

CONDITIONAL RANDOM FIELDS FOR LIDAR POINT CLOUD CLASSIFICATION IN COMPLEX URBAN AREAS

J. Niemeyer*, F. Rottensteiner, U. Soergel

Institute of Photogrammetry and GeoInformation, Leibniz Universität Hannover
D-30167 Hannover, Germany, (niemeyer, rottensteiner, soergel)@ipi.uni-hannover.de

Commission III/4

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

ABSTRACT:

In this paper, we investigate the potential of a Conditional Random Field (CRF) approach for the classification of an airborne LiDAR (Light Detection And Ranging) point cloud. This method enables the incorporation of contextual information and learning of specific relations of object classes within a training step. Thus, it is a powerful approach for obtaining reliable results even in complex urban scenes. Geometrical features as well as an intensity value are used to distinguish the five object classes *building*, *low vegetation*, *tree*, *natural ground*, and *asphalt ground*. The performance of our method is evaluated on the dataset of Vaihingen, Germany, in the context of the 'ISPRS Test Project on Urban Classification and 3D Building Reconstruction'. Therefore, the results of the 3D classification were submitted as a 2D binary label image for a subset of two classes, namely *building* and *tree*.

1 INTRODUCTION

The automated detection of urban objects from remote sensing data has become an important topic in research in photogrammetry and remote sensing. The results are necessary to derive object information for applications such as the generation of 3D city models for visualization and simulations. Data acquired by an airborne LiDAR sensor are especially suitable for this task because they directly describe the object surface in 3D. The core task resolved in this context is the classification of such a point cloud. However, in dense urban areas reliable point labeling is a challenging task due to complex object structure and the high variability of object classes (e.g. buildings, roads, trees, and low vegetation) appearing in this kind of scene. In such a man-made environment different objects usually have a specific kind of relations, which can be utilized in a contextual classification to distinguish multiple object classes simultaneously. Thus, the quality of classification can be improved by incorporating context information compared to approaches only classifying each point independently from its neighbors.

A Conditional Random Field (CRF) is a generalization of a Markov Random Field (MRF) and provides a flexible statistical classification framework for modeling local context. Originally introduced by Lafferty et al. (2001) for labeling one-dimensional sequence data, it has become increasingly popular in computer vision since the approach was adapted to the classification of images by Kumar and Hebert (2006). Some investigations were carried out in the field of remote sensing. For instance, Wegner (2011) analyzes the potential of CRF for building detection in optical airborne images. However, there is only a small amount of work dealing with LiDAR point clouds.

Some initial applications of contextual classification can be found related to robotics and mobile laser scanning utilizing terrestrial LiDAR point clouds. Angelov et al. (2005) performed a point-wise classification on terrestrial laser scans using a simplified subclass of Markov Models, called Associative Markov Networks (AMNs) (Taskar et al., 2004), which encourage neighboring points to belong to the same object class. However, AMNs have some

drawbacks: Small objects cannot be detected correctly due to over-smoothing. In the approach presented by Lim and Suter (2009) a terrestrial point cloud acquired by mobile mapping is processed in two steps. Firstly, they over-segment the data and secondly they perform a segment-wise CRF classification. Where as 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. Shapovalov et al. (2010) investigated the potential of CRFs for the classification of airborne LiDAR point clouds. They improved the drawbacks of an AMN by applying a non-associative Markov Network, which is able to model typical class relations such as 'tree is likely to be above ground'. These interactions represent additional context knowledge and may improve classification results. Similar to the approach described before, the authors applied a segment-based classification; hence, it is limited due to its dependence on the segmentation.

In our previous work (Niemeyer et al., 2011) we introduced a point-wise CRF classification of airborne full-waveform LiDAR data. This approach requires higher computational costs, but enables the correct detection of smaller objects. Each point is linked to points in its geometrical vicinity, which influence each other in the classification. All labels of the entire point cloud are determined simultaneously. The three object classes *ground*, *building*, and *vegetation* were distinguished. However, we applied a simple model for the interactions, penalizing the occurrence of different class labels of neighboring points, if the data did not indicate a change. Moreover, the decision surface in feature space was modeled by a linear function, which may not separate the clusters in a good way.

These aspects are improved in this work. We present a CRF, which is able to learn and model interactions of all object classes and utilize a non-linear decision surface to separate the object clusters in feature space reliably. As CRFs are probabilistic frameworks, each 3D point of a LiDAR point cloud is assigned to the most probable object label given the data. Furthermore, we improve the model for interactions to cope with specific configurations in feature space for neighboring classes. No preliminary segmentation is performed to enable detection of even small objects. In order to obtain objects from the classified point cloud,

*Corresponding author

a binary label image is generated by rasterization for each class of interest. The boundaries of objects are smoothed by applying a morphological filter afterwards. Our results are evaluated in the context of the 'ISPRS Test Project on Urban Classification and 3D Building Reconstruction' hosted by ISPRS WG III/4, a benchmark comparing various classification methods based on aerial images and/or LiDAR data of the same test site.

The next section presents our methodology including a brief description of the CRF. After that, Section 3 comprises the evaluation of our approach as well as discussions. The paper concludes with Section 4.

2 METHODOLOGY

2.1 Conditional Random Fields

It is the goal of point cloud classification to assign an object class label to each 3D point. Conditional Random Fields (CRFs) provide a powerful probabilistic framework for contextual classification and belong to the family of undirected graphical models. Thus, data are represented by a graph $G(\mathbf{n}, \mathbf{e})$ consisting of nodes \mathbf{n} and edges \mathbf{e} . For usual classification tasks, spatial units such as pixels in images or points of point clouds can be modeled as nodes. The edges \mathbf{e} link pairs of adjacent nodes n_i and n_j , and thus enable modeling contextual relations. In our case, each node $n_i \in \mathbf{n}$ corresponds to a 3D point and we assign an object class label y_i based on observed data \mathbf{x} in a simultaneous process. The vector \mathbf{y} contains the labels y_i for all nodes and hence has the same number of elements like \mathbf{n} .

In contrast to generative Markov Random Fields (MRFs), CRFs are discriminative and model the posterior distribution $P(\mathbf{y}|\mathbf{x})$ directly. This leads to a reduced model complexity compared to MRFs (Kumar and Hebert, 2006). Moreover, MRFs are restricted by the assumption of conditionally independent features of different nodes. Theoretically, only links to nodes in the immediate vicinity are allowed in this case. These constraints are relaxed for CRFs, consequently more expressive contextual knowledge represented by long range interactions can be considered in this frameworks. Thus, CRFs are a generalization of MRFs and model the posterior by

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

In Eq. 1, \mathbf{N}_i is the neighborhood of node n_i represented by the edges linked to this particular node. Two potential terms, namely the *association potential* $A_i(\mathbf{x}, y_i)$ computed for each node and the *interaction potential* $I_{ij}(\mathbf{x}, y_i, y_j)$ for each edge are modeled in the exponent; they are described in Section 2.4 in more detail. The partition function $Z(\mathbf{x})$ acts as normalization constant turning potentials into probability values.

2.2 Graphical Model

In our framework, each graph node corresponds to a single LiDAR point, which is assigned to an object label in the classification process. Consequently, the number of nodes in \mathbf{n} is identical to the number of points in the point cloud. The graph edges are used to model the relations between nodes. In our case, they link to the local neighbor points in object space. A nearest neighbor search is performed for each point n_i to detect adjacent nodes $n_j \in \mathbf{N}_i$. Each point is linked to all its neighbors inside a vertical cylinder of a predefined radius r . Using a cylindrical neighborhood enables to model important height differences, which

are not captured in a spherical vicinity (Shapovalov et al., 2010; Niemeyer et al., 2011). The cylinder radius r determines the size of \mathbf{N}_i and thus should be selected related to the point density; a tradeoff between number of edges and computational effort must be chosen (Section 3.1).

2.3 Features

In this research, only LiDAR features are utilized for the classification of the point cloud. A large amount of LiDAR features is presented in Chehata et al. (2009). They analyzed the relative importance of several features for classifying urban areas by a random forest classifier. Based on this research, we defined 13 features for each point and partially adapted them to our specific problem.

Apart from the *intensity value*, all features describe the local geometry of point distribution. An important feature is *distance to ground* (Chehata et al., 2009; Mallet, 2010), which models each point's elevation above a generated or approximated terrain model. In the original work, this value is approximated by using the difference of the trigger point and the lowest point elevation value within a large cylinder. The advantage of this approximation is that no terrain model is necessary, but this assumption is only valid for a flat terrain and not for hilly ground. We adapted the feature, but computed a terrain model (DTM) by applying the filtering algorithm proposed in Niemeyer et al. (2010). The additional features consider a local neighborhood. On the one hand, a spherical neighborhood search captures all points within a 3D-vicinity; on the other hand, points within a vertical cylinder centered at the trigger point model the point distribution in a 2D environment. By utilizing the *point density ratio* ($n_{\text{pts,sphere}}/n_{\text{pts,cylinder}}$), a useful feature indicating vegetation can be computed. Points within a sphere are used to estimate a local robust 3D plane. The *sum of the residuals* is a measurement of the local scattering of the points and serves as feature in our experiment. The *variance of the point elevations* within the current spherical neighborhood indicates whether significant height differences exist, which is typical for vegetated areas and buildings. The remaining features are based on the covariance matrix set up for each point and its three eigenvalues λ_1 , λ_2 , and λ_3 . The local point distribution is described by eigenvalue-based features such as *linearity*, *planarity*, *anisotropy*, *sphericity*, and *sum of the eigenvalues*. Equations are provided in Table 1.

Eigenvalues	$\lambda_1 \geq \lambda_2 \geq \lambda_3$
Planarity	$(\lambda_1 - \lambda_2)/\lambda_1$
Anisotropy	$(\lambda_2 - \lambda_3)/\lambda_1$
Sphericity	$(\lambda_1 - \lambda_3)/\lambda_1$
Sum of eigenvalues	$\sum_{i=1}^3 \lambda_i$

Table 1: Eigenvalue-based features

After computation, the feature values are not homogeneously scaled due to the different physical interpretation and units of these features. In order to align the ranges, a normalization of each feature for all points of a given point cloud is performed by subtracting the mean value and dividing by its standard deviation. Many features depend on the eigenvalues, which leads to a high correlation. Thus, a principal component analysis (PCA) is performed. Nine of the obtained uncorrelated 13 features (maintaining 95 % of information) are then used for CRF-based classification.

In a CRF, each node as well as each edge is associated to a feature vector. Since each 3D LiDAR point represents a single node n_i , we use the 9 features described above to define that node's feature vector $\mathbf{h}_i(\mathbf{x})$. Edges model the relationship between each point n_i and $n_j \in \mathbf{N}_i$. We use a standard approach of an element-wise,

absolute difference between the feature vectors of both adjacent nodes to compute the interaction feature vector

$$\mu_{ij}(\mathbf{x}) = |\mathbf{h}_i(\mathbf{x}) - \mathbf{h}_j(\mathbf{x})|. \quad (2)$$

2.4 Potentials

The exponent of Eq. 1 consists of two terms modeling different potentials. $A_i(\mathbf{x}, y_i)$ links the data to the class labels and hence it is called *association potential*. Note that A_i may potentially depend on the entire dataset \mathbf{x} instead of only the features observed at node n_i in contrast to MRFs. Thus, global context can be integrated into the classification process. $I_{ij}(\mathbf{x}, y_i, y_j)$ represents the *interaction potential*. It models the dependencies of a node n_i on its adjacent nodes n_j by comparing both node labels and considering all the observed data \mathbf{x} . This potential acts as data-dependent smoothing term and enables to model the incorporation of contextual relations explicitly. For both, A_i and I_{ij} , arbitrary classifiers can be used. We utilize a generalized linear model (GLM) for all potentials. A feature space mapping based on a quadratic feature expansion for node feature vectors $\mathbf{h}_i(\mathbf{x})$ as well as the interaction feature vectors $\mu_{ij}(\mathbf{x})$ is performed as described by Kumar and Hebert (2006) in order to introduce a more accurate non-linear decision surface for classes within the feature space. This is indicated by the feature space mapping function Φ and improves the classification compared to our previous work (Niemeyer et al., 2011), in which linear surfaces were used for discriminating object classes. Each vector $\Phi(\mathbf{h}_i(\mathbf{x}))$ and $\Phi(\mu_{ij}(\mathbf{x}))$ consists of 55 elements after expansion in this way.

The association potential A_i determines the most probable label for a single node and is given by

$$A_i(\mathbf{x}, y_i = l) = \mathbf{w}_l^T \cdot \Phi(\mathbf{h}_i(\mathbf{x})), \quad (3)$$

where $\mathbf{h}_i(\mathbf{x})$ is a feature vector of node n_i computed from data \mathbf{x} and \mathbf{w}_l is a vector of feature weights for a certain class l . In the training step there is one such weight vector learned per class. The interaction potential I_{ij} models the influence of each point's neighborhood to the classification process. Therefore, a feature vector $\mu_{ij}(\mathbf{x})$ is computed for each edge in the graph and multiplied to an edge weight vector $\mathbf{v}_{l,k}$, which depends on the label configuration of both adjacent nodes n_i and n_j :

$$I_{ij}(\mathbf{x}, y_i = l, y_j = k) = \mathbf{v}_{l,k}^T \cdot \Phi(\mu_{ij}(\mathbf{x})). \quad (4)$$

There is one weight vector $\mathbf{v}_{l,k}$ for each combination of classes l and k , which are also determined in the training stage. In our previous work (Niemeyer et al., 2011), different class labels l and k were penalized, because $I_{ij}(\mathbf{x}, y_i, y_j)$ was modeled to be proportional to the probability $P(y_i = y_j | \mu_{ij}(\mathbf{x}))$. For the current research, this potential is enhanced and is proportional to the joint probability $P((y_i, y_j) | \Phi(\mu_{ij}(\mathbf{x})))$. Thus, there is a probability for each label configuration. Trained in a learning step, certain class relations are modeled to be more probable than others are. For instance, a point assigned to class *building* is not likely to be surrounded by points belonging to *water*. This advanced information is used to improve the quality of classification. Moreover, the degree of smoothing depends on the feature vector $\mu_{ij}(\mathbf{x})$. 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 (Niemeyer et al., 2011).

2.5 Inference and Training

Exact inference and training is intractable in our application due to the fact that our graph $G(\mathbf{n}, \mathbf{e})$ is large and contains cycles. Thus, approximate methods have to be applied in both cases.

Loopy Belief Propagation (LBP) (Frey and MacKay, 1998) is a standard iterative message passing algorithm for graphs with cycles. We use it for inference, which corresponds to the task of determining the optimal label configuration based on maximizing $P(\mathbf{y}|\mathbf{x})$ for given parameters. The result is a probability value per class for each node. The highest probability is selected according to the maximum a posteriori (MAP) criterion; the corresponding object class label obtaining the maximum probability is assigned to the point afterwards. For the classification, the two vectors \mathbf{w}_l and $\mathbf{v}_{l,k}$ are required, which contain the feature weights for association potential and interaction potential respectively. They are determined in a training step by utilizing a fully labeled part of a point cloud. All weights of \mathbf{w}_l and $\mathbf{v}_{l,k}$ are concatenated in a single parameter vector $\theta = [\mathbf{w}_1, \dots, \mathbf{w}_c; \mathbf{v}_{1,1}, \dots, \mathbf{v}_{c,c}]^T$ with c classes, in order to ease notation. For obtaining the best discrimination of classes, a non-linear numerical optimization is performed by minimizing a certain cost function. Following Vishwanathan et al. (2006), we use the limited-memory Broyden-Fletcher-Goldfarb-Shanno (L-BFGS) method for optimization of the objective function $f = -\log P(\theta|\mathbf{x}, \mathbf{y})$. For each iteration f , its gradient as well as an estimation of the partition function $Z(\mathbf{x})$ is required. Thus, we use L-BFGS combined with LBP as implemented in M. Schmidt's Matlab Toolbox (Schmidt, 2011).

2.6 Object generation

Since the result of the CRF classification is a 3D point cloud, we performed a conversion to raster space in order to obtain 2D objects for each class of interest. If any point was assigned to the trigger class, the certain pixel is labeled by value one and zero otherwise. A morphological opening is carried out to smooth the object boundaries in a post-processing step.

3 EVALUATION

3.1 Benchmark-Dataset

The performance of our method is evaluated on the LiDAR dataset of Vaihingen, Germany (Cramer, 2010), in the context of the 'ISPRS Test Project on Urban Classification and 3D Building Reconstruction'. The dataset 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 %.



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., 2011)

For the benchmark, three test sites with different scenes were considered (Fig. 1). 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. The average point density in the test areas is approximately 8 points/m². Multiple echoes and intensities were recorded. Since we present a supervised classification approach, a training step is necessary in order

to learn the parameter weights (Section 2.5). For our experiments, a fully and manually labeled part of the point cloud with 93,504 points in the east of Area 1 was used.

According to the point density of this scene and based on the knowledge obtained in Niemeyer et al. (2011), we set the parameter radius r for graph construction to 0.75 m for the following experiments, which results in 5 edges on average for each point.

3.2 Result CRF Classification

The original result of CRF classification is the assignment of an object class label to each LiDAR point. Hence, we obtain a labeled 3D point cloud. Five classes were trained for this research: *natural ground*, *asphalt ground*, *building*, *low vegetation*, and *trees*. The results are depicted in Fig. 2.

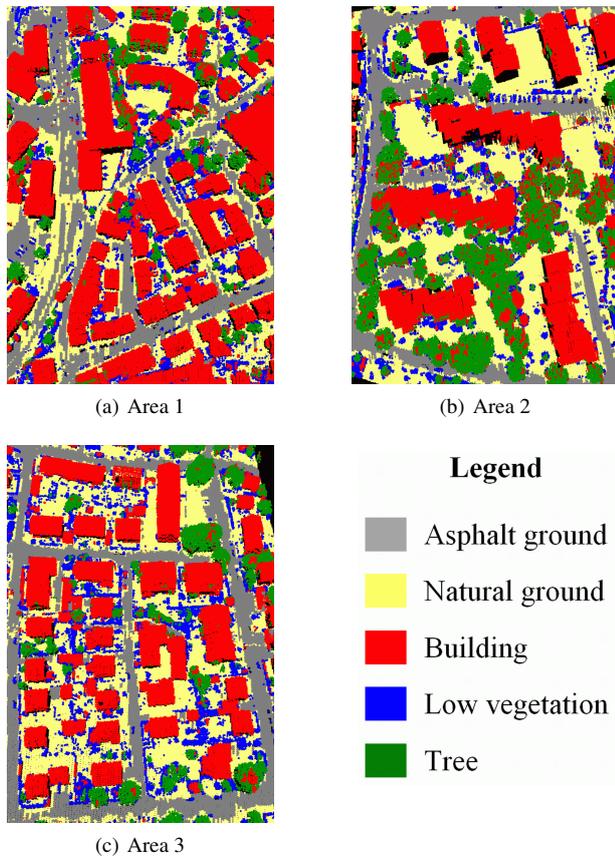


Figure 2: 3D CRF-Classification results.

In general, our approach achieved good classification results. Visual inspection reveals that the most obvious error source are points on trees incorrectly classified as *building*. This is because most of the 3D points are located on the canopy and show similar features to building roofs: the dataset was acquired in August under leaf-on conditions, that is why almost the whole laser energy was reflected from the canopy, whereas second and more pulses within trees were recorded only very rarely (about 2%). Most of our features consider the point distribution within a local neighborhood. However, for tall trees with large diameters the points on the canopy seem to be flat and horizontally located some meters above the ground, which usually hints for points belonging to building roofs. That is the reason for misclassification of some tree points to *buildings*.

Due to missing complete 3D reference data, a quantitative evaluation of the entire point classification result cannot be performed.

Thus, we carry out an object-based evaluation in 2D as described in the Section 3.4.

3.3 Improvements of interactions & feature space mapping

In this research, two improvements for the CRF-framework are implemented: Firstly, the feature space mapping enables a non-linear decision surface in classification. Secondly, weights for all object class relations are learned. In order to evaluate the impact of these changes compared to the approach proposed in Niemeyer et al. (2011), we manually labeled the point cloud of Area 3 and carried out a point-wise validation based on this reference data. Therefore, the results are compared to a classification penalizing different classes in the interaction model and without feature space mapping.

The enhanced CRF increases the overall accuracy by 5.38 percentages and hence yields considerably more sophisticated results. Table 2 shows the differences of completeness and correctness values in percentages obtained by the enhanced CRF classification and the simpler model. Positive values indicate a higher accuracy of results obtained with the complex edge potential and feature space mapping. It can easily be seen, that all values are improved by introducing the complex method. Especially details such as small buildings or hedges are preserved (Fig. 3). To sum up, feature space mapping and modeling of all specific class relations considerably improve the quality. Thus, it is worth the effort of learning more parameters in training.

Classes	Completeness	Correctness
asphalt ground	6.04	6.55
natural ground	7.57	3.68
building	0.60	6.89
low vegetation	10.59	7.90
tree	4.81	1.39

Table 2: Comparison of enhanced CRF results and simpler model (differences of completeness and correctness values).

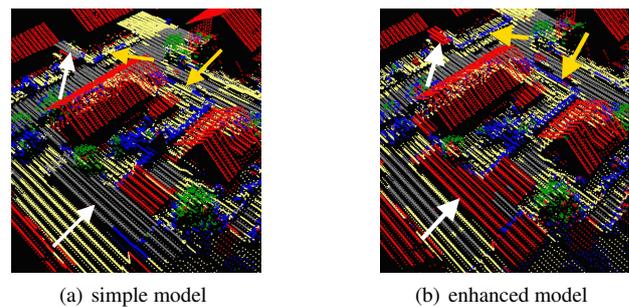


Figure 3: Comparison of the results obtained by a simple and an enhanced CRF model. For instance, garages (white arrows) and hedges (yellow arrows) are better detected with the complex model.

3.4 Evaluation of 2D objects

For the benchmark, a 2D binary label image was required in order to compare the results. Since our result is a classified 3D point cloud, we performed a conversion to raster space with 26 cm pixel size. Using a relatively small pixel size compared to the point density of about 8 points/m² ensures the geometrical correctness of the objects being preserved in binarization step. Label images are generated only for classes *building* and *tree* because the other classes are not in the focus of the benchmark. Thus, we obtained two images for each of the three test sites, which were post-processed by a morphological opening with a 5x5-circle kernel in

order to smooth the fringed boundaries of the objects. These label images are then evaluated in the context of the ISPRS Test Project with the method described in Rutzinger et al. (2009) based on a reference for each class. The results are discussed in the following part.

3.4.1 Building: The first object class being analyzed is *building*. Completeness and correctness values per-area are between 87.0 % and 93.8 %. The quality, which is defined as

$$\text{Quality} = 1 / (1 / \text{Completeness} + 1 / \text{Correctness} - 1), \quad (5)$$

reaches values from 79.4 % to 88.3 % (Table 3). It can be seen, that the quality of Area 1 is about 7 percentages below those obtained for the other two test sites. The corresponding classification results are depicted in Fig. 4. Yellow highlighted areas represent the true positives (TP), blue are false negative (FN) and red are false positive detections (FP). It becomes evident, that the majority of the building shapes were detected correctly. Thus, the approach works well for buildings.

	Area 1	Area 2	Area 3
Completeness	87.0%	93.8%	93.8%
Correctness	90.1%	91.4%	93.7%
Quality	79.4%	86.2%	88.3%

Table 3: Per-area accuracies of class *building*

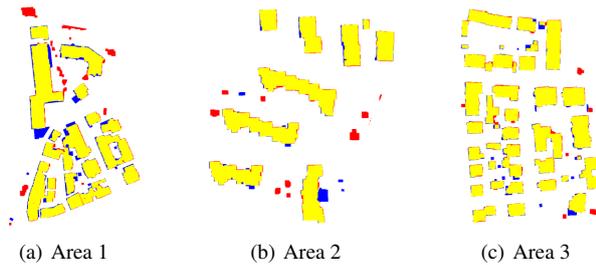


Figure 4: Pixel-wise result of class *buildings* (yellow = TP, red = FP, blue = FN)

Most of the FPs are caused by tree areas wrongly misclassified as *building*. This is due to the similar features, as mentioned before: the LiDAR points covering trees are mainly distributed on the canopy and not within the trees, which leads to a nearly horizontal and planar point distribution.

A very important feature for discriminating several classes such as *ground* or *building roof* is 'height above ground' (Mallet, 2010). By utilizing a DTM for generating this feature in our application, *buildings* and *ground* can be distinguished much better because of hilly ground. However, there are still some larger FN areas, e.g. in the center part of Area 1. This is due to the filtering process of the terrain model generation for this feature, because some buildings and garages situated next to other ground height levels could not be eliminated correctly. Thus, the height differences to ground are very small for these points and lead to a misclassification with object class *ground*.

The evaluation on a per-object level for buildings in the three test sites is depicted in Fig. 5. It is a cumulative histogram showing the quality numbers for all buildings larger than the area corresponding to the respective bin. Here, completeness, correctness and quality are computed separately for all area intervals of the abscissa. As shown in the figure, the accuracies of our method increase with larger object size. All buildings of Area 2 and 3 with an area larger than 87.5 m² are detected correctly, whereas the quality in Area 1 reaches 100 % at objects larger

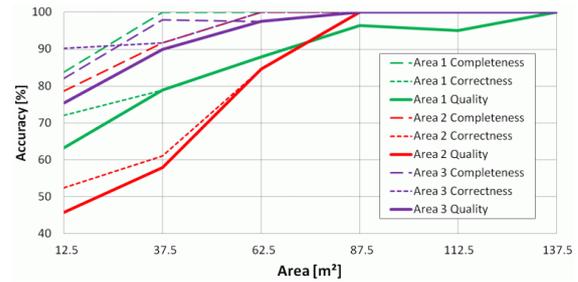


Figure 5: Evaluation of buildings on a per-object level as a function of the object area (cumulative)

than 137.5 m². The above-mentioned misclassification of small regions with class *tree* causes low correctness (and thus quality) values especially for Area 2.

3.4.2 Trees: The results of class *tree* vary considerably. Correctness values are between 69.2 % and 77.0 %, whereas completeness underperforms this with 41.4 % to 74.0 %. This finds expression in only reasonable quality values (Table 4). Especially in Area 1 low completeness values are obtained. As shown in Fig. 6, there are many undetected trees represented as FNs. These are mainly small trees, which were misclassified as *low vegetation* due to the indistinct class definition of *tree* and *low vegetation* in the benchmark, which do not coincide in this case. In general, larger trees could be detected more reliably (Fig. 7). However, especially in Area 1 the cumulative histogram decreases for trees with a larger area due to misclassifications as *building*.

	Area 1	Area 2	Area 3
Completeness	41.4%	74.0%	55.9%
Correctness	69.2%	73.1%	77.0%
Quality	35.0%	58.2%	47.9%

Table 4: Per-area accuracies of class *trees*

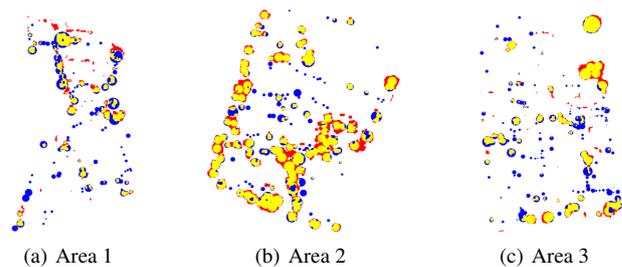


Figure 6: Pixel-wise result of class *tree*

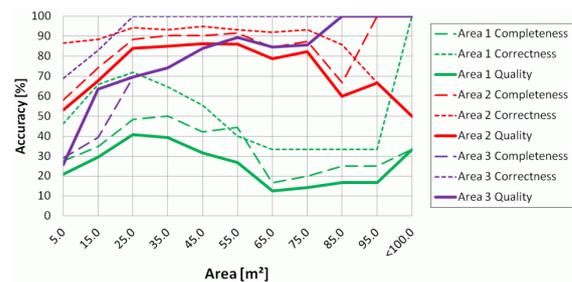


Figure 7: Evaluation of trees on a per-object level as a function of the object area (cumulative)

4 CONCLUSION AND OUTLOOK

In this work, we presented a context-based Conditional Random Field (CRF) classifier applied to complex urban area LiDAR scenes. Representative relations between object classes are learned in a training step and utilized in the classification process. Evaluation is performed in the context of the 'ISPRS Test Project on Urban Classification and 3D Building Reconstruction' hosted by ISPRS WG III/4 based on three test sites with different settings. The result of our classification is a labeled 3D point cloud; each point assigned to one of the object classes *natural ground*, *asphalt ground*, *building*, *low vegetation* or *tree*. A test has shown that feature space mapping and a complex modeling of specific class relations considerably improve the quality. After obtaining a binary grid image with morphological filtering for classes *building* and *tree*, which were taken into account in the test, the evaluation was carried out. In this case, the best results were obtained for *buildings*, which could be detected very reliably. The classification results of *tree* vary considerably in quality depending on the three test sites. Especially in the inner city scene many misclassifications occur, whereas the scenes consisting of high-rising buildings and residential areas could be classified reliably. Some of the small trees are classified as *low vegetation* and hence are not considered in the benchmark evaluation leading to attenuated completeness values. In summary, it can be stated that CRFs provide a high potential for urban scene classification.

In future work we want to integrate new features to obtain a robust classification even in scenes with low point densities in vegetated areas in order to allow a robust classification result and decrease the number of confusion errors with buildings. Moreover, multi-scale features taking larger parts of the environment into account might improve results. Finally, the classified points are grouped to objects in a simple way by 2D rasterization. They should be combined to 3D objects in the next steps by applying a hierarchical CRF.

ACKNOWLEDGMENTS

The authors would like to thank Clément Mallet for computing the point features used for classification. This work is funded by the German Research Foundation (DFG).

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> (in German).

References

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 Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), Vol. 2, IEEE Computer Society, pp. 169–176.

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.

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, pp. 479–485.

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

Lafferty, J., McCallum, A. and Pereira, F., 2001. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. In: *Proceedings of the 18th International Conference on Machine Learning*, pp. 282–289.

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.

Mallet, C., 2010. Analysis of Full-Waveform Lidar Data for Urban Area Mapping. PhD thesis, Télécom ParisTech.

Niemeyer, J., Rottensteiner, F., Kuehn, F. and Soergel, U., 2010. Extraktion geologisch relevanter Strukturen auf Rügen in Laserscanner-Daten. In: *Proceedings 3-Ländertagung 2010, DGPF Tagungsband, Vol. 19, Vienna, Austria*, pp. 298–307. (in German).

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 Berlin / Heidelberg*, pp. 233–244.

Rottensteiner, F., Baillard, C., Sohn, G. and Gerke, M., 2011. ISPRS Test Project on Urban Classification and 3D Building Reconstruction. <http://www.commission3.isprs.org/wg4/>.

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.

Schmidt, M., 2011. UGM: A Matlab toolbox for probabilistic undirected graphical models. <http://www.di.ens.fr/~mschmidt/Software/code.html> (Nov. 15, 2011).

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.

Taskar, B., Chatalbashev, V. and Koller, D., 2004. Learning associative markov networks. In: *Proceedings of the 21st International Conference on Machine Learning, ACM*, p. 102.

Vishwanathan, S., Schraudolph, N., Schmidt, M. and Murphy, K., 2006. Accelerated training of conditional random fields with stochastic gradient methods. In: *Proceedings of the 23rd International Conference on Machine Learning, ACM*, pp. 969–976.

Wegner, J., 2011. Detection and height estimation of buildings based on fusion of one SAR acquisition and an optical image. PhD thesis, Institute of Photogrammetry and GeoInformation, Leibniz University Hannover. In: *Wissenschaftliche Arbeiten der Fachrichtung Geodäsie und Geoinformatik der Leibniz Universität Hannover, Nr. 295, ISSN 0174-1454*.