

# COMPARISON OF BELIEF PROPAGATION AND GRAPH-CUT APPROACHES FOR CONTEXTUAL CLASSIFICATION OF 3D LIDAR POINT CLOUD DATA

*L. Landrieu, C. Mallet*

Univ. Paris-Est, LASTIG MATIS,  
IGN, ENSG,  
73 avenue de Paris, 94160 Saint-Mandé, France

*M. Weinmann*

Karlsruhe Institute of Technology (KIT),  
Institute of Photogrammetry and Remote Sensing,  
Englerstraße 7, 76131 Karlsruhe, Germany

## ABSTRACT

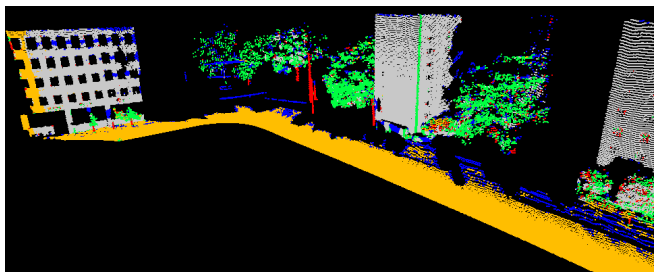
In this paper, we focus on the classification of lidar point cloud data acquired via mobile laser scanning, whereby the classification relies on a context model based on a Conditional Random Field (CRF). We present two approximate inference algorithms based on belief propagation, as well as a graph-cut-based approach not yet applied in this context. To demonstrate the performance of our approach, we present the classification results derived for a standard benchmark dataset. These results clearly indicate that the graph-cut-based method is able to retrieve a labeling of higher likelihood in only a fraction of the time needed for the other approaches. The higher likelihood, in turn, translates into a significant gain in the accuracy of the obtained classification.

**Index Terms**— Point cloud, classification, spatial regularization, belief propagation, graph-cut

## 1. INTRODUCTION

The automated analysis of acquired lidar point cloud data is a crucial task in photogrammetry, remote sensing, computer vision and robotics. In particular, the semantic classification of lidar point clouds acquired with terrestrial or mobile laser scanning systems in urban areas has received significant attention in recent years [1, 2, 3]. The main objective of this task is to associate a semantic label (e.g. *Building*, *Ground*, *Tree Trunk* or *Foliage*) to each 3D point of the considered point cloud and the main challenges are typically represented by the irregular distribution of 3D points, a high complexity of the observed scene and very different objects of interest (Fig. 1).

As the number of 3D points far exceeds the number of semantic objects in the scene, the semantic point cloud classification can be expected to display a high spatial regularity. Furthermore, transitions between different classes are not equally likely, e.g. *Tree Trunk* to *Ground* is more probable than *Foliage* to *Building*. Graphical probabilistic models are widely used in the literature to model the influence of these observations [4, 5]. In [4], for instance, a context model based on a Conditional Random Field (CRF) [6] is used, which is



**Fig. 1.** Labeled point cloud (*Wire*: blue; *Pole/Trunk*: red; *Façade*: gray; *Ground*: brown; *Vegetation*: green).

then solved with a belief propagation algorithm. In this paper, we argue that graph-cut-based approaches allow us to achieve a better approximation of the most likely configuration in a much shorter time.

## 2. RELATED WORK

The task of semantic point cloud classification typically relies on the use of geometric features like the ones presented in [2, 3], but some investigations additionally involve complementary types of features such as echo-based features or full-waveform features [7]. These features, in turn, are provided as input to a classification pipeline. For the latter, standard approaches for supervised classification such as a Support Vector Machine (SVM) classifier or a Random Forest (RF) classifier are typically applied. A comprehensive study focusing on the classification of mobile laser scanning data with a variety of such standard approaches relying on different learning principles can be found in [2] and concludes that a RF classifier provides a good trade-off between classification accuracy and computational effort. However, such standard classification approaches are limited by the fact that the derived labeling is typically not spatially regular, although it should be taken into account that class labels of neighboring 3D points tend to be correlated.

To derive a labeling with a higher spatial regularity, a contextual classification is typically considered for which many approaches rely on the use of a Conditional Random Field

(CRF) [4]. However, retrieving the classification of highest likelihood with such a model is a complicated combinatorial task, which can only be approximately solved. The authors of [4] use loopy belief propagation (LBP), a message-passing algorithm [8] to perform marginal inference, and then apply the maximum-a-posteriori (MAP) principle point by point to compute a smooth labeling.

As suggested in [9] for stereo shape estimation, graph-cut based approaches such as  $\alpha$ -expansion [10] can retrieve a labeling of higher likelihood and in shorter time. In this paper, we propose comparing these in terms of the likelihood of the produced labeling, computation time and quality of the derived labeling.

### 3. METHODOLOGY

We consider a point cloud  $V$ , whose local neighborhood structure is encoded in the graph  $G = (V, E)$ , where the edge set  $E$  encodes the adjacency between nodes. For each point, we observe a feature vector. Our objective is to find a labeling  $x \in \mathcal{K}^V$ , where  $\mathcal{K}$  is the finite set of semantic labels. We denote the set of features for all points in  $V$  by  $f$ , and the feature vector associated with the point  $i$  by  $f_i$ .

#### 3.1. Feature extraction

To adequately describe each considered 3D point  $i$  with geometric features, the local 3D structure at  $i$  is considered which is given in the spatial arrangement of 3D points within the local neighborhood of  $i$ . Accordingly, the local neighborhood of  $i$  has to be recovered first and, subsequently, all 3D points within that local neighborhood can be used to derive respective geometric features describing the local 3D structure at  $i$ .

The recovery of local neighborhoods can generally rely on different neighborhood definitions such as spherical or cylindrical neighborhoods which, in turn, can be parameterized with a specific scale parameter (e.g. a radius or the number of considered nearest neighbors). In this regard, the selection of the scale parameter often remains a crucial step since (1) the selection typically involves prior knowledge about the scene and/or the data and (2) typically an identical value of the scale parameter is used for all 3D points of the considered point cloud. For the consideration of general scenes with very different types of objects and classes of interest, however, a data-driven selection of the scale parameter seems to be desirable. Furthermore, it should be taken into account that 3D points belonging to different classes might favor a different neighborhood size [2], and therefore an approach is to be preferred that allows for locally adapting the size of the local neighborhood for each individual 3D point. In the scope of our work, we apply the definition of a local neighborhood that is formed by the  $k$  nearest neighbors of  $i$ . To select the optimal scale parameter  $k_{\text{opt}}$  for each 3D point  $i$ , we focus on eigenentropy-based scale selection [2] which has proven

to be favorable compared to other alternatives. Accordingly, we first calculate the eigenentropy for different scale parameters  $k$  and then select the optimal scale parameter  $k_{\text{opt}}$  that corresponds to the minimal eigenentropy (i.e. to the minimal disorder of 3D points).

Based on the recovered local neighborhoods, a variety of geometric features can be extracted. In particular, the *local 3D shape features* are widely used as they provide a rather intuitive description of the local 3D structure with a single value per feature. These local 3D shape features are derived from the eigenvalues of the 3D structure tensor and represented by linearity, planarity, sphericity, omnivariance, anisotropy, eigenentropy, sum of eigenvalues and change of curvature [2]. Furthermore, we extract *geometric 3D properties* of the local neighborhood that are given by the height of the 3D point  $i$ , the radius of the local neighborhood, the local point density, the verticality, and the maximum difference as well as the standard deviation of the height values of those points within the local neighborhood [2]. As urban environments are typically composed of a variety of man-made objects with almost perfectly vertical structures, we also consider a 2D projection of the point  $i$  and its  $k_{\text{opt}}$  nearest neighbors onto a horizontally oriented plane. Based on this 2D projection, *local 2D shape features* are defined by the sum and the ratio of the eigenvalues of the 2D structure tensor [2]. Finally, we also define *geometric 2D properties* which are represented by the radius of the local 2D neighborhood in the projection plane and the point density in 2D.

#### 3.2. Classification

To derive the initial labeling, we use a standard Random Forest (RF) classifier [11] that also allows for a soft labeling in terms of class probabilities. Subsequently, we use a CRF-based context model to impose spatial regularity on the initial labeling. More specifically, a CRF-based model like the one presented in [4] leverages contextual information to improve the semantic classification of 3D point clouds. In this model, the logarithm of the conditional likelihood of a labeling  $x$  given the observed features  $f$  is expressed as follows:

$$\ell(x | f) \propto \sum_{i \in V} \psi(x_i, f) + \sum_{(i,j) \in E} \phi(x_i, x_j, f). \quad (1)$$

This posterior probability is split into two parts: a data term and an interaction potential. The function  $\psi(x_i, f)$  encodes the influence of the observed feature on the label, without contextual information. We take  $\psi(x_i, f)$  as the logarithm of the probability given by the RF classifier.

The interaction potential  $\phi(x_i, x_j, f)$  encodes the adequacy between the labels of two adjacent nodes  $i$  and  $j$ . In [4], a classifier is trained to predict the probability of each transition given the difference of feature vectors. We argue here that the training of such a classifier poses certain difficulties as the number of transition classes is quadratic in the number

of semantic classes, and hence would require a much larger training set to obtain a prediction of reasonable quality, as well as more intensive computations. Furthermore, this would require a *point-perfect* precise annotation of the training set, which is difficult as irregularities in the density of the acquisition makes the precise localization of the interface between classes a complicated task. In this paper, we chose to forego the influence of features when it comes to transitions, and take  $\phi(x_i = k, x_j = l, f) = \log(M_{k,l})$ , where  $M$  is the matrix of empirical transition probabilities in our training set. This model allows us to assess the adequation of a given labeling with the observed features while taking contextual information into account. MAP inference, i.e. retrieving the labeling of highest likelihood, is a complicated task and in fact NP-hard. However, approximate solutions can be provided using message-passing algorithms or graph-cut-based methods.

Loopy belief propagation (LBP), which is an extension of the exact Viterbi algorithm [12] from trees to general graphs, is highly popular for its robustness and the simplicity of its implementation. However, it does not provide any proof of convergence or guarantees with respect to the global optimum. In [4], marginal inference is performed first using the sum-product LBP algorithm, and then each point is assigned the label of highest probability. In this paper, we also present the max-product version of the LBP algorithm, which allows us to directly compute the MAP inference.

Graph-cut-based algorithms, such as  $\alpha$ -expansion, offer guarantees that the results will be *close* to the global optimum. Furthermore,  $\alpha$ -expansion is also known for its impressive speed, in part due to the max-flow implementation [13].

#### 4. EXPERIMENTAL RESULTS

We evaluate the performance of different approaches on the *Oakland 3D Point Cloud Dataset* [1], a labeled benchmark dataset which has been acquired with a mobile laser scanning system in Oakland, USA. The ground truth (GT) labeling is provided with respect to five semantic classes given by *Wire*, *Pole/Trunk*, *Façade*, *Ground* and *Vegetation*. Furthermore, a split into training data (about 36.9k labeled 3D points), validation data (about 91.5k labeled 3D points) and test data (about 1.3M labeled 3D points) is already provided.

First, we derive an initial labeling by using low-level geometric 3D and 2D features (cf. Section 3.1) as input for a standard Random Forest classifier. For training, we take into account that an unbalanced distribution of training samples per class might have a detrimental effect on the classification results and hence use a reduced training set comprising 1,000 randomly selected training examples per class [2]. For testing, we evaluate the classification result achieved for the test set with respect to the GT labeling, where we obtain an overall accuracy of  $OA = 92.3\%$ , a kappa value of  $\kappa = 83.0\%$  and an unweighted average of the  $F_1$ -score of  $\bar{F}_1 = 63.5\%$ . The class-wise evaluation metrics of recall, precision and  $F_1$ -score

Class	Recall [%]	Precision [%]	$F_1$ -score [%]
Wire	84.6	10.2	18.2
Pole/Trunk	77.4	25.2	38.0
Façade	68.2	86.5	76.3
Ground	98.7	96.6	97.6
Vegetation	80.6	94.9	87.2

**Table 1.** Classification results derived with the initial labeling ( $OA = 92.3\%$ ,  $\kappa = 83.0\%$ ,  $\bar{F}_1 = 63.5\%$ ).

Class	Recall [%]	Precision [%]	$F_1$ -score [%]
Wire	75.6	15.4	25.5
Pole/Trunk	63.6	61.2	62.4
Façade	79.1	92.9	85.5
Ground	99.3	97.3	98.2
Vegetation	90.0	96.5	93.1

**Table 2.** Classification results derived with the sum-product version of the LBP algorithm ( $OA = 95.4\%$ ,  $\kappa = 89.8\%$ ,  $\bar{F}_1 = 71.1\%$ ,  $\ell = -2.08 \cdot 10^6$ ,  $t = 85s$  for 30 iterations).

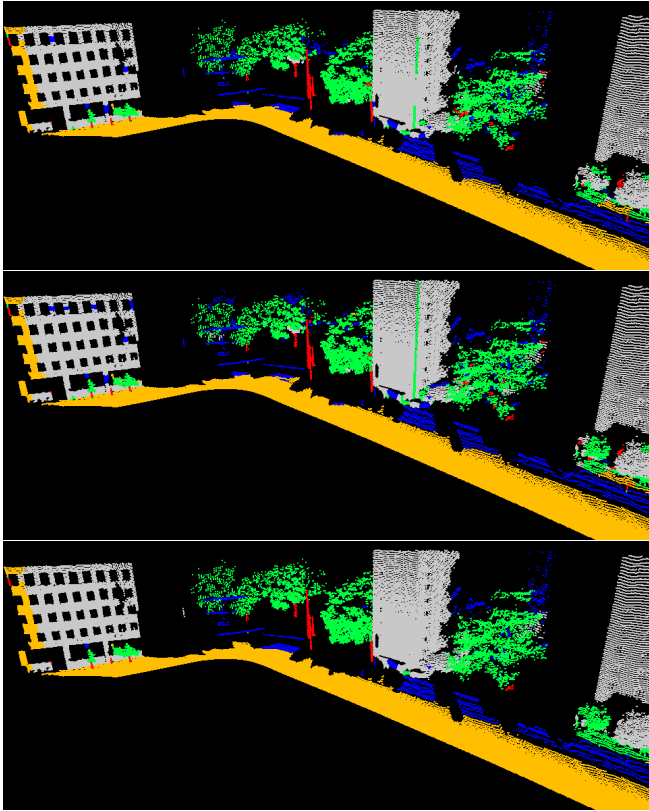
Class	Recall [%]	Precision [%]	$F_1$ -score [%]
Wire	62.7	16.0	25.5
Pole/Trunk	49.9	65.6	56.7
Façade	72.3	92.8	81.2
Ground	99.4	96.0	97.7
Vegetation	88.1	94.0	90.9

**Table 3.** Classification results derived with the max-product version of the LBP algorithm ( $OA = 94.4\%$ ,  $\kappa = 87.3\%$ ,  $\bar{F}_1 = 70.4\%$ ,  $\ell = -2.09 \cdot 10^6$ ,  $t = 95s$  for 30 iterations).

Class	Recall [%]	Precision [%]	$F_1$ -score [%]
Wire	64.7	37.8	47.7
Pole/Trunk	38.2	84.8	52.6
Façade	74.5	92.1	82.3
Ground	99.6	96.7	98.1
Vegetation	95.8	97.4	96.6

**Table 4.** Classification results derived with  $\alpha$ -expansion MAP inference ( $OA = 96.2\%$ ,  $\kappa = 91.4\%$ ,  $\bar{F}_1 = 75.5\%$ ,  $\ell = -1.39 \cdot 10^6$ ,  $t = 9s$ ).

are provided in Table 1 and a visualization of the classified point cloud is given in Fig. 1. For the three MAP inference algorithms presented in Section 3.2, we use MEX implementations provided by [14]. The respective results as well as the required processing times are provided in Fig. 2 and Tables 2-4, respectively. It can be observed that the use of contextual classification generally improves the classification results in terms of  $OA$ ,  $\kappa$  and  $\bar{F}_1$ . As expected, the  $\alpha$ -expansion algorithm reaches a higher likelihood in a much shorter time. We remark that this improvement translates into a classification of higher quality, validating the relevance of the model.



**Fig. 2.** Classification results derived with the sum-product version of LBP (top), the max-product version of LBP (center) and the  $\alpha$ -expansion MAP inference (bottom).

## 5. CONCLUSIONS

In this paper, we have focused on the CRF-based classification of lidar point cloud data. We have presented two approximate inference algorithms based on belief propagation and a graph-cut-based approach represented by  $\alpha$ -expansion. The derived results reveal that the graph-cut-based approach is able to retrieve a labeling of higher likelihood and thus a higher classification accuracy, while it is also significantly faster than the other approaches. In future work, we intend to investigate further approaches for spatial regularization.

## 6. REFERENCES

- [1] D. Munoz, J.A. Bagnell, N. Vandapel, and M. Hebert, "Contextual classification with functional max-margin Markov networks," in *Proc. CVPR*. IEEE, 2009, pp. 975–982.
- [2] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet, "Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 105, pp. 286–304, 2015.
- [3] T. Hackel, J.D. Wegner, and K. Schindler, "Fast semantic segmentation of 3d point clouds with strongly varying density," *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. III-3, pp. 177–184, 2016.
- [4] J. Niemeyer, F. Rottensteiner, and U. Soergel, "Classification of urban lidar data using conditional random field and random forests," in *Proc. JURSE*. IEEE, 2013, pp. 139–142.
- [5] R. Shapovalov, A. Velizhev, and O. Barinova, "Non-associative Markov networks for 3d point cloud classification," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVIII-3A, pp. 103–108, 2010.
- [6] J.D. Lafferty, A. McCallum, and F.C.N. Pereira, "Conditional random fields: probabilistic models for segmenting and labeling sequence data," in *Proc. ICML*. Morgan Kaufmann, 2001, pp. 282–289.
- [7] N. Chahata, L. Guo, and C. Mallet, "Airborne lidar feature selection for urban classification using random forests," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. XXXVIII-3/W8, pp. 207–212, 2009.
- [8] J.S. Yedidia, W.T. Freeman, and Y. Weiss, "Understanding belief propagation and its generalizations," *Exploring Artificial Intelligence in the New Millennium*, vol. 8, pp. 236–239, 2003.
- [9] M.F. Tappen and W.T. Freeman, "Comparison of graph cuts with belief propagation for stereo, using identical MRF parameters," in *Proc. ICCV*. IEEE, 2003, pp. 900–906.
- [10] Y. Boykov, O. Veksler, and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1222–1239, 2001.
- [11] L. Breiman, "Random forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [12] G.D. Forney, "The Viterbi algorithm," *Proceedings of the IEEE*, vol. 61, no. 3, pp. 268–278, 1973.
- [13] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 9, pp. 1124–1137, 2004.
- [14] M. Schmidt, "A Matlab toolbox for probabilistic undirected graphical models," <http://www.cs.ubc.ca/~schmidtm/Software/UGM.html>, 2007.