WEAKLY SUPERVISED SEGMENTATION-AIDED CLASSIFICATION OF URBAN SCENES FROM 3D LIDAR POINT CLOUDS

Stéphane Guinard*, Loïc Landrieu*

*IGN/LASTIG MATIS, Université Paris Est,
73 avenue de Paris, 94160 Saint-Mandé, France
(stephane.guinard, loic.landrieu)@ign.fr

KEY WORDS: Classification, Segmentation, Regularization, LiDAR, Urban, Point Cloud, Random Forest

ABSTRACT:

We consider the problem of the semantic classification of 3D LiDAR point clouds obtained from urban scenes when the training set is limited. We propose a non-parametric segmentation model for urban scenes composed of anthropic objects of simple shapes, partionning the scene into geometrically-homogeneous segments which size is determined by the local complexity. This segmentation can be integrated into a conditional random field classifier (CRF) in order to capture the high-level structure of the scene. For each cluster, this allows us to aggregate the noisy predictions of a weakly-supervised classifier to produce a higher confidence data term. We demonstrate the improvement provided by our method over two publicly-available large-scale data sets.

INTRODUCTION

Automatic interpretation of large 3D point clouds acquired from terrestrial and mobile LiDAR scanning systems has become an important topic in the remote sensing community (Munoz et al., 2009; Weinmann et al., 2015), yet it presents numerous technical challenges. Indeed, the high volume of data and the irregular structure of LiDAR point clouds make assigning a semantic label to each point a difficult endeavor. Furthermore the production of a precise ground truth is particularly difficult and time-consuming. However, LiDAR scans of urban scenes display some form of regularity and a specific structure can then be exploited to improve the accuracy of a noisy semantic labeling.

Foremost, the high precision of LiDAR acquisition methods implies that the number of points far exceeds the number of objects in a scene. Consequently, the sought semantic labeling can be expected to display high spatial regularity. Although the method presented in (Weinmann et al., 2015) relies on the computation of local neighborhood, the resulting classification is not regular in general, as observed in Figure 1b. This regularity prior has been incorporated into context-based graphical models (Anguelov et al., 2005; Shapovalov et al., 2010; Niemeyer et al., 2014) and a structured regularization framework (Landrieu et al., 2017a), significantly increasing the accuracy of input pointwise classifications.

Pre-segmentations of the point cloud have been used to model long-range interactions and to decrease the computational burden of the regularization. The segments obtained can then be incorporated into multi-scale graphical models to ensure a spatially-regular classification. However, the existing models require setting the parameters of the segments in advance, such as their maximum radius (Niemeyer et al., 2016; Golovinskiy et al., 2009), the maximum number of points in each segment (Lim and Suter, 2009), or the total number of segment (Shapovalov et al., 2010).

The aim of our work is to leverage the underlying structure of the point cloud to improve a weak classification obtained from very few annotated points, with a segmentation that requires no preset size parameters. We observe that the structure of urban scenes is mostly shaped by man-made objects (roads, façades, cars...), which are geometrically simple in general. Consequently, well-chosen geometric features associated to their respective points can be expected to be spatially regular. However the extent and number of points of the segments can vary a lot depending on the nature of the corresponding objects. We propose a formulation of the segmentation as a structured optimization problem in order to retrieve geometrically simple super-voxels. Unlike other presegmentation approaches, our method allows the segments' size to be adapted to the complexity of the local geometry, as illustrated in Figure 1c.

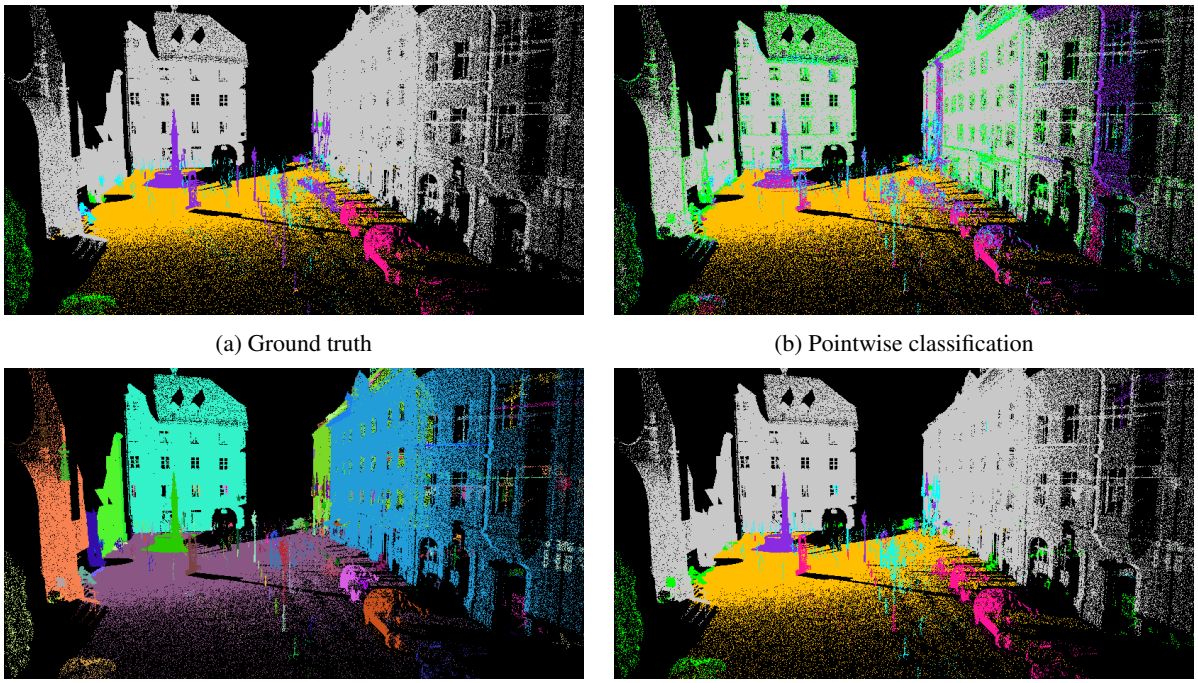
Following the machine-learning principle that an ensemble of weak classifiers can perform better than a strong one (Opitz and Maclin, 1999), a consensus prediction is obtained from the segmentation by aggregating over each segment the noisy predictions of its points obtained from a weakly-supervised classifier. The structure induced by the segmentation and the consensus prediction can be combined into a conditional random field formulation to directly classify the segments, and reach state-of-the-art performance from a very small number of hand-annotated points.

Related Work

Point-wise classification: Weinmann et al. (2015) propose a classification framework based on 3D geometric features which are derived from local neighborhood of optimal size.

Context-based graphical models: the spatial regularity of a semantic labeling can be enforced by graphical models such as random Markov fields (Anguelov et al., 2005; Shapovalov et al., 2010), and its discriminative counterpart, the conditional random field (Niemeyer et al., 2014; Landrieu et al., 2017b). The unary terms are computed by a point-wise classification with a random forest classifier, while the pairwise terms encode the probability of transition between the semantic classes.

Pre-segmentation approaches: A pre-segmentation of the point cloud can be leveraged to improve the classification. Lim and Suter (2009) propose defining each segment as a node in a multi-scale CRF. The super-voxels are defined by a growing region



(a) Ground truth

(b) Pointwise classification

(c) Geometrically homogeneous segmentation

(d) Segmentation-aided regularization

Figure 1. Illustration of the different steps of our method: the pointwise, irregular classification 1b is combined with the geometrically homogeneous segmentation 1c to obtain a smooth, objects-aware classification 1d. In Figures 1a, 1b , 1d, the semantic classes are represented with the following color code: vegetation, façades, hardscape, acquisition artifacts, cars, roads. In Figure 1c, each segment is represented by a random color.

method based on a predefined number of points in each pixel, and a color homogeneity prior. In Niemeyer et al. (2016), the segments are determined using a prior pointwise-classification. A multi-tier CRF is then constructed containing both points and voxels nodes. An iterative scheme is then performed, which alternates between inference in the multi-tier CRF and the computation of the semantically homogeneous segments with a maximum radius constraint. In Shapovalov et al. (2010) the presegmentation is obtained through the k-means algorithm, which requires defining the number of clusters in the scene in advance. Furthermore k-means produces isotropic clusters whose size doesn't adapt to the geometrical complexity of the scene. In Dohan et al. (2015), a hierarchical segmentation is computed using the foreground/background segmentation of Golovinskiy et al. (2009), which uses a preset horizontal and vertical radius as parameters. The segments are then hierarchically merged then classified.

Problem formulation

We consider a 3D point cloud V corresponding to a LiDAR acquisition in an urban scene. Our objective is to obtain a classification of the points in V between a finite set of semantic classes $\mathcal{K}$. We consider that we only have a small number of hand-annotated points as a ground truth from a similar urban scene. This number must be small enough that it can be produced by an operator in a reasonable time, i.e. no more than a few dozen per class.

We present the consituent elements of our approach in this section, in the order in which they are called.

Feature and graph computation: For each point, we compute a vector of geometrical features, described in Section 2.1. In Section 2.3 we present how the adjacency relationship between points is encoded into a weighted graph.

Segmentation in geometrically homogeneous segments: The segmentation problem is formulated as a structured optimization problem presented in Section 3.1, and whose solution can be approximated by a greedy algorithm. In section 3.2, we describe how the higher-level structure of the scene can be captured by a graph obtained from the segmentation.

Contextual classification of the segments: In Section 4, we present a CRF which derived its structure from the segmentation, and its unary parameter from the aggregation of the noisy prediction of a weakly supervised classifier. Finally, we associate the label of the corresponding segment to each point in the point cloud.

FEATURES AND GRAPH COMPUTATION

In this section, we present the descriptors chosen to represent the local geometry of the points, and the adjacency graph capturing the spatial structure of the point cloud.

With a view that the training set is small, and to keep the computational burden of the segmentation to a minimum, we voluntarily limit the number of descriptors used in our pointwise classification. We insist on the fact that the segmentation and the classification do not necessarily use the same descriptors.

Local descriptors

In order to describe the local geometry of each point we define four descriptors: linearity, planarity, scattering and verticality, which we represent in Figure 4.

The features are defined from the local neighborhood of each point of the cloud. For each neighborhood, we compute the eigenvalues $\lambda_1 \geq \lambda_2 \geq \lambda_3$ of the covariance matrix of the positions

of the neighbors. The neighborhood size is chosen such that it minimizes the eigentropy E of the vector $(\lambda_1/\Lambda, \lambda_2/\Lambda, \lambda_3/\Lambda)$ with $\Lambda = \sum_{i=1}^{3} \lambda_i$, in accordance with the optimal neighborhood principle advocated in Weinmann et al. (2015):

$$E = -\sum_{i=1}^{3} \frac{\lambda_i}{\Lambda} \log(\frac{\lambda_i}{\Lambda}).$$

As presented in Demantké et al. (2011), these eigenvalues allows us to qualify the shape of the local neighborhood by deriving the following vectors:

$$Linearity = \frac{\lambda_1 - \lambda_2}{\lambda_1}$$
$$Planarity = \frac{\lambda_2 - \lambda_3}{\lambda_1}$$
$$Scattering = \frac{\lambda_3}{\lambda_1}.$$

The linearity describes how *elongated* the neighborhood is, while the planarity assesses how well it is fitted by a plane. Finally, high-scattering values correspond to an isotropic and spherical neighborhood. The combination of these three features is called dimensionality.

In our experiments, the vertical extent of the optimal neighborhood proved crucial for discriminating roads and façades, and between poles and electric wires, as they share similar dimensionality. To discriminate this class, we introduce a novel descriptor called *verticality* also obtained from the eigen vectors and values defined above. Let u_1, u_2, u_3 be the three eigenvectors associated with $\lambda_1, \lambda_2, \lambda_3$ respectively. We define the unary vector of principal direction in $\mathbb{R}^3_+$ as the sum of the absolute values of the coordinate of the eigenvectors weighted by their eigenvalues:

$$[\hat{u}]_i \propto \sum_{j=1}^{3} \lambda_j \, |[u_j]_i|, \text{ pour } i = 1, 2, 3 \text{ et } \|\hat{u}\| = 1$$

We argue that the vertical component of this vector characterizes the verticality of the neighborhood of a point. Indeed it reaches its minimum (equal to zero) for an horizontal neighborhood, and its maximum (equal to 1) for a linear vertical neighborhood. A vertical planar neighborhood, such as a façade, will have an intermediary value (around 0.7). This behavior is illustrated at Figure in Figure 4.

To illustrate the weak number of features selected, we represent their respective value and range in Figure 2.

Non-local descriptors

Although the neighborhoods' shape of 3D points determine their local geometry, and allows us to compute a geometrically homogeneous segmentation, this not sufficient for classification. Consequently, we use two descriptors of the global position of points: elevation and position with respect to the road.

Computing those descriptors first requires determining the extent of the road with high precision. A binary road/non-road classification is performed using only the local geometry descriptors and a random forest classifier, which achieves very high accuracy and a F-score over 99.5%. From this classification a simple elevation model is computed, allowing us to associate a normalized height with respect to the road to each 3D point.

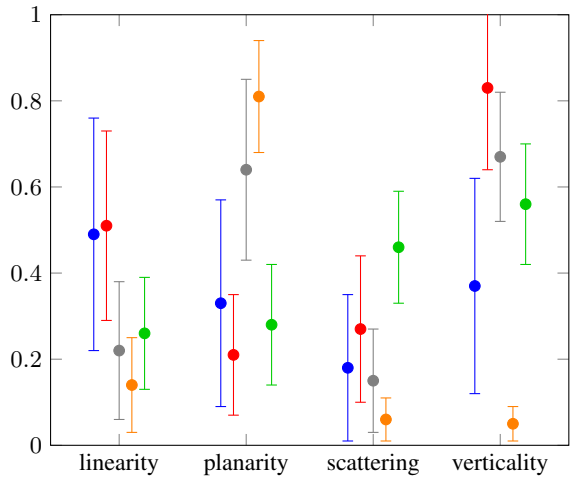


Figure 2. Means and standard deviations of the local descriptors in the Oakland dataset for the following classes: wires, poles, façades, roads, vegetation.

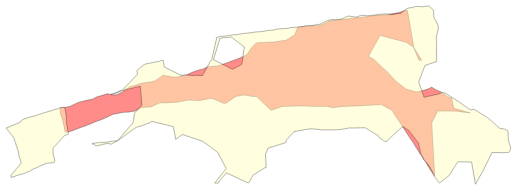


Figure 3. α-shape of the road on our Semantic3D example. In red, the horizontal extent of the road; in yellow, the extent of the non-road class.

To estimate the position with respect to the road we compute the two-dimensional α-shape (Akkiraju et al., 1995) of the points of the road projected on the zero elevation level, as represented in Figure 3. This allows us to compute the *position with respect to the road* descriptor, equal to 1 if a point is outside the extent of the road, 0.5 if the point is close to the edge of the α-shape with a tolerance of 1m, and 0 otherwise.

Adjacency graph

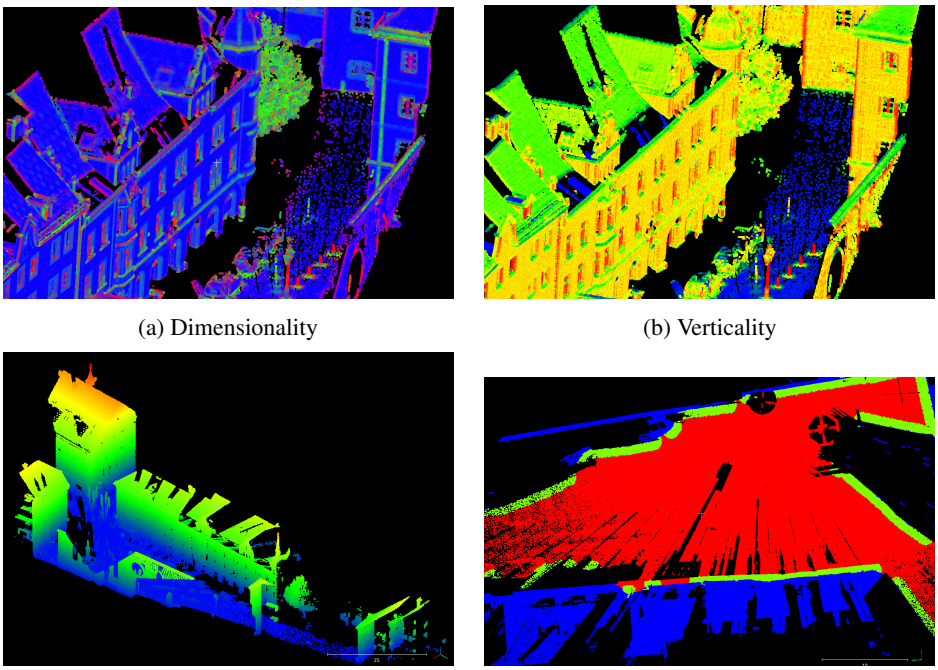
The spatial structure of a point cloud can be represented by an unoriented graph $G = (V, E)$, in which the nodes represent the points of the cloud, and the edges encode their *adjacency relationship*. We compute the 10-nearest neighbors graph, as advocated in (Niemeyer et al., 2011). Remark that this graph defines a symmetric graph-adjacency relationship which is *different* from the optimal neighborhood used in Section 2.1.

SEGMENTATION INTO HOMOGENEOUS SEGMENTS

Potts energy segmentation

To each point, we associate its local geometric feature vector $f_i \in \mathbb{R}^4$ (dimensionality and verticality), and compute a piecewise constant approximation $g^\star$ of the signal $f \in \mathbb{R}^{V \times 4}$ structured by the graph G. $g^\star$ is defined as the vector of $\mathbb{R}^{V \times 4}$ minimizing the following *Potts segmentation energy*:

$$g^\star = \arg\min_{g \in \mathbb{R}^{4 \times V}} \sum_{i \in V} \|g_i - f_i\|^2 + \rho \sum_{(i,j) \in E} \delta(g_i - g_j \neq 0),$$



(a) Dimensionality (b) Verticality

(c) Elevation (d) Position with respect to the road

Figure 4. Representation of the four local geometric descriptors as well as the two global descriptors. In (a), the dimensionality vector [linearity, planarity, scattering] is color-coded by a proportional [red, green, blue] vector. In (b), the value of the verticality is represented with a color map going from blue (low verticality - roads) to green/yellow (average verticality - roofs and façades) to red (high verticality - poles). In (c), is represented the elevation with respect to the road. In (d), the position with respect to the road is represented with the following color-code: inside the road α-shape in red, bordering in green, and outside in blue.

with $\delta(\cdot \neq 0)$ the function of $\mathbb{R}^4 :\mapsto \{0,1\}$ equal to 0 in 0 and 1 everywhere else. The first part of this energy is the fidelity function, ensuring that the constant components of $g^\star$ correspond to homogeneous values of f. The second part is the regularizer which adds a penalty for each edge linking two components with different values. This penalty enforces the simplicity of the shape of the segments. Finally ρ is the regularization strength, determining the trade off between fidelity and simplicity, and implicitly determining the number of clusters.

This structured optimization problem can be efficiently approximated with the greedy graph-cut based ℓ_0-cut pursuit algorithm presented in Landrieu and Obozinski (2016). The segments are defined as the constant connected components of the piecewise constant signal obtained.

The benefit of this formulation is that it does not require defining a maximum size for the segments in terms of extent or points. Indeed large segments of similar points, such as roads or façades, can be retrieved. On the other hand, the granularity of the segments will increase where the geometry gets more complex, as illustrated in Figure 1c.

For the remainder of the article we denote $\mathbb{S} = (S_1, \cdots, S_k)$ the non-overlapping segmentation of V obtained when approximately solving the optimization problem.

Segment-graph

We argue that since the segments capture the objects in the scene, the segmentation represents its underlying high-level structure. To obtain the relationship between objects, we build the *segment-graph*, which is defined as $\mathcal{G} = (\mathbb{S}, \mathbb{E}, w)$ in which the segments of $\mathbb{S}$ are the nodes of $\mathcal{G}$. $\mathbb{E}$ represents the adjacency relationship

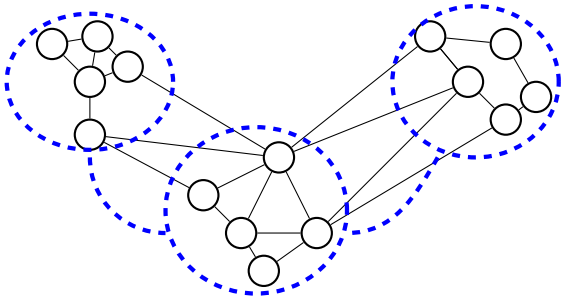


Figure 5. Adjacency structure of the segment-graph. The edges between points are represented in black —— , the segmentation and the adjacency of its components in blue: - - - .

between segments, while w encodes the weight of their boundary, as represented in Figure 5. We define two segments as adjacent if there is an edge in E linking them, and w as the total weight of the edges linking those segments:

$$\begin{cases} \mathbb{E} & = \{(s,t) \in \mathbb{S}^2 \mid \exists (i,j) \in E \cap (s \times t)\} \\ w_{s,t} & = |E \cap (s \times t)|, \quad \forall (s,t) \in \mathbb{S}^2. \end{cases}$$

CONTEXTUAL CLASSIFICATION OF THE SEGMENTS

To enforce spatial regularity, Niemeyer et al. (2014) defines the optimal labeling $l^\star$ of a point cloud as maximizing the posterior distribution $p(l \mid f')$ in a conditional random field model structured by an adjacency graph G, with f' the vector of local and global features. We denote a labeling of V by a vector of

$\Delta(V,\mathcal{K}) = \{l \in \{0,1\}^{V\times\mathcal{K}} \mid \sum_{k\in\mathcal{K}} l_{i,k} = 1, \forall i \in V\}$ (the corners of the simplex) such that $l_{i,k}$ is equal to one if the point i of V is labelled as $k \in \mathcal{K}$, and zero else. For a point i of V, l_i is considered as a vector of $\mathbb{R}^{\mathcal{K}}$. This allows us to define $l^\star$ as the maximizing argument of the following energy:

$$l^\star = \underset{l\in\Delta(V,\mathcal{K})}{\arg\max} \sum_{i\in V} l_i^\intercal p^i + \sum_{(i,j)\in E} l_i^\intercal M_{i,j} l_j, \tag{1}$$

with $p_{i,k} = \log(p(l_i = k \mid f_i'))$ the entrywise logarithm of the probability of node i being in state k, and $M_{(i,j),(k,l)} = \log(p(l_i = k, l_j = l \mid f_i', f_j'))$ the entrywise logarithm of the probability of observing the transition (k,l) at (i,j).

As advocated in Niemeyer et al. (2014), we can estimate $p(l_i = k \mid f_i')$ with a random forest probabilistic classifier p_{RF}. To avoid infinite values, the probability p_{RF} is smoothed by taking a linear interpolation with the constant probability: $p(k \mid f_i) = (1-\alpha)p_{\text{RF}}(k \mid f'_i) + \alpha/|\mathcal{K}|$ with $\alpha = 0.01$ and $|\mathcal{K}|$ the cardinality of the class set. The authors also advocate learning the transition probability from the difference of the features vectors. However, our weak supervision hypothesis prevents us from learning the transitions, as it would require annotations covering the $|\mathcal{K}|^2$ possible combinations extensively. Furthermore the annotation would have to be very precise along the transitions, which are often hard to distinguish in point clouds. We make the simplifying hypothesis that M is of the following form :

$$M_{(i,j),(k,l)} = \begin{cases} 0 & \text{if } k = l \\ \sigma & \text{else,} \end{cases} \tag{2}$$

with σ a non-negative value, which can be determined by cross-validation.

Leveraging the hypothesis that the segments obtained in in Section 3.1 correspond to semantically homogeneous objects, we can assume that the optimal labeling will be constant over each segment of $\mathbb{S}$. In that regard, we propose a formulation of a CRF structured by the segment-graph $\mathcal{G}$ to capture the organization of the segments. We denote $L^\star$ the labeling of $\mathbb{S}$ defined as:

$$L^\star = \underset{L\in\Delta(\mathbb{S},\mathcal{K})}{\arg\max} \sum_{s\in\mathbb{S}} L_s^\intercal P^s + \sum_{(s,t)\in\mathbb{E}} w_{s,t} L_s^\intercal M L_t^\intercal,$$

with $P_k^s = |s| \log(p(L_s = k \mid \{f_i'\}_{i\in s}))$ the logarithm of the probability of segment s being in state k multiplied by the cardinality of s. We define this probability as the average of the probability of each point contained in the segment:

$$p(L_s = k \mid \{f_i'\}_{i\in s}) = \frac{1}{|s|} \sum_{i\in s} p(l_i = k \mid f_i').$$

Note that the influence of the data term of a segment is determined by its cardinality, since the classification of the points remains the final objective. Likewise, the cost of a transition between two segments is weighted by the total weight of the edges at their interface $w_{s,t}$, and represents the magnitude of the interaction between those two segments.

Following the conclusions of Landrieu et al. (2017b), we approximate the labelling maximizing the log-likelihood with the maximum-a-priori principle using the α-expansion algorithm of Boykov et al. (2001), with the implementation of Schmidt (2007).

It is important to remark that the segment-based CRF only involves the segment-graph $\mathcal{G}$, which can be expected to be much smaller than G, making inference potentially much faster.

NUMERICAL EXPERIMENTS

We now demonstrate advantages of our approach through numerical experiments on two public data sets. First, we introduce the data and our evaluation metric, then present the classification results compared to state-of-the-art methods.

Data

To validate our approach, we consider two publicly available data sets.

We first consider the urban part of the Oakland benchmark introduced in Munoz et al. (2009), comprised of 655.297 points acquired by mobile LiDAR. Some classes have been removed from the acquisition (i.e. cars or pedestrians) so that there are only 5 left: electric wires, poles/trunks, façcades, roads and vegetation. We choose to exclude the tree-rich half of the set as the segmentation results are not yet satifying at the trunk-tree interface.

We also consider one of the urban scenes in the Semantic3D benchmark[1], downsampled to 3.5 millions points for memory reasons. This scene, acquired with a fixed LiDAR, contains 6 classes : road, façade, vegetation, car, acquisition artifacts and hardscape.

For each class we hand-pick a small number of representative points such that the discriminative nature of our features illustrated in Figure 2 is represented. We select 15 points per classes for Oakland and 25 to 35 points for semantic3D, for respective totals of 75 and 180 points.

Metric

To take into account the imbalanced distribution of each class (roads and façades comprise up to 80% of the points), we use the unweighted average of the F-scores to evaluate the classification results. Consequently, a classification with decent accuracy over all classes will have a higher score than a method with high accuracy over some classes but poor results for others.

Competing methods

To compare the efficiency of our implementation to the state-of-the-art we have implemented the following methods:

- **Pointwise:** we implemented the pointwise classification relying on optimal neighborhoods of Weinmann et al. (2015), with a random forest (Breiman, 2001) and restricted ourselves to the six geometric features presented in Section 2.1.
- **CRF regularization:** we implemented the CRF defined in (1) without aid from the segmentation.

Results

In Tables 1 and 2, we represent the classification results of our method and the competing methods for both datasets. We observe that both the CRF and the presegmentation approach significantly improve the results compared to the point-wise classification. Although the improvement in term of global accuracy

[1] `http://www.semantic3d.net/`

	pointwise classification			CRF-regularization			our method		
classes	precision	recall	FScore	precision	recall	FScore	precision	rappel	Fscore
wires	4.2	37.1	7.5	**87.8**	32.1	**47.0**	51.2	**35.4**	41.9
poles	9.0	**67.7**	15.9	**78.6**	37.7	51.0	66.1	48.3	**55.8**
façades	57.5	74.9	65.1	79.2	**98.0**	87.6	**91.0**	96.5	**93.6**
road	**99.9**	86.7	92.8	99.6	95.2	97.4	99.6	**99.1**	**99.4**
vegetation	85.5	82.8	84.1	93.5	93.1	93.3	**95.5**	**94.4**	**95**
total	51.2	69.8	53.1	**87.7**	71.2	75.2	80.7	**74.7**	**77.1**

Table 1. Precision, recall and FScore in % for the Oakland benchmark. The global accuracy are respectively 85.2%, 94.8% et 97.3%. In bold, we represent the best value in each category.

	pointwise classification			CRF-regularization			our method		
classes	precision	recall	FScore	precision	recall	FScore	precision	rappel	Fscore
road	**98.7**	96.8	97.7	97.6	**99.0**	**98.3**	97.5	98.7	98.1
vegetation	14.2	82.9	24.2	49.7	84.7	62.6	**52.1**	**93.7**	**67.0**
façade	99.6	88.1	93.5	99.5	97.9	98.7	**99.7**	**98.2**	**98.8**
hardscape	74.2	71.4	73.1	**93.7**	88.7	91.2	92.7	**90.4**	**91.5**
artifacts	18.3	37.5	24.6	**77.9**	**42.1**	**54.7**	73.8	39.3	51.3
cars	28.6	54.8	37.6	66.5	86.2	75.1	**84.0**	**90.0**	**82.3**
total	55.7	71.9	58.4	80.8	83.1	80.1	**83.3**	**85.0**	**82.3**

Table 2. Precision, recall and FScore in % for the Semantic3D benchmark. The global accuracy are respectively 88.4%, 96.9% et 97.2%. In bold, we represent the best value in each category.

of our method compared to the CRF-regularization is limited (a few % at best), the quality of the classification is improved significantly for some hard-to-retrieve classes such as poles, wires, and cars. Furthermore, our method provides us with a object-level segmentation as well.

CONCLUSION

In this article, we presented a classification process aided by a geometric pre-segmentation capturing the high-level organization of an urban scene. We showed that this segmentation allowed us to formulate a CRF to directly classify the segments, improving the results over the CRF-regularization Further developments should focus on improving the quality of the segmentation near loose and scattered acquisition such as foliage. Another possible improvement would be to better exploit the context of the transition. Indeed the form of the transition matrix in (2) is too restrictive, as it does not take into account rules such as "the road is below the façcade" or the "tree-trunk is more likely than foliage-road". Although the weakly-supervised context excludes learning the transition, it would nonetheless be beneficial to incorporate the expertise of the operator.

References

Akkiraju, N., Edelsbrunner, H., Facello, M., Fu, P., Mücke, E. and Varela, C., 1995. Alpha shapes: definition and software. In: *Proceedings of the 1st International Computational Geometry Software Workshop*, Vol. 63, p. 66.

Anguelov, D., Taskarf, 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: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2, IEEE, pp. 169–176.

Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (11), pp. 1222–1239.

Breiman, L., 2001. Random forests. *Machine learning* 45(1), pp. 5–32.

Demantké, J., Mallet, C., David, N. and Vallet, B., 2011. Dimensionality based scale selection in 3d lidar point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(Part 5), pp. W12.

Dohan, D., Matejek, B. and Funkhouser, T., 2015. Learning hierarchical semantic segmentations of lidar data. In: *3D Vision (3DV), 2015 International Conference on*, IEEE, pp. 273–281.

Golovinskiy, A., Kim, V. G. and Funkhouser, T., 2009. Shape-based recognition of 3d point clouds in urban environments. In: *Computer Vision, 2009 IEEE 12th International Conference on*, IEEE, pp. 2154–2161.

Landrieu, L. and Obozinski, G., 2016. Cut pursuit: fast algorithms to learn piecewise constant functions on general weighted graphs.

Landrieu, L., Raguet, H., Vallet, B., Mallet, C. and Weinmann, M., 2017a. A structured regularization framework for spatially smoothing semantic labelings of 3d point clouds.

Landrieu, L., Weinmann, M. and Mallet, C., 2017b. Comparison of belief propagation and graph-cut approaches for contextual classification of 3d lidar point cloud data.

Lim, E. H. 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.

Munoz, D., Bagnell, J. A., Vandapel, N. and Hebert, M., 2009. Contextual classification with functional max-margin markov networks. In: *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, IEEE, pp. 975–982.

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., Rottensteiner, F., Soergel, U. and Heipke, C., 2016. Hierarchical higher order crf for the classification of airborne lidar point clouds in urban areas. In: *23rd International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Congress, ISPRS 2016, 12–19 July 2016, Prague, Czech Republic*, Göttingen: Copernicus GmbH.

Niemeyer, J., Wegner, J. D., Mallet, C., Rottensteiner, F. and Soergel, U., 2011. Conditional random fields for urban scene classification with full waveform lidar data. In: *Photogrammetric Image Analysis*, Springer, pp. 233–244.

Opitz, D. and Maclin, R., 1999. Popular ensemble methods: An empirical study. *Journal of Artificial Intelligence Research* 11, pp. 169–198.

Schmidt, M., 2007. A Matlab toolbox for probabilistic undirected graphical models. `http://www.cs.ubc.ca/~schmidtm/Software/UGM.html`.

Shapovalov, R., Velizhev, E. and Barinova, O., 2010. Non-associative markov networks for 3d point cloud classification. In: *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII, Part 3A*, Citeseer.

Weinmann, M., Urban, S., Hinz, S., Jutzi, B. and Mallet, C., 2015. Distinctive 2d and 3d features for automated large-scale scene analysis in urban areas. *Computers & Graphics* 49, pp. 47–57.