Classification of Urban LiDAR data using Conditional Random Field and Random Forests

Joachim Niemeyer, Franz Rottensteiner, Uwe Soergel
Institute of Photogrammetry and GeoInformation
Leibniz Universität Hannover, Germany
{niemeyer, rottensteiner, soergel}@ipi.uni-hannover.de

***Abstract*— In this work we address the task of contextual classification of an airborne LiDAR point cloud. For that purpose, we integrate a Random Forest classifier into a Conditional Random Field (CRF) framework. A CRF has been shown to deliver good results discerning multiple classes. It is a flexible approach for obtaining a reliable classification even in complex urban scenes. The incorporation of multi-scale features improves the results further. Based on the classification results, 2D building and tree objects are generated and evaluated by the benchmark of ISPRS WG III/4.**

I. Introduction

The automated classification of airborne LiDAR data is a challenging task, especially in dense urban areas. In such a man-made environment different objects usually have certain relations, which can be utilized as context for the classification in order to distinguish multiple object classes more reliably. Conditional Random Fields (CRF) provide a flexible statistical classification framework which is capable of modeling context [7]. As a consequence, CRF have become increasingly popular in computer vision and remote sensing.

However, there is only a small amount of work dealing with LiDAR point clouds and CRF. Some initial applications of contextual classification are related to robotics and mobile laser scanning, dealing with terrestrial LiDAR point clouds. Anguelov et al. [2] performed a point-wise classification of terrestrial laser scans using a subclass of Markov Models, called Associative Markov Networks (AMNs), which encourage neighboring points to belong to the same object class. However, AMNs have a tendency to over-smooth the results. Thus, small objects often cannot be detected correctly.

Shapovalov et al. [11] investigated the potential of CRFs for the classification of airborne LiDAR point clouds. They improved the drawbacks of AMN by applying a non-associative Markov Network, which is able to model typical class relations such as 'a tree is likely to be above ground'. These interactions represent additional context knowledge and may improve classification results. The algorithm consists of two steps. Firstly, they over-segment the data and secondly a segment-wise CRF classification is performed. Whereas this aspect helps to cope with noise and computational complexity, the result heavily depends on the segmentation, and small objects with sub-segment size cannot be detected.

The authors of [11] use a Random Forest (RF) [3] classifier to model the association potential of the CRF. RF are able to handle a large amount of data and features, and they have shown good classification results for LiDAR point clouds [4]. Therefore, we incorporate a RF classifier into our CRF framework. This relaxes one of the restrictions in our previous work [9], where we used generalized linear models (GLM) for the potentials, and thus assumed that the classes can be separated by linear functions. Moreover, many weights had to be trained.

In contrast to [11], we do not perform a preliminary segmentation, but we classify each 3D point in order to enable the detection of small objects such as cars in the scene. As we want to utilize contextual information, all interactions are learned in a training step. For the 3D classification we discern six classes. Finally, building and tree objects are generated from the labeled 3D point cloud. We evaluate our approach on data of the ISPRS test project on urban classification and 3D building reconstruction [12].

II. Methodology

A. Conditional Random Fields

It is the goal of point cloud classification to assign an object class label to each 3D point. CRF provide a powerful probabilistic framework for contextual classification. They belong to the family of undirected graphical models. The underlying graph $G(n, e)$ consists of nodes n and edges e. In our case, each node $n_i \in n$ corresponds to a 3D point and we assign class labels y_i to all points simultaneously based on observed data $\mathbf{x}$. The vector $\mathbf{y}$ contains the labels y_i for all nodes, and hence has the same number of elements as n. The graph edges e are used to model the relations between pairs of adjacent nodes n_i and n_j, and thus enable modeling contextual relations. Therefore, each point n_i is linked to its *k* nearest neighbors ($n_j \in N_i$) in 2D by edges. In contrast to generative Markov Random Fields (MRF), CRF are discriminative classifiers that model the posterior distribution $P(\mathbf{y}|\mathbf{x})$ directly [7]:

$$P(\boldsymbol{y}|\mathbf{x}) = \frac{1}{Z(\boldsymbol{x})} \exp\left(\sum_{i \in n} A_i(\boldsymbol{x}, y_i) + \sum_{i \in n} \sum_{j \in N_i} I_{ij}(\boldsymbol{x}, y_i, y_j)\right). \quad (1)$$

In (1), N_i is the neighborhood of node n_i, corresponding to the edges linked to this particular node. The two terms in

the exponent are called the association potential $A_i(\mathbf{x}, y_i)$ and the interaction potential $I_{ij}(\mathbf{x}, y_i, y_j)$, respectively; they are described in the subsequent sections. The partition function $Z(\mathbf{x})$ acts as normalization constant, turning potentials into probabilities.

Inference is the task of determining the optimal label configuration based on maximizing $P(\mathbf{y}|\mathbf{x})$ for given parameters. We use the message passing algorithm Loopy Belief Propagation [10]. The result is a probability value per class for each node. Class labels are assigned based on the MAP method.

B. Association Potential

The association potential $A_i(\mathbf{x}, y_i)$ links the data to the class labels and determines the most probable label for a single node. In contrast to MRF, A_i may potentially depend on the entire dataset $\mathbf{x}$ instead of only the features observed at node n_i. We apply a RF classifier [3] for defining the potential A_i. In the training step, the number of votes per class in each leaf of the decision trees is determined and can be used to obtain a classification posterior for each sample. A_i is modeled to be proportional to the logarithm of this posterior.

There are some advantages of this classifier compared to a GLM we used in [9]. Firstly, a feature space mapping was applied to discern the object clusters by a linear function. This is not needed any more, because in RF the classes need not to be linearly separable. Secondly, RF are able to handle many more features as they automatically select the most important ones. Hence, a better performance is yielded because only a subset of features is used, whereas GLM have to train weights for each feature, a situation which is aggravated by the feature space mapping. This allows us to utilize multi-scale features. We adapted some of the LiDAR features proposed in [4] and used the following features for classification: 1) Intensity (*I*); 2) height above an approximated DTM; 3) approximated plane: sum, mean and standard deviation of residuals, direction and variance of normal vector; 4) variance of point elevations in a cylinder and a sphere; 5) ratio of point density in a cylinder and a sphere; 6) eigenvalue based features: 3 eigenvalues (λ_1,λ_2,λ_3), omnivariance (*O*), planarity (*P*), anisotropy (*A*), sphericity (*S*), eigenentropy (*E*), scatter (λ_1/λ_3); 7) variation of *I*, *O* , *P*, *A*, *S*, and point density.

For the features of groups 3-7 the local point distribution is considered within a sphere, and for groups 4+5 additionally within a cylinder. We computed all of them for multiple scales with radii $r = 1, 2$ and 3 m. This allows overcoming restrictions of local point features, which improves the classification. 71 features are determined in total for each point and represent the feature vector $\mathbf{h}_i(\mathbf{x})$ for node n_i.

We balance the data by sampling the same number of points for each class in training step in order to obtain an unbiased classification of RF [6]. The RF implementation for Matlab [1] is used, which considers the Gini-Index of the features for training the trees.

C. Interaction Potential

The term $I_{ij}(\mathbf{x}, y_i, y_j)$ in (1) represents the interaction potential and incorporates the contextual relations explicitly in the classification. 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 $\mathbf{x}$. An interaction feature vector $\boldsymbol{\mu}_{ij}(\mathbf{x})$ is computed for each edge in the graph based on the element-wise differences of both adjacent node feature vectors, $\boldsymbol{\mu}_{ij}(\mathbf{x}) = \mathbf{h}_i(\mathbf{x}) - \mathbf{h}_j(\mathbf{x})$. The interaction potential is modeled as the logarithm of the joint probability of two node labels y_i and y_j, $\log P(y_i, y_j|\boldsymbol{\mu}_{ij}(\mathbf{x}))$. Thus, a probability for each label relation has to be determined using a discriminative classifier. As for the association potential, we again apply RF for I_{ij}. The RF classifiers for A_i and I_{ij} are trained separately. In the case of the interaction potential, each pair of classes is considered as a single class by the RF, which results in 36 relations for six object classes. This information modeling the context is utilized to improve the quality of classification by supporting more probable class interactions given the data. The degree of smoothing the labels depends on the feature vector $\boldsymbol{\mu}_{ij}(\mathbf{x})$. For this reason, small objects are better preserved if there is sufficient evidence in the data, which is a major advantage in complex urban areas [8].

D. Object Generation

Since the result of the CRF classification is a labeled 3D point cloud, we project all points belonging to *building* and *vegetation*, respectively, to a 2D label image defined in object space in order to obtain binary object masks for both classes. Morphological opening is carried out to smooth the object boundaries in a post-processing step.

III. Experiments

The performance of our method is evaluated on the LiDAR dataset of Vaihingen, Germany [5], in the context of the ISPRS test project on urban classification and 3D building reconstruction [12]. The point density in the test areas is approximately 8 points/m^2. Multiple echoes and intensities were recorded. Data acquisition was under leaf-on conditions. Since we present a supervised classification approach, a training step is necessary in order to learn the RF. For this purpose, a fully labeled part of the point cloud with 105655 points is used.

Three test sites were considered for the benchmark. Area 1 is situated in the center of the city of Vaihingen. Dense, complex buildings and some trees characterize this test site. Area 2 consists of a few high-rising residential buildings surrounded by trees. In contrast, Area 3 is a purely residential area with small, detached houses.

A. 3D Classification

By applying the CRF-classification, a labeled 3D point cloud is obtained. Six classes, namely *natural ground* (*Nat.*), *asphalt* (*As.*), *building* (*Build.*), *vegetation* (*Veg.*), *fence* and *car*, are discerned for this work. The results are depicted in Fig. 1. All point labels are inferred simultaneously, so we get the most probable label configuration. Each point was linked to its 4 nearest neighbors by edges. The computation performance mainly depends on the number of trees which is evaluated in both RF classifiers, and on the depth of the trees *f*. We used

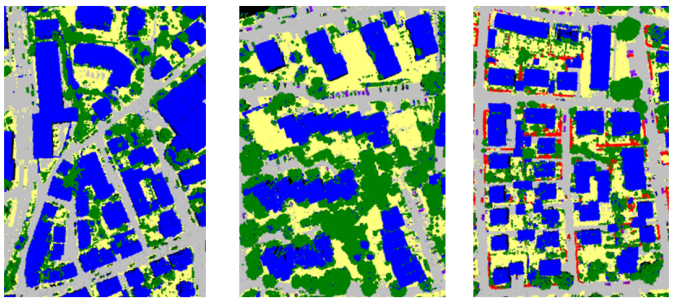

Fig. 1. 3D classification result of the three test areas. (gray: *asphalt*, yellow: *natural ground*, green: *vegetation*, blue: *building*, red: *fence*, purple: *car*)

500 trees and set f corresponding to the square root of the number of inputs ($n_f = 71$) to 9, as recommended in [1].

One important advantage of RF is the possibility to use a small subset of data for training, which allows a significant speed-up. However, the selected data must be sufficient to describe the classes in an appropriate way. Considering the expensive manual effort needed for generating a labeled reference point cloud for training, this aspect is important since only smaller subsets with discriminative features and the requested class relations are necessary. We used 3000 samples per class and per class relation, respectively. Hence the training data are balanced by sub- and oversampling to obtain an unbiased classification result.

In our previous approach utilizing a GLM classifier the weights for both potentials are optimized simultaneously in the training step [9]. For this reason, one fully labeled point cloud including all classes and class interactions is needed. Moreover, many parameters have to be learned, because each feature in $\mathbf{h}_i(\mathbf{x})$ and $\boldsymbol{\mu}_{ij}(\mathbf{x})$ is considered in the classification. For 10 features, six classes and a quadratic feature space mapping this results in 1890 weights, which have to be determined. Thus, the amount of the training data must be sufficient to learn such a large number of parameters.

We compare the RF classification to the GLM approach. However, due to other class definitions we trained and classified the test areas again, but used the same parameters as described in [9]. On a Pentium IV, 16 GB RAM computer training the RF CRF took only 22.3 min, whereas classification was faster with 1.8-2.8 min per area. Thus, training takes significantly less time compared to a linear model classifier. Here parameter learning based on a the entire training dataset took 7.5 hours and classification took 6 min per test area. In summary it can be said, that there are several advantages of RFs implemented in a CRF.

Fig. 1 shows the results for the three test areas obtained by the RF CRF classification. Visual inspection indicates that most objects are detected correctly. Compared to our previous work [9], especially the tree regions in Area 2 and 3 are much better classified. Due to flat point distributions on the canopies and low point densities with only few multiple returns, the canopies of larger trees used to be erroneously classified as *building*. Most of these errors are eliminated by exploiting multi-scale features. A 3D ground truth is available only for Area 3. In comparison to the results with GLM CRF the overall accuracy increases from 81.7 % to 83.7 %, and the kappa index from 0.76 to 0.79. As demonstrated in Table I, the results are comparable for both ground classes. Many confusion errors between *vegetation* and *building* are correct now, which shows much better correctness value for *vegetation* (+9.6 %) and completeness for *buildings* (+9.5 %). A significant improvement is in particular obtained for the object classes with fewer points such as *car* and *fence*, which benefit notably from the RF classifier. Here, completeness and correctness values for *car* improve by 38.6 % and 48.4 %, respectively, whereas the correctness of *fence* increase by 46.2 %. The RF CRF is faster, is able to handle more features, and yields mostly better results.

TABLE I

DIFFERENCES [Δ%] OF RF AND LINEAR MODELS (AREA 3)

Class	Nat.	As.	Build.	Veg.	Car	Fence
Completeness	3.1	-0.8	9.5	-0.4	38.6	-6.3
Correctness	-3.7	2.4	-3.1	9.6	48.4	46.2

TABLE II

EVALUATION RESULTS [%]: COMPLETENESS, CORRECTNESS, QUALITY

Building	object	object $\geq 50m^2$	per area
Area	A1 / A2 / A3	A1 / A2 / A3	A1 / A2 / A3
Compl.	83.8 / 85.7 / 85.7	100 / 100 / 100	89.9 / 88.2 / 92.5
Corr.	72.7 / 54.5 / 81.7	100 / 100 / 100	90.2 / 95.2 / 94.3
Qual.	63.8 / 50.0 / 71.9	100 / 100 / 100	81.9 / 84.4 / 87.6

Tree	object	object $\geq 25m^2$
Area	A1 / A2 / A3	A1 / A2 / A3
Compl.	50.5 / 74.1 / 57.4	42.3 / 95.2 / 77.4
Corr.	46.0 / 63.2 / 55.2	64.7 / 78.8 / 83.9
Qual.	31.7 / 51.7 / 39.2	34.4 / 75.8 / 67.4

B. Evaluation of 2D Objects

2D objects are generated by projecting the points assigned to *building* and *vegetation*, respectively, into a label image with 26 cm pixel size. In order to separate the trees (required for the benchmark) from low vegetation, which is also included in the *vegetation* class, only points with a height above DTM of more than 1.5 m are considered. Object boundaries are smoothed by a morphological closing with a 3x3 kernel. The resulting object masks are evaluated by the benchmark hosts based on reference data, and the results, shown in Table II and in Fig. 2, are determined according to [12].

Fig. 2 depicts the evaluation of the classification for buildings and trees on a per-area level. Yellow pixels are true positives (TP), red ones false positives (FP) and blue ones false negatives (FN). It can be seen that the buildings are detected reliably. This can also be seen in Table II with quality values between 81.9 % and 87.6 %. There are just a few FN which are caused by confusion with vegetation, in particular at building façades, and dormers. The larger FN and FP areas in Area 1

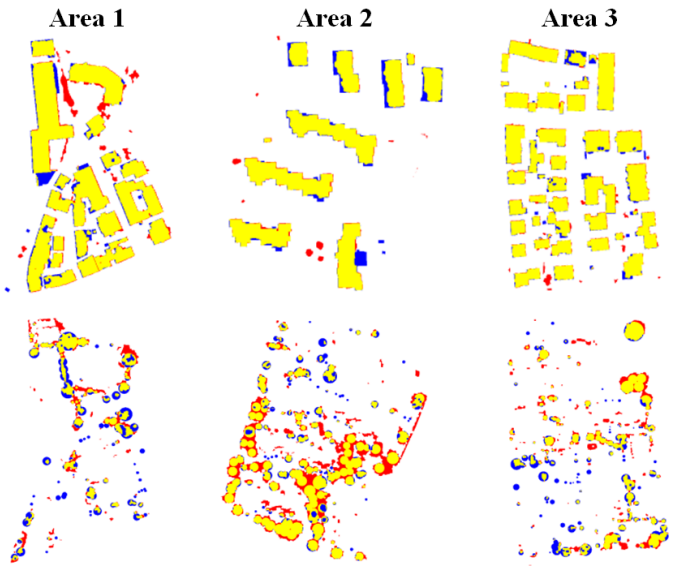

Fig. 2. Area-wise results of buildings (upper row) and trees (lower row) (yellow=TP, red=FP, blue=FN)

TABLE III
DIFFERENCES OF QUALITY [Δ%]: RF - LINEAR MODELS

	Building			Tree		
Area	A1	A2	A3	A1	A2	A3
per area	2.5	-1.8	-0.7	6.1	-2.2	2.4
per object	0.5	4.2	-3.5	10.8	-1.5	13.5
large objects	12.1	15.4	2.5	-4.9	-9.2	-6.8

are due to the feature 'height above DTM', which has by far the most important influence to the classification, as indicated by the RF. In this complex area the DTM could not describe the sloped terrain in an appropriate way. However, the per-object evaluation for buildings larger than 50 m^2 reveals that all objects were detected correctly.

The classification results of trees are a bit less accurate. Trees are approximated by 2D circles in the reference used for the evaluation. This would lead to errors in the per-area analysis [12]. Thus, we focus only on the evaluation of tree objects, which is more expressive compared to the area-based evaluation. Area 2 consists of many large trees, which were detected correctly in most cases. Smaller trees are undetected *e.g.* in Area 1. It can be seen that the simple model of selecting all vegetation points higher than 1.5 m is not sufficient for obtaining reliable tree objects. An improved tree model might solve this problem in future.

Table III shows the differences of the quality values obtained by the proposed approach and our previous work [9]. For positive values the RF method performed better. Especially large building objects $\geq$50 m^2 are detected more accurately; the quality increased by 2.5-15.4 %. The classification of trees mainly benefits on a per pixel and per object level. In particular more small trees in Area 1 and 3 are correctly detected with the new approach, whereas the larger trees are a bit less accurate.

IV. CONCLUSION AND OUTLOOK

We propose a classification method for LiDAR data that incorporates a Random Forest classifier in a Conditional Random Field framework. It could be shown that this approach has many advantages compared to the models previously used. They are faster and can handle more features. This allows the use of multi-scale features. The classes can be discerned more reliably; especially the number of confusions between building and larger trees is reduced in this way. This approach outperforms a CRF with linear models and features computed for only one scale by 2 % considering the overall accuracies (83.7 % compared to 81.7 %). Especially small objects such as cars and fences benefit of this classifier. On top of that, fewer training data are needed.

The evaluation of the 2D objects shows that the buildings are classified very well, whereas a more sophisticated model is needed for reliably detecting trees in the points assigned as *vegetation*. We will investigate this problem in our future work.

ACKNOWLEDGMENT

The Vaihingen data set was provided by the German Society for Photogrammetry, Remote Sensing and Geoinformation (DGPF) [5]. This work is funded by the German Research Foundation (DFG), grant SO935/3-1.

REFERENCES

[1] J. Abhishek, "Classification and Regression by randomForest-matlab", http://code.google.com/p/randomforest-matlab, 2009.
[2] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, A. Ng, "Discriminative Learning of Markov Random Fields for Segmentation of 3d Scan Data". In: Proceeding of CVPR, Vol. 2, IEEE Computer Society, 2005, pp. 169-176.
[3] L. Breiman, "Random Forests", Machine Learning, Vol. 45, 2001, pp. 5-32.
[4] N. Chehata, L. Guo, C. Mallet, "Airborne Lidar Feature Selection for Urban Classification using Random Forests", In: The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. 38(3), 2009, 207-212.
[5] M. Cramer, "The DGPF-Test on Digital Airborne Camera Evaluation - Overview and Test Design", Photogrammetrie-Fernerkundung-Geoinformation 2010(2), pp. 73-82.
[6] C. Elkan, "The Foundations of cost-sensitive Learning", In: International Joint Conference on Artificial Intelligence, Seattle, 2001, pp. 973-978.
[7] Kumar, S. Hebert, M., 2006. "Discriminative Random Fields", International Journal of Computer Vision, 68(2), pp. 179-201.
[8] J. Niemeyer, J. Wegner, C. Mallet, F. Rottensteiner, U. Soergel, "Conditional Random Fields for Urban Scene Classification with Full Waveform Lidar Data", In: Photogrammetric Image Analysis '11, Lecture Notes in Computer Science, Vol. 6952, Springer, 2011, pp. 233-244.
[9] J. Niemeyer, F. Rottensteiner, U. Soergel, "Conditional Random Fields for Lidar Point Cloud Classification in Complex Urban Areas", In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3, 2012, pp. 263-268.
[10] M. Schmidt, "A Matlab Toolbox for Probabilistic Undirected Graphical Models", http://www.di.ens.fr/~mschmidt/Software/UGM.html, 2012.
[11] R. Shapovalov, A. Velizhev, O. Barinova, "Non-Associative Markov Networks for 3D Point Cloud Classification", In: Proceedings of the ISPRS Comm. III symposium - PCV , 2010, Vol. 38, pp. 103-108.
[12] F. Rottensteiner, G. Sohn, J. Jung, M. Gerke, C. Baillard, S. Benitez, S, U. Breitkopf, "The ISPRS benchmark on urban object classification and 3D building reconstruction". In: ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences I-3, 2012, pp. 293-298.