A structured regularization framework for spatially smoothing semantic labelings of 3D point clouds
Loic Landrieu, Hugo Raguet, Bruno Vallet, Clément Mallet, Martin Weinmann

To cite this version:

HAL Id: hal-01505245
https://hal.archives-ouvertes.fr/hal-01505245v2
Submitted on 4 May 2017

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

Loïc Landrieu\textsuperscript{a,}\textsuperscript{*}, Hugo Raguet\textsuperscript{b}, Bruno Vallet\textsuperscript{a}, Clément Mallet\textsuperscript{a}, Martin Weinmann\textsuperscript{a,c,}\textsuperscript{*}

\textsuperscript{a}Université Paris-Est, LASTIG MATIS IGN, ENSG, 94160 Saint-Mandé, France
\textsuperscript{b}Institut de Mathématiques de Marseille, Aix-Marseille Université, 39 rue Frédéric Joliot-Curie, 13453 Marseille, France
\textsuperscript{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.raguet@gmail.com (Hugo Raguet), 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
labeling are therefore desirable. To foster research regarding semantic labeling of 3D point cloud data, a variety of benchmark datasets acquired within urban environments have been released (Munoz et al., 2009; Serna et al., 2014; Vallet et al., 2015; Hackel et al., 2016b).

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

To impose spatial smoothness on this classification result, contextual information among neighboring 3D points is typically taken into account. For this purpose, the spatial structure of a 3D point cloud can be captured by a graph encoding the adjacency relationship between 3D points. Thereby, the adjacency relationship can be derived from the local neighborhood of each 3D point (Weinmann et al., 2015b), pre-segmentations (Niemeyer et al., 2016), or super-voxel structures (Lim and Suter, 2009). Based on the defined adjacency relationship, a context model is typically derived in the form of a graphical model, e.g. a Markov random field (MRF) (Munoz et al., 2009; Shapovalov et al., 2010; Lu and Rasmussen, 2012; Najafi et al., 2014) or its discriminative counterpart, the conditional random field (CRF) (Niemeyer et al., 2011, 2014; Schmidt et al., 2014; Weinmann et al., 2015b). As a consequence of imposing spatial smoothness on the derived labeling, the corresponding classification results are often significantly improved as can be observed in Figure 1c. However, the choice of the inference strategy (marginal, maximum-a-posteriori) will have a profound impact on the precision and nature of the solution (probabilistic or labeling), as well as the computation
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).

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 \(^1\),

\(^1\)Notably, 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 $\ell_0$-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 $K$ the finite set of potential semantic labels for each 3D point. Throughout this paper, we make...
an important distinction between hard and soft labelings. In that regard, we denote by \( \mathcal{S} \) the simplex:

\[
\mathcal{S} = \{ p \in [0, 1]^K | \sum_{k \in K} p_k = 1 \}. \tag{1}
\]

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

\[
S = \{ p \in \{0, 1\}^K | \sum_{k \in K} p_k = 1 \}. \tag{2}
\]

Elements of \( S \) are called hard labelings. For a labeling \( p \in \mathcal{S} \) or \( S \) 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 \( S^V \), \( i \in V \) 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:

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

where $\Phi$ is the *fidelity term*, $\Psi$ the *regularizer*, $\lambda > 0$ the *regularization strength*, and $\Omega$ 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 $\Psi$ favors solutions that are spatially smooth, in the sense that most adjacent nodes share the same label. The regularization strength $\lambda$ 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;
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 (Demanthké 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-
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.

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.

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.
ularity on the semantic labelings of 3D point clouds. In general, context models are based on the construction of an adjacency graph defining the extent to which interactions in a local neighborhood are considered, i.e., it is possible to take into account short-, mid- and long-range dependencies. It is important to note that this adjacency graph is in general not the same as the neighborhood graph used to compute local geometric features.

There are numerous approaches to obtain such adjacency graphs. The most common is to derive the graph from a neighborhood relationship, from nearest neighbors graph (Shapovalov et al., 2010) to cylindrical (Filin and Pfeifer, 2005; Niemeyer et al., 2014) or adaptive neighborhoods (Demantké et al., 2011; Weinmann et al., 2015b). Thereby, a simplifying assumption is typically made by only considering short-range dependencies. This is motivated by the observation that the quality of derived classification results reveals a saturation effect when increasing the scale parameter of the local neighborhood used for defining the adjacency graph. In this regard, the average number of involved neighbors was 7 in an investigation focusing on the classification of airborne laser scanning data (Niemeyer et al., 2011, 2014). Similar observations have been made in an investigation focusing on the classification of mobile laser scanning data (Weinmann et al., 2015b), where the consideration of short-range dependencies already delivers classification results of high quality. However, the latter investigation also indicates that adapting the size of the local neighborhood used for defining the adjacency graph with respect to the locally-adaptive neighborhood used for feature extraction is favorable in comparison to fixed neighborhoods. In these experiments, the average number of involved neighbors was between 15 and 21, depending on the approach used for deriving locally-adaptive neighborhoods.

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

2.3. Spatially smooth labeling

To derive a labeling with a higher spatial regularity, smooth labeling techniques (Schindler, 2012) or approaches for contextual classification can be used. The latter consider an initial labeling derived with a standard classifier and use a statistical context model to increase spatial regularity. Thereby, the classification of a given point does not only take into account
the feature vector corresponding to this considered point, but also the labels corresponding to neighboring points as well.

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

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

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

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

Lellmann et al. (2013) propose a variational approach, in which the discrete label set is relaxed into a continuous one. This approach permits the use of convex optimization algorithms, and yields results with arguably less artifacts than the solutions of combinatorial optimization techniques. The article shows that, with reasonable assumptions, the solution of the optimization can easily be discretized into a spatially smooth labeling.
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 eigentestropy-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 $T \in \mathbb{R}^{3 \times 3}$ as a function of the scale parameter $s$. The 3D structure tensor $T$ 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:

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

(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 $s_{\text{opt}}$ by minimizing the eigenentropy $E_{\mu}(s)$ across varying values of the scale parameter $s$:

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

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 $\mu_j$ 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 $s_{\text{opt}}$ 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 normal-
ization 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 fo-
cus 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 bag-
ging (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, re-
sults 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 computa-
tional 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 $P \in S^n$ 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
form derived from an adjacency graph $G = (V, E, w)$, defined in Section 4.1. In the following, we present and discuss the respective properties of different options for the search space (Section 4.2), the fidelity term (Section 4.3), and the regularizer (Section 4.4).

### 4.1. Adjacency graph of point clouds

In this paper, we focus on the regularization of semantic labelings, while investigating the respective benefits of different adjacency graphs is beyond the scope of our work. We chose a symmetrized 10-neighborhood adjacency graph with constant edge weight for its simplicity of implementation (Indyk and Motwani, 1998). However, our framework can naturally handle graphs with weighted edges, and we directly incorporate edge weights into our objective function.

### 4.2. Search space for probabilistic labelings

In this paper, we restrict the choice of $\Omega$ to hard or soft labelings: $\Omega = \mathcal{S}$ or $\Omega = \mathcal{S}^n$. While hard labelings assign a unique class, soft labelings assign a probability for each class, and consequently contain more information. However, producing a hard labeling for each point remains the main objective of semantic classification.

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

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

The main advantage of choosing $\Omega = \mathcal{S}$ over $\mathcal{S}^n$ is that a soft labeling allows the confidence assessment of the associated hard labeling through the computation of its entropy. This can be useful when the focus of the classification is precision rather than the full coverage of the point cloud. In such circumstances, the global labeling can be sorted by increasing entropy, ensuring that the first points have higher confidence.
4.3. Fidelity terms for regularizing distributions

The fidelity term \( \Phi(P, \cdot) : S^n \mapsto \mathbb{R} \) of the optimization problem defined in (3) enforces the influence of the soft labeling \( P \) in \( S^n \). 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 \( \phi(p, \cdot) : S \mapsto \mathbb{R} \) 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 \( S \), but can be restricted to \( \Omega = S \) since \( S \subset S \).

In this section, we present four different choices for \( \phi \). 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 \( p \):

\[
\phi_{\text{linear}}(p, q) = -\langle p, q \rangle = -\sum_{k \in K} p_k q_k.
\]

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. \( S \), 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 \( p \):

\[
\phi_{\text{log}}(p, q) = -\langle q, \log(\hat{p}) \rangle = -\sum_{k \in K} q_k \log(\hat{p}_k),
\]

\[4.3.3. 
\]
where \( \log \) denotes the entrywise logarithm and \( \tilde{p} \) is a version of the observed probability which is smoothed to prevent numerical issues: \( \tilde{p}_k = \frac{p_k}{\mathcal{K}} + \alpha p_k \) 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 \( \alpha \) 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:

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

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 \( q \):

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

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 \( \alpha \in [0,1] \):

\[
\phi_{\text{KL}}(p, q) = - \sum_{k \in \mathcal{K}} \tilde{p}_k \log (\tilde{q}_k), \tag{11}
\]
with \( \hat{p}_k = \alpha K + (1 - \alpha) p_k \) and likewise \( \hat{q}_k = \alpha K + (1 - \alpha) q_k \).

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 \( \Psi(\cdot) : \mathcal{S}^n \rightarrow \mathbb{R} \) 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:

\[
\Psi(Q) = \sum_{(i,j) \in E} w_{ij} \psi(Q_i - Q_j),
\]

with \( \psi : \mathbb{R}^K \rightarrow \mathbb{R} \) 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
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 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 \( \Omega = \mathcal{S} \) and a functional \( \psi_{\text{Potts}} \) equal to 0 in 0 and 1 everywhere else:

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

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 \( \psi \) as a norm over \( \mathbb{R}^K \).

Over vector spaces, the use of an Euclidean norm is often considered, enjoying theoretical isotropy. Here, however, \( \psi \) 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 \( \ell_1 \)-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:

\[
\psi_{\text{TV}}(q) = \sum_{k \in \mathcal{K}} |q_k|. \tag{14}
\]

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 \( \Omega = \mathcal{S} \) 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 \( \alpha \)-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 \( \ell_0 \)-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 \( \ell_0 \)-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
cuts, and enumerating their connected components. A backward-step is then performed to check if merging existing adjacent components can decrease the objective function.

However, in (Landrieu and Obozinski, 2016b), \( \ell_0 \)-cut pursuit is only defined in the one-dimensional setting in which only one scalar value is associated with each node. In this paper, we extend this algorithm to a multi-dimensional setting in which we associate a multi-dimensional, simplex-constrained value to each node. This extension is made easy by the separability hypothesis of the fidelity term, which ensures that, given a partition of \( V \), the associated optimal distribution can be computed independently for each component by minimizing the sum of the associated fidelity terms.

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

Computing such piecewise-constant labelings is the critical step in each of the three main steps of \( \ell_0 \)-cut pursuit, namely the computation of the optimal binary cuts and associated optimal distribution, and the backward step. Consequently the extension of \( \ell_0 \)-cut pursuit to multi-dimensional, simplex-bound data can be implemented easily and remains very efficient.

5.3. Continuous space and convex functional

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

Given the level of uncertainty over the data and over the parameters of our regularization framework, high precision is not required for the minimization of the objective functional. We thus resort to first-order proximal splitting algorithms, well-adapted to such large-scale situations where the functional is a sum of simple terms. This approach has been considered for

\(^3\)A C++ implementation can be downloaded at www.loiclandrieu.com/.
instance in Lellmann et al. (2009), where the authors use a Douglas–Rachford splitting algorithm to solve a specific instance of (3). Since this publication, more powerful splitting schemes have been developed. Although primal-dual schemes are popular thanks to their generality (Chambolle and Pock, 2011), the preconditioned generalized forward-backward splitting algorithm (Raguet and Landrieu, 2015) is more suited to our graph-structured problem, while taking full advantage of the smoothness of the data-fidelity term. We refer to the latter and references therein for more details.

6. Experimental Results

In this section, we first present the involved benchmark datasets (Section 6.1), the considered evaluation metrics (Section 6.2) and the competing methods (Section 6.3). Subsequently, we present the derived results (Section 6.4). The experimental framework will be available at www.loiclandrieu.com/.

6.1. Datasets

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

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

The Oakland 3D Point Cloud Dataset (Munoz et al., 2009) is a labeled benchmark dataset which has often been used to evaluate approaches focusing on a semantic labeling of 3D point clouds. This dataset has been acquired in the vicinity of the CMU campus in Oakland, USA, with a moving platform equipped with a side-looking Sick laser scanner used in push-broom mode (Munoz et al., 2008). During data acquisition, the speed of the platform reached up to 20km/h, and the acquired 3D point clouds reveal a point density with significant variation. A separation of the dataset into training

---

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

5 The 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).
data (about 37k labeled 3D points) and test data (about 1.32M labeled 3D points) is already provided.

The Oakland-5C Dataset refers to the provided reference labeling with respect to five semantic classes. These classes are defined as Wire, Pole/Trunk, Façade, Ground and Vegetation.

The Oakland-3C Dataset refers to the provided reference labeling with respect to three structural classes. These classes are defined as Linear Structures, Planar Structures and Volumetric Structures.

For both the Oakland-5C Dataset and the Oakland-3C Dataset, the number of training examples per class is very unbalanced which can have a detrimental effect on the training process (Chen et al., 2004; Criminisi and Shotton, 2013). To avoid such effects, we introduce a class re-balancing which relies on randomly selecting 1,000 labeled 3D points per class as new training set and discarding all other points.

6.1.2. Paris-rue-Cassette Database

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

In our experiments, we consider the seven dominant classes defined as Façade, Ground, Cars, 2-Wheelers, Road Inventory, Pedestrians and Vegetation. All 3D points belonging to the other classes are removed as these

\(^6\)The Paris-rue-Cassette Database is publicly available at \url{http://data.ign.fr/benchmarks/UrbanAnalysis/} (last access: 17 November 2016).
classes are not considered as representative (Weinmann et al., 2015c). To separate the dataset into training data and test data, we randomly select 1,000 labeled 3D points per class as training set and all remaining labeled 3D points as test set.

6.2. Evaluation metrics

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

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

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

6.3. Competing methods

In this subsection, we briefly summarize the benchmarked algorithms (i.e. the considered configurations of our framework) and some state-of-the-art methods that are involved for comparison. In Section 4 we listed two
search spaces ($\Omega = S$ or $\tilde{S}$), four fidelity functions (linear, linear-logarithmic, quadratic and Kullback-Leibler), as well as the two regularizers (Potts penalty, total variation). Of the 16 possible combinations, only 8 are unique and relevant. Indeed, when considering fidelity functions that induce a hard assignment (namely linear and linear-logarithmic) with the Potts penalty, the choice of $\Omega$ is irrelevant as all solutions belong to $S$. Consequently, $\ell_0$-cut pursuit and $\alpha$-expansion will minimize the same functional. As both approaches approximate the global solution, one could expect different results. However in all our numerical experiments the difference in the final value of the functional was small enough that its corresponding assignment was almost identical.

Some simple calculus shows that for the Potts penalty, the values of the linear, quadratic and Kullback-Leibler fidelities at the corners of the simplex amount to the same penalty with a different regularization strength. Similarly, we do not consider the total variation regularizer with a discrete search space, as it is redundant with the Potts penalty. We list the 8 combinations of regularizers and fidelity functions that correspond to unique algorithms in Table 1.

To compare with the graphical models approach advocated by Niemeyer et al. (2014) and based on a graphical model in the form of a CRF, we compute the solutions provided by loopy belief propagation, both for the marginal inference (LBP) and MAP-inference (LBP_MAP). The MAP-inference can also be advantageously computed with $\alpha$-expansion, as mentioned by Landrieu et al. (2017), and corresponds to the log_potts shorthand in Table 1. The implementations of the inference algorithms were obtained at (Schmidt, 2012).

The regularizing approach proposed by Lellmann et al. (2009) for image labeling corresponds to lin_TV. However, we use the preconditioned generalized forward-backward splitting algorithm (PFDR) algorithm which is much faster.

### 6.4. Experimental results

In Table 2, we provide the results of the full classifications for the methods invoked in Section 6.3, and the accuracy/coverage plot is represented in Figure 3. The classwise results are displayed in Tables 4, 5, and 6. The computation time, referenced in Table 3, is mainly dependent on the chosen regularizer as it dictates the used algorithm.
<table>
<thead>
<tr>
<th>Method</th>
<th>Fidelity</th>
<th>Regularizer</th>
<th>Domain</th>
<th>Minimizing algorithm</th>
</tr>
</thead>
<tbody>
<tr>
<td>lin_potts</td>
<td>linear</td>
<td>Potts penalty</td>
<td>S</td>
<td>α-expansion</td>
</tr>
<tr>
<td>log_potts</td>
<td>linear-logarithmic</td>
<td>Potts penalty</td>
<td>S</td>
<td>α-expansion</td>
</tr>
<tr>
<td>lin_TV</td>
<td>linear</td>
<td>total variation</td>
<td>S</td>
<td>PFDR</td>
</tr>
<tr>
<td>log_TV</td>
<td>linear-logarithmic</td>
<td>total variation</td>
<td>S</td>
<td>PFDR</td>
</tr>
<tr>
<td>l22_TV</td>
<td>quadratic</td>
<td>total variation</td>
<td>S</td>
<td>PFDR</td>
</tr>
<tr>
<td>KL_TV</td>
<td>Kullback-Leibler</td>
<td>total variation</td>
<td>S</td>
<td>PFDR</td>
</tr>
<tr>
<td>l22_bound</td>
<td>quadratic</td>
<td>Potts penalty</td>
<td>S</td>
<td>ℓ₀-cut pursuit</td>
</tr>
<tr>
<td>KL_bound</td>
<td>Kullback-Leibler</td>
<td>Potts penalty</td>
<td>S</td>
<td>ℓ₀-cut pursuit</td>
</tr>
</tbody>
</table>

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).
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 $F_1$-scores over all classes.

<table>
<thead>
<tr>
<th>Method</th>
<th>Oakland-3C Dataset</th>
<th>Oakland-5C Dataset</th>
<th>Paris-rue-Cassette Database</th>
</tr>
</thead>
<tbody>
<tr>
<td>pointwise</td>
<td>93.8 71.4</td>
<td>92.3 63.5</td>
<td>81.0 41.0</td>
</tr>
<tr>
<td>LBP</td>
<td>95.8 73.6</td>
<td>94.7 69.8</td>
<td>82.6 45.6</td>
</tr>
<tr>
<td>LBP_MAP</td>
<td>95.5 74.9</td>
<td>94.5 69.8</td>
<td>82.6 45.6</td>
</tr>
<tr>
<td>log_potts</td>
<td>97.3 78.2</td>
<td>95.7 74.4</td>
<td>92.5 65.4</td>
</tr>
<tr>
<td>lin_potts</td>
<td>97.5 78.9</td>
<td>95.8 75.1</td>
<td>92.3 61.3</td>
</tr>
<tr>
<td>l22_bound</td>
<td>97.2 78.3</td>
<td>95.7 73.6</td>
<td>93.6 65.6</td>
</tr>
<tr>
<td>KL_bound</td>
<td>97.0 78.3</td>
<td>95.7 74.7</td>
<td>93.7 64.4</td>
</tr>
<tr>
<td>lin_TV</td>
<td>97.5 78.3</td>
<td>95.7 74.9</td>
<td>92.4 61.2</td>
</tr>
<tr>
<td>log_TV</td>
<td>97.3 78.2</td>
<td>95.7 74.4</td>
<td>93.1 60.9</td>
</tr>
<tr>
<td>l22_TV</td>
<td>97.5 78.7</td>
<td>95.7 74.8</td>
<td>92.4 61.2</td>
</tr>
<tr>
<td>KL_TV</td>
<td>97.5 78.7</td>
<td>95.9 75.3</td>
<td>91.4 61.2</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Method</th>
<th>Oakland-3C Dataset (1.3M points)</th>
<th>Oakland-3C Dataset (1.3M points)</th>
<th>Paris-rue-Cassette Database (12M points)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LBP</td>
<td>45</td>
<td>35</td>
<td>720</td>
</tr>
<tr>
<td>LBP_MAP</td>
<td>54</td>
<td>31</td>
<td>800</td>
</tr>
<tr>
<td>α-expansion</td>
<td>24</td>
<td>15</td>
<td>400</td>
</tr>
<tr>
<td>ℓ₀-cut pursuit</td>
<td>27</td>
<td>21</td>
<td>600</td>
</tr>
<tr>
<td>PFDR</td>
<td>72</td>
<td>43</td>
<td>2100</td>
</tr>
</tbody>
</table>

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.

<table>
<thead>
<tr>
<th>Method</th>
<th>OA ($\bar{F}_1$)</th>
<th>$F_1$ (Linear Structures)</th>
<th>$F_1$ (Planar Structures)</th>
<th>$F_1$ (Volumetric Structures)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pointwise</td>
<td>93.8 71.4</td>
<td>28.7 97.1</td>
<td>98.1 92.6</td>
<td>88.5</td>
</tr>
<tr>
<td>LBP</td>
<td>95.8 75.6</td>
<td>36.2 98.1</td>
<td>98.0 91.9</td>
<td></td>
</tr>
<tr>
<td>LBP_MAP</td>
<td>95.5 74.9</td>
<td>34.8 98.8</td>
<td>98.0 91.9</td>
<td></td>
</tr>
<tr>
<td>log_potts</td>
<td>97.3 78.2</td>
<td>40.1 98.8</td>
<td>98.0 91.9</td>
<td></td>
</tr>
<tr>
<td>lin_potts</td>
<td>97.5 78.9</td>
<td>41.6 98.9</td>
<td>97.7 95.5</td>
<td></td>
</tr>
<tr>
<td>l22_bound</td>
<td>97.2 78.3</td>
<td>40.5 98.8</td>
<td>97.7 95.0</td>
<td></td>
</tr>
<tr>
<td>KL_bound</td>
<td>97.0 78.3</td>
<td>41.1 98.9</td>
<td>96.0</td>
<td></td>
</tr>
<tr>
<td>lin_TV</td>
<td>97.5 78.7</td>
<td>41.0 98.9</td>
<td>96.0</td>
<td></td>
</tr>
<tr>
<td>log_TV</td>
<td>97.3 78.2</td>
<td>40.0 98.9</td>
<td>95.7</td>
<td></td>
</tr>
<tr>
<td>l22_TV</td>
<td>97.5 78.7</td>
<td>41.0 98.9</td>
<td>96.1</td>
<td></td>
</tr>
<tr>
<td>KL_TV</td>
<td>97.5 78.7</td>
<td>40.9 99.0</td>
<td>96.2</td>
<td></td>
</tr>
</tbody>
</table>
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.

<table>
<thead>
<tr>
<th>Method</th>
<th>OA</th>
<th>$F_1$</th>
<th>$F_1$ (Wire)</th>
<th>$F_1$ (Pole/Trunk)</th>
<th>$F_1$ (Facade)</th>
<th>$F_1$ (Ground)</th>
<th>$F_1$ (Vegetation)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pointwise</td>
<td>92.3</td>
<td>63.5</td>
<td>18.2</td>
<td>38</td>
<td>76.3</td>
<td>97.6</td>
<td>87.2</td>
</tr>
<tr>
<td>LBP</td>
<td>94.7</td>
<td>70.9</td>
<td>22.2</td>
<td>59.8</td>
<td>81.2</td>
<td>98.0</td>
<td>93.4</td>
</tr>
<tr>
<td>LBP_MAP</td>
<td>94.5</td>
<td>69.8</td>
<td>21.2</td>
<td>55.6</td>
<td>81.0</td>
<td>98.0</td>
<td>93.0</td>
</tr>
<tr>
<td>lin_potts</td>
<td>95.8</td>
<td>74.4</td>
<td>29.1</td>
<td>65.7</td>
<td>83.4</td>
<td>98.2</td>
<td>95.7</td>
</tr>
<tr>
<td>l22_bound</td>
<td>95.7</td>
<td>73.6</td>
<td>26.5</td>
<td>64.7</td>
<td>83.1</td>
<td>98.1</td>
<td>95.9</td>
</tr>
<tr>
<td>KL_bound</td>
<td>95.7</td>
<td>74.7</td>
<td>29.4</td>
<td>67.3</td>
<td>82.9</td>
<td>98.2</td>
<td>95.7</td>
</tr>
<tr>
<td>lin_TV</td>
<td>95.7</td>
<td>74.9</td>
<td>27.5</td>
<td>69.3</td>
<td>83.4</td>
<td>98.1</td>
<td>95.9</td>
</tr>
<tr>
<td>log_TV</td>
<td>95.7</td>
<td>74.4</td>
<td>28.4</td>
<td>66.5</td>
<td>83.1</td>
<td>98.1</td>
<td>95.7</td>
</tr>
<tr>
<td>l22_TV</td>
<td>95.7</td>
<td>74.8</td>
<td>27.8</td>
<td>68.9</td>
<td>83.2</td>
<td>98.1</td>
<td>95.8</td>
</tr>
<tr>
<td>KL_TV</td>
<td>95.9</td>
<td>75.3</td>
<td>29.1</td>
<td>69.6</td>
<td>83.2</td>
<td>98.1</td>
<td>96.3</td>
</tr>
</tbody>
</table>

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 ($F_1$), and the classwise $F_1$-scores.

<table>
<thead>
<tr>
<th>Method</th>
<th>OA</th>
<th>$F_1$</th>
<th>$F_1$(F)</th>
<th>$F_1$(G)</th>
<th>$F_1$(C)</th>
<th>$F_1$(2W)</th>
<th>$F_1$(RI)</th>
<th>$F_1$(P)</th>
<th>$F_1$(V)</th>
</tr>
</thead>
<tbody>
<tr>
<td>pointwise</td>
<td>81.0</td>
<td>41.0</td>
<td>85.4</td>
<td>96.8</td>
<td>45.4</td>
<td>10.6</td>
<td>10.7</td>
<td>5.0</td>
<td>33.3</td>
</tr>
<tr>
<td>LBP</td>
<td>83.2</td>
<td>44.3</td>
<td>87.0</td>
<td>97.6</td>
<td>55.5</td>
<td>14.3</td>
<td>12.8</td>
<td>6.4</td>
<td>36.4</td>
</tr>
<tr>
<td>LBP_MAP</td>
<td>82.6</td>
<td>43.6</td>
<td>86.6</td>
<td>97.4</td>
<td>54.3</td>
<td>13.9</td>
<td>12.7</td>
<td>5.6</td>
<td>35.0</td>
</tr>
<tr>
<td>log_potts</td>
<td>92.5</td>
<td>65.4</td>
<td>94.7</td>
<td>95.1</td>
<td>82.2</td>
<td>48.7</td>
<td>16.5</td>
<td>65.9</td>
<td>54.9</td>
</tr>
<tr>
<td>lin_potts</td>
<td>92.3</td>
<td>61.3</td>
<td>94.5</td>
<td>95.8</td>
<td>78.9</td>
<td>48.1</td>
<td>17.7</td>
<td>43.7</td>
<td>50.6</td>
</tr>
<tr>
<td>l22_bound</td>
<td>93.6</td>
<td>65.6</td>
<td>95.7</td>
<td>97.4</td>
<td>83.5</td>
<td>66.7</td>
<td>19.5</td>
<td>46.6</td>
<td>50.0</td>
</tr>
<tr>
<td>KL_bound</td>
<td>93.7</td>
<td>64.4</td>
<td>95.7</td>
<td>98.1</td>
<td>82.4</td>
<td>46.9</td>
<td>32.7</td>
<td>44.1</td>
<td>51.2</td>
</tr>
<tr>
<td>lin_TV</td>
<td>92.4</td>
<td>61.2</td>
<td>94.4</td>
<td>97.4</td>
<td>81.8</td>
<td>42.3</td>
<td>28.7</td>
<td>35.7</td>
<td>48.0</td>
</tr>
<tr>
<td>log_TV</td>
<td>93.1</td>
<td>60.9</td>
<td>95.0</td>
<td>98.2</td>
<td>82.9</td>
<td>41.2</td>
<td>29.8</td>
<td>29.2</td>
<td>50.1</td>
</tr>
<tr>
<td>l22_TV</td>
<td>92.4</td>
<td>61.2</td>
<td>94.4</td>
<td>96.5</td>
<td>80.5</td>
<td>44.4</td>
<td>23.1</td>
<td>38.9</td>
<td>50.5</td>
</tr>
<tr>
<td>KL_TV</td>
<td>91.4</td>
<td>61.2</td>
<td>93.7</td>
<td>93.9</td>
<td>76.4</td>
<td>48.3</td>
<td>18.6</td>
<td>43.8</td>
<td>53.6</td>
</tr>
</tbody>
</table>
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.

(a) Ground truth

(b) Pointwise classification

(c) Regularized classification

(d) Confidence map

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).
6.5. Discussion

From Table 2 we can observe that regularization does indeed improve both the accuracy and the $F_1$-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).
In Tables 4, 5 and 6, we observe that although our framework barely improves the $F_1$-score of easy-to-classify classes such as Ground or Façade, our methods display significant improvement over hard classes, such as Wire, Pole/Trunk, Cars, 2-Wheelers and Pedestrians. Indeed, the $F_1$-score of Wire (18.2% to 29.4%) and Pole/Trunk (38% to 70%) is almost doubled for the Oakland-5C Dataset for the best performing methods. More impressively, in the Paris-rue-Cassette Database our methods are able to retrieve decent classifications for classes that were mostly mislabeled by the pointwise labeling. For example, 2-Wheelers were initially classified with a 10.6% $F_1$-score, but the best regularized labeling boasts a classification score of 66.7%. Likewise, regularization was able to improve the classification of the class Pedestrians from 5.0% to 65.9%, and the classification of the class Cars from 45.4% to 83.5%.

We explain this large improvement over hard classes because regularization removes isolated misclassified points scattered over the point cloud. Those classes are also the least represented, their recall is particularly sensitive to such misclassifications. The regularization increases the precision as well by enforcing homogeneity of a tightly connected set of points, which often belong to the same class. Remark that this will only improve the classification if the initial labeling is mostly right to begin with.

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

6.6. Choosing the fidelity

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

For example, when combined with either the TV or boundary penalty, the linear and linear-logarithmic fidelity yield hard smoothed assignments, while the quadratic and KL fidelity yield probabilistic assignments. Both the linear-logarithmic and Kullback-Leibler fidelity involve entrywise logarithms of probabilities, which can induce numerical issues, and consequently require a supplementary smoothing parameter. However, a smoothing parameter of $\alpha = 0.05$ seems to yield good results in general, and does not require extensive cross-validation. Finally, both quadratic and linear fidelities have fewer
parameters and easier computation since both their gradient and proximal
operator are very straightforward to obtain.

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

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

• **Hard assignment expected**: When only the 100% coverage is rele-
vant for the application, both linear and linear-logarithmic fidelity can
be employed.

• **Soft assignment expected**: The quadratic and KL fidelity provide a
probabilistic output when combined with the TV or boundary penalty.

### 6.7. Choosing the penalty

The influence of the penalty is more drastic, as it impacts both the nature
of the output and the computational efficiency as well. The Potts penalty,
combined with the $\alpha$-expansion algorithm, is the fastest of all approaches. In
terms of accuracy and $F_1$-score, it offers excellent performances as well, mak-
ing it a solid choice when only a hard assignment is expected. The boundary
penalty, when solved with the $\ell_0$-cut pursuit algorithm, is slightly slower to
compute, for comparable performance in terms of classification. However,
when combined with the quadratic or KL fidelity, it provides a probabilistic
classification as output which allows us to evaluate the confidence of each
assignment. Finally, the TV penalty reaches excellent performance, with a
probabilistic output as well.

However, solving the convex problem takes more time than the other ap-
proaches. We observe that, when combined with linear or linear-logarithmic
fidelity, most of the assignments of the regularized solution lie within a corner
of the simplex, in accordance with the principles of linear programming. In
this case, the probabilistic nature of the solution cannot be exploited beyond
removing a small proportion (about 2-4%) of uncertain points. Furthermore,
the accuracy of assignments obtained with the TV regularizer seems to evolve
smoothly and monotonically with respect to the coverage. On the other hand,
the boundary size seems to induce an accuracy-coverage plot which is more
subject to sharp breaks and irregularities. This can be explained by the non-
convexity of the associated objective function, and the greedy nature of its
solving algorithm.

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

- **Hard assignment expected**: When only the 100% coverage is relevant for the application, the Potts penalty should be used, with α-expansion as a solver, for its speed and the quality of its output.

- **Soft assignment expected**:
  - **Speed is the priority**: In this case, the boundary size penalty combined with the ℓ₀-cut pursuit algorithm is more advantageous.
  - **Quality is the priority**: The total variation penalty offers excellent precision, both for partial and complete coverage, at the price of a longer computation time.

### 6.8. Extension

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

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

We did not benchmark the effect of choosing a different adjacency tree
structure, nor the different weighting schemes that can be applied. When
using varying edge weights with the TV penalty, preconditioning strategies
such as the one used by PFDR are an absolute must to avoid drastic convergence speed increase.
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 $\ell_0$-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


Landrieu, L., Obozinski, G., 2016b. Cut pursuit: fast algorithms to learn piecewise constant functions on general weighted graphs. HAL hal-01306779, 1–35.


40


Weinmann, M., 2016. Reconstruction and analysis of 3D scenes – From irregularly distributed 3D points to object classes. Springer, Cham, Switzerland.


