



**HAL**  
open science

## Segmentation et classification de points 3D obtenus à partir de relevés laser terrestres : une approche par super-voxels

Ahmad Kamal Aijazi, Paul Checchin, Laurent Trassoudaine

### ► To cite this version:

Ahmad Kamal Aijazi, Paul Checchin, Laurent Trassoudaine. Segmentation et classification de points 3D obtenus à partir de relevés laser terrestres : une approche par super-voxels. RFIA 2012 (Reconnaissance des Formes et Intelligence Artificielle), Jan 2012, Lyon, France. pp.978-2-9539515-2-3. hal-00656538

**HAL Id: hal-00656538**

**<https://hal.science/hal-00656538>**

Submitted on 17 Jan 2012

**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.

# Segmentation et classification de points 3D obtenus à partir de relevés laser terrestres : une approche par super-voxels

Ahamad Kamal Aijazi

Paul Checchin

Laurent Trassoudaine

Clermont Université, Université Blaise Pascal, Institut Pascal, BP 10448, F-63000 CLERMONT-FD  
CNRS, UMR 6602, Institut Pascal, F-63171 AUBIERE

kamalaijazi@gmail.com, {paul.checchin, laurent.trassoudaine}@univ-bpclermont.fr

## Résumé

*Une méthode de classification de données 3D éparses obtenues à partir de relevés laser terrestres en environnement urbain est présentée. Elle est fondée sur une technique de segmentation manipulant des super-voxels. Le nuage de points 3D est tout d'abord divisé en voxels caractérisés par plusieurs attributs. Ils deviennent des super-voxels qui sont liés entre eux pour former des objets par une méthode de chaînage plutôt que par une croissance de régions classique. Ces objets sont ensuite classifiés à partir de modèles géométriques et de descripteurs locaux. Afin d'évaluer les résultats, une nouvelle métrique est introduite. Elle caractérise à la fois les résultats de la segmentation et de la classification. L'influence sur les résultats de classification de la taille du voxel et de l'incorporation des informations couleur RVB et intensité de réflectance dans le super-voxel est aussi analysée.*

## Mots Clef

Segmentation, point 3D, image 3D, super-voxel, classification, scènes urbaines.

## Abstract

*A method to classify urban scenes based on a super-voxel segmentation of sparse 3D data is presented. The 3D point cloud is first segmented into voxels which are then characterized by several attributes transforming them into super-voxels. These are joined together by using a link-chain method rather than the usual region growing algorithm to create objects. These objects are then classified using geometrical models and local descriptors. In order to evaluate the results, a new metric is presented which combines both segmentation and classification results simultaneously. The effects of voxel size and incorporation of RGB color and reflectance intensity on the classification results are also discussed.*

## Keywords

Segmentation, 3D Point Cloud, Super-Voxel, Classification, Urban Scene.

## 1 Introduction

La segmentation et la classification automatiques de données 3D issues de scènes urbaines suscitent un grand intérêt dans la communauté scientifique en raison de la demande croissante d'analyses et de cartographies de ces situations. Par ailleurs, les progrès de la technologie d'acquisition de ces données ne sont pas étrangers à cette évolution. L'extraction automatique (ou partiellement supervisée) d'importantes structures urbaines telles que les routes, la végétation, les lampadaires et les bâtiments à partir de ces données 3D est une approche intéressante pour l'analyse de ces scènes car cela peut considérablement réduire les ressources nécessaires à l'analyse des données et à la modélisation 3D des villes. Une façon courante de recueillir rapidement des données 3D d'environnements urbains est d'utiliser un LiDAR aéroporté [23], [27]. Cette solution génère un scan 3D en un temps très court mais l'angle d'observation n'est pas adapté à l'analyse de ces scènes. L'utilisation d'un LiDAR embarqué sur un véhicule terrestre permet d'obtenir des données denses avec un point de vue plus proche des paysages urbains. Les données sont donc plus riches, notamment sur les façades d'immeubles. Toutefois des inconvénients persistent avec notamment la nécessité d'une puissance de traitement accrue pour gérer l'augmentation du volume de données 3D. En contrepartie, il est avantageux de disposer d'une discrétisation plus détaillée des vues latérales des objets pour aboutir à un modèle plus complet des structures urbaines, en particulier des façades de bâtiments, des lampadaires, etc.

Le travail présenté traite de la segmentation et de la classification de données 3D de scènes urbaines<sup>1</sup>. L'objectif est de fournir un pré-traitement efficace pour des algorithmes plus avancés ou un accélérateur pour des techniques de classification plus spécifiques.

## 2 Etat de l'art

L'étude bibliographique est scindée en deux parties relatives à la segmentation et à la classification. Les approches au sein de celles-ci sont regroupées par tendances.

<sup>1</sup>Ce travail est supporté par l'Agence Nationale de la Recherche (ANR CONTINT iSpace & Time – ANR-10-CONT-23) et par le Conseil Général de l'Allier.

## 2.1 Dans le domaine de la segmentation

Afin d'exploiter pleinement les nuages de points 3D, une segmentation efficace s'avère être une étape de pré-traitement nécessaire et essentielle dans un certain nombre de tâches de perception autonome.

### **Descripteurs spécialisés et discontinuités surfaciques.**

Parmi les premiers travaux sur le sujet, les auteurs de [1], [12] et [17] ont utilisé de petits ensembles de caractéristiques spécialisées telles que la densité locale des points ou la hauteur au sol pour discriminer quelques catégories d'objets dans les scènes d'extérieur, ou pour séparer les objets de l'arrière-plan. Dans la littérature, on trouve aussi certaines méthodes de segmentation fondées sur les discontinuités de surface telle que celle de Moosman et al. [16] qui ont utilisé, pour un maillage du terrain, la convexité des surfaces comme séparateur entre les objets.

**Graph clustering.** Dernièrement, la phase de segmentation est communément formulée comme un problème de regroupements à l'aide de graphes. Les exemples de telles approches sont les « Graph-Cuts » avec notamment ses versions dites normalisées ou les « Min-Cuts ». Golovinskiy et Funkhouser [5] ont étendu la segmentation des nuages de points 3D par « Graph-Cuts » en utilisant les k-plus proches voisins (k-NN) pour construire un graphe 3D. Dans leurs travaux, ils affectent une pondération aux arcs calculée selon la décroissance exponentielle de la distance. Une limite de cette méthode est qu'elle nécessite la connaissance préalable de l'emplacement de l'objet à segmenter. Un autre algorithme de segmentation pour des images d'espaces naturels, récemment introduit par Felzenszwalb et Huttenlocher (FH) [4], est largement utilisé pour plusieurs applications robotiques en raison de son efficacité. Zhu et al. [30] ont présenté une méthode dans laquelle un graphe 3D est construit par une technique k-NN tout en faisant l'hypothèse d'un monde plan pour effectuer des suppressions lors du pré-traitement (nous avons effectué la même hypothèse). Le partitionnement 3D est ensuite obtenu avec l'algorithme FH. Triebel et al. [26] ont modifié l'algorithme FH pour les images de profondeur en proposant une technique de segmentation non supervisée probabiliste. Dans cette approche, les données 3D sont dans un premier temps sur-segmentées lors du pré-traitement. Schoenberg et al. [22] ont appliqué l'algorithme FH aux données 3D couleur obtenues à partir d'une paire caméra-laser. Le poids des arcs est calculé comme une combinaison pondérée des distances euclidiennes, des différences d'intensité des pixels et des angles entre les normales aux surfaces estimées en chaque point 3D. L'algorithme FH est alors exécuté sur le graphe de l'image pour fournir la segmentation définitive. L'évaluation de l'algorithme est uniquement fait sur des portions de route. Strom et al. [24] ont proposé une approche similaire, mais l'algorithme FH est modifié pour intégrer les différences entre les angles des normales aux surfaces en plus des différences de couleur. L'évaluation de la segmentation a été faite visuellement sans vérité de

terrain.

Notre approche diffère des méthodes mentionnées ci-dessus car au lieu d'utiliser les propriétés de chaque point pour la segmentation, ce qui conduit souvent à une sur-segmentation, nous avons regroupé les points 3D en voxels à partir de la distance euclidienne puis nous avons assigné des propriétés normalisées à ces voxels pour les transformer en super-voxels. Cette solution permet d'éviter la sur-segmentation et réduit de manière conséquente l'ensemble des données ainsi que le temps de post-traitement.

**Primitives géométriques.** Une approche de segmentation de points 3D à base d'arbres a été proposée dans [19]. Les noeuds du graphe correspondent aux primitives géométriques qui sont ici des ellipsoïdes gaussiens. Ces ellipsoïdes sont ensuite fusionnés en utilisant un algorithme de croissance d'arbres. Le résultat de la segmentation est similaire à celui obtenu à partir de super-voxels avec des voxels de formes ellipsoïdales et de tailles variables. Contrairement à cette méthode, notre approche utilise des parallélépipèdes de différentes tailles comme entités géométriques et une méthode de chaînage pour les regrouper.

## 2.2 Dans le domaine de la classification

Auparavant, la recherche liée à la classification et l'analyse de scènes 3D urbaines était principalement menée en utilisant soit les données recueillies par LiDAR aéroporté pour l'extraction des bâtiments du sol nu [15] [29], soit les données 3D issues de télémètres statiques pour l'extraction des caractéristiques des bâtiments comme les murs et les fenêtres [20].

**Modèles pour discriminer.** Dans [13], une méthode multi-échelle de champs aléatoires conditionnels est proposée pour classer des données laser 3D d'extérieur en introduisant des arcs régionaux en plus des arcs locaux et des noeuds existants. Le processus se poursuit par l'ajustement de portions planaires sur les objets tels que le sol ou les bâtiments en utilisant un algorithme de RANSAC comme post-traitement, pour aboutir à un modèle géométrique de la scène. Dans [11], les auteurs extraient les routes et les objets à proximité tels que les panneaux. Ils utilisent une méthode des moindres carrés puis un RANSAC pour ajuster et extraire des plans à partir des données. Un filtre de Kalman permet l'extraction des routes en milieu urbain. Enfin, Douillard et al. [2] ont présenté une méthode dans laquelle les points 3D sont projetés sur l'image pour trouver des régions d'intérêt pour la classification.

**Descripteurs.** Une méthode de classification exploitant des caractéristiques globales est présentée dans [6] : une seule image globale panoramique ou sphérique (« spin image ») est utilisée pour chaque objet afin de détecter les voitures dans la scène alors que dans [21], une technique à base d'histogramme, « Fast Point Feature Histogram » (FPFH), permet d'accéder à des caractéristiques globales pour une identification de l'objet et une détection du point de vue simultanées. La littérature propose également des méthodes de classification utilisant des descripteurs et des

caractéristiques locales via les « spin images » [7], des descripteurs harmoniques sphériques [8], des signatures à base de diffusion de la chaleur [25], des distributions de formes [18], et des SURF 3D [10]. Il y a également un troisième type de classification à base de « Bag Of Features (BOF) » tel que le propose [14].

Dans le travail proposé, nous utilisons des modèles géométriques construits à partir des caractéristiques locales et des descripteurs pour classer les différents objets segmentés d'une scène urbaine représentés par des groupes de voxels. Le terrain est supposé plan et utilisé comme un séparateur d'objets. La technique de segmentation est décrite dans la section 3. La section 4 traite de la classification de ces objets segmentés. Dans la section 5, une nouvelle métrique d'évaluation est introduite pour évaluer à la fois la segmentation et la classification, tandis que dans la section 6, nous présentons les résultats de l'algorithme. Enfin, nous concluons à la section 7.

### 3 Segmentation des voxels

#### 3.1 Voxélisation des données

Lorsque l'on traite de grands ensembles de données 3D, le coût du traitement de tous les points est très élevé et le processus devient inadapté pour des applications temps réel. Il est donc nécessaire de chercher à réduire le nombre de ces points par regroupement ou en retirant les points redondants ou inutiles. Ainsi, les différents points 3D sont ici regroupés pour former une représentation de niveau supérieur ou voxel tel que le présente la figure 1.

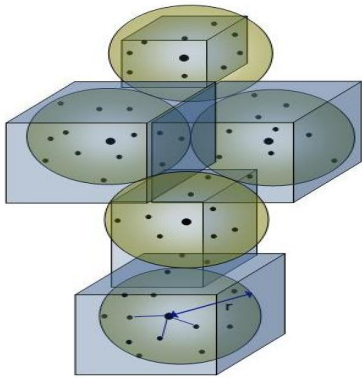


FIGURE 1 – Plusieurs points 3D sont regroupés pour former des voxels parallélépipédiques de taille maximale  $2r$ . La taille des voxels est définie par les valeurs extrémales des coordonnées des points rassemblés au sein du voxel.

Pour  $p$  points 3D,  $s$  voxels, avec  $s \ll p$ , sont constitués en utilisant les  $k$ -NN et le poids  $w = 1/d$  pour chaque voxel voisin ( $d$  est la distance euclidienne au voisin). La taille maximale d'un voxel  $2r$ , où  $r$  est le rayon de la sphère incluant les points voisins, dépend de la densité du nuage de points 3D. Dans [13], les valeurs des couleurs sont aussi prises en compte mais on observe que pour des tailles de voxels relativement petites, la variation des propriétés des voxels est faible alors que le coût calculatoire augmente significativement. Ainsi, la couleur n'est pas retenue pour la construction des voxels et seule la distance intervient à

ce stade. En revanche, d'autres paramètres seront pris en compte lors de la segmentation des voxels qui aboutira aux objets. Par ailleurs, on s'assure que chaque point 3D qui appartient à un voxel ne peut être considéré pour une voxélisation ultérieure. Cela réduit le temps de calcul et évite une sur-segmentation. Nous utilisons des parallélépipèdes pour les voxels car leur symétrie évite des problèmes d'ajustement lors des regroupements et minimise l'impact de la forme des voxels pendant l'extraction des caractéristiques. Bien que la taille maximale d'un voxel soit prédéfinie, les tailles réelles varient en fonction des coordonnées extrémales des points du voisinage pour respecter le profil de la structure.

Une fois que les voxels sont créés, il est nécessaire de trouver les propriétés de chaque voxel. Ces propriétés comprennent les normales aux surfaces, la couleur RVB, l'intensité, les primitives géométriques telles que le barycentre, le centre géométrique, les valeurs minimales et maximales selon chaque axe, etc. Certaines de ces propriétés sont des valeurs moyennées et normalisées. Les normales aux surfaces sont calculées en utilisant l'ACP (Analyse en Composante Principale). Dans [9], il a été montré que l'ACP donnait de meilleurs résultats que la méthode de l'aire moyenne pour estimer les normales. Etant donné un nuage de points 3D  $\mathcal{D} = \{x_i\}_{i=1}^n$ , l'approximation, par ACP, de la normale en un point  $p \in \mathcal{D}$  donné, est typiquement calculée en déterminant les  $k$ -NN  $x_k \in \mathcal{D}$  de  $p$ . Etant donné les  $K$  voisins, la valeur approchée de la normale est égale au vecteur propre associé à la plus petite valeur propre de la matrice semi-définie positive :

$$\mathbf{P} = \sum_{k=1}^K (x_k - \bar{p})^T (x_k - \bar{p}) \quad (1)$$

où  $\bar{p}$  est le barycentre des points du voisinage :  $\bar{p} = \frac{1}{K} \sum_{j=1}^K x_j$ . La normale de la surface estimée est ambiguë en termes de signe ; pour tenir compte de cette ambiguïté, une phase d'homogénéisation est effectuée par un simple produit scalaire. Toutefois, le signe du vecteur normal n'est pas important à ce stade de l'algorithme où seule l'orientation est pertinente. Une normale à la surface est estimée pour l'ensemble des points appartenant au voxel. Elle est ensuite associée à ce voxel particulier avec ses autres caractéristiques.

Toutes ces propriétés sont ensuite utilisées dans le regroupement des voxels devenus super-voxels (et appelés à présent  $s$ -voxels) en objets, puis lors de la classification de ces objets. L'avantage de cette approche est que plutôt que d'utiliser des milliers de points pour l'ensemble des données à classifier, on peut maintenant s'appuyer sur un nombre réduit de  $s$ -voxels pour obtenir des résultats similaires. Dans notre cas, des ensembles de données de 110392, 53676 et 27396 points ont été réduits à respectivement 18541, 6928 et 7924  $s$ -voxels avant de subir un autre traitement.

#### 3.2 Groupement par chaînage

L'étape qui fait suite à la conversion des données en  $s$ -voxels est le regroupement de ceux-ci afin d'obtenir une

segmentation en objets. Habituellement, pour ce type de tâche, un algorithme de croissance de région [16] [28] est utilisé dans lequel les propriétés de toute la région influencent les conditions aux frontières. Cela peut parfois conduire à une segmentation erronée. Il est aussi commun dans ce type de méthodes d'utiliser une approche fondée sur la fusion des nœuds dans laquelle, à partir de chaque nœud, les conditions de similarité sont vérifiées dans les 5 directions possibles (on ne revient pas en arrière). L'algorithme proposé ici introduit une méthode de chaînage au lieu de fusionner les  $s$ -voxels pour obtenir des objets segmentés.

Dans cette méthode, chaque  $s$ -voxel est considéré comme un maillon d'une chaîne. Tous les maillons secondaires, attachés à chacun des maillons principaux, sont déterminés. A la fin, tous les maillons principaux sont reliés entre eux, en enlevant les maillons secondaires redondants, pour former une chaîne continue comme le montre la figure 2.

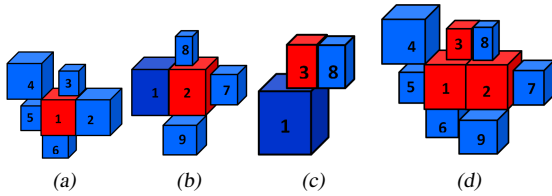


FIGURE 2 – Regroupement de  $s$ -voxels par chaînage. (a) Le  $s$ -voxel 1 est utilisé comme maillon principal (en rouge), les maillons secondaires associés sont en bleu. (b) et (c) représentent une configuration équivalente pour les  $s$ -voxels principaux 2 et 3. (d) Liaison entre les maillons principaux pour former une chaîne en éliminant les maillons secondaires redondants.

Soit  $\mathbf{V}_P$  un maillon principal et  $\mathbf{V}_n$  le  $n^{\text{ième}}$  maillon secondaire. Chaque  $\mathbf{V}_n$  est lié à  $\mathbf{V}_P$  si et seulement si les trois conditions suivantes sont remplies :

$$|\mathbf{V}_{P_{X,Y,Z}} - \mathbf{V}_{n_{X,Y,Z}}| \leq (w_D + c_D) \quad (2)$$

$$|\mathbf{V}_{P_{R,V,B}} - \mathbf{V}_{n_{R,V,B}}| \leq 3\sqrt{w_C} \quad (3)$$

$$|\mathbf{V}_{P_I} - \mathbf{V}_{n_I}| \leq 3\sqrt{w_I} \quad (4)$$

où, pour le maillon principal et secondaire respectivement :

- $\mathbf{V}_{P_{X,Y,Z}}, \mathbf{V}_{n_{X,Y,Z}}$  sont les centres géométriques ;
- $\mathbf{V}_{P_{R,V,B}}, \mathbf{V}_{n_{R,V,B}}$  sont les valeurs moyennes de RVB ;
- $\mathbf{V}_{P_I}, \mathbf{V}_{n_I}$  sont les valeurs moyennes d'intensité ;
- $w_C$  est le coefficient de pondération de la couleur, égal à la valeur maximum des variances  $Var(R, V, B)$  ;
- $w_I$  est le coefficient de pondération de l'intensité, égal à la valeur maximum des variances  $Var(I)$ .

$w_D$  est le poids en distance donné par  $\frac{(\mathbf{V}_{P_{s_{X,Y,Z}}} + \mathbf{V}_{n_{s_{X,Y,Z}}})}{2}$ .  $s_{X,Y,Z}$  représente la taille d'un  $s$ -voxel selon les axes  $X, Y$  et  $Z$ .

$c_D$  est une constante d'inter-distance qui tient compte de la densité des points mais aussi des éventuelles erreurs de mesures, occlusions et trous. Sa valeur doit être définie avec

précaution en fonction des données.

L'orientation de la normale n'est pas prise en compte à ce stade afin de permettre la segmentation d'objets complets et non celle de simples plans.

Cette méthode de segmentation garantit que seules les conditions des frontières adjacentes sont prises en compte, sans influence des propriétés d'un voisin éloigné. Cela peut se révéler plus adapté aux forts changements structurels d'un environnement urbain. Le principe de la segmentation est résumé dans l'algorithme 1.

#### Algorithme 1 Segmentation

- 1: **répéter**
- 2: Sélectionner un point 3D pour la voxélisation
- 3: Trouver tous les points du voisinage à inclure dans le voxel en utilisant les k-NN en spécifiant la taille maximale du voxel.
- 4: Trouver toutes les caractéristiques du  $s$ -voxel y compris la normale à la surface en utilisant l'ACP
- 5: **jusqu'à ce que** tous les points 3D appartiennent à un voxel
- 6: **répéter**
- 7: Choisir un  $s$ -voxel comme maillon principal
- 8: Trouver tous les maillons secondaires attachés à un maillon principal
- 9: **jusqu'à ce que** tous les  $s$ -voxels soient utilisés
- 10: Lier tous les maillons principaux pour former une chaîne en éliminant les maillons redondants dans le processus

Avec cette méthode 18541, 6928 et 7924  $s$ -voxels obtenus à partir du traitement de 3 jeux de données différents ont pu être respectivement segmentés en 237, 75 et 41 objets distincts.

## 4 Classification des objets

Afin de classer ces objets, nous supposons que le sol est un plan. Il est vu comme un séparateur entre les objets. Pour ce faire, le sol est classifié et éliminé de la scène afin que seuls les autres objets subsistent comme s'ils étaient suspendus dans l'espace. Ils sont ainsi bien identifiés ce qui les rend plus facile à classer comme le montre la figure 3. Le sol ou la route puis les objets restants sont classifiés

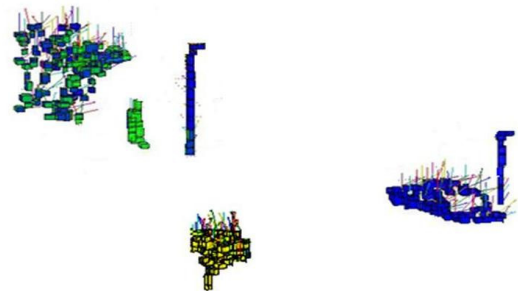


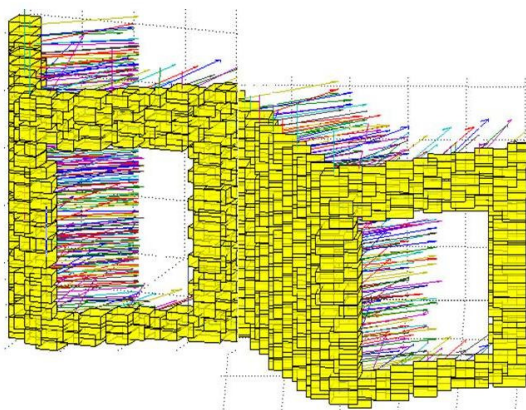
FIGURE 3 – Objets de la scène segmentés après retrait du sol.

en utilisant leurs caractéristiques géométriques et locales. Elles incluent principalement :

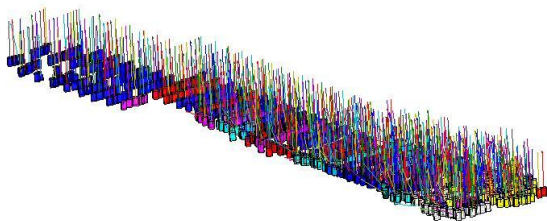
- a. **la normale à la surface** : l'orientation des normales à la surface est essentielle pour la classification du sol et des façades de bâtiments. Pour les objets liés au sol,

les normales sont colinéaires à l'axe Z (axe de l'altitude) alors que pour les façades, les normales sont parallèles à l'axe XY (plan du sol), voir figure 4 ;

- b. **le barycentre et le centre géométrique** : la différence de hauteur entre le centre géométrique et le barycentre est un paramètre très performant pour distinguer les objets comme les arbres et la végétation, où  $h(\text{barycentre} - \text{centre géométrique}) > 0$  avec  $h$  étant la fonction hauteur ;
- c. **la couleur et l'intensité** : couleur et intensité sont également des facteurs discriminants entre plusieurs objets ;
- d. **la forme** : la forme des objets joue naturellement un rôle important puisque les piétons, longs et étroits, de même que les poteaux, encore plus longs, se distinguent des voitures et des arbres plus larges. De même, alors que le sol est un plan à une altitude basse, les bâtiments sont représentés par de grands plans verticaux.



(a) Normales aux bâtiments.



(b) Normales à la route.

FIGURE 4 – (a) La normale à la surface des  $s$ -voxels appartenant aux bâtiments sont parallèles au plan du sol. (b) La normale à la surface des  $s$ -voxels appartenant à la route sont bien perpendiculaires au plan du sol.

L'utilisation de ces descripteurs permet la classification des scènes urbaines en 5 classes différentes : bâtiment, route, voiture, poteau et arbre. Les résultats de classification et une nouvelle métrique d'évaluation sont discutés dans les sections suivantes.

## 5 Evaluation

Dans des travaux antérieurs, différents paramètres d'évaluation ont été introduits pour quantifier indépendamment les résultats de segmentation et ceux de classification. Nous

proposons ici une nouvelle métrique qui permet une évaluation simultanée. Elle est fondée sur la comparaison du pourcentage total de  $s$ -voxels correctement classés pour un objet donné. Soit  $T_i$ ,  $i \in \{1, \dots, N\}$ , le nombre total de  $s$ -voxels distribués dans les objets appartenant aux  $N$  classes ; il s'agit de la vérité de terrain. Soit  $t_j$ ,  $i \in \{1, \dots, N\}$ , le nombre total de  $s$ -voxels classés dans une classe particulière  $j$  et distribués dans les objets appartenant aux  $N$  différentes classes (par exemple un  $s$ -voxel classifié comme appartenant à un bâtiment peut en fait appartenir à un arbre) alors le rapport  $S_{jk}$  ( $j$  est le numéro de la classe qui apparaît sur une ligne de la matrice, voir table 2 par exemple, et  $k \in \{1, \dots, N\}$ ) est donné par  $S_{jk} = \frac{t_{jk}}{T_k}$ . Ces valeurs  $S_{jk}$  sont calculées pour chaque classe et sont utilisées pour remplir une matrice de confusion. Chaque ligne de la matrice représente une classe particulière. Ainsi pour une classe de type 1, on définit les formules suivantes :

- **Vrai positif (True Positive) TP** =  $S_{11}$  (i.e. la diagonale de la matrice représente les TPs)
- **Faux positif (False Positive) FP** =  $\sum_{m=2}^N S_{1m}$
- **Vrai négatif (True Negative) TN** =  $(1 - \text{FP})$
- **Faux négatif (False Negative) FN** =  $(1 - \text{TP})$

La diagonale de la matrice (ou TPs) indique la précision de la segmentation (Segmentation ACCuracy) **SACC**, de façon similaire au score récemment introduit par Douillard et al. [3]. Les effets des  $s$ -voxels non classifiés sont automatiquement incorporés dans la précision de la segmentation. En utilisant les valeurs ci-dessus, la précision de la classification (Classification ACCuracy) **CACC** est donnée par :

$$\text{CACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \quad (5)$$

La valeur de **CACC** est calculée pour les  $N$  classes d'objets présentes dans la scène. La précision globale de la classification (Overall Classification ACCuracy) **OCACC** est calculée comme suit :

$$\text{OCACC} = \frac{1}{N} \sum_{i=1}^N \text{CACC}_i \quad (6)$$

De la même manière, la précision globale de segmentation (Overall Segmentation ACCuracy) **OSACC** peut être calculée.

Les valeurs  $T_i$  et  $t_j$  sont laborieusement saisies à la main en mettant en correspondance les données voxelisées, les  $s$ -voxels et les points finalement classifiés.

## 6 Résultats

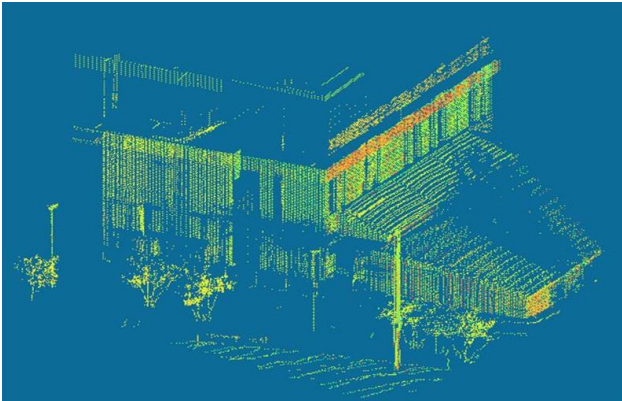
L'algorithme a été validé sur des données 3D acquises pour différentes scènes urbaines sur le campus de l'Université Blaise Pascal de Clermont-Ferrand à l'aide d'un scanner laser 3D LEICA HDS-3000. Ces points 3D sont aussi caractérisés par leur intensité de réflectance et leurs valeurs RVB. Les résultats de trois ensembles de données sont discutés ici. Les informations relatives à ces trois jeux sont résumés dans la table 1. Les résultats de segmentation et de classification sont présentés sur les figures 5, 6 et 7 et dans les tables 2, 3 et 4. Ils ont été obtenus en utilisant une taille maximale de voxel de  $0,3 \text{ m}$  et  $c_D = 0,25 \text{ m}$ .

TABLE 1 – Résultats sur les données 3D Université Blaise Pascal.

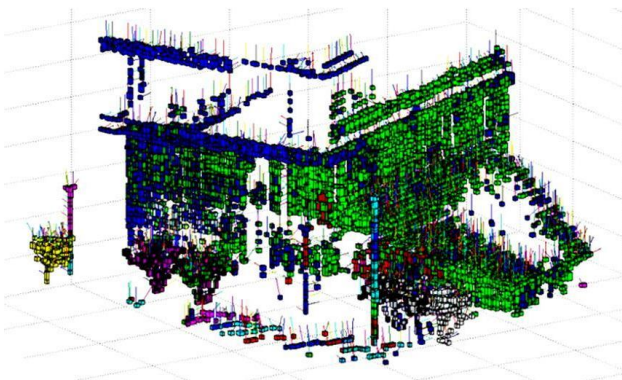
Jeu de données #	Nombre de points 3D	Nombre de $s$ -voxels segmentés	Nombre d'objets segmentés
# 1	27,396	7,924	41
# 2	53,676	6,928	75
# 3	110,392	18,541	237

TABLE 2 – Résultat de classification du 1<sup>er</sup> jeu de données dans la nouvelle métrique d'évaluation.

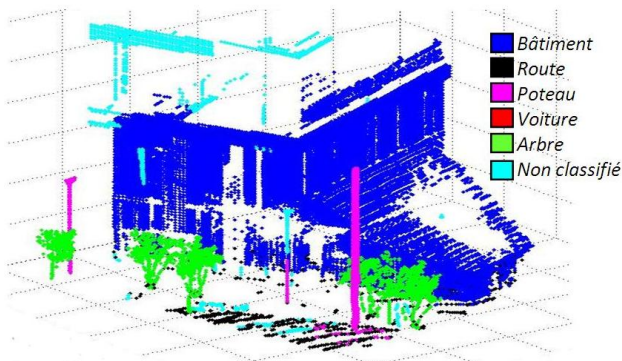
	Bâtiment	Route	Arbre	Poteau	Voiture	CACC
Bâtiment	0,943	0,073	0	0	0	0,935
Route	0,007	0,858	0,015	0,008	0	0,914
Arbre	0	0,025	0,984	0	0	0,979
Poteau	0	0,049	0	0,937	0	0,944
Voiture	–	–	–	–	–	–
Précision globale de segmentation : OSACC						0,930
Précision globale de classification : OCACC						0,943



(a) Points 3D.



(b) Voxélisation et segmentation en objets.



(c) Points labélisés.

FIGURE 5 – (a) Les points 3D du premier jeu de données. (b) La voxélisation des points 3D. (c) Le résultat de classification (points 3D labélisés).

## 6.1 Effet de la taille des voxels sur la précision de la classification

Comme les propriétés d'un  $s$ -voxel sont calculées en fonction de tous les attributs de ses éléments et qu'elles sont RFIA 2012

TABLE 3 – Résultat de classification du 2<sup>e</sup> jeu de données dans la nouvelle métrique d'évaluation.

	Bâtiment	Route	Arbre	Poteau	Voiture	CACC
Bâtiment	0,996	0,007	0	0	0	0,995
Route	0	0,906	0,028	0,023	0,012	0,921
Arbre	0	0,045	0,922	0	0	0,938
Poteau	0	0,012	0	0,964	0	0,976
Voiture	0	0,012	0	0	0,907	0,947
Précision globale de segmentation : OSACC						0,939
Précision globale de classification : OCACC						0,955

TABLE 4 – Résultat de classification du 3<sup>e</sup> jeu de données dans la nouvelle métrique d'évaluation.

	Bâtiment	Route	Arbre	Poteau	Voiture	CACC
Bâtiment	0,901	0,005	0,148	0	0	0,874
Route	0,003	0,887	0,011	0,016	0,026	0,916
Arbre	0,042	0,005	0,780	0	0	0,867
Poteau	0	0,002	0	0,966	0	0,982
Voiture	0	0,016	0,12	0	0,862	0,863
Précision globale de segmentation : OSACC						0,879
Précision globale de classification : OCACC						0,901

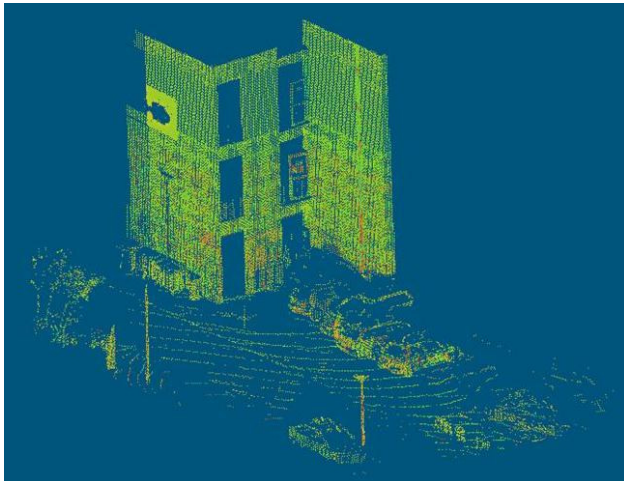
ensuite utilisées pour la segmentation puis la classification, la taille des voxels a une incidence sur le processus de classification. Par ailleurs, comme la taille des voxels change, la constante d'interdistance  $c_D$  doit également être ajustée. L'effet de la taille des voxels sur la classification est étudié en faisant varier cette taille ainsi que  $c_D$  entre 0,1  $m$  et 1,0  $m$  sur le premier jeu de données. La précision de la classification est présentée sur la figure 8a. Ensuite, pour ces mêmes valeurs, le temps de traitement est indiqué sur la figure 8b.

Une valeur arbitraire de temps  $T_a$  est choisie pour la comparaison (sur l'axe Z, le temps varie de 0 à 200  $T_a$ ). Le résultat de comparaison est ainsi indépendant du processeur utilisé, inchangé ici.

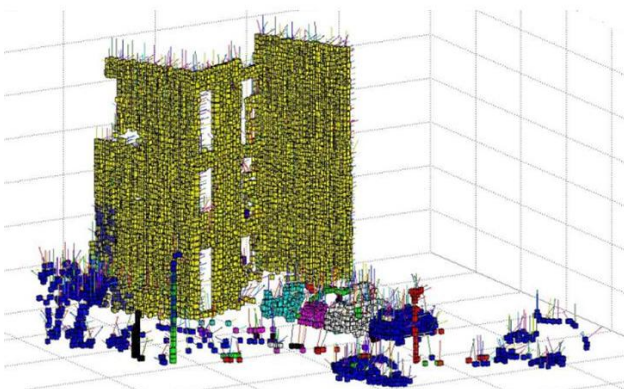
Les résultats montrent qu'avec une taille de voxel plus petite, la segmentation et la classification sont meilleures (avec une valeur convenable de  $c_D$ ) mais que le coût calculatoire augmente. Il est évident que la variation de  $c_D$  n'a pas d'impact significatif sur le temps  $t$ . On observe également que la qualité de la classification ne s'améliore plus beaucoup lorsque la taille a atteint une certaine valeur, mais que le coût du calcul continue de se dégrader. Comme OCACC et le temps (tous deux tracés sur l'axe Z) sont indépendants, en combinant les deux tracés 3D de la figure 8, on trouve les valeurs optimales qui dépendent de l'exigence de l'application. Dans notre cas, on prend une taille maximale de voxel de 0,3  $m$  et  $c_D = 0,25 m$ .

## 6.2 Couleur RVB et intensité de réflectance

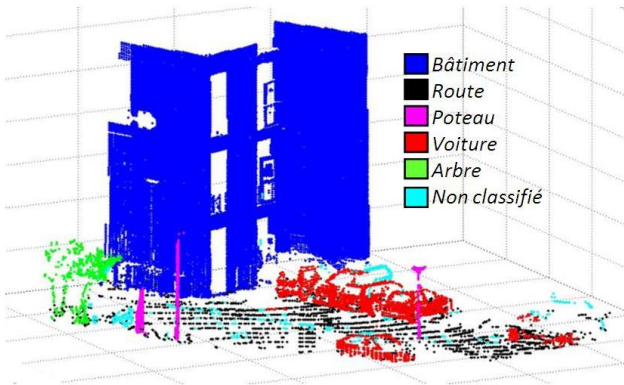
L'effet de l'introduction de la couleur et de l'intensité de réflectance est également étudié. Les résultats sont présentés dans la table 5. On observe que l'incorporation de la couleur RVB seule n'est pas suffisante en environnement urbain du fait que la perception de ce dernier est fortement impactée par les variations d'éclairage (une partie d'un objet peut être dans l'ombre ou réfléchir la lumière du soleil), y compris dans la même scène. Le processus de segmentation est détérioré ainsi que la classification. C'est peut-être la cause de la moins bonne classification qui est observable dans la première partie de la table 5. C'est la raison pour laquelle l'intensité (du signal réfléchi) est introduite. Elle est indépendante de l'illumination et ainsi plus consistante. Les résultats de la classification ainsi améliorés sont présentés dans la deuxième partie de la table 5.



(a) Points 3D.



(b) Voxélisation et segmentation en objets.

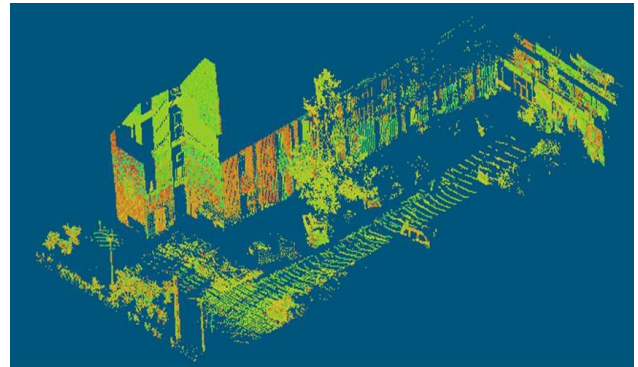


(c) Points labélisés.

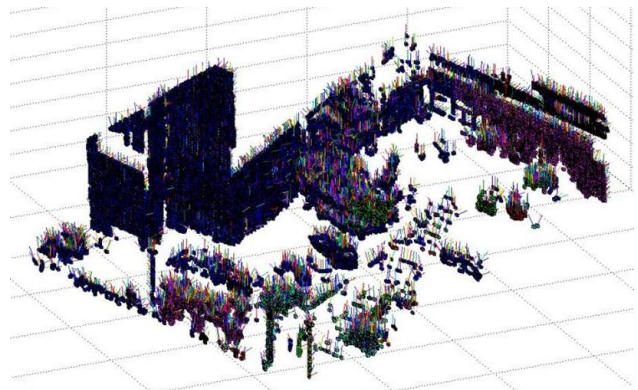
FIGURE 6 – (a) Les points 3D du deuxième jeu de données. (b) La voxélisation des points 3D. (c) Le résultat de classification (points 3D labélisés).

## 7 Conclusion

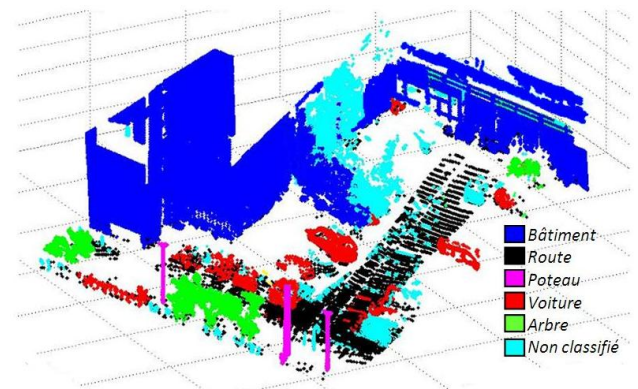
Nous avons présenté une méthode de segmentation et de classification des scènes urbaines 3D fondée sur une approche à base de super-voxels. Pour la segmentation, une méthode de groupement par chaînage est proposée. Elle est suivie par une classification des objets en utilisant des descripteurs locaux et des modèles géométriques. Afin d'évaluer les résultats, nous avons introduit une nouvelle métrique qui quantifie à la fois les résultats de segmentation et de classification. Les scores obtenus atteignent une pré-



(a) Points 3D.



(b) Voxélisation et segmentation en objets.



(c) Points labélisés.

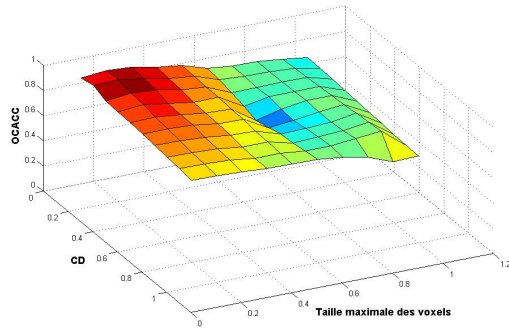
FIGURE 7 – (a) Les points 3D du troisième jeu de données. (b) La voxélisation des points 3D. (c) Le résultat de classification (points 3D labélisés).

cision globale de segmentation de 87 % et une précision de classification d'environ 90 %. Notre étude montre que la précision de classification est accrue si la taille des voxels est diminuée mais au détriment du coût calculatoire. Aussi, une solution pour déterminer une valeur optimale est présentée. Ce travail montre également l'importance d'utiliser des valeurs d'intensité ainsi que des couleurs RVB dans la segmentation et la classification. La méthode proposée peut aussi être utilisée comme une aide (« add-on boost ») pour d'autres algorithmes de classification.

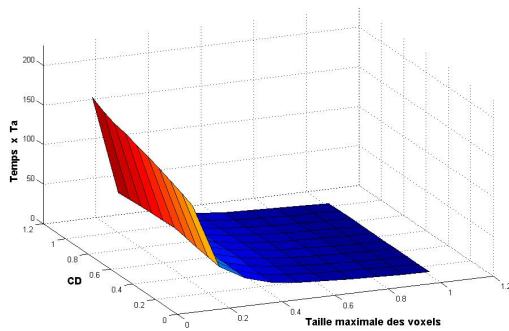
## Références

- [1] D. Anguelov, B. Taskar, V. Chatalbashev, D. Koller, D. Gupta, G. Heitz, and A. Ng. Discriminative Learning of Markov Random





(a) Influence de la taille des voxels sur OCACC.



(b) Influence de la taille des voxels sur le temps de traitement.

FIGURE 8 – (a) L'effet de la taille maximale des voxels sur la variation de OCACC. (b) L'effet de la taille maximale des voxels sur la variation du temps de traitement. En utilisant les deux graphiques, on trouve facilement la valeur optimale pour ces 2 paramètres (taille maximale et  $CD$ ).

TABLE 5 – Précision globale de la segmentation et de la classification avec les caractéristiques couleur et intensité de réflectance.

Jeu de données #	Couleur seule		Intensité et couleur	
	OSACC	OCACC	OSACC	OCACC
# 1	0,660	0,772	0,930	0,943
# 2	0,701	0,830	0,939	0,955
# 3	0,658	0,766	0,879	0,901

Fields for Segmentation of 3D Scan Data. In *Conf. on Computer Vision and Pattern Recognition*, Los Alamitos, CA, USA, 2005.

[2] B. Douillard, A. Brooks, and F.T. Ramos. A 3D Laser and Vision Based Classifier. In *5th Int. Conf. on Intelligent Sensors, Sensor Networks and Information Processing*, Melbourne, Australia, 2009.

[3] B. Douillard, J. Underwood, N. Kuntz, V. Vlaskine, A. Quadros, P. Morton, and A. Frenkel. On the Segmentation of 3D LIDAR Point Clouds. In *IEEE Int. Conf. on Robotics and Automation*, Shanghai, China, May 2011.

[4] P. Felzenszwalb and D. Huttenlocher. Efficient Graph-Based Image Segmentation. *Int. Journal of Computer Vision*, 59 :167–181, 2004.

[5] A. Golovinskiy and T. Funkhouser. Min-Cut Based Segmentation of Point Clouds. In *IEEE Workshop on Search in 3D and Video (S3DV) at ICCV*, September 2009.

[6] A. Halma, F. ter Haar, E. Bovenkamp, P. Eendebak, and A. van Eekeren. Single spin image-ICP matching for efficient 3D object recognition. In *Proceedings of the ACM workshop on 3D object retrieval*, 3DOR '10, pages 21–26, New York, NY, USA, 2010.

[7] A. Johnson. *Spin-Images : A Representation for 3-D Surface Matching*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, August 1997.

[8] M. Kazhdan, T. Funkhouser, and S. Rusinkiewicz. Rotation Invariant Spherical Harmonic Representation of 3D Shape Descriptors. In *Symposium on Geometry Processing*, June 2003.

[9] K. Klasing, D. Althoff, D. Wollherr, and M. Buss. Comparison of Surface Normal Estimation Methods for Range Sensing Applications. In *IEEE Int. Conf. on Robotics and Automation*, Kobe, Japan, May 2009.

[10] J. Knopp, M. Prasad, and L. Van Gool. Orientation invariant 3D object classification using hough transform based methods. In *Proceedings of the ACM workshop on 3D object retrieval*, 3DOR '10, pages 15–20, New York, NY, USA, 2010.

[11] J. Lam, K. Kusevic, P. Mrstik, R. Harrap, and M. Greenspan. Urban Scene Extraction from Mobile Ground Based LiDAR Data. In *Int. Symposium on 3D Data Processing Visualization and Transmission*, Paris, France, 2010.

[12] E. Lim and D. Suter. Conditional Random Field for 3D Point Clouds with Adaptive Data Reduction. In *Int. Conf. on Cyberworlds*, pages 404–408, Hannover, October 2007.

[13] E. Lim and D. Suter. Multi-scale Conditional Random Fields for Over-Segmented Irregular 3D Point Clouds Classification. In *Computer Vision and Pattern Recognition Workshop*, pages 1–7, Anchorage, AK, USA, June 2008.

[14] Y. Liu, H. Zha, and H. Qin. Shape Topics-A Compact Representation and New Algorithms for 3D Partial Shape Retrieval. In *IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 2025–2032, New York, NY, USA, June 2006.

[15] W. L. Lu, K. Okuma, and J. J. Little. A Hybrid Conditional Random Field for Estimating the Underlying Ground Surface from Airborne LiDAR Data. *IEEE Transactions on Geoscience and Remote Sensing*, 47(8) :2913–2922, 2009.

[16] F. Moosmann, O. Pink, and C. Stiller. Segmentation of 3D Lidar Data in non-flat Urban Environments using a Local Convexity Criterion. In *Proc. of the IEEE Intelligent Vehicles Symposium (IV)*, pages 215–220, Nashville, Tennessee, USA, June 2009.

[17] D. Munoz, N. Vandapel, and M. Hebert. Onboard contextual classification of 3-D point clouds with learned high-order Markov Random Fields. In *IEEE Int. Conf. on Robotics and Automation*, pages 2009 – 2016, Kobe, Japan, May 2009.

[18] R. Osada, T. Funkhouser, B. Chazelle, and D. Dobkin. Shape distributions. *ACM Trans. Graph.*, 21 :807–832, October 2002.

[19] F. Pauling, M. Bosse, and R. Zlot. Automatic Segmentation of 3D Laser Point Clouds by Ellipsoidal Region Growing. In *Australasian Conf. on Robotics & Automation*, Sydney, Australia, 2009.

[20] S. Pu and G. Vosselman. Building Facade Reconstruction by Fusing Terrestrial Laser Points and Images. *Sensors*, 9(6) :4525–4542, 2009.

[21] R.B. Rusu, G. Bradski, R. Thibaux, and J. Hsu. Fast 3D Recognition and Pose Using the Viewpoint Feature Histogram. In *IEEE/RSJ Int. Conf. on Intellig. Robots and Systems*, Taipei, Taiwan, October 2010.

[22] J. Schoenberg, A. Nathan, and M. Campbell. Segmentation of dense range information in complex urban scenes. In *IEEE/RSJ Int. Conf. on Intellig. Robots and Systems*, Taipei, Taiwan, October 2010.

[23] G. Sithole and G. Vosselman. Experimental comparison of filter algorithms for bare-Earth extraction from airborne laser scanning point clouds. *ISPRS Journal of Photogrammetry and Remote Sensing*, 59(1-2) :85 – 101, 2004.

[24] J. Strom, A. Richardson, and E. Olson. Graph-based Segmentation for Colored 3D Laser Point Clouds. In *Proceedings of the IEEE/RSJ Int. Conf. on Intellig. Robots and Systems*, October 2010.

[25] J. Sun, M. Ovsjanikov, and L. Guibas. A Concise and Provably Informative Multi-Scale Signature Based on Heat Diffusion. In *Proceedings of the Symposium on Geometry Processing*, pages 1383–1392, Aire-la-Ville, Switzerland, 2009.

[26] R. Triebel, J. Shin, and R. Siegwart. Segmentation and Unsupervised Part-based Discovery of Repetitive Objects. In *Proceedings of Robotics : Science and Systems*, Zaragoza, Spain, June 2010.

[27] V. Verma, R. Kumar, and S. Hsu. 3D building detection and modeling from aerial lidar data. In *Conf. on Computer Vision and Pattern Recognition*, New York, USA, June 2006.

[28] M. Vieira and K. Shimada. Surface mesh segmentation and smooth surface extraction through region growing. *Computer Aided Geometric Design*, 22(8) :771 – 792, 2005.

[29] G. Vosselman, P. Kessels, and B. Gorte. The utilisation of airborne laser scanning for mapping. *Int. Journal of Applied Earth Observation and Geoinformation*, 6(3-4) :177 – 186, 2005.

[30] X. Zhu, H. Zhao, Y. Liu, Y. Zhao, and H. Zha. Segmentation and classification of range image from an intelligent vehicle in urban environment. In *IEEE/RSJ Int. Conf. on Intellig. Robots and Systems*, Taipei, Taiwan, October 2010.