

Descripteurs sphériques multi-échelles pour la classification sémantique de nuages de points 3D

H. Thomas¹

J.E. Deschaud¹

B. Marcotegui¹

F. Goulette¹

¹ Mines Paristech

hugues.thomas@mines-paristech.fr

1 Introduction

Un algorithme de classification de points 3D transforme des informations géométriques en informations sur la nature des objets d'une scène. Deux paradigmes se démarquent dans la littérature pour atteindre cet objectif. Le premier vise à assurer la cohérence des objets en les segmentant avant de choisir une classe pour chacun d'eux [1]. Le deuxième cherche à se prémunir des erreurs de segmentation en choisissant une classe pour chacun des points indépendamment [6, 2]. Notre algorithme appartient à ce deuxième paradigme (voir figure 1).

Sans segmentation préalable, la classification de chaque point s'appuie uniquement sur l'apparence géométrique de son voisinage. Afin de caractériser cette apparence, il existe plusieurs descripteurs géométriques. Nous avons choisi d'utiliser les valeurs issues de la matrice de covariance locale proposées par Weinmann et al. [6] car elles caractérisent très bien la géométrie locale, tout en étant rapide à calculer. D'après leurs travaux exhaustifs, il est possible d'obtenir une bonne classification à partir de ces simples descripteurs en utilisant des algorithmes d'apprentissage classiques, notamment les *forêts d'arbres aléatoires* qui offrent les meilleures performances dans ce cas.

Les deux voisinages communément utilisés dans le domaine des nuages de points sont les *voisinages sphériques* et les *k plus proches voisins*. Les premiers sont définis par un rayon fixé tandis que les seconds sont définis par un nombre de voisins fixé. Quelle que soit la définition choisie, fixer ce paramètre contrôlant la taille du voisinage de manière unique sur tout le nuage n'est pas adapté à la diversité des objets rencontrés dans une scène 3D. L'approche multi-échelles offre de meilleures performances au prix d'un temps de calcul plus élevé. Cependant Hackel et al. [2] ont proposé une astuce très simple pour accélérer drastiquement les temps de calculs sans diminuer les performances, par sous-échantillonnages successifs du nuage de points à chaque échelle de voisinage. Malgré leurs résultats au-dessus de l'état de l'art, leur définition de voisinages multi-échelles utilise les plus proches voisins, qui manquent de cohérence géométrique. Nous introduisons, dans le paragraphe 2.1, une nouvelle définition de voisinage multi-échelles, qui assure l'uniformité des voisinages en les choisissant sphériques. Pour compléter notre méthode de classification, nous décrirons en détails nos descripteurs et notre stratégie d'apprentissage dans les paragraphes 2.2 et 2.3 respectivement.

Enfin, dans la section 3, nous proposerons deux expériences visant à évaluer les performances de notre méthode face à l'état de l'art, et sa capacité à passer à l'échelle sur une base de données plus large.

2 Description de l'algorithme de classification

2.1 Définition des voisinages sphériques multi-échelles

La définition du voisinage sphérique d'un point $\mathbf{p}_0 \in \mathbb{R}^3$ dans un nuage $\mathcal{C} \subset \mathbb{R}^3$ et de rayon $r \in \mathbb{R}$ est :

$$\mathfrak{S}_r(\mathbf{p}_0, \mathcal{C}) = \{ \mathbf{p} \in \mathcal{C} \mid \|\mathbf{p} - \mathbf{p}_0\| \leq r \} \quad (1)$$

contrairement à la définition de k plus proches voisins, elle correspond donc à une partie fixée de l'espace, mais peut contenir un nombre de points très variable. Pour limiter cette variation, nous adoptons la même stratégie que Hackel et al. [2], en sous-échantillonnant le nuage avant de calculer les voisinages. Plus précisément, à chaque échelle, les voisinages sont calculés sur un nuage sous échantillonné proportionnellement au rayon. La méthode de sous-échantillonnage choisie consiste à garder le barycentre des points tombant dans la même case d'une grille 3D. La taille des cases de cette grille est choisie en divisant le rayon du voisinage par un facteur $\rho \in \mathbb{R}$. Ce paramètre contrôle donc la densité des voisinages, plus il sera grand, plus ceux-ci contiendront de points. Trois autres paramètres caractérisent notre définition de voisinages multi-échelles : le plus petit rayon de voisinages $r_0 \in \mathbb{R}$, le nombre d'échelles $S \in \mathbb{N}$, et le rapport de taille d'une échelle à la suivante $\varphi \in \mathbb{R}$. La

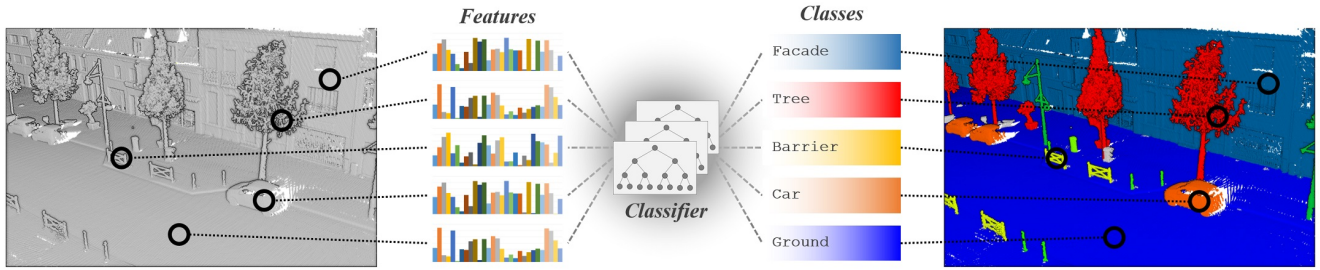


FIGURE 1 – Notre algorithme de classification

définition de notre voisinage à chaque échelle $s \in \{0, \dots, S - 1\}$ est alors :

$$\mathcal{N}_s(\mathbf{p}_0) = \mathfrak{S}_{r_s}(\mathbf{p}_0, \mathcal{C}_s) \quad (2)$$

avec $r_s = r_0 * \varphi^s$ étant le rayon de l'échelle s , et \mathcal{C}_s étant le nuage sous-échantillonné par un grille de taille r_s/ρ .

2.2 Choix des descripteurs géométriques

Pour rester dans les mêmes conditions expérimentales que [6, 2], nos descripteurs n'utilisent pas d'informations additionnelles comme la couleur ou l'intensité. La majorité d'entre eux découlent des valeurs propres $\lambda_1 \geq \lambda_2 \geq \lambda_3 \in \mathbb{R}$ et des vecteurs propres $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3 \in \mathbb{R}^3$ de la matrice de covariance du voisinage. Ils sont définis dans le tableau 1. A la différence de Hackel et al. [2], nous gardons la verticalité du premier et du dernier vecteur propre. De plus, nous utilisons la valeur absolue des moments autour des vecteurs propres car ceux-ci n'ont pas d'orientation définie. Enfin le nombre de points dans un voisinage est gardé comme mesure de l'occupation volumique, portant le nombre total de descripteurs à 18 par échelle.

2.3 Stratégie d'apprentissage

La classification de nuages de points pose deux problèmes. D'abord, la taille des données est généralement énorme, ensuite, les classes sont fortement déséquilibrées. Pour un apprentissage efficace, une solution consiste à choisir un sous ensemble du nuage de points comme jeux d'apprentissage, assez petit pour autoriser des temps d'entraînement raisonnables et dans lequel les classes sont équilibrées. Pour se comparer aux travaux antérieurs [6, 2], nous suivons cette stratégie et nous utilisons le même type d'algorithme de classification (une forêt d'arbres aléatoires). La seule différence entre notre algorithme de classification et ceux de [6] et [2] réside donc dans le choix du type de voisinages et des descripteurs. Dans la suite nous appellerons notre algorithme RF_MSSF pour "Multi-Scale Spherical Features with a Random Forest".

3 Résultats

3.1 Comparaison à l'état de l'art

Comme expliqué précédemment, dans cette expérience, nous nous plaçons dans les mêmes conditions que [6, 2]. Nous allons donc utiliser les jeux de données Paris-Rue-Madame [4] et Paris-Rue-Cassette [5]. Ces deux nuages de points représentent respectivement une rue de 160 mètres avec 20 millions de points et une rue de 200 mètres avec 12 millions de points. Pour chacun des deux nuages, nous utiliserons un jeu d'apprentissage de 1000 points par classes choisis aléatoirement. Nous sommes allés plus loin que les travaux précédentes en répétant l'expérience 100 fois sur des jeux d'apprentissages différents. En effet, on ne pourrait pas assurer la validité de la comparaison sans vérifier que le choix aléatoire du jeu d'apprentissage n'influe que peu sur les résultats de classification.

Les valeurs des paramètres de notre algorithme sont : $S = 8$, $r_0 = 0.1$ m, $\varphi = 2$ et $\rho = 5$. Les trois premiers paramètres ont été choisis afin que les tailles des différents voisinages aillent du plus petit objet à l'échelle d'une façade. La valeur du dernier paramètre ρ a été choisie empiriquement afin d'optimiser la qualité des résultats sous la contrainte d'un temps de calcul maximum. Nous choisissons d'afficher les résultats en mesure F_1 . Le tableau 2 montre la moyenne et l'écart type de nos performances de classification sur les 100 essais, face aux résultats avancés dans les articles [6] et [2]. En moyenne, notre algorithme dépasse les précédents résultats de 15 points sur Rue Madame, et de 4 points sur Rue Cassette. Il est intéressant de constater que l'écart type de nos résultats est limité à quelques points même pour les classes les plus difficiles. Au vu de ces résultats, nous pouvons conclure que la comparaison des algorithmes est valable, et que les performances de nos descripteurs multi-échelles sont largement au-dessus de l'état de l'art.

3.2 Expérimentation sur des données à grande échelle

Cette dernière expérience vise à montrer que notre algorithme de classification se généralise à des jeux de données de plus grande envergure. Pour cela, nous utilisons le jeu de données Paris-Lille-3D [3], qui contient plus de 140 millions de points

TABLE 1 – Descripteurs utilisés pour la classification

Descripteurs	Définitions
Somme des valeurs propres	$\sum \lambda_i$
Omnivariance	$\left(\prod \lambda_i\right)^{\frac{1}{3}}$
Entropie propre	$-\sum \lambda_i \ln(\lambda_i)$
Linéarité	$(\lambda_1 - \lambda_2)/\lambda_1$
Planéité	$(\lambda_2 - \lambda_3)/\lambda_1$
Sphéricité	λ_3/λ_1
Variation de courbure	$\lambda_3/(\lambda_1 + \lambda_2 + \lambda_3)$
Verticalité (x2)	$\left \frac{\pi}{2} - \text{angle}(\mathbf{e}_i, \mathbf{e}_z)\right _{i \in (0,2)}$
Moments absolus (x6)	$\frac{1}{ \mathcal{N} } \left \sum \langle \mathbf{p} - \mathbf{p}_0, \mathbf{e}_i \rangle^k \right _{i \in (0,1,2)}$
Moments verticaux (x2)	$\frac{1}{ \mathcal{N} } \sum \langle \mathbf{p} - \mathbf{p}_0, \mathbf{e}_z \rangle^k$
Nombre de points	$ \mathcal{N} $

TABLE 2 – Résultats sur Rue Madame et Rue Cassette

Classes	RF_MSSF	[2]	[6]
Façade	99, 10% (0, 06)	98, 51%	95, 73%
Sol	98, 28% (0, 09)	98, 11%	91, 82%
Voitures	97, 63% (0, 22)	94, 23%	71, 37%
Motos	76, 20% (1, 70)	64, 35%	17, 25%
Panneaux	80, 55% (6, 40)	50, 70%	9, 34%
Piétons	87, 55% (2, 11)	38, 88%	3, 21%
Moyenne	89, 89%	74, 13%	48, 12%
Façade	98, 61% (0, 10)	96, 85%	92, 85%
Sol	98, 87% (0, 10)	98, 47%	97, 83%
Voitures	91, 85% (0, 90)	89, 43%	64, 23%
Motos	74, 15% (2, 60)	67, 84%	29, 23%
Panneaux	22, 35% (5, 89)	31, 34%	25, 01%
Piétons	52, 07% (4, 18)	39, 60%	16, 61%
Végétation	83, 36% (1, 32)	67, 90%	39, 52%
Moyenne	74, 47%	70, 20%	52, 18%

sur 2 km de rues à Paris et à Lille. A la différence de Rue Cassette et Rue Madame, ce jeu de données a été entièrement classifié à la main. Il est donc plus grand et de meilleure qualité. Nous utilisons les classes grossières fournies avec les données et qui définissent des ensembles d'objets cohérents (voir tableau 3). Parmi ces 10 classes, les points "non classifiés" seront ignorés pour l'apprentissage et dans la mesure des résultats.

Le jeu de données contient 3 nuages, *Lille1*, *Lille2* et *Paris* et nous avons choisi de séparer le nuage *Lille1* en 2 parties pour obtenir 4 nuages de tailles à peu près homogènes. Cette séparation en 4 parties nous a permis d'entraîner notre algorithme sur 3 nuages et de le tester sur un dernier, à la manière d'une validation croisée, produisant ainsi des résultats de classification plus généraux. Puisque l'environnement est du même type que dans l'expérience précédente nous gardons les mêmes paramètres et choisissons 50000 points par nuage à l'aide d'une méthode d'apprentissage actif comme jeu d'entraînement. Le tableau 3 montre des résultats de bonne qualité, même dans le cas où l'algorithme a été entraîné sur Lille et testé sur Paris.

TABLE 3 – Résultats sur chaque partie Paris-Lille-3D

Nuage	Sol	Bâtiments	Signalisation	Poteaux	Poubelles	Barrières	Piétons	Voitures "	Végétation "	Moyenne
Lille1_1	99, 47%	96, 95%	83, 01%	82, 76%	83, 97%	49, 34%	14, 80%	95, 24%	83, 09%	76, 51%
Lille1_2	99, 73%	94, 89%	78, 38%	72, 79%	0, 34%	57, 09%	18, 88%	97, 33%	86, 36%	67, 31%
Lille2	99, 32%	98, 30%	70, 32%	80, 63%	44, 36%	42, 63%	38, 11%	96, 50%	74, 77%	71, 66%
Paris	99, 55%	93, 84%	77, 24%	71, 59%	5, 18%	59, 93%	0, 32%	89, 98%	89, 47%	65, 23%

4 Conclusion

Notre méthode a prouvé être à la fois supérieure à l'état de l'art et généralisable à de larges bases de données. De plus les résultats de notre algorithme avec entraînement et test sur deux villes différentes sont très prometteurs car ils montrent sa robustesse face au sur-apprentissage. Enfin nous constatons que la différence du type de voisinage utilisé entre [2] et notre méthode, qui pourrait passer pour un simple détail d'implémentation, a un grand effet sur les résultats. En effet, le type de caractérisation local d'un voisinage va grandement changer la capacité des descripteurs à être robuste.

Références

- [1] A. Golovinskiy, V. G. Kim, and T. Funkhouser. Shape-based recognition of 3d point clouds in urban environments. In *Computer Vision, 2009 IEEE 12th International Conference on*, pages 2154–2161. IEEE, 2009.
- [2] T. Hackel, J. D. Wegner, and K. Schindler. Fast semantic segmentation of 3d point clouds with strongly varying density. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Prague, Czech Republic*, 3 :177–184, 2016.
- [3] X. Roynard, J.-E. Deschaud, and F. Goulette. Paris-Lille-3D : a large and high-quality ground truth urban point cloud dataset for automatic segmentation and classification. *arXiv preprint arXiv :1712.00032*, 2017.
- [4] A. Serna, B. Marcotegui, F. Goulette, and J.-E. Deschaud. Paris-rue-Madame database : a 3d mobile laser scanner dataset for benchmarking urban detection, segmentation and classification methods. In *4th International Conference on Pattern Recognition, Applications and Methods ICPRAM 2014*, 2014.
- [5] B. Vallet, M. Brédif, A. Serna, B. Marcotegui, and N. Paparoditis. Terramobilita/iqmulus urban point cloud analysis benchmark. *Computers & Graphics*, 49 :126–133, 2015.
- [6] M. Weinmann, B. Jutzi, S. Hinz, and C. Mallet. Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers. *ISPRS Journal of Photogrammetry and Remote Sensing*, 105 :286–304, 2015.