



Pré-segmentation pour la classification faiblement supervisée de scènes urbaines à partir de nuages de points 3D LIDAR

Stéphane Guinard, Loic Landrieu, Bruno Vallet

► To cite this version:

Stéphane Guinard, Loic Landrieu, Bruno Vallet. Pré-segmentation pour la classification faiblement supervisée de scènes urbaines à partir de nuages de points 3D LIDAR. 2017. hal-01499571v1

HAL Id: hal-01499571

<https://hal.science/hal-01499571v1>

Preprint submitted on 3 Apr 2017 (v1), last revised 13 Apr 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.

Pré-segmentation pour la classification faiblement supervisée de scènes urbaines à partir de nuages de points 3D LIDAR

Weakly supervised segmentation-aided classification of urban scenes from 3D LIDAR point clouds

Stéphane Guinard*

Loïc Landrieu*

Bruno Vallet*

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

Résumé

Nous traitons le problème de la classification sémantique de nuages de points 3D LIDAR pour les scènes urbaines à partir d'un jeu d'apprentissage limité. Nous introduisons un modèle de segmentation non paramétrique pour les scènes urbaines formées par des objets anthropiques de formes simples. Notre modèle segmente la scène en régions géométriquement homogènes dont la taille est automatiquement déterminée par la complexité de la géométrie locale. Cette segmentation peut ensuite être intégrée dans un CRF (Conditional Random Field) afin de modéliser la structure globale de la scène. Cela nous permet d'aggréger, pour chaque groupe de points, les prédictions bruitées d'une classification faiblement supervisée. Nous démontrons l'efficacité de notre méthode sur deux jeux de données en accès libre.

Mots Clés

classification, segmentation, régularisation, LIDAR, scènes urbaines, nuages de points.

Abstract

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

Keywords

classification, segmentation, regularization, LIDAR, urban scenes, point cloud.

1 Introduction

L'interprétation automatique de nuages de points 3D, acquis par LIDAR terrestre ou mobile, est devenue un enjeu important dans le monde de la télédétection [10, 19]. Cependant, cette tâche présente encore de nombreux défis techniques. En effet, la volumétrie et l'irrégularité de la structure des nuages de points LIDAR rendent difficile l'association d'une classe à un point. De plus, la production d'une vérité terrain précise est particulièrement longue et difficile. Cependant, les nuages de points LIDAR de scènes urbaines montrent une certaine forme de régularité et une structure qui peuvent être exploitées afin d'améliorer la précision d'une classification sémantique bruitée.

Tout d'abord, la densité de points obtenue lors d'une acquisition LIDAR signifie que le nombre de points acquis est nettement supérieur au nombre d'objets sémantiques dans la scène. Par conséquent, la classification recherchée se doit de présenter une forte régularité spatiale, comme illustré dans la Figure 1a. Bien que la méthode présentée dans [19] repose sur l'utilisation de voisinages locaux, la classification produite n'est pas régulière, comme illustré dans la Figure 1b. Cet a priori de régularité a été utilisé dans de nombreux modèles graphiques modélisant l'influence du contexte [2, 15, 11], améliorant nettement la qualité de la classification initiale.

L'approche de pré-segmentation des nuages de points a déjà été utilisée afin de modéliser les interactions lointaines entre points, mais aussi pour diminuer la complexité du calcul. Les segments obtenus peuvent ensuite être ajoutés dans un modèle graphique à étages pour assurer la régularité spatiale de la classification. En revanche, les modèles existants nécessitent de déterminer à l'avance les paramètres des segments, tels que leur rayon maximum [12, 21], le nombre maximum de points les composant [9] ou le nombre total de segments [15].

L'objectif de notre travail est d'utiliser la structure sous-jacente des nuages de points pour améliorer une classification bruitée obtenue avec une vérité terrain très limitée. Dans une scène urbaine, la majorité des objets (routes,

façades, voitures ...) sont d'origine humaine et ont une structure géométrique relativement simple. Par conséquent, les caractéristiques géométriques associées à chaque point doivent être régulières. Cependant, la taille et le nombre de points de chaque segment peuvent énormément varier en fonction des objets concernés. Nous proposons de considérer le problème de segmentation comme un problème d'optimisation structurée afin d'obtenir des super-voxels géométriquement simples. A la différence des autres méthodes de présegmentation, notre méthode choisit automatiquement la taille des segments de sorte qu'elle s'adapte à la complexité de la géométrie locale, comme illustré par la Figure 1c.

Suivant le principe qu'un ensemble de classifieurs faibles donne de meilleurs résultats qu'un seul classifieur fort [18], une prédiction peut être obtenue par consensus en agrégeant pour chaque segment les prédictions bruitées issues d'une classification faiblement supervisée de chaque point. La structure induite par la segmentation et le consensus de prédictions peuvent être combinés au sein d'un CRF (Conditional Random Field) pour classifier directement les segments. Cela permet d'atteindre les performances de l'état de l'art avec un nombre de points annotés très restreint.

1.1 Travaux connexes

Classification point par point : Weinmann et al. [19] ont proposé un cadre pour la classification de nuages de points reposant sur des caractéristiques géométriques 3D obtenues à partir d'un voisinage local de taille optimale.

Modèles graphiques : la régularité spatiale d'une classification sémantique peut être obtenue à partir de modèles graphiques modélisant l'influence du contexte, comme les champs de Markov aléatoires [2, 15], ou leur équivalent discriminatif, les CRF [11, 7]. Les termes unaires sont obtenus grâce à une classification point par point utilisant les Random Forest, alors que les termes binaires représentent la probabilité de transition entre classes.

Approches utilisant une pré-segmentation : la pré-segmentation d'un nuage de points peut être utilisée pour améliorer une classification de ses points. Lim and Suter [9] proposent de considérer chaque segment comme un nœud dans un CRF multi-échelle. Des super-voxels sont obtenus par croissance de région nécessitant de définir le nombre de points maximal pour chaque voxel et un critère d'homogénéité colorimétrique. Dans Niemeyer et al. [12], les segments sont déterminés grâce à une première classification des points sans tenir compte du contexte : les segments sont définis comme les composantes connexes sémantiquement homogènes, avec une contrainte de diamètre maximal. L'inférence peut ensuite être calculée dans un CRF à deux étages, mêlant les points et les super-voxels. Un processus itératif est mis en place, alternant l'inférence et le calcul de segments. Shapovalov et al. [15] propose d'obtenir une présegmentation à l'aide de l'algorithme k-means, qui nécessite de définir à l'avance le nombre de seg-

ments. De plus, k-means produit des segments isotropiques ne tenant pas compte de la complexité locale de la scène. Dohan et al. [20] propose une segmentation hiérarchique calculée à partir d'une segmentation avant plan / arrière plan, et qui nécessite de définir à l'avance les dimensions horizontales et verticales des segments. Les segments sont ensuite fusionnés hiérarchiquement puis classifiés.

1.2 Présentation du problème

Nous considérons V un nuage de points 3D acquis par LIDAR dans une scène urbaine. Notre objectif est d'obtenir une classification des points de V parmi un ensemble fini de classes sémantiques $\mathcal{K}$.

Nous faisons l'hypothèse que nous ne disposons que d'une vérité terrain composée d'un faible nombre de points, obtenus sur une scène similaire. Ces points doivent être suffisamment peu nombreux pour qu'ils puissent être saisis rapidement par un opérateur, c'est-à-dire pas plus de quelques dizaines de points par classe.

Les éléments de la structure algorithmique de notre méthode sont présentés dans la suite de cette section dans l'ordre où ils doivent être appelés.

Calcul des descripteurs et du graphe d'adjacence : Pour chaque point, nous calculons un vecteur de descripteurs locaux et non-locaux, comme décrits respectivement dans les Sections 2.1 et 2.2. Dans la Section 2.3 nous présentons la manière dont la relation d'adjacence entre points est capturée par un graphe.

Segmentation géométriquement homogène : La segmentation est considérée comme un problème d'optimisation structurée présenté dans la Section 3.1, et dont la solution peut être approximée par un algorithme glouton. Dans la Section 3.2, nous présentons la manière dont la structure haut-niveau de la scène peut être capturée par un graphe obtenu à partir de la segmentation.

Classification contextuelle des segments : Dans la Section 4, nous utilisons un CRF dont la structure est déterminée par la segmentation, et ses termes unaires obtenus par agrégation des prédictions bruitées d'une classification faiblement supervisée. Enfin, nous associons à chaque point du nuage la classe qui a été associée au segment auquel il appartient.

2 Calcul des descripteurs et du graphe d'adjacence

Dans cette section, nous présentons les descripteurs choisis pour représenter la géométrie locale et la position globale des points, ainsi que le graphe d'adjacence, qui capture la structure spatiale du nuage de points.

Partant du principe que le jeu d'apprentissage est limité, et dans un souci de réduire la complexité du calcul de la segmentation, nous nous attachons à choisir un nombre restreint de descripteurs. Il est important de remarquer que la segmentation et la classification des points n'utilisent pas nécessairement les mêmes descripteurs.

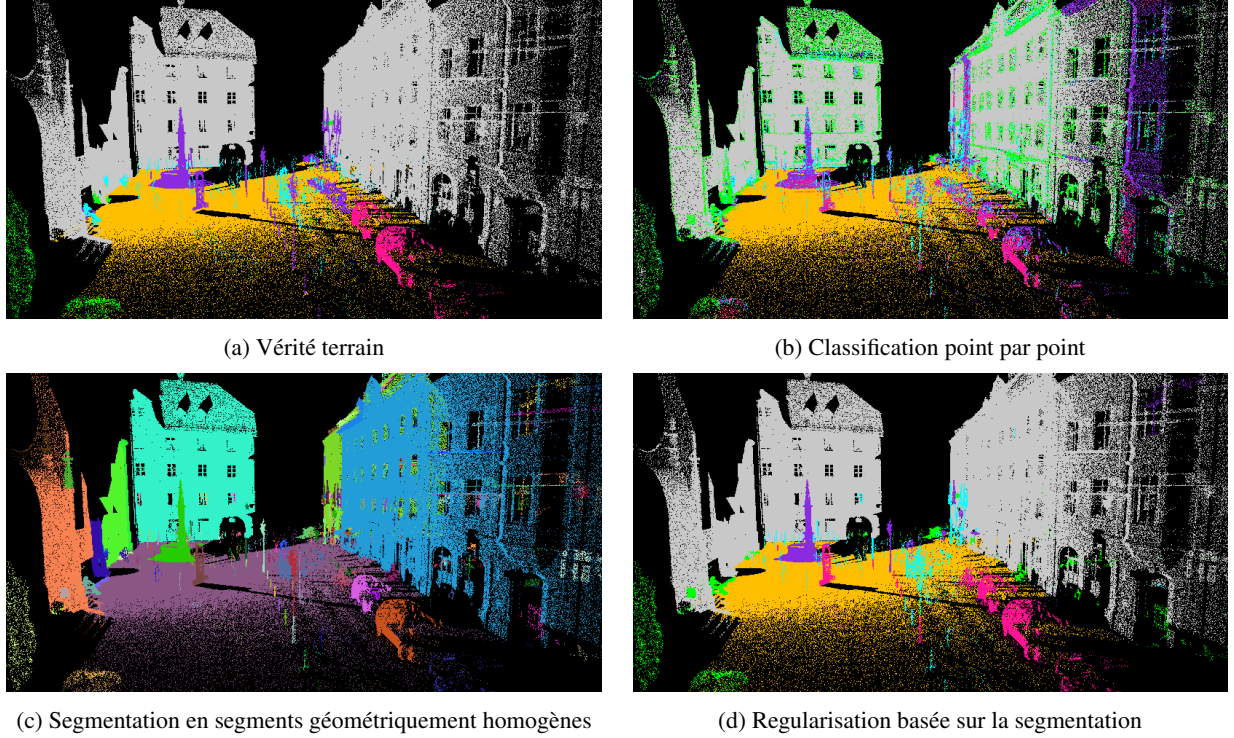


FIGURE 1 – Comparaison des résultats obtenus avec une classification point par point et la méthode proposée de régularisation basée sur un CRF obtenu à partir d’une pré-segmentation. Dans 1a, 1b, 1d, sont représentées les classes sémantiques selon le code couleur suivant : végétation, façades, aménagements, artefacts, voitures, routes. Dans la Figure 1c, chaque segment est représenté par une couleur aléatoire.

2.1 Descripteurs locaux

Pour décrire la géométrie locale du voisinage de chaque point nous définissons quatre descripteurs : la linéarité, la planéité, la dispersion et la verticalité, représentés dans la Figure 2.

Ces descripteurs sont calculés à partir d’un voisinage local pour chaque point du nuage. Pour chaque voisinage, nous calculons les valeurs propres $\lambda_1 > \lambda_2 > \lambda_3$ de la matrice de covariance de la position des voisins. La taille du voisinage est choisie de façon à ce que l’*eigentropy* E du vecteur $(\lambda_1/\Lambda, \lambda_2/\Lambda, \lambda_3/\Lambda)$, avec $\Lambda = \sum_{i=1}^3 \lambda_i$, soit minimale, en accord avec la définition de voisinage optimal présentée dans Weinmann et al. [19]. Comme présenté dans Demantké et al. [4], ces valeurs propres nous permettent également de décrire la forme des voisinages locaux. La linéarité décrit le degré d’*elongation* du voisinage. La planéité évalue à quel point le voisinage peut être approximé par un plan. Enfin, une dispersion élevée correspond à un voisinage sphérique et isotrope. La combinaison de ces trois descripteurs est appelée dimensionnalité.

Durant nos expériences, la dimension verticale du voisinage optimal s’est montrée cruciale pour pouvoir distinguer les routes des façades et les fils électriques des poteaux, classes qui partagent la même dimensionnalité. Par conséquent nous avons mis au point un descripteur capturant la *verticalité* du voisinage local d’un point à partir

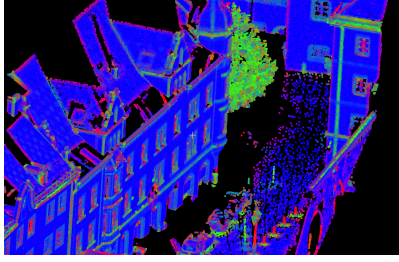
des vecteurs et valeurs propres définis ci-dessus. Notons u_1, u_2, u_3 les trois vecteurs propres associés à $\lambda_1, \lambda_2, \lambda_3$ respectivement. Nous définissons le vecteur unitaire de direction principale dans $\mathbb{R}_+^3$ comme la somme des valeurs absolues des coordonnées des vecteurs propres pondérées par les valeurs propres :

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

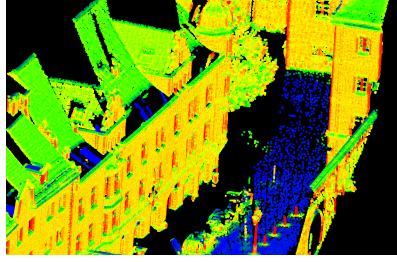
Nous avançons que la composante verticale de ce vecteur caractérise la verticalité d’un voisinage, étant minimale (égale à 0) pour un voisinage horizontal, et maximale (égale à 1) pour un voisinage linéaire vertical. Un voisinage planaire selon la verticale et une autre direction, comme le présentent les points de façades, a une valeur intermédiaire (environ 0.7). Ce comportement est illustré à la Figure 2 et quantifié à la Figure 3.

2.2 Descripteurs non-locaux

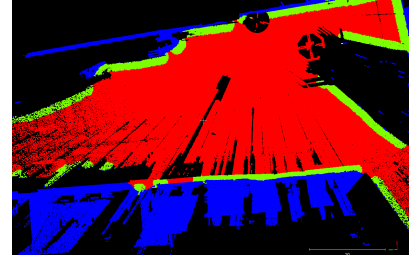
Si la forme des voisinages des points permet de renseigner sur leur géométrie locale, et donc de segmenter en éléments géométriquement simples, cela s’avère insuffisant pour classifier les points eux-mêmes. Aussi nous utilisons deux descripteurs qui décrivent la position globale des points considérés : l’élévation et la position par rapport à la route.



(a) Dimensionnalité



(b) Verticalité



(c) Position par rapport à la route

FIGURE 2 – Représentation des quatre descripteurs géométriques locaux et de la position par rapport à la route. Dans 2a, le vecteur de dimensionnalité [Linéarité, Planéité, Dispersion] est représenté par le vecteur [rouge, vert, bleu] qui lui est proportionnel. Dans 2b, les valeurs de la verticalité sont représentées par un code de couleur allant du bleu (faible verticalité : routes), au vert/jaune (verticalité moyenne : façades et toits), au rouge (forte verticalité : poteaux). Dans 2c, la position par rapport à la route est représentée avec le code de couleur suivant : bleu pour les points dans l’ α -shape de la route, vert pour ceux qui sont au bord, et rouge pour ceux qui sont en-dehors.

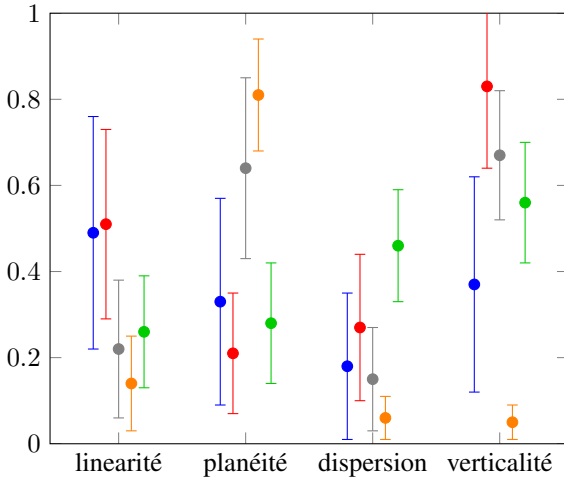


FIGURE 3 – Moyenne et écart type des descripteurs locaux dans le jeu de données Oakland pour les classes suivantes : fils électriques, poteaux, façades, routes, végétation.

Le calcul de ces valeurs nécessite tout d’abord de délimiter avec précision la route. Un pré-classement reposant uniquement sur les descripteurs géométriques et utilisant un classifieur par forêts aléatoires parvient à une classification pratiquement parfaite des points entre route et non-route, avec un F-score supérieur à 99.5%. À partir de cette classification, un modèle d’élévation peut être calculé et la hauteur normalisée par rapport à la route peut être facilement obtenue.

Pour estimer la position par rapport à la route nous calculons l’ α -shape en deux dimensions [1] des points de la route projetés sur le niveau d’élévation nul. Cela permet de calculer le descripteur *position par rapport à la route*, qui vaut 0 pour un point appartenant à la route, entre 0 et 1 si le point se situe vers la frontière et 1 en dehors de la route.

2.3 Graphe d’adjacence

La structure spatiale d’un nuage de point peut être représentée par un graphe non-orienté $G = (V, E)$, dans lequel chaque noeud représente un point du nuage, et les arêtes encodent leur *relation d’adjacence*. Nous calculons pour chaque point ses 10 plus proches voisins, comme recommandé par [13]. Il est important de noter que ce graphe définit une relation d’adjacence différente du voisinage optimal évoqué dans la Section 2.1.

3 Segmentation géométriquement homogène

3.1 Segmentation avec l’énergie de Potts

Nous associons à chaque point un vecteur contenant les quatre descripteurs géométriques locaux (linéarité, planéité, dispersion et verticalité) : $f_i \in \mathbb{R}^4$. L’objectif de la segmentation est de calculer une approximation constante par partie g^* du signal $f \in \mathbb{R}^{V \times 4}$ structurée par le graphe G . g^* est défini comme le vecteur de $\mathbb{R}^{V \times 4}$ minimisant l’énergie de segmentation de Potts suivante :

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

avec $\delta(\cdot)$ la fonction de $\mathbb{R}^4 \rightarrow \{0, 1\}$ égale à 0 en 0 et 1 partout ailleurs. La première partie de cette énergie est la fonction de fidélité, garantissant que les segments de g^* correspondent à des valeurs homogènes de f . La seconde partie est le terme de régularisation, qui ajoute une pénalité pour chaque arête reliant deux segments ayant des valeurs différentes. Cette pénalité incite les segments à avoir des formes simples. Enfin, ρ est la force de régularisation, qui arbitre le compromis entre la fidélité et la simplicité, et détermine implicitement le nombre de segments.

Ce problème d’optimisation structurée peut être efficacement approximé par l’algorithme glouton ℓ_0 -cut poursuivi

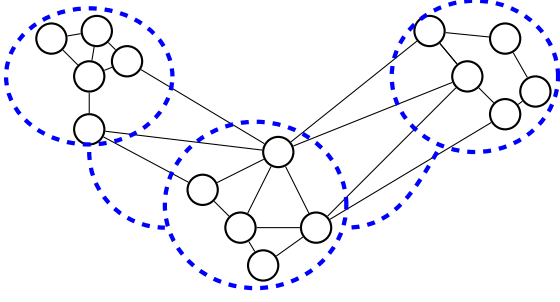


FIGURE 4 – Structure d’adjacence du graphe des segments. Les arêtes entre les points sont représentées par un trait plein noir —, la segmentation et l’adjacence entre ses segments par un trait pointillé bleu : - - -.

[5], se reposant sur le calcul efficace de coupes minimales. Les segments sont définis comme les composantes constantes et connectées du signal constant par morceaux obtenu.

L’avantage de cette formulation est qu’elle ne nécessite pas de définir une taille maximale pour les segments, que ce soit en superficie ou en nombre de points. Lorsque la structure est simple, comme pour les segments de route ou les façades, les segments peuvent contenir des dizaines de milliers de points. À l’inverse, la granularité de la segmentation augmente là où la géométrie locale est plus complexe, comme illustré par la Figure 1c. Pour le reste de cet article, nous notons $\mathbb{S} = (S_1, \dots, S_k)$ la segmentation sans chevauchement de V , obtenue en résolvant approximativement le problème d’optimisation.

3.2 Graphe des segments

Comme les segments représentent les objets de la scène, la segmentation produite capture sa structure haut-niveau. Pour obtenir les relations entre objets, nous construisons un *graphe des segments*, défini par $\mathcal{G} = (\mathbb{S}, \mathbb{E}, w)$, dans lequel les segments de $\mathbb{S}$ sont les nœuds de $\mathcal{G}$. $\mathbb{E}$ représente les relations d’adjacence entre segments, tandis que w correspond au poids de leurs frontières, comme illustré dans la Figure 4. Nous considérons deux segments comme adjacents s’il existe une arête de E les reliant, et w est le poids total des arêtes liant ces deux segments :

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

4 Classification contextuelle des segments

Pour garantir la régularité spatiale, Niemeyer et al. [11] définissent la classification optimale l^* d’un nuage de points comme celle maximisant la distribution postérieure $p(l \mid f')$ dans un CRF structuré par un graphe d’adjacence G , avec f' le vecteur des descripteurs d’un segment. Nous

représentons une classification de V par un vecteur de $\Delta(V, \mathcal{K}) = \{l \in \{0, 1\}^{V \times \mathcal{K}} \mid \sum_{k \in \mathcal{K}} l_{i,k} = 1, \forall i \in V\}$ (les coins du simplex), tel que $l_{i,k}$ est égal à 1 si le point i de V est classifié comme $k \in \mathcal{K}$ et 0 partout ailleurs. Pour un point i de V , l_i est considéré comme un vecteur de $\mathbb{R}^{\mathcal{K}}$. Cela nous permet de définir l^* comme l’argument qui maximise l’énergie suivante :

$$l^* = \arg \max_{l \in \Delta(V, \mathcal{K})} \sum_{i \in V} l_i^T p_i + \sum_{(i,j) \in E} l_i^T M_{i,j} l_j, \quad (1)$$

avec $p_{i,k} = \log(p(l_{i,k} = 1 \mid f'_i))$ le logarithme terme à terme de la probabilité du nœud i d’être associé à la classe k , et $M_{(i,j),(k,l)} = \log(p(l_{i,k} = 1, l_{j,l} = 1 \mid f'_i, f'_j))$ le logarithme terme à terme de la probabilité que les nœuds i et j présentent la transition (k, l) .

Comme préconisé dans Niemeyer et al. [11], il est possible d’estimer $p(l_{i,k} = 1 \mid f'_i)$ avec un classifieur probabiliste par forêt aléatoire : p_{RF} . Afin de ne pas avoir de valeurs nulles, nous lisons p_{RF} par une interpolation linéaire sur les zones de probabilité constante : $p(k \mid f'_i) = (1 - \alpha)p_{\text{RF}}(k \mid f'_i) + \alpha/|\mathcal{K}|$ avec $\alpha = 0.01$ et $|\mathcal{K}|$ la cardinalité du jeu de données. Les auteurs recommandent aussi d’apprendre les probabilités de transition à partir du vecteur des différences de descripteurs. Cependant, le fait de réaliser une classification faiblement supervisée nous empêche d’apprendre ces transitions, car cela nécessiterait d’avoir des annotations suffisantes pour les $|\mathcal{K}|^2$ transitions possibles. De plus, cela impliquerait de localiser de façon précise les transitions entre classes, ce qui s’avère délicat dans un nuage de points. Nous faisons donc l’hypothèse simplificatrice que M soit nulle sur sa diagonale et égale à σ , une valeur positive pouvant être déterminée par cross-validation :

$$M_{(i,j),(k,l)} = \begin{cases} 0 & \text{si } k = l \\ \sigma & \text{sinon.} \end{cases} \quad (2)$$

Faisant l’hypothèse que les segments obtenus dans la Section 3.1 correspondent à des objets sémantiquement homogènes, nous pouvons affirmer que la classification optimale sera constante sur chaque segment de $\mathbb{S}$. Nous proposons donc la formulation suivante d’un CRF structuré par le graphe des segments $\mathcal{G}$ afin de retrouver l’organisation des segments. Nous notons L^* appartenant à $\{0, 1\}^{\mathbb{S} \times \mathcal{K}}$ la classification de $\mathbb{S}$ définie comme :

$$L^* = \arg \max_{L \in \Delta(\mathbb{S}, \mathcal{K})} \sum_{s \in \mathbb{S}} L_s^T P_s + \sum_{(s,t) \in \mathbb{E}} w_{s,t} L_s^T M L_t^T,$$

avec $P_{s,k} = |s| \log(p(L_s = k \mid \{f_i\}_{i \in s}))$ le logarithme de la probabilité du segment s d’être associé à la classe k , multiplié par la cardinalité de s . Nous définissons cette probabilité comme la moyenne des probabilités des points appartenant au segment s .

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

Remarquons que l'influence d'un segment sur L^* est déterminée par sa cardinalité, ceci est dû au fait que la classification des points reste l'objectif final. De même, le coût de la transition entre deux segments est pondéré par le poids total des arêtes à leur frontière $w_{s,t}$ et représente l'importance de leur frontière. D'après les conclusions de Landrieu et al. [7], nous approximations la classification optimale avec l'algorithme α -expansion de Boykov et al. [3], en utilisant l'implémentation de Schmidt [14].

Il est important de noter que le CRF basé sur les segments repose sur le graphe des segments $\mathcal{G}$, dont nous supposons qu'il est significativement plus petit que G , accélérant l'inférence. Nous attirons l'attention du lecteur sur le fait qu'il n'est nullement nécessaire que les descripteurs utilisés pour la classification à l'aide de forêts aléatoires soient les mêmes que ceux utilisés pour la segmentation.

5 Expériences

Nous démontrons les avantages de notre méthode au travers d'expériences numériques sur deux jeux de données en libre accès. Nous présentons dans un premier temps les données et notre métrique d'évaluation, puis une comparaison entre notre approche et d'autres méthodes issues de l'état de l'art.

5.1 Données

Pour valider notre approche, nous utilisons deux jeux de données, que nous appellerons respectivement Oakland et Semantic3D.

Le jeu de données Oakland est présenté par Munoz et al. [10]. Il contient au total 1.3 millions de points acquis par LIDAR mobile, mais nous nous restreignons à la partie urbaine du nuage de points, qui contient 655 297 points. Plusieurs classes, telles que les voitures ou les piétons ont été filtrées au préalable, laissant cinq classes dans le nuage : les fils électriques, les poteaux et troncs d'arbres, les façades, les routes et la végétation. Nous avons choisi d'exclure la partie la plus fournie en arbres car les segmentations obtenues sur cette partie n'étaient pas satisfaisantes au niveau des liaisons tronc/feuillage.

Les cinq classes présentes dans le nuage ne sont pas également réparties. Les routes sont en effet majoritairement représentées car elles sont plus proches du capteur au moment de l'acquisition. À l'inverse, les fils électriques, poteaux et troncs d'arbres ne représentent que quelques milliers de points dans le nuage.

Nous présentons aussi dans cet article nos résultats pour une des scènes constituant le benchmark Semantic3D¹. La scène initiale contient 10 millions de points acquis par LIDAR terrestre, mais pour des raisons de complexité de calcul nous nous sommes restreints à un sous-échantillon du nuage, de 3 millions de points. Cette scène urbaine acquise par LIDAR fixe comprend six classes : route, façade, végétation, voitures, artefacts d'acquisition, aménagement ur-

bain. Les routes et façades y sont, là encore, majoritairement représentées.

Pour chaque classe, nous sélectionnons manuellement un nombre identique de points représentatifs de chaque classe, respectivement 15 points par classe pour Oakland et 25 pour Semantic3D. Ces points constituent nos jeux d'apprentissage (respectivement 75 points au total pour Oakland et 150 pour Semantic3D).

5.2 Métrique

Afin de prendre en compte les effectifs déséquilibrés de chaque classe, nous avons choisi d'utiliser la moyenne non pondérée des F-scores de chaque classe pour évaluer nos classifications. Cette mesure permet d'associer un meilleur score à une classification qui aurait une précision moyenne sur toutes les classes, par rapport à une classification qui serait très performante sur certaines classes et mauvaise sur d'autres. Nous ajoutons à titre indicatif le taux de bonnes classifications pour évaluer la performance globale des méthodes.

5.3 Méthodes comparées

Pour comparer l'efficacité de notre méthode à l'état de l'art, nous avons implémenté les méthodes suivantes :

- **Classification point par point** : nous avons implémenté la méthode de classification point par point utilisant des voisinages optimaux proposée par Weinmann et al. [19] et avons utilisé l'algorithme des forêts aléatoires à partir des descripteurs mentionnés dans la Section 2.1.
- **Régularisation par un CRF** : nous avons implémenté un CRF, tel que défini dans (1) sans aide de la segmentation.

5.4 Résultats

Les résultats de la classification de notre méthode et des méthodes comparées apparaissent dans le Tableau 1 pour Oakland et dans le Tableau 2 pour Semantic3D. Notre méthode et la méthode de régularisation améliorent toutes deux les résultats de façon significative par rapport à la classification point par point. Si l'amélioration du taux de classification de notre méthode n'est que de quelques pourcents par rapport à la régularisation par CRF, la qualité de la classification est notablement améliorée sur les fils électriques, les poteaux, ou les aménagements urbains, qui sont particulièrement difficiles à classer.

6 Conclusion

Dans cet article, nous avons présenté une méthode de classification aidée par une pré-segmentation géométrique capturant la structure haut-niveau d'une scène urbaine. Nous avons montré que cette segmentation permettait d'utiliser un CRF pour classer directement les segments, ce qui améliore la qualité des résultats par rapport à une régularisation par CRF classique. Les recherches futures devraient se concentrer sur l'amélioration de la qualité de la segmentation dans les parties peu denses et irrégulières du nuage,

1. <http://www.semantic3d.net/>

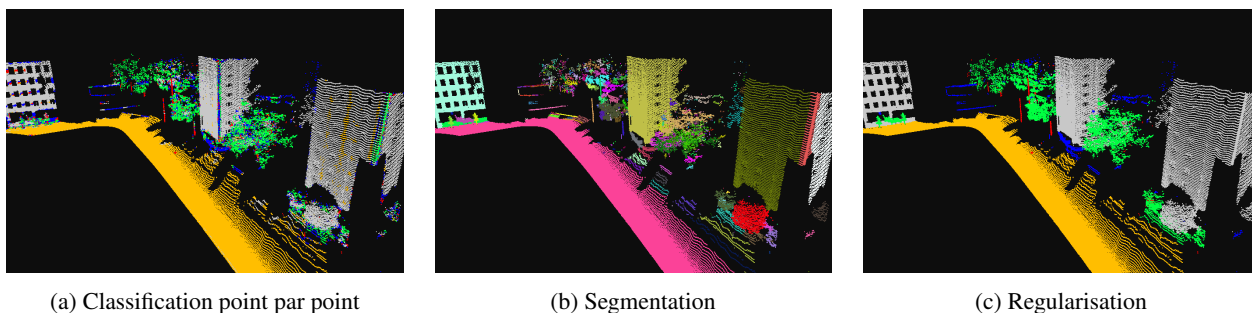


FIGURE 5 – résultats de la classification sur un détail de la base de données Oakland. En 5b est représentée la segmentation géométriquement homogène en associant à chaque segment une couleur aléatoire. En 5a et 5c est représentée la classification par notre méthode avec le code couleur suivant : végétation, façades, poteaux, fils électriques, routes.

	classification point par point			régularisation par CRF			notre méthode		
classe	précision	rappel	FScore	précision	rappel	FScore	précision	rappel	Fscore
fils électriques	4.2	37.1	7.5	87.8	32.1	47.0	51.2	35.4	41.9
poteaux/troncs	9.0	67.7	15.9	78.6	37.7	51.0	66.1	48.3	55.8
façades	57.5	74.9	65.1	79.2	98.0	87.6	91.0	96.5	93.6
routes	99.9	86.7	92.8	99.6	95.2	97.4	99.6	99.1	99.4
végétation	85.5	82.8	84.1	93.5	93.1	93.3	95.5	94.4	95
total	51.2	69.8	53.1	87.7	71.2	75.2	80.7	74.7	77.1

TABLE 1 – Précision, rappel et FScore calculés en % pour chaque classe sur Oakland. Les taux de classifications corrects sont respectivement de 85.2%, 94.8% et 97.3%. Les valeurs en gras sont les meilleures dans chaque catégorie.

	classification point par point			régularisation par CRF			notre méthode		
classe	précision	rappel	FScore	précision	rappel	FScore	précision	rappel	Fscore
routes	98.7	96.8	97.7	97.6	99.0	98.3	97.5	98.7	98.1
vegetation	14.2	82.9	24.2	49.7	84.7	62.6	52.1	93.7	67.0
façades	99.6	88.1	93.5	99.5	97.9	98.7	99.7	98.2	98.8
aménagement	74.2	71.4	73.1	93.7	88.7	91.2	92.7	90.4	91.5
artifacts	18.3	37.5	24.6	77.9	42.1	54.7	73.8	39.3	51.3
voitures	28.6	54.8	37.6	66.5	86.2	75.1	84.0	90.0	82.3
total	55.7	71.9	58.4	80.8	83.1	80.1	83.3	85.0	82.3

TABLE 2 – Précision, rappel et FScore calculés en % pour chaque classe pour Semantic3D. Les taux de classifications corrects sont respectivement de 88.4%, 96.9% et 97.2%. Les valeurs en gras sont les meilleures dans chaque catégorie.

comme le feuillage. Une autre amélioration possible est l'exploitation du contexte de la scène au niveau des transitions. En effet, la matrice de transition que nous utilisons est, dans son état actuel (2), trop limitée, car elle ne prend pas en compte des règles de transition comme "la route est sous les façades" ou "une transition tronc/feuillage est plus probable que route/feuillage". Bien qu'une classification faiblement supervisée exclu d'apprendre les transitions, l'expertise d'un opérateur pourrait aider à les estimer.

Références

- [1] Akkiraju, N. and Edelsbrunner, H. and Facello, M. and Fu, P. and Mücke, EP. and Varela, C., 1995. Alpha shapes : definition and software. Proceedings of the 1st International Computational Geometry Software Workshop.
- [2] Anguelov, D., Taskarf, B., Chatalbashev, V., Koller, D., Gupta, D., Heitz, G. and Ng, A., 2005. Discriminative learning of Markov random fields for segmentation of 3d scan data. In : *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, Vol. 2, IEEE, pp. 169–176.
- [3] Boykov, Y., Veksler, O. and Zabih, R., 2001. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23 (11), pp. 1222–1239.
- [4] Demantké, J., Mallet, C., David, N. and Vallet, B., 2011. Dimensionality based scale selection in 3d lidar point clouds. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 38(Part 5), pp. W12.

- [5] Landrieu, L. and Obozinski, G., 2016. Cut pursuit : fast algorithms to learn piecewise constant functions on general weighted graphs. *Preprint*.
- [6] Landrieu, L., Raguét, H., Vallet, B., Mallet, C. and Weinmann, M., 2017a. A structured regularization framework for spatially smoothing semantic labelings of 3d point clouds. *Preprint*.
- [7] Landrieu, L., Weinmann, M. and Mallet, C., 2017b. Comparison of belief propagation and graph-cut approaches for contextual classification of 3d lidar point cloud data. Geoscience and Remote Sensing Symposium, 2017. IGARSS'2017
- [8] Lee, D.-T. and Schachter, B. J., 1980. Two algorithms for constructing a Delaunay triangulation. *International Journal of Computer & Information Sciences* 9(3), pp. 219–242.
- [9] Lim, E. H. and Suter, D., 2009. 3d terrestrial lidar classifications with super-voxels and multi-scale conditional random fields. *Computer-Aided Design* 41(10), pp. 701–710.
- [10] Munoz, D., Bagnell, J. A., Vandapel, N. and Hebert, M., 2009. Contextual classification with functional max-margin Markov networks. In : *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009.*, IEEE, pp. 975–982.
- [11] Niemeyer, J., Rottensteiner, F. and Soergel, U., 2014. Contextual classification of lidar data and building object detection in urban areas. *ISPRS journal of photogrammetry and remote sensing* 87, pp. 152–165.
- [12] Niemeyer, J., Rottensteiner, F., Soergel, U. and Heipke, C., 2016. Hierarchical higher order crf for the classification of airborne lidar point clouds in urban areas. *23rd International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences Congress, ISPRS'2016*.
- [13] Niemeyer, J., Wegner, J. D., Mallet, C., Rottensteiner, F. and Soergel, U., 2011. Conditional random fields for urban scene classification with full waveform lidar data. In : *Photogrammetric Image Analysis*, Springer, pp. 233–244.
- [14] Schmidt, M., 2007. A Matlab toolbox for probabilistic undirected graphical models. <http://www.cs.ubc.ca/~schmidtm/Software/UGM.html>.
- [15] Shapovalov, R., Velizhev, E. and Barinova, O., 2010. Nonassociative Markov networks for 3d point cloud classification. In : *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XXXVIII, Part 3A*, Citeseer.
- [16] Weinmann, M., Jutzi, B. and Mallet, C., 2013. Feature relevance assessment for the semantic interpretation of 3d point cloud data. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 5, pp. W2.
- [17] Weinmann, M., Jutzi, B. and Mallet, C., 2014. Describing paris : automated 3d scene analysis via distinctive low-level geometric features. In : *Proceedings of the IQmulus workshop on processing large geospatial data*, pp. 1–8.
- [18] Opitz, D., Maclin, R., 1999. Popular ensemble methods : An empirical study *Journal of Artificial Intelligence Research* 11, pp. 169–198.
- [19] Weinmann, M., Urban, S., Hinz, S., Jutzi, B. and Mallet, C., 2015. Distinctive 2d and 3d features for automated large-scale scene analysis in urban areas. *Computers & Graphics* 49, pp. 47–57.
- [20] Dohan, D., Matejek, B., Funkhouser, T., 2015. Learning hierarchical semantic segmentations of LiDAR data. *2015 International Conference on 3D Vision (3DV)*, pp. 273–281.
- [21] Golovinskiy, A., Kim, V., Funkhouser, T., 2009. Shape-based recognition of 3D point clouds in urban environments. *2009 International Conference on 3D Vision (3DV)* 49, pp. 2154–2161.