



HAL
open science

A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds

Loic Landrieu, Hugo Raguete Raguete, Bruno Vallet, Clément Mallet, Martin Weinmann

► **To cite this version:**

Loic Landrieu, Hugo Raguete Raguete, Bruno Vallet, Clément Mallet, Martin Weinmann. A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds. 2017. hal-01505245v1

HAL Id: hal-01505245

<https://hal.science/hal-01505245v1>

Preprint submitted on 11 Apr 2017 (v1), last revised 4 May 2017 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Manuscript Number:

Title: A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds

Article Type: Research Paper

Section/Category: Lidar/laser scanning

Keywords: semantic labeling; point cloud; classification; regularization; structured optimization; graphical models

Corresponding Author: Dr. Martin Weinmann,

Corresponding Author's Institution: Karlsruhe Institute of Technology (KIT)

First Author: Loic Landrieu

Order of Authors: Loic Landrieu; Hugo Raguét; Bruno Vallet; Clément Mallet; Martin Weinmann

KIT-Campus South | IPF | Englerstr. 7 | 76131 Karlsruhe, Germany

Prof Derek Lichti
Department of Geomatics Engineering
The University of Calgary
2500 University Dr NW
Calgary, AB Canada T2N 1N4

Dr. Martin Weinmann
Karlsruhe Institute of Technology
Institute of Photogrammetry and Remote Sensing
Englerstraße 7, 76131 Karlsruhe, Germany
Email: martin.weinmann@kit.edu

Dr. Loic Landrieu
Institut National de l'Information Géographique et
Forestiere
LASTIG MATIS
73 avenue de Paris, 94160 Saint-Mandé, France
Email: loic.landrieu@ign.fr

Date: 6 April 2017

Paper Submission to ISPRS Journal of Photogrammetry and Remote Sensing

Dear Prof Lichti, dear editors,

We would like to submit the following manuscript to the ISPRS Journal of Photogrammetry and Remote Sensing:

"A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds"
(Authors: Loic Landrieu, Hugo Rague, Bruno Vallet, Clément Mallet and Martin Weinmann)

In our manuscript, we consider the problem of obtaining spatially smooth semantic labelings of 3D LiDAR point clouds. We view the problem as an optimization problem structured by an adjacency graph, and propose a modular optimization framework for which we present state-of-the-art solving algorithms. The main contributions of this paper are:

- we developed a rich framework which comprises the state-of-the-art of spatial regularization such as graphical models like CRFs;
- we list 2 search spaces (hard and soft labelings), 4 fidelity functions, and 2 spatial regularity penalizers which can be combined into our framework to obtain 8 different spatially-regularizing functions with very different properties;
- we introduce a modified version of the greedy L0-cut pursuit algorithm modified to function on multi-dimensional simplex-constrained data for the continuous space / non-convex case;
- we obtain both significant speed-ups and precision increase over state-of-the-art spatial smoothing approaches on 3 open source datasets, in particular for hard-to-retrieve classes;
- we propose to compare soft labelings with accuracy/coverage plots.

We confirm that the work is original. A partial result on the choice of the algorithm for solving the CRFs formulations in remote sensing was accepted at IGARSS'2017.

Among the editors we may suggest Franz Rottensteiner and Konrad Schindler for the manuscript.

We hope that the contents of our manuscript are acceptable for publication in the ISPRS Journal of Photogrammetry and Remote Sensing. Please feel free to contact us at loic.landrieu@ign.fr or martin.weinmann@kit.edu for any questions regarding the submission.

Thank you for your time and consideration. We look forward to hearing from you soon.

Sincerely,

Loic Landrieu and Martin Weinmann

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds

Loïc Landrieu^{a,*}, Hugo Rague^b, Bruno Vallet^a, Clément Mallet^a, Martin
Weinmann^{a,c,*}

^a*Université Paris-Est, LASTIG MATIS IGN, ENSG, 94160 Saint-Mandé, France*

^b*Institut de Mathématiques de Marseille, Aix-Marseille Université, 39 rue Frédéric
Joliot-Curie, 13453 Marseille, France*

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

Abstract

In this paper, we introduce a mathematical framework for obtaining spatially smooth semantic labelings of 3D point clouds from a pointwise classification. We argue that structured regularization offers a more versatile alternative to the standard graphical model approach. Indeed, our framework allows us to choose between a wide range of fidelity functions and regularizers, influencing the properties of the solution. In particular, we investigate the conditions under which the smoothed labeling remains probabilistic in nature, allowing us to measure the uncertainty associated with each label. Finally, we present efficient algorithms to solve the corresponding optimization problems.

To demonstrate the performance of our approach, we present classification results derived for standard benchmark datasets. We demonstrate that the structured regularization framework offers higher accuracy at a lighter computational cost in comparison to the classic graphical model approach.

Keywords: Semantic labeling, point cloud, classification, regularization, scene interpretation, structured optimization, graphical models

*Corresponding authors

Email addresses: loic.landrieu@ign.fr (Loïc Landrieu), hugo.rague@gmail.com (Hugo Rague), bruno.vallet@ign.fr (Bruno Vallet), clement.mallet@ign.fr (Clément Mallet), martin.weinmann@kit.edu (Martin Weinmann)

Preprint submitted to ISPRS Journal of Photogrammetry and Remote Sensing April 7, 2017

1. Introduction

Due to the steadily increasing availability of geospatial information, their automated analysis has become a topic of major interest in photogrammetry, remote sensing, robotics, and computer vision. In particular, the representation of a scene in the form of a 3D point cloud and a subsequent semantic interpretation of this point cloud serve as the basis for many applications, such as scene modeling, autonomous navigation, or object detection. For instance, the analysis of 3D point cloud data acquired within urban environments benefits from a semantic labeling since the latter can be exploited for the creation of large-scale city models (Lafarge and Mallet, 2012) or urban accessibility diagnosis (Serna and Marcotegui, 2013).

The semantic interpretation typically consists in assigning a semantic label (e.g. *building*, *ground* or *vegetation*) to each point of the considered 3D point cloud, as shown in Figures 1a-1c. This assignment can be accompanied by an estimation of the confidence of the labeling of each point in the form of a probability distribution over the labels, as illustrated in Figure 1d. Such a certainty assessment can prove useful when either the precision or the recall of the classification is more crucial for a given application. In the case of autonomous navigation for example, merely the possibility of an obstacle can be enough to alter course, and a probabilistic occupancy map is preferred to a binary one (Moravec and Elfes, 1985; Hornung et al., 2013). In the case of reconstruction tasks which necessitate the removal of a specific semantic class beforehand (Clode et al., 2004), precision is the focus in order to not accidentally remove relevant information. In a context of active learning, an assessment of the labeling certainty can guide an operator to the areas of the point cloud in which the classification is least certain, as they are more prone to be labeled incorrectly and might require manual re-labeling (Jing et al., 2004). The nature of the assignment, either a probability or a label, depends on the choice of the method used for inference.

The semantic labeling of 3D point clouds has been addressed by numerous investigations (Munoz et al., 2009; Shapovalov et al., 2010; Mallet et al., 2011; Niemeyer et al., 2014; Xu et al., 2014; Guo et al., 2015; Weinmann et al., 2015a; Weinmann, 2016; Hackel et al., 2016b). However, this problem remains challenging due to the irregular point sampling, occlusions, and the complexity of the considered scenes, which induce a loose yet meaningful structure to the data. Furthermore, the consideration of larger scenes typically results in a huge amount of data and efficient techniques for semantic

1
2
3
4
5
6
7
8
9 labeling are therefore desirable. To foster research regarding semantic label-
10 ing of 3D point cloud data, a variety of benchmark datasets acquired within
11 urban environments have been released (Munoz et al., 2009; Serna et al.,
12 40 2014; Vallet et al., 2015; Hackel et al., 2016b).

13
14 The straightforward approach for semantically labeling a considered 3D
15 point cloud consists in extracting a variety of features for all points, con-
16 catenating these features to a feature vector, which is then classified with
17 a classifier trained on representative training examples. This strategy has
18 45 for instance been followed in the framework introduced by Weinmann et al.
19 (2015a), in which a diversity of distinctive geometric low-level features serve
20 as input for a standard supervised classification scheme. While this rather
21 simple approach already yields good classification rates due to the use of
22 distinctive features, the visualization of the classified 3D point cloud reveals
23 a noisy behavior as each point is treated individually by only considering the
24 50 respective feature vector for classification. To illustrate this effect, we pro-
25 vide a ground truth labeling for a considered 3D point cloud in Figure 1a and
26 a pointwise classification based on distinctive geometric low-level features in
27 Figure 1b. Considering the ground truth labeling, one can observe a high
28 spatial regularity of the labeling. Indeed, as the number of 3D points far
29 exceeds the number of objects in the scene, it is reasonable to assume that
30 most 3D points are surrounded by points of the same label.
31 55

32 To impose spatial smoothness on this classification result, contextual in-
33 formation among neighboring 3D points is typically taken into account. For
34 this purpose, the spatial structure of a 3D point cloud can be captured by a
35 graph encoding the adjacency relationship between 3D points. Thereby, the
36 adjacency relationship can be derived from the local neighborhood of each 3D
37 point (Weinmann et al., 2015b), pre-segmentations (Niemeyer et al., 2016),
38 60 or super-voxel structures (Lim and Suter, 2009). Based on the defined ad-
39 jacency relationship, a context model is typically derived in the form of a
40 graphical model, e.g. a Markov random field (MRF) (Munoz et al., 2009;
41 Shapovalov et al., 2010; Lu and Rasmussen, 2012; Najafi et al., 2014) or its
42 discriminative counterpart, the conditional random field (CRF) (Niemeyer
43 65 et al., 2011, 2014; Schmidt et al., 2014; Weinmann et al., 2015b). As a
44 consequence of imposing spatial smoothness on the derived labeling, the cor-
45 responding classification results are often significantly improved as can be ob-
46 served in Figure 1c. However, the choice of the inference strategy (marginal,
47 maximum-a-posteriori) will have a profound impact on the precision and na-
48 70 ture of the solution (probabilistic or labeling), as well as the computation
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

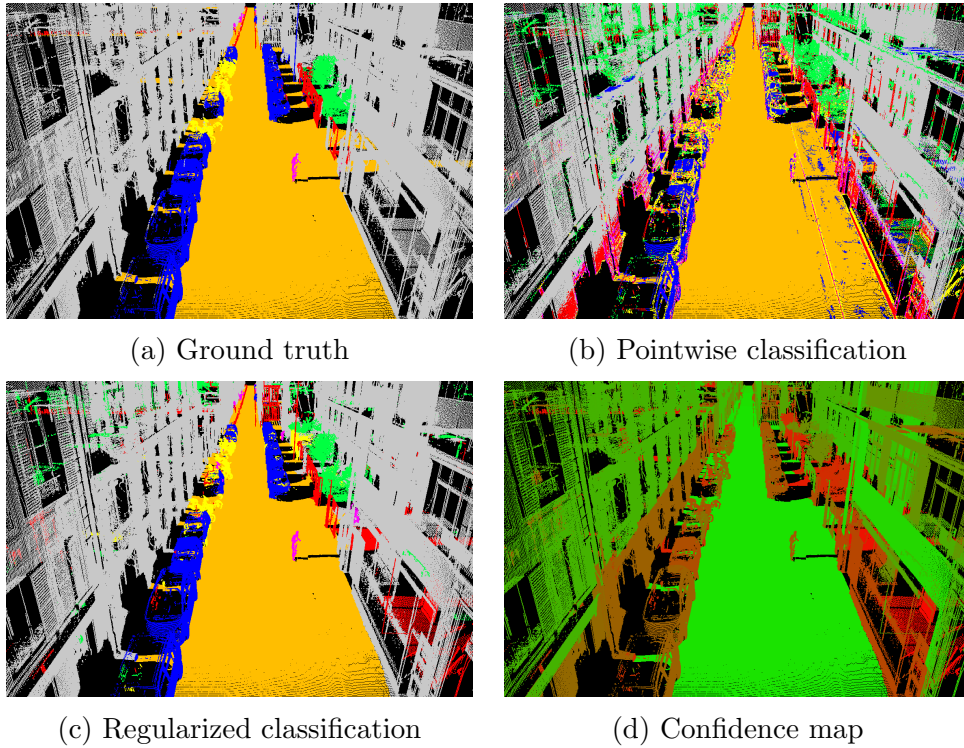



Figure 1: Visualization of a 3D point cloud labeling for a part of the Paris-rue-Cassette Database (Vallet et al., 2014). In (a), (b), and (c), the color encoding addresses the classes *Façade* (gray), *Ground* (orange), *Cars* (blue), *2-Wheelers* (yellow), *Road Inventory* (red), *Pedestrians* (magenta) and *Vegetation* (green). In (d), the confidence is represented from green to red: confident  uncertain. Remark that misclassifications in (c) correspond to the least confident area in (d).

times.

In this paper, we propose to consider the problem of spatially smoothing semantic labelings of 3D point clouds from a structured regularization perspective. While using such a model results in a loss of interpretability compared to a probabilistic approach, it offers several advantages. In particular, the structured regularization approach allows:

- the choice from a wide range of fidelity functions and regularizers ¹,

¹Notably, this framework allows us to express the graphical model approach as a special

- the choice to retain or not the probabilistic aspect of the input labeling, and
- the use of fast solving algorithms, compared to slow and memory-intensive message-passing algorithms.

After briefly introducing the used notation and the formal description of the considered problem in Section 1, we summarize related work in Section 2. Subsequently, in Section 3, we outline the main idea of the fundamental framework for pointwise point cloud classification, which follows the work presented in (Weinmann et al., 2015a; Weinmann, 2016) but additionally allows the choice between *hard labelings*, in which each 3D point is assigned one unique class label, and *soft labelings* in the form of class probabilities.

As the labelings obtained via pointwise classification are not generally spatially smooth, we present a general regularization framework that takes into account the fact that class labels of neighboring 3D points tend to be correlated. In Section 4, we present the label-smoothing problem as a regularization problem structured by an adjacency graph, and we present four fidelity functions as well as two graph-structured regularizers and two possible search spaces. The choice of a fitting minimizing algorithm hinges on the respective properties of the fidelity and regularizing functions. In Section 5, we present efficient, state-of-the-art algorithms to solve the different cases encountered. Furthermore, we present a novel extension of the ℓ_0 -cut pursuit algorithm presented by Landrieu and Obozinski (2016a), allowing the input to take the form of multi-dimensional probabilities instead of one-dimensional values only. To demonstrate the performance of our methodology, we present in Section 6 experimental results for different configurations of our framework as well as for state-of-the-art approaches. We focus on the quality of the resulting classifications, and to stress the advantages of probabilistic labelings, we present the partial coverage classification as well. Finally, in Section 7, we provide concluding remarks and suggestions for future work.

1.1. Notation

We denote V the finite set of 3D points to label, and \mathcal{K} the finite set of potential semantic labels for each 3D point. Throughout this paper, we make

instance.

an important distinction between hard and soft labelings. In that regard, we denote by \mathcal{S} the simplex:

$$\mathcal{S} = \{p \in [0, 1]^{\mathcal{K}} \mid \sum_{k \in \mathcal{K}} p_k = 1\}. \quad (1)$$

Elements of \mathcal{S} are called *soft labelings*. Similarly, we denote \mathbf{S} the corners of the simplex:

$$\mathbf{S} = \{p \in \{0, 1\}^{\mathcal{K}} \mid \sum_{k \in \mathcal{K}} p_k = 1\}. \quad (2)$$

Elements of \mathbf{S} are called *hard labelings*. For a labeling $p \in \mathcal{S}$ or \mathbf{S} and a semantic class $k \in \mathcal{K}$, we denote the probability associated with class k by p_k , and consider the probability p as a vector of size $|\mathcal{K}|$.

Throughout this paper, we denote labelings of a single 3D point in lowercase, and global labelings, relative to the entire point cloud V , in uppercase. For such a global labeling $P \in \mathcal{S}^V$ or \mathbf{S}^V , $i \in V$ we denote the labeling of a point $i \in V$ by P_i and its probability for class k by $P_{i,k}$.

1.2. Problem statement

We consider a set of 3D points V for which we have a soft labeling $P \in \mathcal{S}^V$ obtained via a classification algorithm which does not directly account for spatial smoothness. The goal is to find P^* , an improved labeling with increased spatial smoothness while remaining as close as possible to the input labeling P . Our proposed approach is to define P^* as the solution of a well-chosen optimization problem, whose *objective functional* is structured by an adjacency graph capturing the spatial relationship between the 3D points. P^* can be either a soft labeling or a hard labeling depending on the parameterization of the regularization problem. This process can be broken down into three parts:

- **Computing the initial labeling:** The proposed regularization framework is not affected by the choice of the method used to obtain the initial classification. However, our approach is more suited when the initial labeling is probabilistic. In this paper, we use a classification framework which is described in Section 3 and relies on the use of a diversity of low-level geometric 3D and 2D features as input for a standard random forest classifier (Weinmann et al., 2015a).

- **Parameterizing the regularization problem:** We define P^* as the result of an optimization problem with the following structure:

$$P^* \in \arg \min_{Q \in \Omega^V} \{ \Phi(P, Q) + \lambda \Psi(Q) \}, \quad (3)$$

where Φ is the *fidelity term*, Ψ the *regularizer*, $\lambda > 0$ the *regularization strength*, and Ω the search space. The fidelity term $\Phi(P, Q)$ enforces the influence of the initial labeling P , in the sense that it decreases as Q is closer to P . The regularizer Ψ favors solutions that are spatially smooth, in the sense that most adjacent nodes share the same label. The regularization strength λ is a user-defined parameter which dictates the influence of the regularization with respect to the fidelity term. In Section 4, we present the respective advantages of four fidelity terms, two regularizers, and two different search spaces.

- **Solving the optimization problem:** The choice of the minimizing algorithm to solve the regularization problem (3) hinges on the respective properties of the fidelity term and regularizer functions. In particular, we distinguish three settings: combinatorial, convex continuous and non-convex continuous problems. In Section 5, we present efficient, state-of-the-art algorithm for each case.

2. Related work

In recent years, a lot of attention has been paid to the semantic classification of 3D point clouds. Many investigations focus on a pointwise classification (Section 2.1) which serves as initial labeling for our framework. Spatially smooth labelings are subsequently obtained by computing an adjacency structure (Section 2.2) which allows for contextual classification (Section 2.3).

2.1. Semantic classification of 3D point clouds

The classic approach for point cloud classification is to treat each point individually by extracting a set of handcrafted features describing that point and using the respective feature vector as input for a standard supervised classification algorithm. Consequently, much effort has been spent on feature extraction and the classification procedure itself. In the following subsections, we summarize the main ideas behind both aspects.

A variety of handcrafted 3D shape features derived from the 3D structure tensor have been presented in different investigations (West et al., 2004;

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Pauly et al., 2003). Those features are advantageously completed by further characterization of the local 3D structure, e.g. in terms of angular statistics (Munoz et al., 2009), height and plane characteristics (Mallet et al., 2011; Guo et al., 2015), low-level 3D and 2D features (Weinmann et al., 2015a), or moments and height features (Hackel et al., 2016b).

The definition of an appropriate local neighborhood that comprises the local 3D structure is a crucial issue as it serves as the basis for feature extraction. Such neighborhoods are parameterized with a single parameter, commonly referred to as the *scale* and typically represented by the radius (Lee and Schenk, 2002; Filin and Pfeifer, 2005) or the number of nearest neighbors considered (Linsen and Prautzsch, 2001). To avoid invoking prior knowledge about the scene, a data-driven solution for selecting the optimal neighborhood size of each point is desirable. Respective approaches are for instance based on the local surface variation (Pauly et al., 2003; Belton and Lichti, 2006) and the combined consideration of curvature, point density and noise of normal estimation (Mitra and Nguyen, 2003; Lalonde et al., 2005). Further approaches have been presented with dimensionality-based scale selection (Demantké et al., 2011), and eigenentropy-based scale selection (Weinmann et al., 2015a; Weinmann, 2016).

Other approaches focus on the computation of local 3D features at different scales. In this regard, it has been proposed to consider a collection of spherical neighborhoods (Brodu and Lague, 2012), a collection of cylindrical neighborhoods (Niemeyer et al., 2014), a combination of cylindrical and spherical neighborhoods (Blomley et al., 2016), a combination of neighborhoods in the form of voxels, blocks and pillars (Hu et al., 2013), or a combination of neighborhoods in the form of spatial bins, planar segments and local neighborhoods (Gevaert et al., 2016). A different strategy has been followed with the generation of a scale pyramid by repeated downsampling via a voxel-grid filter (Hackel et al., 2016b), which allows for calculating features based on a fixed, small number of nearest neighbors for each of these scales.

Depending on the system used for data acquisition, further types of data can be recorded in addition to spatial coordinates. Accordingly, complementary types of features can be derived based on the additional data, e.g. echo-based features (Chehata et al., 2009; Mallet et al., 2011; Waldhauser et al., 2014), full-waveform features (Chehata et al., 2009; Mallet et al., 2011) or radiometric features (Niemeyer et al., 2014; Schmidt et al., 2014).

The extracted features are concatenated to feature vectors that are pro-

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

vided as input to a classifier. In most cases, the focus is put on supervised classification. Accordingly, representative training data is required to train the involved classifier so that it is afterwards able to generalize to unseen data and thus able to assign a (semantic) class label to each point of the point cloud.

210 A variety of techniques may be applied for supervised classification, such as random forest classifiers (Chehata et al., 2009), support vector machine classifiers (Mallet et al., 2011), or Bayesian discriminant analysis classifiers (Khoshelham and Oude Elberink, 2012). Those classifiers are rather easy to use and meanwhile available in numerous software tools. A variety of such standard classifiers has recently been involved in a comprehensive study focusing on the classification of mobile laser scanning data (Weinmann et al., 2015a; Weinmann, 2016), where the derived results reveal that a random forest classifier provides a good trade-off between classification accuracy and computational efficiency. However, due to the pointwise consideration relying only on a feature vector per point, the labeling derived with such standard classifiers typically lacks spatial regularity, i.e. the classified point cloud typically reveals a “noisy” behavior, although it should be taken into account that class labels of neighboring 3D points tend to be correlated.

220
225 Since the regularization framework proposed in this paper is independent of the choice of the method used for the initial labeling, we focus on the use of standard techniques. We use an existing classification framework² presented by Weinmann et al. (2015a), which is based on the use of a variety of low-level geometric 3D and 2D features as input for a standard random forest classifier. The choice of this classification framework is motivated by the fact that it (1) focuses on a data-driven neighborhood recovery and is thus applicable for different point clouds without involving prior knowledge about the scene and/or the data, (2) exploits a set of informative features that are still interpretable, (3) already provides a reasonable initial labeling, and (4) can easily be adapted to produce a soft labeling rather than a hard one.

2.2. Graph structure of point clouds

Statistical context models are commonly used for modeling the relationship between neighboring points, and, consequently, imposing spatial reg-

²Respective implementations can be found at <http://www.ipf.kit.edu/code.php>.

1
2
3
4
5
6
7
8
9
240 ularity on the semantic labelings of 3D points clouds. In general, context
10 models are based on the construction of an adjacency graph defining the extent
11 to which interactions in a local neighborhood are considered, i.e. it is
12 possible to take into account short-, mid- and long-range dependencies. It
13 is important to note that this adjacency graph is in general not the same as
14 the neighborhood graph used to compute local geometric features.
15
245

16
17 There are numerous approaches to obtain such adjacency graphs. The
18 most common is to derive the graph from a neighborhood relationship, from
19 nearest neighbors graph (Shapovalov et al., 2010) to cylindrical (Filin and
20 Pfeifer, 2005; Niemeyer et al., 2014) or adaptive neighborhoods (Demantké
21 et al., 2011; Weinmann et al., 2015b). Thereby, a simplifying assumption is
22 typically made by only considering short-range dependencies. This is moti-
23 vated by the observation that the quality of derived classification results
24 reveals a saturation effect when increasing the scale parameter of the local
25 neighborhood used for defining the adjacency graph. In this regard, the av-
26 erage number of involved neighbors was 7 in an investigation focusing on the
27 classification of airborne laser scanning data (Niemeyer et al., 2011, 2014).
28 Similar observations have been made in an investigation focusing on the clas-
29 sification of mobile laser scanning data (Weinmann et al., 2015b), where the
30 consideration of short-range dependencies already delivers classification re-
31 sults of high quality. However, the latter investigation also indicates that
32 adapting the size of the local neighborhood used for defining the adjacency
33 graph with respect to the locally-adaptive neighborhood used for feature
34 extraction is favorable in comparison to fixed neighborhoods. In these ex-
35 periments, the average number of involved neighbors was between 15 and 21,
36 depending on the approach used for deriving locally-adaptive neighborhoods.
37
265

38
39 Instead of defining context models on the basis of neighboring points,
40 different entities may be used as well. In this regard, several investigations
41 advocate a super-voxel-based approach to represent the higher order struc-
42 ture (Lim and Suter, 2009; Niemeyer et al., 2016; Guignard and Landrieu,
43 2017).
44
270

45 46 47 48 49 *2.3. Spatially smooth labeling*

50
51 To derive a labeling with a higher spatial regularity, smooth labeling
52 techniques (Schindler, 2012) or approaches for contextual classification can
53 be used. The latter consider an initial labeling derived with a standard
54 classifier and use a statistical context model to increase spatial regularity.
55
275
56 Thereby, the classification of a given point does not only take into account
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9 the feature vector corresponding to this considered point, but also the labels
10 corresponding to neighboring points as well.

11
12 Respective approaches have for instance been used in the form of associa-
13 tive Markov networks (Munoz et al., 2009), non-associative Markov networks
14 (Shapovalov et al., 2010), conditional random fields (Niemeyer et al., 2014;
15 Schmidt et al., 2014; Weinmann et al., 2015b; Landrieu et al., 2017), simpli-
16 fied Markov random fields (Lu and Rasmussen, 2012), multi-stage inference
17 procedures focusing on point cloud statistics and relational information over
18 different scales (Xiong et al., 2011), and spatial inference machines modeling
19 mid- and long-range dependencies inherent in the data (Shapovalov et al.,
20 2013).
21
22

23 Some of the presented approaches rely on the consideration of point cloud
24 segments, e.g. (Shapovalov et al., 2010; Xiong et al., 2011), whereas others
25 directly classify points, e.g. (Niemeyer et al., 2014). In this paper, we take
26 into account that segment-based methods heavily depend on the quality of
27 the results of the segmentation algorithm, and we therefore focus on the
28 regularization of point clouds without the use of pre-segmentations or super-
29 voxels.
30
31

32 In pairwise context models, such as CRFs or MRFs, retrieving the most
33 likely spatially smooth labeling is referred to as maximum-a-posteriori in-
34 ference. In practice, this labeling can only be approximated, using efficient
35 combinatorial optimization techniques, notably based on graph-cuts. This
36 form of inference produces a hard labeling, and hence loses the probabilistic
37 nature of the initial labeling.
38
39

40 Alternatively the label distribution can be computed for each node with
41 marginal inference, which allows us to keep the probabilistic nature of the
42 classification. However, marginal inference is typically approximated with
43 message-passing algorithms such as loopy belief propagation (Niemeyer et al.,
44 2011, 2014; Weinmann et al., 2015b), which are significantly slower and lead
45 to classifications of lower likelihood and less accurate classification results,
46 as detailed in (Landrieu et al., 2017).
47
48

49 Lellmann et al. (2013) propose a variational approach, in which the dis-
50 crete label set is relaxed into a continuous one. This approach permits the
51 use of convex optimization algorithms, and yields results with arguably less
52 artifacts than the solutions of combinatorial optimization techniques. The
53 article shows that, with reasonable assumptions, the solution of the opti-
54 mization can easily be discretized into a spatially smooth labeling.
55
56
57
58

3. Probabilistic classification of point clouds

Our framework requires first computing a pointwise probabilistic classification $P \in \mathcal{S}^V$. While the exact method used to obtain said classification does not impact its smoothing, provided it is reasonably good, we present the fundamentals of pointwise semantic labeling for the sake of completeness. We follow a slightly modified version of the framework presented by Weinmann et al. (2015a), in which the output is a soft labeling rather than a hard one. In the following, we briefly address the recovery of a local neighborhood for each 3D point (Section 3.1) which allows to compute low-level geometric 2D and 3D features based on the point statistics within (Section 3.2), and finally the supervised classification based on the extracted features (Section 3.3).

3.1. Recovery of local neighborhoods

To appropriately describe the local 3D structure at a considered point of the point cloud, we take into account that we are dealing with mobile laser scanning data acquired in urban environments. In such scenarios, the point density varies strongly according to the distance of the target such that it is advisable to use a spherical neighborhood definition relying on a scale parameter in the form of either a radius or the number of nearest neighbors that are considered. To allow for flexibility with respect to the given data, we focus on a data-driven approach to determine neighborhood size by selecting the number of nearest neighbors in the local 3D neighborhood of each individual point with *eigenentropy-based scale selection* (Weinmann et al., 2015a). This approach has proven to compare favorably to a variety of other approaches, and neither involves parameter tuning nor prior knowledge about the scene.

More specifically, for varying values of the scale parameter $s \in \mathbb{N}$, we use the 3D coordinates of each point and its s nearest neighbors to derive the respective 3D structure tensor $\mathbf{T} \in \mathbb{R}^{3 \times 3}$ as a function of the scale parameter s . The 3D structure tensor \mathbf{T} is a symmetric positive semi-definite matrix, i.e. its three eigenvalues exist, are non-negative and correspond to an orthogonal system of eigenvectors. Once normalized by their sum, the eigenvalues $\mu_1(s) \geq \mu_2(s) \geq \mu_3(s) \geq 0$ can be considered as “quasi-probabilities”, allowing us to define an energy function for optimal neighborhood size selection on the basis of the Shannon entropy:

$$E_\mu(s) = - \sum_{j=1}^3 \mu_j(s) \ln \{\mu_j(s)\}. \quad (4)$$

This energy function of the scale parameter s is known as the *eigenentropy* and describes the order/disorder of 3D points within the local 3D neighborhood. We select the parameter s_{opt} by minimizing the eigenentropy $E_{\mu}(s)$ across varying values of the scale parameter s :

$$s_{\text{opt}} = \arg \min_{s \in \mathcal{K}} E_{\mu}(s). \quad (5)$$

In the scope of our work, we test different values of s within a predefined set, with a lower boundary of $s_{\text{min}} = 10$ neighbors to remain statistically meaningful (Demantké et al., 2011; Weinmann et al., 2015a; Weinmann, 2016) and an upper boundary of $s_{\text{max}} = 100$ to limit the computational effort.

3.2. Extraction of low-level geometric 3D and 2D features

In the scope of this work, we describe each 3D point by considering all points within its local neighborhood of optimal size and calculating the respective values for a set of handcrafted geometric features proposed in (Weinmann et al., 2015a; Weinmann, 2016). These features are rather intuitive, and each feature is only represented by a single value.

The considered feature set comprises 14 low-level geometric 3D features. Eight of them are derived from the normalized eigenvalues μ_j and represented by linearity, planarity, sphericity, omnivariance, anisotropy, eigenentropy, sum of eigenvalues and change of curvature (West et al., 2004; Pauly et al., 2003). The other features are derived from the optimal neighborhood itself and given by the height of the considered point, 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 corresponding to those points within the local neighborhood.

To take into account the particular role played by the vertical dimension in urban environments, we also consider 2D features defined in analogy to the 3D case. More specifically, we use the normalized eigenvalues of the 2D structure tensor derived from the 2D projections of a considered point and its s_{opt} nearest neighbors onto a horizontally oriented plane. We consider the 2D features determined by the sum and ratio of these normalized eigenvalues, as well as the radius of the local neighborhood and the local point density in the 2D projection.

The derived values for all features extracted for a point are finally concatenated to a feature vector. Taking into account that the defined features

correspond to different quantities with different units, we introduce a normalization across all feature vectors which maps each dimension onto the interval $[0, 1]$. Thereby, the mapping function is defined based on the training data.

3.3. Supervised classification

To obtain an initial labeling based on the derived feature vectors, we focus on *ensemble learning* which relies on the idea of strategically generating a set of weak learners and combining them in order to create a single strong learner. A rather intuitive combination of weak learners is realized via bagging (Breiman, 1996), where bootstrapped replica of the training data (i.e. randomly drawn subsets) are used to train a set of weak learners of the same type. As each of the weak learners is trained on an independent subset, the weak learners are all randomly different from one another. This, in turn, results in a de-correlation between individual hypotheses and thus an improved generalization and robustness may be expected when taking the respective majority vote over all hypotheses (Criminisi and Shotton, 2013).

The most popular example for bagging is represented by a random forest classifier (Breiman, 2001) which relies on a set of N_T decision trees as weak learners and typically yields a good trade-off between accuracy and computational effort (Weinmann et al., 2015a; Weinmann, 2016). To obtain an initial global soft labeling, we use a random forest classifier whose parameters are cross-validated, and define the probability that a point i belongs to the class k as:

$$P_{i,k} = \frac{N_k}{N_T},$$

where N_k is the number of decision trees having voted for class k . This soft labeling in the form of classwise probabilities may not be spatially regular, and hence may be used as the basis for a subsequent regularization as we explain in the following section.

4. Regularizing soft labelings on a weighted graph

We consider $P \in \mathcal{S}^n$ a global soft labeling of a 3D point cloud and seek an alternate labeling P^* with increased spatial regularity. In that regard, we define P^* as the solution of the structured optimization problem (3) with well-chosen search space, fidelity terms, and regularizing functions. This problem is said to be *structured*, as both fidelity terms and regularizers have a specific

1
2
3
4
5
6
7
8
9 form derived from an adjacency graph $G = (V, E, w)$, defined in Section 4.1.
10
11 400 In the following, we present and discuss the respective properties of different
12 options for the search space (Section 4.2), the fidelity term (Section 4.3), and
13 the regularizer (Section 4.4).
14

15 4.1. Adjacency graph of point clouds

16
17 In this paper, we focus on the regularization of semantic labelings, while
18 405 investigating the respective benefits of different adjacency graphs is beyond
19 the scope of our work. We chose a symmetrized 10-neighborhood adjacency
20 graph with constant edge weight for its simplicity of implementation (Indyk
21 and Motwani, 1998). However, our framework can naturally handle graphs
22 with weighted edges, and we directly incorporate edge weights into our ob-
23 jective function.
24
25 410

26 4.2. Search space for probabilistic labelings

27
28 In this paper, we restrict the choice of Ω to hard or soft labelings: $\Omega = \mathcal{S}$
29 or $\Omega = \mathbf{S}$. While hard labelings assign a unique class, soft labelings assign a
30 probability for each class, and consequently contain more information. How-
31 415 ever, producing a hard labeling for each point remains the main objective of
32 semantic classification.
33

34 The most straightforward way to produce a hard labeling from a soft
35 labeling is to assign the label which has the highest probability for each
36 point independently, assuming it is unique. If it is not, the label can be
37 chosen arbitrarily from the classes of highest probability, with the lowest
38 index for example. In other words, for a soft labeling $P \in \mathcal{S}^n$, we define the
39 associated hard labeling $\hat{P} \in \mathbf{S}^n$ such that for all nodes $i \in V$:
40
41
42

$$43 \hat{P}_{i,k} = \begin{cases} 1 & \text{if } k = \min \arg \max_{l \in \mathcal{K}} P_{i,l} \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

44
45
46
47
48 The main advantage of choosing $\Omega = \mathcal{S}$ over \mathbf{S} is that a soft labeling
49 allows the confidence assessment of the associated hard labeling through
50 the computation of its entropy. This can be useful when the focus of the
51 420 classification is precision rather than the full coverage of the point cloud. In
52 such circumstances, the global labeling can be sorted by increasing entropy,
53 ensuring that the first points have higher confidence.
54
55
56
57
58

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

4.3. Fidelity terms for regularizing distributions

The fidelity term $\Phi(P, \cdot) : \mathcal{S}^n \mapsto \mathbb{R}$ of the optimization problem defined in (3) enforces the influence of the soft labeling P in \mathcal{S}^n . In the scope of this paper, we focus on fidelity terms that are separable with respect to V :

$$\Phi(P, Q) = \sum_{i \in V} \phi(P_i, Q_i),$$

where $\phi(p, \cdot) : \mathbf{S} \mapsto \mathbb{R}$ is a smooth and convex function called the *fidelity function*. Such a function $\phi(p, q)$ must be minimal for $q = p$, and increases as q differs from p . It is important to note that the fidelity function must be defined on the convex domain \mathcal{S} , but can be restricted to $\Omega = \mathbf{S}$ since $\mathbf{S} \subset \mathcal{S}$.

In this section, we present four different choices for ϕ . This list is not comprehensive, and could be extended with the Riemannian distance (Aström et al., 2016), or non-differentiable norms such as the L_1 -norm or its variant presented in (Huber, 1964).

4.3.1. Linear fidelity

The *linear fidelity* is traditionally used as a convex relaxation of unary potentials for labeling problems. We define it here as the opposite of the scalar product with the observed probability p :

$$\phi_{\text{linear}}(p, q) \doteq -\langle p, q \rangle = -\sum_{k \in \mathcal{K}} p_k q_k. \quad (7)$$

In accordance with general results of linear programming, the linear fidelity function encourages solutions that lie in the corner of the feasibility set, i.e. \mathbf{S} , as illustrated in Figure 2a. Although the choice of the regularizer can alter this behavior, this fidelity function should be used when a hard labeling is preferred.

The main advantage of this fidelity function is its simplicity: it is a simple scalar product and its gradient is constant. However linearly factoring the observed probability might be too simplistic, in particular when it comes to low observed probability. For example, the penalty for choosing two labels with probability 0.5 is the same than choosing one label with probability 0.

4.3.2. Linear-logarithmic fidelity

The *linear-logarithmic fidelity* is defined as the opposite of the scalar product with the logarithm of the observed probability p :

$$\phi_{\log}(p, q) \doteq -\langle q, \log(\hat{p}) \rangle = -\sum_{k \in \mathcal{K}} q_k \log(\hat{p}_k), \quad (8)$$

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

where \log denotes the entrywise logarithm and \hat{p} is a version of the observed probability which is smoothed to prevent numerical issues: $\hat{p}_k = \frac{\alpha}{K} + \alpha p_k$ with $\alpha \in [0, 1]$. Due to its linearity, this function tends to induce a hard labeling as well, as illustrated in Figure 2a. We remark that the fidelity term of the log-likelihood in graphical models such as MRFs or CRFs corresponds to this fidelity function. However, the probabilistic modeling setting is restricted to hard labelings for q while we extend it to the simplex \mathcal{S} .

This fidelity, while still simple to compute, necessitates the tuning of a supplementary tuning parameter. In our experiments, the influence of α was fairly minimal and we chose $\alpha = 0.05$ across all experiments. The main advantage of this function is that it heavily penalizes choosing labels with low probability.

4.3.3. Quadratic fidelity

The *quadratic fidelity* corresponds to the sum of squared differences (SSD) between distributions:

$$\phi_{\text{quadratic}}(p, q) \doteq \|p - q\|^2 = \sum_{k \in \mathcal{K}} (p_k - q_k)^2. \quad (9)$$

Unlike the two linear functions presented above, this fidelity function does not favor hard labelings, and it hence retains the probabilistic nature of P . The penalty is proportional to the Euclidian distance on the simplex between the observed and the assigned probability, as represented in Figure 2b.

4.3.4. Kullback-Leibler fidelity

The *Kullback-Leibler (KL) fidelity* relies on the Kullback-Leibler divergence $\text{KL}(p, q)$ which has been introduced in (Kullback and Leibler, 1951) as a measure of similarity between two distributions p and q :

$$\text{KL}(p, q) \doteq \sum_{k \in \mathcal{K}} p_k \log \left(\frac{p_k}{q_k} \right) = - \sum_{k \in \mathcal{K}} p_k \log (q_k) + \text{function of } p. \quad (10)$$

Since we are only interested in optimizing the fidelity function with respect to q , we can discard the constant part of the previous equation. As with the linear-logarithmic fidelity, we smooth both p and q using a convex combination with the uniform distribution parameterized by $\alpha \in [0, 1]$:

$$\phi_{\text{KL}}(p, q) \doteq - \sum_{k \in \mathcal{K}} \hat{p}_k \log (\hat{q}_k), \quad (11)$$

with $\hat{p}_k = \frac{\alpha}{K} + (1 - \alpha)p_k$ and likewise $\hat{q}_k = \frac{\alpha}{K} + (1 - \alpha)q_k$.

This fidelity term is better suited for comparing distributions than the quadratic fidelity. Indeed, it penalizes more heavily the disparity between p and q if p is a confident labeling, meaning that one label dominates the other. On the other hand, if p is such that all classes have similar probabilities, then differences between q and p should be less penalized. This property, illustrated in Figure 2c, reflects that the observed probability should be most influential when confident, while regularity should be the deciding factor for ambiguous labelings.

4.4. Penalizers inducing spatial regularity

The regularizer $\Psi(\cdot) : \mathcal{S}^n \mapsto \mathbb{R}$ favors solutions of (3) which are spatially smooth, in the sense that most adjacent nodes in the graph G share the same label. In this section, we present two popular spatial regularity-inducing penalizers and their respective properties. For this purpose, we consider the graph $G = (V, E, w)$ defined in Section 4.1. As with the fidelity terms, all penalizers are defined on the convex domain.

We define a global labeling Q as spatially smooth if the number of non-zeros values in $\{x_i - x_j \mid (i, j) \in E\}$ is small compared to the number of edges. Indeed, for such a labeling, most nodes are surrounded by neighbors of the same label. Such a labeling is constant with respect to a partition of G which is *coarse*, i.e. with a number of constant connected components that is small with respect to the number of nodes.

We restrict ourselves to regularizers that factorize over the edges of G , i.e. that can be written under the following form:

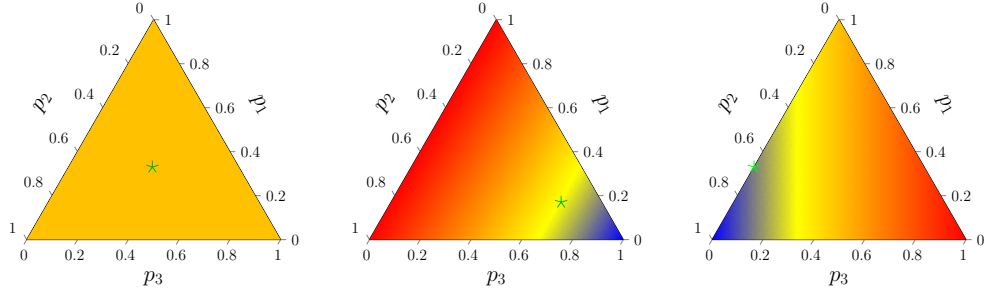
$$\Psi(Q) \doteq \sum_{(i,j) \in E} w_{ij} \psi(Q_i - Q_j), \quad (12)$$

with $\psi : \mathbb{R}^{\mathcal{K}} \mapsto \mathbb{R}$ a functional minimal in 0, encouraging spatially smooth solutions. Regularizers of this form were first introduced by Geman and Reynolds (1992), and include many of the most commonly used spatial regularity-inducing penalties.

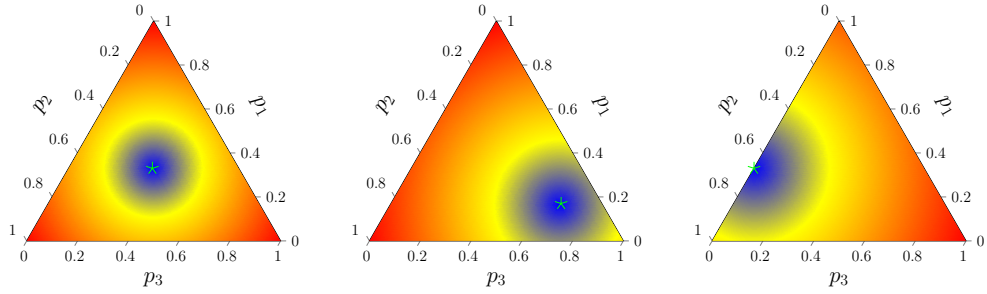
4.4.1. Potts penalty

Pairwise graphical models such as MRFs and CRFs encode the influence of the context with an interaction potential between adjacent nodes whose value is zero when the labels are identical and strictly non-negative when

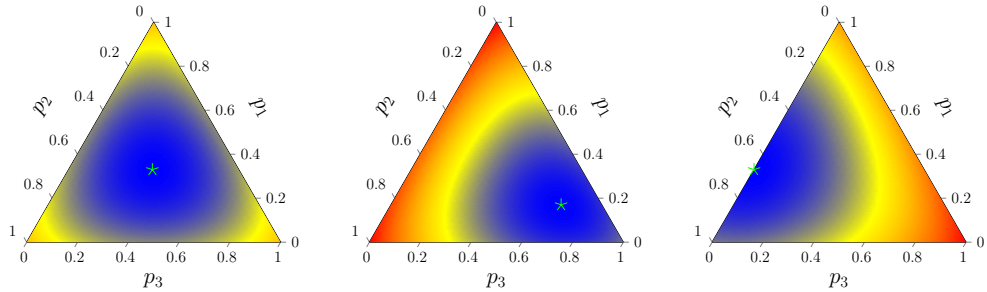
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65




(a) Linear and linear-logarithmic fidelity



(b) Quadratic fidelity



(c) Kullback-Leibler fidelity

Figure 2: Surface plot of the fidelity functions over the simplex for $|\mathcal{K}| = 3$. The observed distribution p is represented by a green star \star , while the value of the fidelity function of the corresponding point of the simplex is represented with the following normalized colormap: low  high. We remark that the quadratic divergence illustrated in (b) only takes the radial distance into account. On the other hand, the linear fidelities illustrated in (a) are minimal at the simplex corner closer to the observed distribution, or constant if the distribution is uniform, as seen on the top left figure. Finally, the confidence of the observed distribution is taken into account when estimating the Kullback-Leibler fidelity illustrated in (c).

they are different (Potts, 1952). This translates into choosing $\Omega = \mathbf{S}$ and a functional ψ_{Potts} equal to 0 in 0 and 1 everywhere else:

$$\psi_{\text{Potts}}(d) = \begin{cases} 0 & \text{if } q = 0 \\ 1 & \text{else.} \end{cases} \quad (13)$$

This functional can naturally be extended to the case where $\Omega = \mathcal{S}$. In this case, the *Potts penalty* corresponds to the total weighted cut between constant components of G , and is referred to as the total boundary size.

4.4.2. Total Variation

The *graph total variation* (TV) can be seen as a convex relaxation of the above Potts penalty, which, to a certain extent, can also enforce piecewise-constant solutions (Rudin et al., 1992). Its definition depends on the context but usually consists in setting ψ as a norm over $\mathbb{R}^{\mathcal{K}}$.

Over vector spaces, the use of an Euclidean norm is often considered, enjoying theoretical isotropy. Here, however, ψ is applied to the difference between two discrete distributions, for which the notion of isotropy is not relevant. Observe that the use of an Euclidean norm would enforce equality of neighboring distributions only as a whole, that is to say in the solution set either the two distributions are exactly the same, or they differ over each label for which the observation P differs. As a consequence, if two neighboring labelings in P disagree over the probability of one given label, and thus their equality is not favored in the solution set, then equality of the discrete probabilities for the other labels would not be favored either.

Although we do not investigate its practical advantage further, we prefer using an ℓ_1 -norm, which is the separable sum of the absolute values of the finite differences, thus favoring the equality of all neighboring discrete probabilities more independently:

$$\psi_{\text{TV}}(q) = \sum_{k \in \mathcal{K}} |q_k|. \quad (14)$$

5. Graph-structured optimization

The choice of a fitting algorithm to minimize objective functionals of the form (3) hinges on the respective properties of the fidelity and regularizing functions as well as the search space. We distinguish three settings, as they necessitate vastly different approaches to be solved:

- combinatorial;
- continuous space and non-convex functional;
- continuous space and convex functional.

5.1. Combinatorial

If the search space $\Omega = \mathbf{S}$ is discrete, the problem is said to be *combinatorial*. The sheer number of combinations and the lack of continuity prevent the retrieval of a global minimizer in general. When considering only two labels however, the objective functional can be solved with graph cuts algorithms due to its submodularity (Boykov et al., 2001).

If the number of labels exceeds two, the functional is no longer submodular and can only be approximately minimized. The α -expansion algorithm introduced in (Boykov and Kolmogorov, 2004) allows us to approximately solve such problems through a sequence of binary labeling problems, which can in turn be solved efficiently with graph cuts. This algorithm is widely used because of its performance, its theoretical guarantees and the availability of its implementation.

5.2. Continuous space and non-convex functional

We consider the case when $\Omega = \mathcal{S}$ is continuous, but the regularizer is non-convex, typically the Potts penalty extended to \mathcal{S} . In this setting, no guarantee on the global optimality can be established, however numerous approximated algorithms exist. A first approach proposed by Ishikawa (2003) is to discretize the search space and to treat the problem as a combinatorial one. A more recent approach proposed by Landrieu and Obozinski (2016b) allows us to keep the continuity of the search space and provides better results with fewer cuts. As this algorithm has only been presented for one-dimensional values, we focus on a natural extension to multi-dimensional, simplex-constrained values in the following.

This greedy algorithm, dubbed ℓ_0 -cut pursuit, exploits the fact that spatially smooth labelings can be broken down into a small number of constant connected components to accelerate the resolution of the corresponding optimization problem. The ℓ_0 -cut pursuit algorithm maintains a current partition of the graph in which the nodes of each component share the same value. This partition is initialized such that all the nodes are in the same component, and is then refined by computing binary partitions, called *optimal binary*

1
2
3
4
5
6
7
8
9
550 *cuts*, and enumerating their connected components. A backward-step is then
10 performed to check if merging existing adjacent components can decrease the
11 objective function.
12

13 However, in (Landrieu and Obozinski, 2016b), ℓ_0 -*cut pursuit* is only de-
14 fined in the one-dimensional setting in which only one scalar value is as-
15 sociated with each node. In this paper, we extend this algorithm to a
16 555 multi-dimensional setting in which we associate a multi-dimensional, simplex-
17 constrained value to each node. This extension is made easy by the separa-
18 bility hypothesis of the fidelity term, which ensures that, given a partition
19 of V , the associated optimal distribution can be computed independently for
20 each component by minimizing the sum of the associated fidelity terms.
21
22 560

23 Furthermore, the four fidelity functions listed in Section 4.3 are such that
24 the constant distribution q_A minimizing the sum of fidelity terms for a subset
25 of nodes $A \subset V$ is also simplex-bound, and easy to compute. Indeed, for the
26 linear and linear-logarithmic fidelity, q_A is the hard labeling corresponding
27 to the class maximizing the sum of the distributions associated to the nodes
28 565 of A . For the quadratic and Kullback-Leibler fidelity, q_A is the average of
29 the distributions of the nodes of A .
30
31

32 Computing such piecewise-constant labelings is the critical step in each
33 of the three main steps of ℓ_0 -cut pursuit, namely the computation of the
34 570 optimal binary cuts and associated optimal distribution, and the backward
35 step. Consequently the extension of ℓ_0 -cut pursuit to multi-dimensional,
36 simplex-bound data can be implemented easily and remains very efficient ³.
37
38

39 5.3. Continuous space and convex functional 40

41 In this last setting, on top of the convexity of the search space $\Omega = \mathcal{S}$, we
42 575 consider a functional ψ (and hence Ψ), which is also convex. In order to favor
43 the sparsity of $\{x_i - x_j \mid (i, j) \in E\}$, and hence a small number of constant
44 connected components in the set of minimizers, the graph total variation is
45 however non-differentiable.
46

47 Given the level of uncertainty over the data and over the parameters of
48 580 our regularization framework, high precision is not required for the mini-
49 mization of the objective functional. We thus resort to first-order *proximal*
50 *splitting algorithms*, well-adapted to such large-scale situations where the
51 functional is a sum of simple terms. This approach has been considered for
52
53
54

55 ³A C++ implementation can be downloaded at www.loiclandrieu.com/.
56
57
58

1
2
3
4
5
6
7
8
9 instance in Lellmann et al. (2009), where the authors use a *Douglas–Rachford*
10 *splitting algorithm* to solve a specific instance of (3). Since this publication,
11 585 more powerful splitting schemes have been developed. Although primal-dual
12 schemes are popular thanks to their generality (Chambolle and Pock, 2011),
13 the *preconditioned generalized forward-backward splitting algorithm* (Raguet
14 and Landrieu, 2015) is more suited to our graph-structured problem, while
15 taking full advantage of the smoothness of the data-fidelity term. We refer
16 590 to the latter and references therein for more details ⁴.
17
18
19
20

21 6. Experimental Results

22
23 In this section, we first present the involved benchmark datasets (Sec-
24 tion 6.1), the considered evaluation metrics (Section 6.2) and the competing
25 595 methods (Section 6.3). Subsequently, we present the derived results (Sec-
26 tion 6.4). The experimental framework will be available at www.loiclandrieu.com/.
27
28
29

30 6.1. Datasets

31
32 Since our main goal consists in testing the applicability of the involved
33 600 methods and the reproducibility of derived results, we want to facilitate an
34 objective comparison to other methodologies. Hence, we test our framework
35 on three publicly available and labeled 3D point cloud datasets which are
36 described in the following subsections.
37
38

39 6.1.1. Oakland-5C Dataset and Oakland-3C Dataset

40
41 605 The *Oakland 3D Point Cloud Dataset*⁵ (Munoz et al., 2009) is a labeled
42 benchmark dataset which has often been used to evaluate approaches focus-
43 ing on a semantic labeling of 3D point clouds. This dataset has been acquired
44 in the vicinity of the CMU campus in Oakland, USA, with a moving plat-
45 form equipped with a side-looking Sick laser scanner used in push-broom
46 mode (Munoz et al., 2008). During data acquisition, the speed of the plat-
47 610 form reached up to 20km/h, and the acquired 3D point clouds reveal a point
48 density with significant variation. A separation of the dataset into training
49
50
51

52
53 ⁴Efficient implementations in C++, interfaced with MEX for MATLAB/GNU Octave
54 users, can be found at <https://www.ceremade.dauphine.fr/~raguet/pgfb/>.

55 ⁵The Oakland 3D Point Cloud Dataset is publicly available at http://www.cs.cmu.edu/~vmr/datasets/oakland_3d/cvpr09/doc/ (last access: 17 November 2016).
56
57
58

1
2
3
4
5
6
7
8
9 data (about 37k labeled 3D points) and test data (about 1.32M labeled 3D
10 points) is already provided.

11
12 615 The *Oakland-5C Dataset* refers to the provided reference labeling with re-
13 spect to five semantic classes. These classes are defined as *Wire, Pole/Trunk,*
14 *Façade, Ground* and *Vegetation*.

15
16 The *Oakland-3C Dataset* refers to the provided reference labeling with
17 respect to three structural classes. These classes are defined as *Linear Struc-*
18 *tures, Planar Structures* and *Volumetric Structures*.
19 620

20 For both the Oakland-5C Dataset and the Oakland-3C Dataset, the num-
21 ber of training examples per class is very unbalanced which can have a detri-
22 mental effect on the training process (Chen et al., 2004; Criminisi and Shot-
23 ton, 2013). To avoid such effects, we introduce a class re-balancing which
24 relies on randomly selecting 1,000 labeled 3D points per class as new training
25 625 set and discarding all other points.
26
27

28 6.1.2. *Paris-rue-Cassette Database*

29 To include larger MLS datasets in our experiments, we also make use of
30 the *Paris-rue-Cassette Database*⁶ (Vallet et al., 2014), a point cloud dataset
31 630 which has been acquired in January 2013 with the mobile laser scanning
32 system called Stereopolis II (Paparoditis et al., 2012). This system involves
33 two plane sweep lidars of type *Riegl LMS-Q120i* and a 3D lidar of type
34 *Velodyne HDL-64E* to capture the local 3D geometry of the scene. The
35 Riegl devices are placed on each side of the vehicle and serve for observing
36 the building façades with a centimeter accuracy, whereas the Velodyne de-
37 vice mainly serves for observing the bottom part in between. In total, the
38 635 dataset contains 12M points corresponding to a street section with a length
39 of approximately 200m as well as a reference labeling which includes both
40 pointwise labels and segmented objects. The annotation has been carried
41 out by recovering a regular 2D topology for the point cloud stream during
42 data acquisition and an offline human interaction via a graph editing tool
43 based on standard 2D image segmentation techniques (Brédif et al., 2014).
44 640

45 In our experiments, we consider the seven dominant classes defined as
46 *Façade, Ground, Cars, 2-Wheelers, Road Inventory, Pedestrians* and *Vege-*
47 *tation*. All 3D points belonging to the other classes are removed as these
48 645
49
50
51
52

53
54 ⁶The Paris-rue-Cassette Database is publicly available at [http://data.ign.fr/
55 benchmarks/UrbanAnalysis/](http://data.ign.fr/benchmarks/UrbanAnalysis/) (last access: 17 November 2016).
56
57
58

1
2
3
4
5
6
7
8
9 classes are not considered as representative (Weinmann et al., 2015c). To
10 separate the dataset into training data and test data, we randomly select
11 1,000 labeled 3D points per class as training set and all remaining labeled
12 3D points as test set.
13
14

15 6.2. Evaluation metrics

16
17 To evaluate the performance of the benchmarked approaches, we com-
18 pare the derived labeling to the reference labeling on a per-point basis. For
19 this purpose, we consider both classwise and global evaluation metrics. The
20 classwise evaluation metrics are represented by recall (R), precision (P) and
21 F_1 -score. Whereas recall represents a measure of completeness or quantity,
22 F_1 -score. Whereas recall represents a measure of completeness or quantity,
23 precision represents a measure of exactness or quality. The F_1 -score is a
24 compound metric combining precision and recall with equal weights. The
25 global evaluation metrics are represented by overall accuracy (OA) and the
26 unweighted average of the F_1 -score over all classes (\bar{F}_1). In this regard, it
27 should be taken into account that a consideration of the overall accuracy
28 might not be sufficient if the number of examples per class is very inhomogeneous
29 for the test data. The indicator \bar{F}_1 allows judging about the quality
30 of classification results based on classwise evaluation metrics.
31
32

33 As stated in Section 4.2, the advantage of probabilistic labelings is that
34 their certainty can be estimated. To each point-level assignment we associate
35 a certainty measure by computing its entropy H . A low entropy designates a
36 high confidence assignment (for example $H([1\ 0\ 0]) = 0$), while a high entropy
37 denotes an ambiguous assignment (for example $H([\frac{1}{3}\ \frac{1}{3}\ \frac{1}{3}]) = \log(3)$). We
38 define the partial assignment at coverage $f\%$ as the fraction of an assignment
39 P when only considering the $f\%$ lowest entropy pointwise assignments, i.e.
40 only the most confident points. We can evaluate the \bar{F}_1 -score of such a partial
41 assignment by comparing it against the corresponding partial ground-truth.
42
43

44 To demonstrate the benefit of this confidence assignment, we provide the
45 *accuracy/coverage plots* of the best performing methods in Figure 3. Those
46 plots are obtained by sorting the points by increasing entropy, and computing
47 the accuracy of the partial assignment for coverage going from 70% to full
48 coverage.
49
50

51 6.3. Competing methods

52
53 In this subsection, we briefly summarize the benchmarked algorithms
54 (i.e. the considered configurations of our framework) and some state-of-the-
55 art methods that are involved for comparison. In Section 4 we listed two
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9 search spaces ($\Omega = \mathcal{S}$ or \mathbf{S}), four fidelity functions (*linear*, *linear-logarithmic*,
10 *quadratic* and *Kullback-Leibler*), as well as the two regularizers (*Potts penalty*,
11 *total variation*). Of the 16 possible combinations, only 8 are unique and
12 relevant. Indeed, when considering fidelity functions that induce a hard as-
13 685 signment (namely *linear* and *linear-logarithmic*) with the Potts penalty, the
14 choice of Ω is irrelevant as all solutions belong to \mathbf{S} . Consequently, ℓ_0 -cut
15 pursuit and α -expansion will minimize the same functional. As both ap-
16 proaches approximate the global solution, one could expect different results.
17 However in all our numerical experiments the difference in the final value
18 of the functional was small enough that its corresponding assignment was
19 690 almost identical.

20
21
22
23 Some simple calculus shows that for the Potts penalty, the values of the
24 linear, quadratic and Kullback-Leibler fidelities at the corners of the simplex
25 amount to the same penalty with a different regularization strength. Simi-
26 695 larly, we do not consider the total variation regularizer with a discrete search
27 space, as it is redundant with the Potts penalty. We list the 8 combinations
28 of regularizers and fidelity functions that correspond to unique algorithms in
29 Table 1.
30
31

32 700 To compare with the graphical models approach advocated by Niemeyer
33 et al. (2014) and based on a graphical model in the form of a CRF, we com-
34 pute the solutions provided by loopy belief propagation, both for the marginal
35 inference (LBP) and MAP-inference (LBP_MAP). The MAP-inference can
36 also be advantageously computed with α -expansion, as mentioned by Lan-
37 drieru et al. (2017), and corresponds to the *log_potts* shorthand in Table 1.
38 705 The implementations of the inference algorithms were obtained at (Schmidt,
39 2012).
40
41

42 The regularizing approach proposed by Lellmann et al. (2009) for image
43 labeling corresponds to *lin_TV*. However, we use the preconditioned general-
44 710 ized forward-backward splitting algorithm (PFDR) algorithm which is much
45 faster.
46
47

48 6.4. Experimental results

49
50 In Table 2, we provide the results of the full classifications for the meth-
51 ods invoked in Section 6.3, and the accuracy/coverage plot is represented
52 715 in Figure 3. The classwise results are displayed in Tables 4, 5, and 6. The
53 computation time, referenced in Table 3, is mainly dependent on the chosen
54 regularizer as it dictates the used algorithm.
55
56
57
58

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Method	Fidelity	Regularizer	Domain	Minimizing algorithm
lin_potts	linear	Potts penalty	\mathcal{S}	α -expansion
log_potts	linear-logarithmic	Potts penalty	\mathcal{S}	α -expansion
lin_TV	linear	total variation	\mathcal{S}	PFDR
log_TV	linear-logarithmic	total variation	\mathcal{S}	PFDR
l22_TV	quadratic	total variation	\mathcal{S}	PFDR
KL_TV	Kullback-Leibler	total variation	\mathcal{S}	PFDR
l22_bound	quadratic	Potts penalty	\mathcal{S}	ℓ_0 -cut pursuit
KL_bound	Kullback-Leibler	Potts penalty	\mathcal{S}	ℓ_0 -cut pursuit

Table 1: List of the benchmarked algorithms with their characteristics.

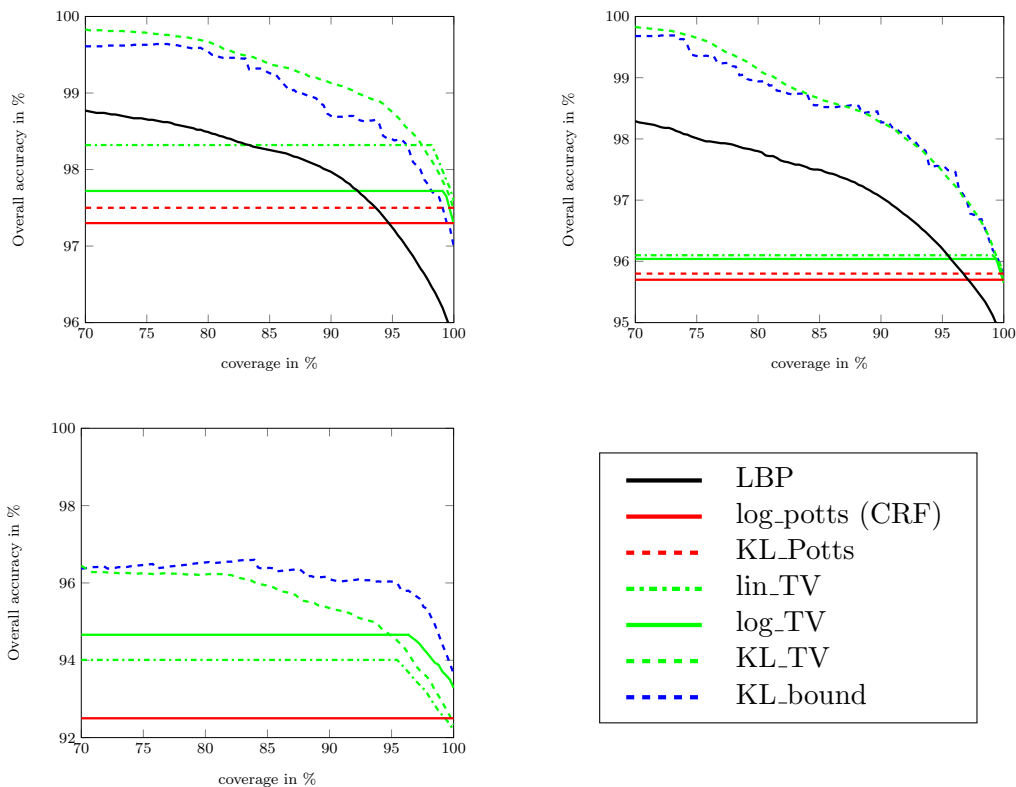


Figure 3: Accuracy/coverage plot for the three datasets and the top performing methods from 70% to 100% coverage for the Oakland-3C Dataset (top left), the Oakland-5C Dataset (top right) and the Paris-rue-Cassette Database (bottom left).

Method	Oakland-3C Dataset		Oakland-5C Dataset		Paris-rue-Cassette Database	
	OA	\bar{F}_1	OA	\bar{F}_1	OA	\bar{F}_1
pointwise	93.8	71.4	92.3	63.5	81.0	41.0
LBP	95.8	75.6	94.7	70.9	83.2	44.3
LBP_MAP	95.5	74.9	94.5	69.8	82.6	43.6
log_potts	97.3	78.2	95.7	74.4	92.5	65.4
lin_potts	97.5	78.9	95.8	75.1	92.3	61.3
l22_bound	97.2	78.3	95.7	73.6	93.6	65.6
KL_bound	97.0	78.3	95.7	74.7	93.7	64.4
lin_TV	97.5	78.7	95.7	74.9	92.4	61.2
log_TV	97.3	78.2	95.7	74.4	93.1	60.9
l22_TV	97.5	78.7	95.7	74.8	92.4	61.2
KL_TV	97.5	78.7	95.9	75.3	91.4	61.2

Table 2: Classification results (in %) derived with the considered methods for the three datasets. OA is the overall accuracy and \bar{F}_1 is the unweighted average of the \bar{F}_1 -scores over all classes.

Method	Oakland-5C Dataset (1.3M points)	Oakland-3C Dataset (1.3M points)	Paris-rue-Cassette Database (12M points)
LBP	45	35	720
LBP_MAP	54	31	800
α -expansion	24	15	400
ℓ_0 -cut pursuit	27	21	600
PFDR	72	43	2100

Table 3: Required time in seconds for solving the optimization problem depending on the dataset and the chosen algorithm on an i7-4790 CPU 3.60GHz with 8GB of RAM.

Method	OA	F_1	F_1 (Linear Structures)	F_1 (Planar Structures)	F_1 (Volumetric Structures)
pointwise	93.8	71.4	28.7	97.1	88.5
LBP	95.8	75.6	36.2	98.1	92.6
LBP_MAP	95.5	74.9	34.8	98.0	91.9
log_potts	97.3	78.2	40.1	98.8	95.6
lin_potts	97.5	78.9	41.6	98.9	96.0
l22_bound	97.2	78.3	40.5	98.8	95.5
KL_bound	97.0	78.3	41.1	98.7	95.0
lin_TV	97.5	78.7	41.0	98.9	96.0
log_TV	97.3	78.2	40.0	98.9	95.7
l22_TV	97.5	78.7	41.0	98.9	96.1
KL_TV	97.5	78.7	40.9	99.0	96.2

Table 4: Classification results (in %) for the Oakland-3C Dataset. We present the overall accuracy (OA), the unweighted average of the F_1 -score over all classes (\bar{F}_1), and the classwise F_1 -scores.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Method	OA	F_1	$F_1(\text{Wire})$	$F_1(\text{Pole/Trunk})$	$F_1(\text{Façade})$	$F_1(\text{Ground})$	$F_1(\text{Vegetation})$
pointwise	92.3	63.5	18.2	38	76.3	97.6	87.2
LBP	94.7	70.9	22.2	59.8	81.2	98.0	93.4
LBP_MAP	94.5	69.8	21.2	55.6	81.0	98.0	93.0
lin_potts	95.8	75.1	28.2	70.0	83.3	98.2	95.8
log_potts	95.7	74.4	29.1	65.7	83.4	98.2	95.7
l22_bound	95.7	73.6	26.5	64.7	83.1	98.1	95.9
KL_bound	95.7	74.7	29.4	67.3	82.9	98.2	95.7
lin_TV	95.7	74.9	27.5	69.3	83.4	98.1	95.9
log_TV	95.7	74.4	28.4	66.5	83.1	98.1	95.7
l22_TV	95.7	74.8	27.8	68.9	83.2	98.1	95.8
KL_TV	95.9	75.3	29.1	69.6	83.2	98.1	96.3

Table 5: Classification results (in %) for the Oakland-5C Dataset. We present the overall accuracy (OA), the unweighted average of the F_1 -score over all classes (\bar{F}_1), and the classwise F_1 -scores.

Method	OA	F_1	$F_1(\text{F})$	$F_1(\text{G})$	$F_1(\text{C})$	$F_1(\text{2W})$	$F_1(\text{RI})$	$F_1(\text{P})$	$F_1(\text{V})$
pointwise	81.0	41.0	85.4	96.8	45.4	10.6	10.7	5.0	33.3
LBP	83.2	44.3	87.0	97.6	55.5	14.3	12.8	6.4	36.4
LBP_MAP	82.6	43.6	86.6	97.4	54.3	13.9	12.7	5.6	35.0
log_potts	92.5	65.4	94.7	95.1	82.2	48.7	16.5	65.9	54.9
lin_potts	92.3	61.3	94.5	95.8	78.9	48.1	17.7	43.7	50.6
l22_bound	93.6	65.6	95.7	97.4	83.5	66.7	19.5	46.6	50.0
KL_bound	93.7	64.4	95.7	98.1	82.4	46.9	32.7	44.1	51.2
lin_TV	92.4	61.2	94.4	97.4	81.8	42.4	28.7	35.7	48.0
log_TV	93.1	60.9	95.0	98.2	82.9	41.2	29.8	29.2	50.1
l22_TV	92.4	61.2	94.4	96.5	80.5	44.4	23.1	38.9	50.5
KL_TV	91.4	61.2	93.7	93.9	76.4	48.3	18.6	43.8	53.6

Table 6: Classification results (in %) for the Paris-rue-Cassette Database with 7 classes represented by *Façade* (F), *Ground* (G), *Cars* (C), *2-Wheelers* (2W), *Road Inventory* (RI), *Pedestrians* (P) and *Vegetation* (V). We present the overall accuracy (OA), the unweighted average of the F_1 -score over all classes (\bar{F}_1), and the classwise F_1 -scores.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

720 To obtain a visual impression about the quality of the derived classification results, a visualization of the classified point clouds is provided for the Oakland-3C Dataset in Figure 4, for the Oakland-5C Dataset in Figure 5 and for the Paris-rue-Cassette Database in Figure 1. All these figures contain an illustration of the ground truth labeling, the initial labeling derived via pointwise classification, the labeling derived via structured regularization relying on the KL_bound method and the confidence of the derived labeling.

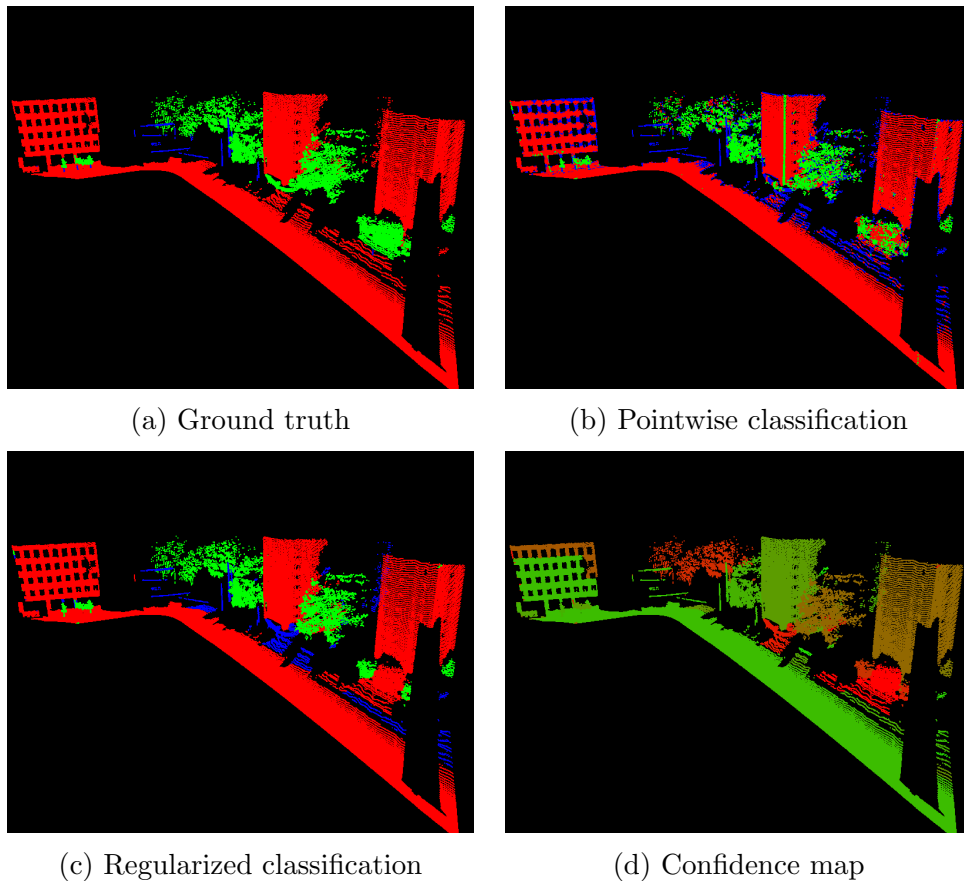



Figure 4: Visualization of a 3D point cloud labeling for a part of the Oakland-3C Dataset. In (a), (b), and (c), the color encoding addresses the classes *Linear Structures* (blue), *Planar Structures* (red) and *Volumetric Structures* (green). In (d), the confidence is represented from green to red: confident  uncertain. Remark that misclassifications in (c) correspond to the least confident area in (d).

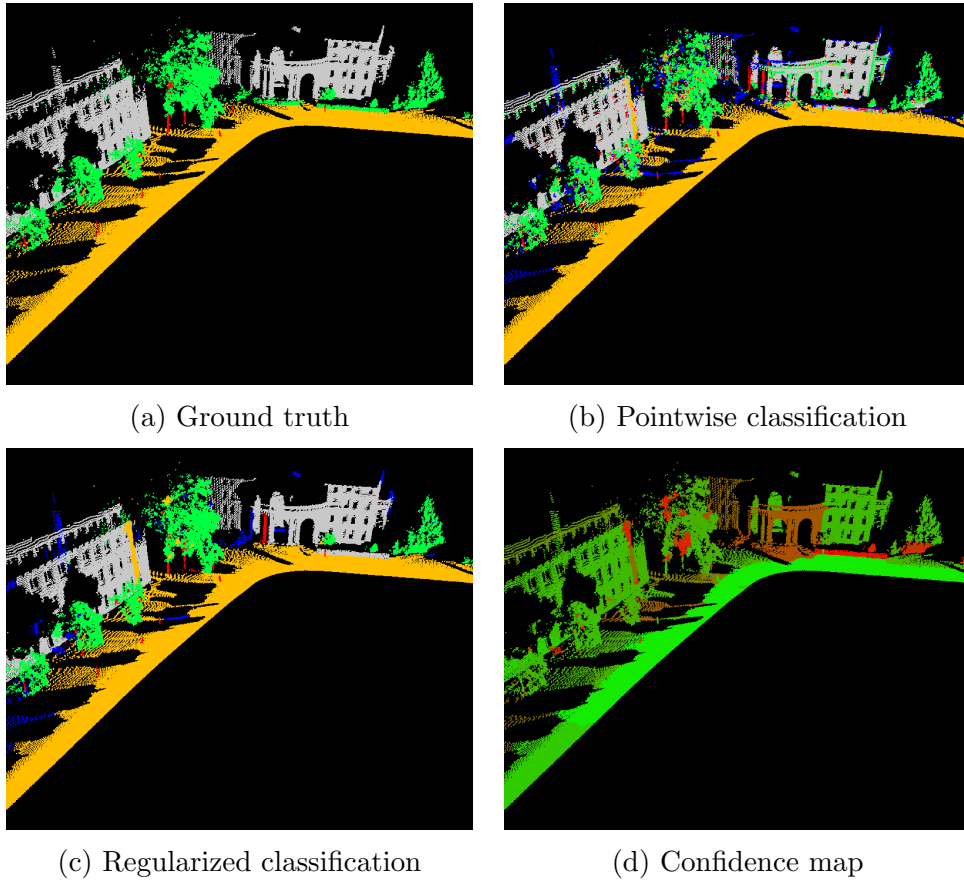



Figure 5: Visualization of a 3D point cloud labeling for a part of the Oakland-5C Dataset. In (a), (b), and (c), the color encoding addresses the classes *Wire* (blue), *Pole/Trunk* (red), *Façade* (gray), *Ground* (orange) and *Vegetation* (green). In (d), the confidence is represented from green to red: confident  uncertain. Remark that misclassifications in (c) correspond to the least confident area in (d).

725 6.5. Discussion

From Table 2 we can observe that regularization does indeed improve both the accuracy and the \bar{F}_1 -score of the pointwise classification. We observe that message-passing algorithms such as LBP and LBP_MAP underperform when compared to other approaches, while their computation time is among the highest, which is in accordance with the observations made by Landrieu et al. (2017).

1
2
3
4
5
6
7
8
9
10 In Tables 4, 5 and 6, we observe that although our framework barely
11 improves the F_1 -score of easy-to-classify classes such as *Ground* or *Façade*,
12 our methods display significant improvement over hard classes, such as *Wire*,
13 *Pole/Trunk*, *Cars*, *2-Wheelers* and *Pedestrians*. Indeed, the F_1 -score of *Wire*
14 (18.2% to 29.4%) and *Pole/Trunk* (38% to 70%) is almost doubled for the
15 Oakland-5C Dataset for the best performing methods. More impressively, in
16 the Paris-rue-Cassette Database our methods are able to retrieve decent clas-
17 sifications for classes that were mostly mislabeled by the pointwise labeling.
18 For example, *2-Wheelers* were initially classified with a 10.6% F_1 -score, but
19 the best regularized labeling boasts a classification score of 66.7%. Likewise,
20 regularization was able to improve the classification of the class *Pedestrians*
21 from 5.0% to 65.9%, and the classification of the class *Cars* from 45.4% to
22 83.5%.
23

24
25
26 We explain this large improvement over hard classes because regular-
27 ization removes isolated misclassified points scattered over the point cloud.
28 Those classes are also the least represented, their recall is particularly sen-
29 sitive to such misclassifications. The regularization increases the precision
30 as well by enforcing homogeneity of a tightly connected set of points, which
31 often belong to the same class. Remark that this will only improve the
32 classification if the initial labeling is mostly right to begin with.
33

34
35 Among the methods implemented in our framework, the difference of
36 performance is rather small, with a difference of less than 1% in accuracy.
37 The benefits in choosing a given configuration lies elsewhere, namely in its
38 computational load and the nature of the obtained smoothed assignment.
39

40 6.6. Choosing the fidelity

41
42 When combined with the same penalizer, the influence of the fidelity
43 function seems limited in terms of accuracy. However, this choice influences
44 qualitative properties of the solution.
45

46 For example, when combined with either the TV or boundary penalty,
47 the linear and linear-logarithmic fidelity yield hard smoothed assignments,
48 while the quadratic and KL fidelity yield probabilistic assignments. Both the
49 linear-logarithmic and Kullback-Leibler fidelity involve entrywise logarithms
50 of probabilities, which can induce numerical issues, and consequently require
51 a supplementary smoothing parameter. However, a smoothing parameter of
52 $\alpha = 0.05$ seems to yield good results in general, and does not require exten-
53 sive cross-validation. Finally, both quadratic and linear fidelities have fewer
54
55
56
57
58

1
2
3
4
5
6
7
8
9 parameters and easier computation since both their gradient and proximal
10 operator are very straightforward to obtain.

11
12 770 The linear-logarithmic fidelity takes into account the observed probabil-
13 ity in a non-linear way, penalizing assignments with low probability much
14 more than it favors assignments with high probability. This penalty should
15 hence be preferred when considering hard assignments. Conversely, the KL
16 fidelity takes the observed probability into account linearly, while penalizing
17 strong confidence outputs. This penalty should hence be preferred when a
18 775 probabilistic output is expected.

19
20 The choice of the fidelity should in general be cross-validated as a meta-
21 parameter. However, depending on the nature of the expected output, this
22 choice can be restricted.
23
24

- 25 780 • **Hard assignment expected:** When only the 100% coverage is rele-
26 vant for the application, both linear and linear-logarithmic fidelity can
27 be employed.
28
- 29 • **Soft assignment expected:** The quadratic and KL fidelity provide a
30 probabilistic output when combined with the TV or boundary penalty.
31
32

33 785 6.7. Choosing the penalty

34
35 The influence of the penalty is more drastic, as it impacts both the nature
36 of the output and the computational efficiency as well. The Potts penalty,
37 combined with the α -expansion algorithm, is the fastest of all approaches. In
38 terms of accuracy and \bar{F}_1 -score, it offers excellent performances as well, mak-
39 790 ing it a solid choice when only a hard assignment is expected. The boundary
40 penalty, when solved with the ℓ_0 -cut pursuit algorithm, is slightly slower to
41 compute, for comparable performance in terms of classification. However,
42 when combined with the quadratic or KL fidelity, it provides a probabilistic
43 classification as output which allows us to evaluate the confidence of each
44 assignment. Finally, the TV penalty reaches excellent performance, with a
45 795 probabilistic output as well.

46
47 However, solving the convex problem takes more time than the other ap-
48 proaches. We observe that, when combined with linear or linear-logarithmic
49 fidelity, most of the assignments of the regularized solution lie within a corner
50 of the simplex, in accordance with the principles of linear programming. In
51 800 this case, the probabilistic nature of the solution cannot be exploited beyond
52 removing a small proportion (about 2-4%) of uncertain points. Furthermore,
53
54
55
56
57
58

1
2
3
4
5
6
7
8
9 the accuracy of assignments obtained with the TV regularizer seems to evolve
10 smoothly and monotonically with respect to the coverage. On the other hand,
11 the boundary size seems to induce an accuracy-coverage plot which is more
12 805 subject to sharp breaks and irregularities. This can be explained by the non-
13 convexity of the associated objective function, and the greedy nature of its
14 solving algorithm.
15

16
17 As for the choice of the fidelity function, this choice can be cross-validated
18 810 as a meta-parameter, as the effect of regularization can vary depending on the
19 dataset and the quality of the initial labeling. However, a general guideline
20 is presented in the following:
21

- 22
23 • **Hard assignment expected:** When only the 100% coverage is rel-
24 evant for the application, the Potts penalty should it be used, with
25 815 α -expansion a a solver, for its speed and the quality of its output.
- 26
27 • **Soft assignment expected:**
 - 28 – **Speed is the priority:** In this case, the boundary size penalty
29 combined with the ℓ_0 -cut pursuit algorithm is more advantageous.
 - 30 – **Quality is the priority:** The total variation penalty offers ex-
31 cellent precision, both for partial and complete coverage, at the
32 price of a longer computation time.
33 820

34 6.8. Extension

35
36
37 It is important to note that the cut pursuit algorithm cannot handle
38 different values for the transition between classes. If such a transition matrix
39 can be either inferred or cross-validated, then the other penalty shall always
40 825 be favored.
41

42
43 The only drawback of the TV penalty in our application is its speed.
44 However, this issue could be addressed by adapting the cut pursuit algorithm
45 to multi-dimensional simplex constrained values. In this case, the TV penalty
46 830 would combine the benefits of both the Potts penalty and the boundary
47 penalty, while retaining the robustness associated with its convex nature.
48

49
50 We did not benchmark the effect of choosing a different adjacency tree
51 structure, nor the different weighting schemes that can be applied. When
52 using varying edge weights with the TV penalty, preconditioning strategies
53 835 such as the one used by PFDR are an absolute must to avoid drastic conver-
54 gence speed increase.
55
56
57
58
59
60
61
62
63
64
65

7. Conclusions

In this article, we presented a regularization framework based on structured optimization to smooth semantic labelings on 3D point clouds. We demonstrated that this approach is superior to the classically used belief propagation algorithm. Furthermore, we presented a family of regularizers and fidelity functions which allows to retain the probabilistic nature of the labeling after smoothing, allowing us to estimate its confidence at each point. We also presented an efficient algorithm to solve the subsequent optimization problem, and extended the existing ℓ_0 -cut pursuit algorithm to our multi-dimensional, simplex-constrained setting.

Besides different extensions of our regularization framework, we also intend to investigate the potential of Convolutional Neural Networks (CNNs) adapted to 3D data. Among different strategies, particularly the one involving a 3D-CNN to classify each 3D point of a point cloud by considering a voxel-occupancy grid corresponding to the respective local neighborhood seems to be promising (Savinov, 2017; Hackel et al., 2016a; Huang and You, 2016). However, 3D-CNNs typically require a large amount of training data, and the network architecture as well as its internal settings are often heuristically defined by the user. The framework presented in the scope of this paper provides a competitive baseline for such approaches, as it represents an important alternative which is given by a theoretically well-founded structured regularization delivering classification results of high quality at a lighter computational cost and also for scenarios, where only smaller amounts of training data are available. Furthermore, although labelings obtained with CNNs tend to be spatially smoother than pointwise labelings, the degree of spatial regularity depends on the width of the convolutional filters, and hence is not easily tunable. Because our framework is agnostic with respect to how the initial labeling is obtained, it could be used to precisely set the level of smoothness in post-processing, at a light computational cost.

References

- Aström, F., Petra, S., Schmitzer, B., Schnörr, C., 2016. Image labeling by assignment. arXiv preprint arXiv:1603.05285, 1–9.
- Belton, D., Lichti, D.D., 2006. Classification and segmentation of terrestrial laser scanner point clouds using local variance information. International

1
2
3
4
5
6
7
8
9 Archives of the Photogrammetry, Remote Sensing and Spatial Information
10 Sciences XXXVI-5, 44–49.

11
12 Blomley, R., Jutzi, B., Weinmann, M., 2016. Classification of airborne laser
13 scanning data using geometric multi-scale features and different neighbour-
14 hood types. ISPRS Annals of the Photogrammetry, Remote Sensing and
15 875 Spatial Information Sciences III-3, 169–176.

16
17
18 Boykov, Y., Kolmogorov, V., 2004. An experimental comparison of min-
19 cut/max-flow algorithms for energy minimization in vision. IEEE Trans-
20 actions on Pattern Analysis and Machine Intelligence 26 (9), 1124–1137.

21
22
23 880 Boykov, Y., Veksler, O., Zabih, R., 2001. Fast approximate energy minimiza-
24 tion via graph cuts. IEEE Transactions on Pattern Analysis and Machine
25 Intelligence 23 (11), 1222–1239.

26
27 Brédif, M., Vallet, B., Serna, A., Marcotegui, B., Paparoditis, N., 2014.
28 TerraMobilita/IQmulus urban point cloud classification benchmark, in:
29 Proceedings of the IQmulus Workshop on Processing Large Geospatial
30 885 Data, Cardiff, UK, 8 July, pp. 1–6.

31
32
33 Breiman, L., 1996. Bagging predictors. Machine Learning 24 (2), 123–140.

34
35 Breiman, L., 2001. Random forests. Machine Learning 45 (1), 5–32.

36
37 Brodu, N., Lague, D., 2012. 3d terrestrial lidar data classification of complex
38 890 natural scenes using a multi-scale dimensionality criterion: applications in
39 geomorphology. ISPRS Journal of Photogrammetry and Remote Sensing
40 68, 121–134.

41
42
43 Chambolle, A., Pock, T., 2011. A first-order primal-dual algorithm for convex
44 problems with applications to imaging. Journal of Mathematical Imaging
45 895 and Vision 40, 120–145.

46
47
48 Chehata, N., Guo, L., Mallet, C., 2009. Airborne lidar feature selection
49 for urban classification using random forests. International Archives of
50 the Photogrammetry, Remote Sensing and Spatial Information Sciences
51 XXXVIII-3/W8, 207–212.

52
53
54 900 Chen, C., Liaw, A., Breiman, L., 2004. Using random forest to learn im-
55 balanced data. Technical Report. University of California, Berkeley, CA,
56 USA.

- 1
2
3
4
5
6
7
8
9 Clode, S., Kootsookos, P.J., Rottensteiner, F., 2004. The automatic extrac-
10 tion of roads from lidar data. *International Archives of the Photogramme-*
11 *try, Remote Sensing and Spatial Information Sciences XXXV-B3*, 231–236.
12 905
- 13
14 Criminisi, A., Shotton, J., 2013. Decision forests for computer vision and
15 medical image analysis. *Advances in Computer Vision and Pattern Recog-*
16 *nition*, Springer, London, UK.
17
- 18
19 Demantké, J., Mallet, C., David, N., Vallet, B., 2011. Dimensionality based
20 910 scale selection in 3d lidar point clouds. *International Archives of the Pho-*
21 *togrammetry, Remote Sensing and Spatial Information Sciences XXXVIII-*
22 *5/W12*, 97–102.
23
- 24
25 Filin, S., Pfeifer, N., 2005. Neighborhood systems for airborne laser data.
26 *Photogrammetric Engineering & Remote Sensing* 71 (6), 743–755.
27
- 28
29 915 Geman, D., Reynolds, G., 1992. Constrained restoration and the recovery
30 of discontinuities. *IEEE Transactions on Pattern Analysis and Machine*
31 *Intelligence* 14 (3), 367–383.
32
- 33
34 Gevaert, C.M., Persello, C., Vosselman, G., 2016. Optimizing multiple kernel
35 learning for the classification of UAV data. *Remote Sensing* 8 (12), 1–22.
36
- 37
38 920 Guignard, S., Landrieu, L., 2017. Weakly supervised segmentation-aided
39 classification of urban scenes from 3d lidar point clouds. *International*
40 *Archives of the Photogrammetry, Remote Sensing and Spatial Information*
41 *Sciences* .
42
- 43
44 925 Guo, B., Huang, X., Zhang, F., Sohn, G., 2015. Classification of airborne
45 laser scanning data using JointBoost. *ISPRS Journal of Photogrammetry*
46 *and Remote Sensing* 100, 71–83.
47
- 48
49 Hackel, T., Savinov, N., Ladicky, L., Wegner, J.D., Schindler, K., Pollefeys,
50 M., 2016a. Large-scale point cloud classification benchmark. [http://www.](http://www.semantic3d.net)
51 [semantic3d.net](http://www.semantic3d.net).
52
- 53
54 930 Hackel, T., Wegner, J.D., Schindler, K., 2016b. Fast semantic segmentation
55 of 3d point clouds with strongly varying density. *ISPRS Annals of the*
56 *Photogrammetry, Remote Sensing and Spatial Information Sciences III-3*,
57 177–184.
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10 Hornung, A., Wurm, K.M., Bennewitz, M., Stachniss, C., Burgard, W., 2013.
11 935 Octomap: An efficient probabilistic 3d mapping framework based on oc-
12 trees. *Autonomous Robots* 34, 189–206.
- 13
14 Hu, H., Munoz, D., Bagnell, J.A., Hebert, M., 2013. Efficient 3-d scene
15 analysis from streaming data, in: *Proceedings of the IEEE International*
16 *Conference on Robotics and Automation*, IEEE, Karlsruhe, Germany, 6-10
17 940 May, pp. 2297–2304.
- 18
19
20 Huang, J., You, S., 2016. Point cloud labeling using 3d convolutional neu-
21 ral network, in: *Proceedings of the International Conference on Pattern*
22 *Recognition*, IEEE, Cancun, Mexico, 4-8 December, pp. 1–6.
- 23
24
25 Huber, P.J., 1964. Robust estimation of a location parameter. *The Annals*
26 945 *of Mathematical Statistics* 35 (1), 73–101.
- 27
28 Indyk, P., Motwani, R., 1998. Approximate nearest neighbors: towards re-
29 moving the curse of dimensionality, in: *Proceedings of the 30th Annual*
30 *ACM Symposium on Theory of Computing*, ACM, Dallas, TX, USA, 24-
31 26 May, pp. 604–613.
- 32
33
34 950 Ishikawa, H., 2003. Exact optimization for Markov random fields with convex
35 priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*
36 25 (10), 1333–1336.
- 37
38
39 Jing, F., Li, M., Zhang, H.J., Zhang, B., 2004. Entropy-based active learning
40 with support vector machines for content-based image retrieval, in: *Pro-*
41 955 *ceedings of the IEEE International Conference on Multimedia and Expo*,
42 IEEE, Taipei, Taiwan, 27-30 June, pp. 85–88.
- 43
44
45 Khoshelham, K., Oude Elberink, S.J., 2012. Role of dimensionality reduction
46 in segment-based classification of damaged building roofs in airborne laser
47 scanning data, in: *Proceedings of the International Conference on Geo-*
48 960 *graphic Object-Based Image Analysis*, Rio de Janeiro, Brazil, 7-9 May,
49 pp. 372–377.
- 50
51
52 Kullback, S., Leibler, R.A., 1951. On information and sufficiency. *The Annals*
53 *of Mathematical Statistics* 22 (1), 79–86.
- 54
55
56
57
58

- 1
2
3
4
5
6
7
8
9
10 Lafarge, F., Mallet, C., 2012. Creating large-scale city models from 3d-
11 965 point clouds: a robust approach with hybrid representation. *International*
12 *Journal of Computer Vision* 99 (1), 69–85.
- 13
14 Lalonde, J.F., Unnikrishnan, R., Vandapel, N., Hebert, M., 2005. Scale
15 selection for classification of point-sampled 3d surfaces, in: *Proceedings of*
16 *the International Conference on 3-D Digital Imaging and Modeling, IEEE,*
17 970 *Ottawa, Canada, 13-16 June*, pp. 285–292.
- 18
19 Landrieu, L., Mallet, C., Weinmann, M., 2017. Comparison of belief propaga-
20 tion and graph-cut approaches for contextual classification of 3d lidar point
21 cloud data, in: *Proceedings of the IEEE International Geoscience and Re-*
22 *remote Sensing Symposium, IEEE, Fort Worth, TX, USA, 23-28 July*, pp.
23 1–4.
24 975
- 25
26 Landrieu, L., Obozinski, G., 2016a. Cut pursuit: fast algorithms to learn
27 piecewise constant functions, in: *Proceedings of the International Confer-*
28 *ence on Artificial Intelligence and Statistics, Microtome Publishing, Cadiz,*
29 *Spain, 9-11 May*, pp. 97–104.
- 30
31 Landrieu, L., Obozinski, G., 2016b. Cut pursuit: fast algorithms to learn
32 piecewise constant functions on general weighted graphs. *HAL hal-*
33 980 *01306779*, 1–35.
- 34
35
36
37 Lee, I., Schenk, T., 2002. Perceptual organization of 3d surface points. *Inter-*
38 *national Archives of the Photogrammetry, Remote Sensing and Spatial*
39 985 *Information Sciences XXXIV-3A*, 193–198.
- 40
41
42 Lellmann, J., Kappes, J., Yuan, J., Becker, F., Schnörr, C., 2009. Convex
43 multi-class image labeling by simplex-constrained total variation, in: *Pro-*
44 *ceedings of the Second International Conference on Scale Space and Vari-*
45 *ational Methods in Computer Vision, Springer, Voss, Norway, 1-5 June,*
46 990 *pp. 150–162.*
- 47
48
49 Lellmann, J., Lellmann, B., Widmann, F., Schnörr, C., 2013. Discrete
50 and continuous models for partitioning problems. *International Journal of*
51 *Computer Vision* 104 (3), 241–269.
- 52
53
54 Lim, E.H., Suter, D., 2009. 3d terrestrial lidar classifications with super-
55 995 *voxels and multi-scale conditional random fields. Computer-Aided Design*
56 *41 (10)*, 701–710.
- 57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9 Linsen, L., Prautzsch, H., 2001. Local versus global triangulations, in: Proceedings of Eurographics, Manchester, UK, 5-7 September, pp. 257–263.
10
11
- 12 Lu, Y., Rasmussen, C., 2012. Simplified Markov random fields for efficient
13 semantic labeling of 3d point clouds, in: Proceedings of the IEEE/RSJ
1000 International Conference on Intelligent Robots and Systems, IEEE, Vil-
14 amoura, Portugal, 7-12 October, pp. 2690–2697.
15
16
17
- 18 Mallet, C., Bretar, F., Roux, M., Soergel, U., Heipke, C., 2011. Relevance
19 assessment of full-waveform lidar data for urban area classification. ISPRS
20 Journal of Photogrammetry and Remote Sensing 66 (6), S71–S84.
1005
21
22
- 23 Mitra, N.J., Nguyen, A., 2003. Estimating surface normals in noisy point
24 cloud data, in: Proceedings of the Annual Symposium on Computational
25 Geometry, ACM, San Diego, CA, USA, 8-10 June, pp. 322–328.
26
27
- 28 Moravec, H., Elfes, A., 1985. High resolution maps from wide angle sonar,
29 in: Proceedings of the IEEE International Conference on Robotics and
1010 Automation, IEEE, St. Louis, MO, USA, 25-28 March, pp. 116–121.
30
31
- 32 Munoz, D., Bagnell, J.A., Vandapel, N., Hebert, M., 2009. Contextual classi-
33 fication with functional max-margin Markov networks, in: Proceedings of
34 the IEEE Conference on Computer Vision and Pattern Recognition, IEEE,
35 Miami, FL, USA, 20-25 June, pp. 975–982.
1015
36
37
- 38 Munoz, D., Vandapel, N., Hebert, M., 2008. Directional associative Markov
39 network for 3-d point cloud classification, in: Proceedings of the Interna-
40 tional Symposium on 3D Data Processing, Visualization and Transmission,
41 Georgia Institute of Technology, Atlanta, GA, USA, 18-20 June, pp. 63–70.
42
43
- 44 Najafi, M., Taghavi Namin, S., Salzmann, M., Petersson, L., 2014. Non-
45 associative higher-order Markov networks for point cloud classification,
46 in: Proceedings of the European Conference on Computer Vision, Zurich,
47 Switzerland, 6-12 September, pp. 500–515.
1020
48
49
- 50 Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification
51 of lidar data and building object detection in urban areas. ISPRS Journal
1025 of Photogrammetry and Remote Sensing 87, 152–165.
52
53
54
55
56
57
58
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- 1030 Niemeyer, J., Rottensteiner, F., Soergel, U., Heipke, C., 2016. Hierarchical higher order CRF for the classification of airborne lidar point clouds in urban areas. *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLI-B3*, 655–662.
- Niemeyer, J., Wegner, J.D., Mallet, C., Rottensteiner, F., Soergel, U., 2011. Conditional random fields for urban scene classification with full waveform lidar data, in: *Proceedings of the ISPRS Conference on Photogrammetric Image Analysis*, Springer, Munich, Germany, 5-7 October, pp. 233–244.
- 1035 Paparoditis, N., Papelard, J.P., Cannelle, B., Devaux, A., Soheilian, B., David, N., Houzay, E., 2012. Stereopolis II: A multi-purpose and multi-sensor 3d mobile mapping system for street visualisation and 3d metrology. *Revue Française de Photogrammétrie et de Télédétection* 200, 69–79.
- 1040 Pauly, M., Keiser, R., Gross, M., 2003. Multi-scale feature extraction on point-sampled surfaces. *Computer Graphics Forum* 22 (3), 281–289.
- Potts, R.B., 1952. Some generalized order-disorder transformations. *Mathematical Proceedings of the Cambridge Philosophical Society* 48, 106–109.
- 1045 Raguet, H., Landrieu, L., 2015. Preconditioning of a generalized forward-backward splitting and application to optimization on graphs. *SIAM Journal on Imaging Sciences* 8 (4), 2706–2739.
- Rudin, L.I., Osher, S., Fatemi, E., 1992. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena* 60, 259–268.
- Savinov, N., 2017. Point cloud semantic segmentation via deep 3D convolutional neural network. <https://github.com/nsavinov/semantic3dnet>.
- 1050 Schindler, K., 2012. An overview and comparison of smooth labeling methods for land-cover classification. *IEEE Transactions on Geoscience and Remote Sensing* 50 (11), 4534–4545.
- 1055 Schmidt, A., Niemeyer, J., Rottensteiner, F., Soergel, U., 2014. Contextual classification of full waveform lidar data in the Wadden Sea. *IEEE Geoscience and Remote Sensing Letters* 11 (9), 1614–1618.
- Schmidt, M., 2012. UGM: Matlab code for undirected graphical models. <http://www.cs.ubc.ca/~schmidtm/Software/UGM.html> (last access: 17 November 2016).

- 1
2
3
4
5
6
7
8
9 Serna, A., Marcotegui, B., 2013. Urban accessibility diagnosis from mobile
10 1060 laser scanning data. *ISPRS Journal of Photogrammetry and Remote Sensing* 84, 23–32.
11
12
13
14 Serna, A., Marcotegui, B., Goulette, F., Deschaud, J.E., 2014. Paris-rue-
15 Madame database: a 3d mobile laser scanner dataset for benchmarking
16 urban detection, segmentation and classification methods, in: *Proceedings*
17 1065 *of the International Conference on Pattern Recognition Applications and*
18 *Methods*, ACM, Angers, France, 6-8 March, pp. 819–824.
19
20
21 Shapovalov, R., Velizhev, A., Barinova, O., 2010. Non-associative Markov
22 networks for 3d point cloud classification. *International Archives of*
23 *the Photogrammetry, Remote Sensing and Spatial Information Sciences*
24 1070 XXXVIII-3A, 103–108.
25
26
27 Shapovalov, R., Vetrov, D., Kohli, P., 2013. Spatial inference machines,
28 in: *Proceedings of the IEEE Conference on Computer Vision and Pattern*
29 *Recognition*, IEEE, Portland, OR, USA, 23-28 June, pp. 2985–2992.
30
31
32 Vallet, B., Brédif, M., Serna, A., Marcotegui, B., Paparoditis, N., 2014. IQ-
33 1075 mulus & TerraMobilita Contest – Analysis of mobile laser scans (MLS) in
34 dense urban environments. MATIS Laboratory, French National Mapping
35 Agency (IGN) and Center for Mathematical Morphology (CMM), MINES
36 ParisTech. <http://data.ign.fr/benchmarks/UrbanAnalysis/> (last access: 17 November 2016).
37
38
39
40 1080 Vallet, B., Brédif, M., Serna, A., Marcotegui, B., Paparoditis, N., 2015.
41 TerraMobilita/iQmulus urban point cloud analysis benchmark. *Computers*
42 *& Graphics* 49, 126–133.
43
44
45 Waldhauser, C., Hochreiter, R., Otepka, J., Pfeifer, N., Ghuffar, S., Ko-
46 rzeniowska, K., Wagner, G., 2014. Automated classification of airborne
47 1085 laser scanning point clouds, in: Koziel, S., Leifsson, L., Yang, X.-S. (Eds.),
48 *Solving Computationally Expensive Engineering Problems: Methods and*
49 *Applications*, Springer, New York, USA, pp. 269–292.
50
51
52 Weinmann, M., 2016. Reconstruction and analysis of 3D scenes – From ir-
53 regularly distributed 3D points to object classes. Springer, Cham, Switzer-
54 1090 land.
55
56
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Weinmann, M., Jutzi, B., Hinz, S., Mallet, C., 2015a. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing* 105, 286–304.

1095 Weinmann, M., Schmidt, A., Mallet, C., Hinz, S., Rottensteiner, F., Jutzi, B., 2015b. Contextual classification of point cloud data by exploiting individual 3d neighborhoods. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences II-3/W4*, 271–278.

1100 Weinmann, M., Urban, S., Hinz, S., Jutzi, B., Mallet, C., 2015c. Distinctive 2d and 3d features for automated large-scale scene analysis in urban areas. *Computers & Graphics* 49, 47–57.

1105 West, K.F., Webb, B.N., Lersch, J.R., Pothier, S., Triscari, J.M., Iverson, A.E., 2004. Context-driven automated target detection in 3-d data, in: *Proceedings of SPIE 5426, Automatic Target Recognition XIV*, SPIE, Orlando, FL, USA, 12 April, pp. 133–143.

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

1110 Xu, S., Vosselman, G., Oude Elberink, S., 2014. Multiple-entity based classification of airborne laser scanning data in urban areas. *ISPRS Journal of Photogrammetry and Remote Sensing* 88, 1–15.