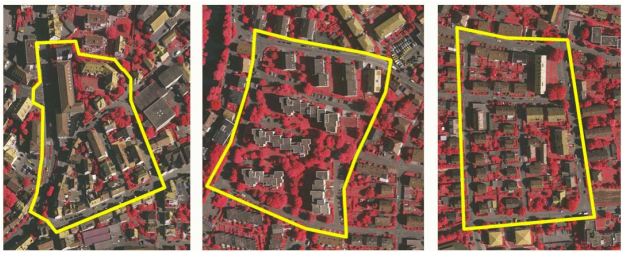


Figure 1. Test sites of scene Vaihingen. 'Inner City' (Area 1, left), 'High-Riser' (Area 2, middle) and 'Residential' (Area 3, right) (Rottensteiner et al., 2014)

For this study, three test sites with different scenes are considered (Fig. 1). Area 1 is situated in the centre of the city of Vaihingen. Dense, complex buildings and some trees characterise this test site. Area 2 consists of a few high-rising residential buildings surrounded by trees. In contrast, Area 3 is a purely residential neighbourhood with small, detached houses.

We manually labelled the point cloud of the three test areas to enable an evaluation of the 3D classification results. The combined point cloud consists of 780,879 points. Additionally, fully labelled training data are required to learn the joint probabilities of the classes. Here, we have two training sets $T1$ and $T2$ located directly in the East of test areas 1 and 3, respectively, with 94,405 points in total. One set, $T1$, is used to train the point-based CRF_P and classify the three test areas as well as the other training set $T2$. The latter is then segmented based on the classification result in a similar way as the test areas. The difference is that points are only combined to one segment if 1) they have the same classification label and 2) the reference label of the training data does not change. Afterwards, the label of the reference is assigned to this segment, preventing ambiguity.

4.2 Parameters

The following section describes the parameters used in this study.

For the construction of the graph in CRF_P, each point is linked to its $k = 4$ nearest neighbours in 2D, which is a good trade-off between accuracy and computational time. In case of the CRF_S, the linking of segments is a bit more complex due to the two interaction potentials. We use a kd-tree searching for neighbouring points within a sphere and having a different segment ID than the currently investigated point. We found $r_{cr} = 1\,m$ to be a good value because only direct neighbours are to be considered for ψ_{cr}. In the training step, it is learned that certain class relations do not appear in this local neighbourhood. For example, there is no link between a *car* segment and a *gable roof* segment, whereas a *car* is likely to be situated close to a *road* or another *car* segment. Long range edges are constructed for segments that have a distance of more than r_{cr} and less than $r_{lr} = 20\,m$, and that consist of at least 10 points, making the edges more robust and expressive. In order to make the obtained potentials of relations to objects located further away comparable to those of the local neighbourhood, a distance-dependent weight is introduced. Starting at r_{cr} with 100 % it linearly decreases in relation to the distance down to 30 % at r_{lr}. These values have been found empirically; they might of course also be learned in the future.

All potentials used in this study are modelled with RF classifiers. As RF optimize the overall error rate, a class with many samples might lead to a bias in the training step. Thus, the training set is balanced by randomly selecting the same number of samples for each class (Chen et al., 2004). In case of the CRF_P, two RF classifiers have to be learned on the training data for the unary and pairwise potentials. The first one distinguishes eight classes and the latter classifies 64 different object class relations. We use 100 trees with 10,000 training samples for each class, and a maximum depth of 25 in both cases. For CRF_S, we train three independent RF classifiers (one for the unary potential ϕ, one for ψ_{cr} and one for ψ_{lr}, respectively). We use 400 trees with a maximum depth of 25 in each case. For the unary potential classifier, 5,000 samples are used per class, whereas both pairwise potential classifiers ψ_{lr} and ψ_{lr} are generated on 2,000 samples to train the 64 class relations. The maximum depth is 25 and the minimal number of samples to perform a split is 2 for each RF. Moreover, the number of the random feature subset is set to the square root of all input features, following Gislason et al. (2006).

In the segmentation process, a threshold parameter for the distance in the Euclidean Clustering algorithm is needed. Points are connected to one segment if $d_{max} \leq 1\,m$. The used distance threshold for the RANSAC plane detection to separate the *gable roof* segments is 0.2 m.

4.3 Preliminary Evaluation

In order to obtain a first qualitative evaluation of our framework, we classified the three test sites with both CRFs as described in Section 3. The final confusion matrix for the point labels after the second, segment-based CRF_S is shown in Tab. 1. Completeness, correctness as well as quality values per class are also reported. Rutzinger et al. (2009) defined the quality as

$$\text{Quality} = \frac{1}{\text{Completeness}^{-1} + \text{Correctness}^{-1} - 1}. \quad (5)$$

The overall accuracy (OA) for the three areas is 80.5 %, which is a reasonable result for the challenging areas and the separation of eight different object classes. The influence of the amount of classes c on the OA is discussed in Section 4.4. It is easy to see that the accuracies in terms of completeness and correctness for the object classes vary significantly.

Best classification results were obtained for *gable roof* and *road* with high completeness ($> 90\,\%$) and correctness ($> 87\,\%$) values, resulting in high quality values. *Natural soil*, buildings with *flat roofs* as well as *trees* are more challenging and achieve quality values of 66-68 %. Most classification errors appear for *low vegetation*, *cars* and *façades*. These classes, which are not so prominent in the data set, have lower quality values between 26-40 %. Since these classes are not represented by many points in the point cloud, a few misclassifications have a strong influence on the completeness and correctness values. In Niemeyer et al. (2014) it was shown for a point-based CRF that these small classes particularly benefit from the context, and were improved compared to a common, non-contextual classifier. However, it is still challenging to detect them reliably.

Table 1 additionally provides a comparison to the point-based classification CRF_P. The difference is shown in brackets for the accuracy values. Positive values correspond to a better result

Reference\Class	Natural Soil	Road	Gable Roof	Low Veg.	Car	Flat Roof	Façade	Tree	Correctness
Natural Soil	18.6	2.3	0.0	1.0	0.0	0.4	0.2	0.1	82.2 (4.2)
Road	3.3	24.1	0.0	0.1	0.1	0.0	0.0	0.0	87.3 (-0.2)
Gable Roof	0.0	0.0	13.7	0.1	0.0	0.8	0.1	0.8	89.2 (0.4)
Low Veg.	0.7	0.1	0.0	2.6	0.1	0.1	0.3	0.4	61.0 (-0.3)
Car	0.1	0.0	0.0	0.1	0.4	0.1	0.0	0.0	54.7 (0.0)
Flat Roof	0.0	0.0	0.1	0.2	0.0	5.5	0.0	0.4	87.7 (0.0)
Façade	0.3	0.1	0.3	1.1	0.0	0.5	1.7	1.3	32.4 (0.8)
Tree	0.3	0.1	0.4	2.5	0.0	0.2	0.5	13.8	77.9 (0.1)
Completeness	80.0 (0.2)	90.4 (3.0)	94.0 (-0.3)	34.8 (0.4)	59.5 (1.0)	73.0 (1.0)	60.1 (-0.5)	82.1 (0.6)	Overall Accuracy: 80.5 (1.0)
Quality	68.2 (3.1)	79.9 (2.4)	84.4 (0.1)	28.4 (0.2)	39.9 (0.5)	66.2 (0.8)	26.6 (0.5)	66.6 (0.4)	

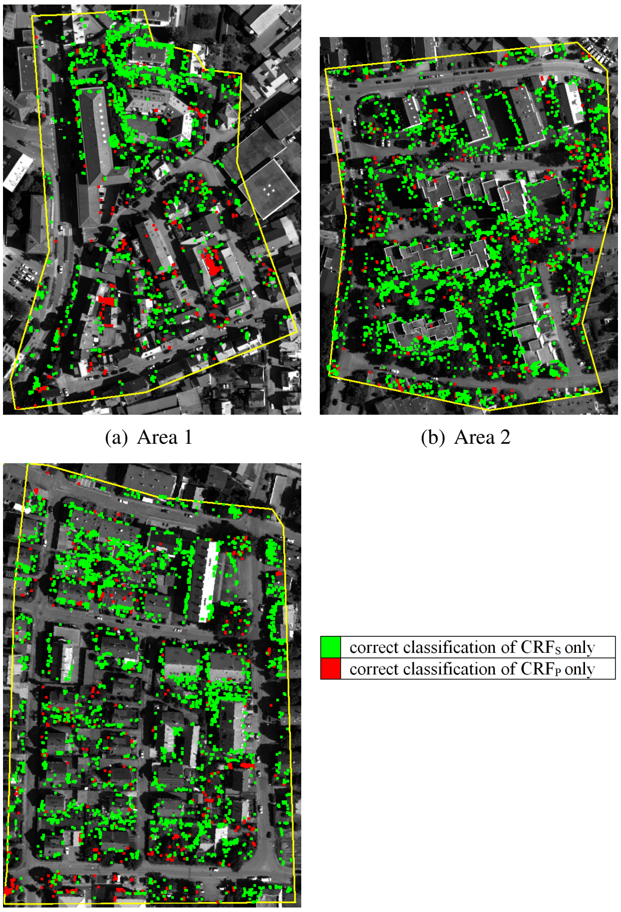
Table 1. Confusion matrix of CRF_S for the three test sites in [%]. The comparison to CRF_P is shown in brackets. Values in green correspond to a better result of CRF_S.

of CRF_S. Only four of the 16 values were decreased slightly by the segments (correctness of *road* and *low vegetation* as well as completeness of *gable roof* and *façade*, respectively). Two more values stayed unchanged and the other ten values were improved. Especially the quality values of *natural soil* and *road* were increased by 3.1 % and 2.4 %. This result shows that both ground classes benefit by the segment classification. The positive influence of the two-stage approach slightly increases the OA by 1.0 % and hence improved the results of CRF_P.

4.4 Discussion

A point-wise comparison of CRF_S and CRF_P is shown in Fig. 2. Points being classified correctly only after the segment-wise approach are highlighted in green, red indicates points that are only correct for the point-wise case. For the three test areas 1.5 % of the points are corrected by CRF_S. Many of these points are located close to buildings, most of them belong to the class *natural soil*. The vicinity to the building makes a correct classification using point-wise CRF_P difficult since only local context is considered. These ground points located next to buildings can show slightly different features (such as the eigenvalue based features and those based on the approximated plane) due to the presence of the vertical *façade* points leading to a three-dimensional point distribution. The same effect appears for *natural soil* points in areas with *low vegetation*. The coarser scale of CRF_S helps to improve these problems in some cases because a more stable mean value for the segment is considered instead. One more example is shown in Fig. 3. Here, some *tree* points were wrongly assigned to the class *gable roof* in CRF_P, but CRF_S was able to correct this mistake. In only 0.3 % of all points CRF_P is better than CRF_S. In this case the distribution of the classes is more homogeneous and comprises mainly *natural soil* (40 %), *low vegetation* and *façade* (19 % each) as well as *gable roof* (12 %).

A relatively large amount of points (≈ 20 %) was not classified correctly in both variants. The misclassifications appear for all reference classes, but this effect particularly arises for *natural soil*, *road*, and *trees*. Moreover, the less frequent class *façades* is often mixed up with *trees*. One main reason is that the features might not be discriminative enough in order to distinguish certain classes reliably. For example a large amount of confusion errors is observed between *natural soil* and *road* with 3.3 % and 2.3 % (Tab. 1). The most important feature to separate these classes is the intensity. However, even in the original data, it is sometimes difficult for a human operator to find the exact class boundaries because the intensity values are very similar. An example is given in Fig. 4 showing an intensity coloured subset of a point cloud as well as the corresponding orthophoto. Although the parking spaces are sealed surfaces and thus belong to the class *road*, the intensity values are clearly different and better point to the class *natural soil*. A solution might be to combine both classes and introduce only one class *ground* instead. This significantly increases the OA from 80.3 % to 86.8 %. Furthermore, a separation of *gable roof* and *flat roof* might be too detailed to be detected correctly based on the data. Additionally merging these classes improves the OA to 87.8 %. Moreover, it is hard to decide whether a point belongs to *low vegetation* or a *tree*. The most important feature is the elevation of the points, but there is no clear definition of the classes. As a consequence a combination of both classes eliminates lots of errors and further improve the OA to 90.2 %. We see that the number of classes has a significant influence on the OA. With only 5 classes *ground*, *roof*, *vegetation*, *car* and *façade* the OA increases by nearly 10 %.

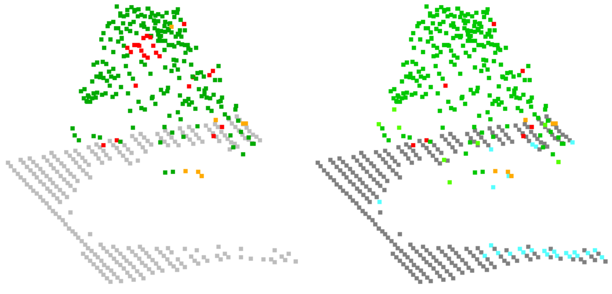

🟩 correct classification of CRFS only
🟥 correct classification of CRFP only

(a) Area 1

(b) Area 2

(c) Area 3

Figure 2. Point-wise comparison of the results obtained by CRF_S and CRF_P.



(a) CRF_P (b) CRF_S

Figure 3. Example of a correction by CRF_S. In 3(a) some *tree* points (green) are wrongly labelled as *gable roof* (red) in CRF_P. In the segment-based case in 3(b) this problem is partly corrected.

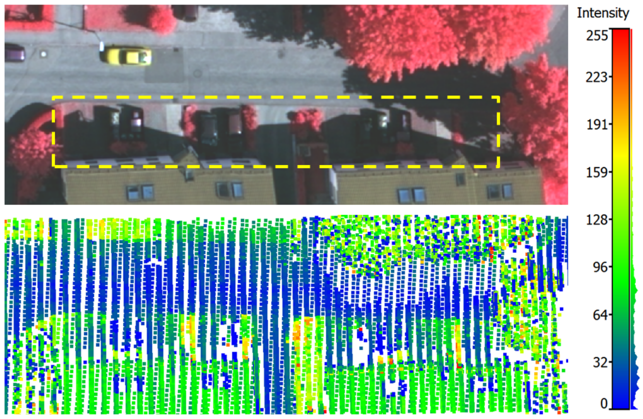


Figure 4. Echo intensity of *road* and *natural soil*. The parking spaces are highlighted in the orthophoto.

The relatively simple segmentation algorithm performs reasonably well when also applying the two post-processing steps for the classes *gable roof* and *façade*. It has turned out that a post-processing for all segments of the class *gable roof* is useful. In many cases both parts of a building are detected as one single segment. In order to obtain expressive features for each segment, these large segment is divided into several parts by a RANSAC plane detection algorithm. After this step each plane of a building is represented as a separate segment, which mainly influences the calculation of the normal vector and the planarity for each segment. The quality of the *façade* segments also benefit from the additional RANSAC post-processing by merging the points with a lower point density. However, there are still some aspects concerning the segmentation that can improve the overall performance in the future. There are currently too many small segments, making it difficult to derive features such as the size of the area, segment normal vector etc. Another example are trees because sometimes only one segment was detected for several trees located closely next to each other. This problem also influences the calculation of the normal vectors per segment. An option to cope with this problem is to apply a single tree detection based on a canopy height model. As *façade* points tend to be classified as *tree*, the generation of larger *façade* segments is more challenging. Here the segmentation with the Conditional Euclidean Clustering in its current state is not sophisticated enough because only points with the same label are connected to one segment. Thus, the final label of CRF_P is a too restricting feature to separate the classes. One solution is to use the beliefs (or the margin between the most and the second likely class) instead to set up a softer decision in the case of two classes being comparably likely. Moreover, other features such as the elevation difference can also be considered in the clustering step. We intend to apply the mentioned aspects in future work in order to improve the segmentation.

The features used in this study are able to improve the results by a 1 % increase in OA. However, most of the segment-based features are simply a mean value and the standard deviation of the corresponding point features up to know. Many more features can be introduced for segments such as area, volume, point density, compactness, etc. A good indicator to eliminate for instance the remaining misclassifications of *gable roof* assigned to *tree* is to investigate the height differences at the boundary of two segments. For future work, it is also planned to introduce additional contextual features which can be computed based on prior knowledge about the segment class from the segmentation. Some options are histograms about the class distribution of neighbouring segments, which might support the classification process. One can think of fixing the *road* points of CRF_P and derive a model to describe the road. In particular the points located in the middle of a road are detected very reliably in CRF_P. A road represents the network structure and characterises an urban area. Each object can than be analysed in relation to the closest street and features such as object orientation and distance to the street can be introduced. Some kind of learned relative location prior (Gould et al., 2008) would be interesting to integrate in this connection. However, the implementation for a point cloud in the airborne case is more challenging than for terrestrial images because of a missing reference direction in the scene. This has to be defined in advance to enable learning of the objects' arrangement. It can be an option to apply a kind of local relative location prior with respect to each road segment and investigate the orientation and distribution from objects in relation to the road. Additionally, also the structure of the graph can be improved by setting long range interactions only parallel to the street, for example. To conclude, the spatial alignment between segments and objects, respectively, should be utilised as a feature.

Nevertheless, we already achieved an improvement of the accuracy by the segment-based classification. We think that this framework provides a good potential and after investigating some of the mentioned issues we expect a further improvement.

5. CONCLUSION AND OUTLOOK

In this study, we have presented a two-stage contextual classification framework for LiDAR point clouds which is able to take into account different scales of a scene to model context information. The first step is a point-based Conditional Random Field (CRF) operating in a local neighbourhood. Based on these results segments are constructed and serve as input for the second CRF classifying the segments. More regional context information can be incorporated by considering the segments. Two different interaction potentials are trained for local and regional scales.

We evaluated the approach and obtained preliminary results which slightly improve a single, point-based CRF in overall accuracy. However, at the current state of the development, the full potential of the framework is not yet exploited. Two main aspects, which will be improved in future work, are the segmentation algorithm and the selection of features for the nodes and the interactions.

ACKNOWLEDGEMENTS

The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) (Cramer, 2010): http://www.ifp.uni-stuttgart.de/dgpf/DKEP-Allg.html. The authors also thank Lena Albert for the fruitful discussions about the graphic models.

References

Albert, L., Rottensteiner, F. and Heipke, C., 2014. Land Use Classification Using Conditional Random Fields for the Verification of Geospatial Databases. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. II-4, Suzhou, China, pp. 1–7.

Anguelov, D., Taskar, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G. and Ng, A., 2005. Discriminative Learning of Markov Random Fields for Segmentation of 3d Scan Data. In: Proceedings of the 2005 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Vol. 2, IEEE Computer Society, 20-26 June, San Diego, USA, pp. 169–176.

Boykov, Y. Y. and Jolly, M.-P., 2001. Interactive Graph Cuts for Optimal Boundary & Region Segmentation of Objects in nd Images. In: Proceedings of the Eighth IEEE International Conference on Computer Vision (ICCV), 2001, Vol. 1, IEEE, 7-14 July,Vancouver, Canada, pp. 105–112.

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

Chehata, N., Guo, L. and Mallet, C., 2009. Airborne Lidar Feature Selection for Urban Classification using Random Forests. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVIII, Part 3/W8, Paris, France, pp. 207–212.

Chen, C., Liaw, A. and Breiman, L., 2004. Using Random Forest to learn Imbalanced Data. Technical report, University of California, Berkeley.

Cramer, M., 2010. The DGPF-Test on Digital Airborne Camera Evaluation – Overview and Test Design. Photogrammetrie-Fernerkundung-Geoinformation 2010(2), pp. 73–82.

Frey, B. and MacKay, D., 1998. A Revolution: Belief Propagation in Graphs with Cycles. In: Advances in Neural Information Processing Systems 10, MIT Press, 1-6 Dec 1997, Denver, USA, pp. 479–485.

Gislason, P. O., Benediktsson, J. A. and Sveinsson, J. R., 2006. Random Forests for Land Cover Classification. Pattern Recognition Letters 27(4), pp. 294–300.

Gould, S., Rodgers, J., Cohen, D., Elidan, G. and Koller, D., 2008. Multi-class segmentation with relative location prior. International Journal of Computer Vision 80(3), pp. 300–316.

Hänsch, R., 2014. Generic object categorization in PolSAR images-and beyond. PhD thesis, Technische Universität Berlin, Germany. Deutsche Geodätische Kommission - Reihe C, no. C726.

He, X., Zemel, R. S. and Carreira-Perpiñán, M., 2004. Multiscale conditional random fields for image labeling. In: Computer vision and pattern recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE computer society conference on, Vol. 2, IEEE, pp. II–695.

Kraus, K. and Pfeifer, N., 1998. Determination of Terrain Models in Wooded Areas with Airborne Laser Scanner Data. ISPRS Journal of Photogrammetry and Remote Sensing 53(4), pp. 193–203.

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

Lim, E. and Suter, D., 2007. Conditional Random Field for 3d Point Clouds with Adaptive Data Reduction. In: International Conference on Cyberworlds, 24-26 Oct., Hannover, Germany, pp. 404–408.

Lim, E. 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.

Luo, C. and Sohn, G., 2014. Scene-layout compatible conditional random field for classifying terrestrial laser point clouds. In: ISPRS Annals of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 1, pp. 79–86.

Najafi, M., Namin, S. T., Salzmann, M. and Petersson, L., 2014. Non-associative higher-order markov networks for point cloud classification. In: Computer Vision–ECCV 2014, Springer, pp. 500–515.

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., Mallet, C., Rottensteiner, F. and Soergel, U., 2011. Conditional Random Fields for Urban Scene Classification with Full Waveform LiDAR Data. In: Photogrammetric Image Analysis (PIA), Lecture Notes in Computer Science, Vol. 6952, Springer, 5-7 Oct., Munich, Germany, pp. 233–244.

Rottensteiner, F., Sohn, G., Gerke, M., Wegner, J. D., Breitkopf, U. and Jung, J., 2014. Results of the isprs benchmark on urban object detection and 3d building reconstruction. ISPRS Journal of Photogrammetry and Remote Sensing 93, pp. 256 – 271.

Rusu, R. B., 2009. Semantic 3D Object Maps for Everyday Manipulation in Human Living Environments. PhD thesis, Computer Science department, Technische Universitaet Muenchen, Germany.

Rusu, R. B. and Cousins, S., 2011. 3D is here: Point Cloud Library (PCL). In: IEEE International Conference on Robotics and Automation (ICRA), Shanghai, China.

Rutzinger, M., Rottensteiner, F. and Pfeifer, N., 2009. A Comparison of Evaluation Techniques for Building Extraction from Airborne Laser Scanning. IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing 2(1), pp. 11–20.

Schindler, K., 2012. An Overview and Comparison of Smooth Labeling Methods for Land-Cover Classification. Transactions on Geoscience and Remote Sensing (TGRS) 50(11), pp. 4534–4545.

Shapovalov, R., Velizhev, A. and Barinova, O., 2010. Non-Associative Markov Networks for 3D Point Cloud Classification. In: Proceedings of the ISPRS Commission III Symposium - PCV 2010, International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38/Part A, ISPRS, Saint-Mandé, France, pp. 103–108.

Shapovalov, R., Vetrov, D. and Kohli, P., 2013. Spatial inference machines. In: IEEE Conference on Computer Vision and Pattern Recognition, 23-28 June, Portland, USA, pp. 1–8.

Xiong, X., Munoz, D., Bagnell, J. A. and Hebert, M., 2011. 3-D scene analysis via sequenced predictions over points and regions. In: Proceedings of IEEE International Conference on Robotics and Automation (ICRA11), 9-13 May, Shanghai, China, pp. 2609 – 2616.