HAL
open science

Reconstruction 3D de scènes dynamiques par segmentation au sens du mouvement

Cansen Jiang, Danda Pani Paudel, Yohan Fougerolle, David Fofi, Cédric Demonceaux

To cite this version:

Cansen Jiang, Danda Pani Paudel, Yohan Fougerolle, David Fofi, Cédric Demonceaux. Reconstruction 3D de scènes dynamiques par segmentation au sens du mouvement. RFIA 2016 - Le 20èmé congrès national sur la Reconnaissance des Formes et l'Intelligence Artificielle , Jun 2016, Clermont-Ferrand, France. hal-01338632

HAL Id: hal-01338632

https://hal.science/hal-01338632v1

Submitted on 28 Jun 2016

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.

Reconstruction 3D de scènes dynamiques par segmentation au sens du mouvement

Jiang Cansen Danda Pani Paudel Yohan Fougerolle David Fofi Cédric Demonceaux
Le2i UMR 6306 CNRS, Univ. Bourgogne Franche-Comté
Cansen.Jiang@u-bourgogne.fr

Résumé

L'objectif de ce travail est de reconstruire les parties statiques et dynamiques d'une scène 3D à l'aide d'un robot mobile équipé d'un capteur 3D. Cette reconstruction nécessite la classification des points 3D acquis au cours du temps en point fixe et point mobile indépendamment du déplacement du robot. Notre méthode de segmentation utilise directement les données 3D et étudie les mouvements des objets dans la scène sans hypothèse préalable. Nous développons un algorithme complet reconstruisant les parties fixes de la scène à chaque acquisition à l'aide d'un RANSAC qui ne requiert que 3 points pour recaler les nuages de points. La méthode a été expérimentée sur de larges scènes en extérieur. Par ailleurs, nous montrons sur les séquences tests KITTI que la prise en compte des données 3D permet d'améliorer les approches 2D en levant les ambiguïtés dues à la perte d'une dimension dans les images.

Mots-Clefs

Estimation de pose, reconstruction 3D, SfM, segmentation au sens du mouvement.

Abstract

This paper aims to build the static-map of a dynamic scene using a mobile robot equipped with 3D sensors. The sought static-map consists of only the static scene parts, which has a vital role in scene understanding and landmark based navigation. We propose a Sparse Subspace Clustering-based Motion Segmentation method that categories the static scene parts and the multiple moving objects using their 3D motion trajectories. Our motion segmentation method uses the raw trajectory data, allowing the objects to move in direct 3D space. We also propose a complete pipeline for static-map building which estimates the inter-frame motion parameters by exploiting the minimal 3-point Random Sample Consensus algorithm on the feature correspondences only from the static scene parts. The proposed method has been especially designed and tested for large scene in real outdoor environments. On one hand, our 3D Motion Segmentation approach outperforms its 2D based counterparts, for extensive experiments on KITTI dataset. On the other hand, separately reconstructed static-maps and moving objects for various dynamic scenes are very satisfactory. This paper is a french version of [9].

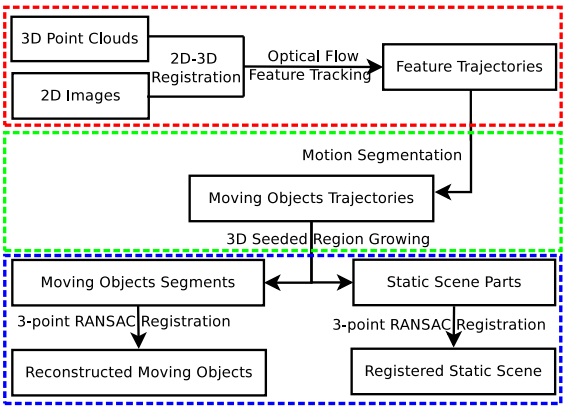


FIGURE 1 – Schéma de l'approche pour décomposer en carte statique et objets mobiles.

Keywords

Mapping ; Motion and Path Planning ; SLAM.

1 Introduction

Les approches de SLAM (*Simultaneous Localization and Mapping*) visuel sont désormais très efficaces lorsque la scène dans laquelle navigue le robot est totalement fixe. Dans le cas où des objets en mouvement sont présents dans l'environnement, le problème reste ouvert. La localisation est alors obtenue en estimant le mouvement de la caméra par des méthodes robustes négligeant les zones mobiles [5], [10]. Les parties dynamiques de la scène sont alors considérées comme des *outliers*. Cependant, lorsqu'un nombre important d'objets mobiles est présent dans la scène, il devient difficile de ne pas en tenir compte car cela dégrade fortement la précision de la localisation [3]. C'est pourquoi, nous proposons une méthode complète pour la construction de la partie fixe de la scène en tenant compte des parties dynamiques qui la composent. Notre méthode peut se décomposer en trois étapes (Fig. 1) : a) construction de la trajectoire 3D des points caractéristiques de la scène ; b) segmentation au sens du mouvement (SM) ; c) recalage 3D de la scène.

Nous nous plaçons dans le cas d'un robot mobile équipé d'une caméra RGB-D fournissant à la fois une information 2D photométrique et une information 3D géométrique de la scène. Nous proposons de calculer le flot optique dans

les images 2D pour en déduire les trajectoires des points 3D associés afin de combiner de manière pertinente les deux jeux de données. En effet, en pratique, le suivi des points 2D est plus stable et robuste que celui de leurs homologues 3D. Par contre, les trajectoires obtenues de cette manière sont, en pratique, numériquement instables à cause d'un fort déséquilibre entre le nombre de points fixes et le nombre de points dynamiques. Pour résoudre ce problème, nous employons une technique d'échantillonnage basée sur la vraisemblance du flot optique qui permet de contre-balancer le nombre de points 3D mobiles et fixes.

Ensuite, à l'aide des trajectoires 3D des points d'intérêt épars, nous proposons un partitionnement en sous-espaces des points 3D (3D-SSC) pour la segmentation du mouvement. Bien que de nombreuses méthodes de SM soient disponibles dans le cadre de la vidéosurveillance, du suivi d'objets et de la reconnaissance d'action [18], elles fournissent des solutions pour des objets mobiles en ne considérant exclusivement que les données 2D ou 3D. La méthode proposée ici estime la segmentation au sens du mouvement 3D à partir des trajectoires brutes guidées par les données 2D. L'algorithme 3D-SSC entend déterminer les sous-espaces minimaux qui représentent le mieux les trajectoires données. Dans ces travaux, nous montrons que notre approche 3D-SSC améliore significativement les méthodes basées 2D.

Pendant l'étape de reconstruction 3D des éléments composant la scène, les objets dynamiques doivent être segmentés puis supprimés de la scène de manière à ce que seuls les éléments fixes composent la carte. Grâce à la segmentation au sens du mouvement, l'ensemble des points d'intérêt épars peut être subdivisé en plusieurs sous-espaces (chaque sous-espace étant affecté à un objet suivant une trajectoire unique). Ensuite, une segmentation dense des points 3D statiques et dynamiques de la scène est obtenue à partir de ces sous-espaces, à l'aide d'une technique de croissance de région. Nous avons développé deux applications, *i.e.* la création d'une carte statique et la reconstruction des objets mobiles. La carte statique est construite en recalant des nuages de points grâce à un RANSAC sur 3 points. Cet algorithme utilise une représentation de Cayley pour la matrice de rotation, ce qui permet d'obtenir la transformation rigide entre deux nuages de points de manière linéaire [4]. La construction de cette carte donne des résultats très satisfaisants sur des scènes d'extérieur complexes. Les objets mobiles sont reconstruits de manière dense par le recalage successif de leurs observations sous différents points de vue.

La contribution majeure de cet article est double :

— un nouveau cadre pour la segmentation au sens du mouvement 3D est proposé ; cette méthode groupe les trajectoires des primitives 3D à partir d'un algorithme de de partitionnement en sous-espaces épars ;

— la chaîne complète de création d'une carte statique, tirant avantage de la méthode SM proposée, est présentée. Notre système fournit conjointement la reconstruction 3D des parties statiques d'une scène et une meilleure reconstruction des objets mobiles qui la composent.

2 Etat de l'art

Pendant des décennies, de très nombreux travaux ont été menés dans le domaine de la segmentation du mouvement à partir d'images (SM) [18]. Dans cet état de l'art, nous nous focaliserons essentiellement sur les méthodes de segmentation du mouvement 3D. Parmi les approches les plus représentatives, l'Analyse en Composantes Principales Généralisée (GPCA) [24], les méthodes basées RANSAC [27], le partitionnement agglomératif en sous-espaces (ASC) [20] ou épars (SCC) [6] ont été intensivement étudiés [18]. Habituellement, le problème de la SM est traité en séparant les mouvements en sous-espaces tels que chaque trajectoire appartienne à son sous-espace correspondant. A cet égard, la GPCA estime les sous-espaces globaux linéaires pour l'agrégation du mouvement, là où le LSA (*Local Subspace Affinity*) opère localement [28]. Bien que ces méthodes fournissent de précieuses informations pour l'agrégation des mouvements, leur usage pratique est limité à cause de leur grande sensibilité au bruit et aux points aberrants ainsi qu'à l'augmentation drastique de leur complexité calculatoire face à l'augmentation du nombre d'objets en mouvement. ASC est une méthode plus robuste qui combine les techniques de compression de données, de minimisation de rang et de représentation éparse. Inspirés par l'ASC, Elhamifar et Vidal [6] ont proposé l'algorithme SSC qui s'appuie sur l'idée d'une représentation éparse autonome. SSC est considérée comme étant la méthode de référence de la littérature [18].

Au-delà des méthodes de SM basées sur des primitives 2D, des méthodes hybrides 2D/3D ou des méthodes 3D ont aussi été développées. Les travaux récents de Stuckler et al. [23] effectuent une segmentation dense du mouvement 3D sur des données RGB-D à l'aide d'un algorithme d'espérance-maximisation. De manière similaire, Papon et al. [15] et Koo et al. [11] font un suivi et une segmentation des objects dans une séquence RGB-D. Toutefois, ces méthodes font l'hypothèse d'une caméra statique et ont été principalement appliquées dans des environnements controlés. Perera et al. [19] utilisent une fonction de distance signée tronquée pour segmenter les objets mobiles en utilisant une représentation volumétrique. Cet algorithme est supporté par une méthode de SM gloutonne basée RANSAC [27] et souffre donc des problèmes mentionnés plus haut. D'une manière différente, Sofer et al. [22] effectuent une segmentation du mouvement 3D à partir d'un apprentissage statistique actif (*active machine learning* ou AML) [21]. Malgré le fait que l'AML fournisse une classification très précise, l'apprentissage qu'il impose le rend fastidieux.

La construction de cartes statiques est un sujet important en vision par ordinateur et en robotique. Wang et al. [26]

ont proposé une méthode qui combine SLAM et suivi d'objets mobiles (SLAM-MOT), en s'appuyant sur un a priori cartographique ou une hypothèse de cohérence du mouvement. Cependant, la méthode échoue à appréhender les mouvements lents et les objets temporairement stationnaires. Pomerleau et al. [17] ont proposé de détecter les objets mobiles en utilisant la technique du lancer de rayons. Les changements spatiaux sont mesurés dans la carte obtenue après estimation du mouvement à partir de l'odométrie et d'un raffinement par ICP. Cette méthode fait également l'hypothèse que les zones dynamiques sont minoritaires dans la scène. De même, Ambrus et al. [1] ont proposé de maintenir et mettre à jour des modèles spatiaux à partir d'un recalage basé sur la distance entre deux 3D-NDT (*Normal Distribution Transforms*), qui peut être vu comme une sous-classe d'un mélange gaussien nécessitant un modèle de référence fidèle qui rend la méthode inadaptée aux environnements dynamiques inconnus a priori.

3 Segmentation 3D au sens du mouvement

La segmentation au sens du mouvement a pour objectif de regrouper les zones de la scène en fonction de leur déplacement. Nous supposons que le robot mobile capture une séquence de nuages de points 3D dans un environnement où plusieurs objets sont en mouvement. Ainsi, les points caractéristiques suivis durant la séquence sont répartis sur n objets mobiles se déplaçant dans la scène à des vitesses différentes que nous devons partager en n sous-espaces distincts regroupant toutes les trajectoires linéairement dépendantes.

3.1 Mouvement 3D

Considérons X et Y deux vecteurs, en coordonnées cartésiennes, correspondants au même point 3D à deux instants distincts. Ces deux vecteurs sont liés par un mouvement rigide décrit par une rotation R et une translation t :

$$\mathsf{Y} = \underbrace{\begin{bmatrix} \mathsf{R} & \mathsf{t} \end{bmatrix}}_{\mathsf{T}\in\mathbb{R}^{3\times 4}} \begin{bmatrix} \mathsf{X} \\ 1 \end{bmatrix}, \tag{1}$$

où T représente la matrice de transformation 3D. Soit $\{\mathsf{X}\}_{i=1}^{P}$ un ensemble de P points d'un même objet rigide défini dans un repère quelconque. Si au cours de la séquence $\{f_j\}_{j=1}^{F}$, les points sont liés par les transformations $\{\mathsf{T}_j\}_{j=1}^{F}$, tous les points caractéristiques Y_{ji} (*i.e.*, j^{ème} point de la i^{ème} image) sont déterminés par :

$$\underbrace{\begin{bmatrix} \mathsf{Y}_{11} & \mathsf{Y}_{12} & \cdots & \mathsf{Y}_{1P} \\ \mathsf{Y}_{21} & \mathsf{Y}_{22} & \cdots & \mathsf{Y}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \mathsf{Y}_{F1} & \mathsf{Y}_{F2} & \cdots & \mathsf{Y}_{FP} \end{bmatrix}}_{Y\in\mathbb{R}^{3F\times P}} = \underbrace{\begin{bmatrix} \mathsf{T}_1 \\ \mathsf{T}_2 \\ \vdots \\ \mathsf{T}_F \end{bmatrix}}_{T\in\mathbb{R}^{3F\times 4}} \underbrace{\begin{bmatrix} \mathsf{X}_1 & \mathsf{X}_2 & \cdots & \mathsf{X}_P \\ 1 & 1 & \cdots & 1 \end{bmatrix}}_{X\in\mathbb{R}^{4\times P}}, \tag{2}$$

où T et X représentent respectivement le mouvement et la structure de l'objet. Les colonnes de la matrice Y correspondent à la trajectoire des points d'intérêts étudiés. Remarquons qu'étant donné que toutes les entrées de la dernière ligne de X sont égales à un, la trajectoire des points (*i.e.* les colonnes de Y) décrit un espace de $\mathbb{R}^{3F}$ de dimension au plus trois et que le rang de Y est de quatre au maximun. Dans le cas où la scène est composée de plusieurs objets en mouvement, notons $\{S_\ell\}_{\ell=1}^{n}$ la collection de n sous-espaces de $\mathbb{R}^{3F}$ de dimension $\{d_\ell\}_{\ell=1}^{n}$. Si $\{Y_\ell\}_{\ell=1}^{n}$ correspond à n différents mouvements inconnus, la matrice $\boldsymbol{Y}$, contenant les m trajectoires peut être exprimée par :

$$\boldsymbol{Y} = \begin{bmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_m \end{bmatrix} = \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \mathsf{C}, \tag{3}$$

où $Y_\ell \in \mathbb{R}^{3F\times m_\ell}$ est une matrice de rang d_ℓ composée de m_ℓ trajectoires de S_ℓ, et $\mathsf{C} \in \mathbb{R}^{m\times m}$ est une matrice de permutation inconnue. L'équation (3) décrit ainsi les trajectoires $\{\boldsymbol{y}_k\}_{k=1}^{m}$ comme l'union de n sous-espaces.

3.2 Partitionnement épars en sous-espaces

D'après l'équation (3), le problème se réduit à la décomposition de $\boldsymbol{Y}$ en $\{Y_\ell\}_{\ell=1}^{n}$ et C. Ce problème est résolu dans [6] à l'aide d'une technique de relaxation solution de :

$$\min \| \mathsf{C} \|_1, \text{ t.q. } \boldsymbol{Y} = \boldsymbol{Y}\mathsf{C}, \ \mathrm{diag}(\mathsf{C}) = 0. \tag{4}$$

Malheureusement, nous ne pouvons appliquer directement la méthode [6] puisque notre formulation implique des modifications importantes dans le cas 3D. En effet, dans notre cas les entrées de C sont forcées pour être non-négatives afin de distinguer les objets se déplaçant à la même vitesse mais dans une direction opposée. C'est pourquoi nous proposons de modéliser le problème comme solution de :

$$\min \| \mathsf{C} \|_1, \text{ t.q. } \boldsymbol{Y} = \begin{bmatrix} \boldsymbol{Y} & \mathsf{I}_d \end{bmatrix} \mathsf{C}, \mathrm{diag}(\mathsf{C}) = 0, c_{ij} \geq 0. \tag{5}$$

où I_d est la matrice identité de dimension $3F\times 3F$, I_d et c_{ij} les entrées de C. Lorsque la représentation éparse de la matrice C est calculée, nous construisons un graphe pondéré $\mathcal{G}$ des poids $\mathcal{W} = \mathsf{C} + \mathsf{C}^{\mathsf{T}}$. La segmentation en trajectoires est obtenue grâce à la méthode de partitionnement spectral [14] sur le graphe $\mathcal{G}$.

L'approche ainsi proposée offre plusieurs avantages :

— Nous analysons le mouvement directement en 3D et évitons ainsi les effets de perspective dus à la projection de la scène 3D sur un imageur 2D.

— L'interprétation des mouvements devient plus précise : la rotation et la translation 3D sont directement déduites et interprétables pour chaque objet en mouvement.

4 Construction de la carte statique

Reconstruire en 3D un environnement dans lequel se déplacent des objets reste encore un problème ouvert. En effet, à chaque instant, le fond fixe est occulté par des objets et ces zones d'occultation. Dans ce travail, nous cherchons à reconstruire de la façon la plus fidèle possible la partie immobile de la scène en bouchant les zones occultées dès que celles-ci sont découvertes, soit grâce aux déplacements

des objets, soit grâce aux changements d'angles de vue de la caméra mobile. Pour ce faire, nous retirons dans un premier temps les zones mobiles de l'environnement à l'aide de l'approche décrite en section 3. Puis nous recalons les points 3D considérés comme fixes à l'aide d'un RANSAC ne nécessitant que 3 points (Fig. 1).

4.1 Construction des trajectoires caractéristiques et segmentation

Les trajectoires caractéristiques sont construites en utilisant à la fois les données 2D et 3D. Le suivi des points étant plus robuste sur les images 2D, il est réalisé à partir d'une estimation du flot optique multi-résolution [12] directement dans les images RGB. Par la suite, les trajectoires 3D sont simplement déduites à partir de la correspondance directe entre les données 2D-3D (supposées calibrées et synchronisées comme présenté dans [16]). En pratique, les zones fixes étant prédominantes, les points 3D suivis appartenant aux objets mobiles sont sous-représentés. Cette sous-représentation induit des instabilités numériques lors de la décomposition de l'espace des points 3D en sous-espaces linéairement indépendants. C'est pourquoi, nous introduisons un échantillonnage basé sur le maximum de vraisemblance. Notons $\{\mathbf{v}_k\}_{k=1}^m$ les vitesses mesurées des m points 3D suivis $\{\boldsymbol{y}_k\}_{k=1}^m$ (Eq. (3)). Si $\{c_k\}_{k=1}^m$ représente la nature des points en fonction de leur trajectoire (mobile=1, et fixe = 0), la fonction de vraisemblance est notée :

$$\mathcal{L}(c_k = 1|Z) = e^{\|\mathbf{v}_k - \bar{\mathbf{v}}\|^2/\sigma^2}, \tag{6}$$

où $\bar{\mathbf{v}}$ et σ sont respectivement la médiane et l'écart type de la vitesse. Un sous-ensemble de points 3D suivis est alors sélectionné uniformément en fonction de leur nature $(c_k = 0, 1)$ (6) pour obtenir un équilibre entre le nombre de points fixes et mobiles. Durant ce processus, nous rejetons également toutes les trajectoires qui ne sont pas suffisamment "lisses" le long de la séquence. La Figure. 2 montre l'efficacité de cet échantillonnage grâce auquel les points caractéristiques en mouvement (van, train, cycliste et piétons) sont sous-échantillonnés.

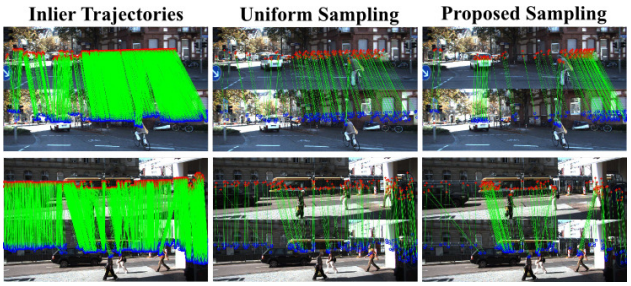


FIGURE 2 – Comparaison entre l'échantillonnage uniforme et l'échantillonnage à l'aide du maximum de vraisemblance : les lignes vertes montrent le suivi des trajectoires .

Une fois les trajectoires 3D éparses segmentées (Eq. (5)), nous appliquons un algorithme de croissance de région [25] sur l'ensemble du nuage de points 3D afin d'obtenir une segmentation dense de la scène.

4.2 Recalage

La carte statique est obtenue par le recalage des nuages de points considérés comme appartenant à la partie fixe de la scène. Ce recalage peut se faire à l'aide de seulement 3 points en correspondance. En effet, soit g la représentation de Gibbs de la matrice de rotation R. $\mathsf{G} = [\mathsf{g}]_\times$ et I_3 sont respectivement la matrice antisymétrique et la matrice identité de taille 3×3. Nous pouvons exprimer R par la transformation de Cayley [2],

$$\mathsf{R} = (\mathsf{I}_3 + \mathsf{G})^{-1}(\mathsf{I}_3 - \mathsf{G}). \tag{7}$$

A partir de l'équation (7), l'équation (1) peut se réécrire :

$$(\mathsf{I}_3 + \mathsf{G})\mathsf{Y} = (\mathsf{I}_3 - \mathsf{G})\mathsf{X} + (\mathsf{I}_3 + \mathsf{G})\mathsf{t}. \tag{8}$$

que l'on peut reformuler de la façon suivante :

$$(\mathsf{Y} - \mathsf{X}) = -(\mathsf{Y} + \mathsf{X})\mathsf{G} + \tilde{\mathsf{t}}. \tag{9}$$

Cette équation est linéaire par rapport à g et $\tilde{\mathsf{t}}$. Ainsi, chaque paire de points en correspondance nous donne 2 équations indépendantes. Le système ayant 6 inconnues, seulement 3 correspondances sont nécessaires pour le résoudre linéairement. Pour rendre la méthode moins sensible aux bruits de mesure, nous proposons d'inclure ce recalage dans un RANSAC tirant aléatoirement les 3 points parmi les points en correspondance des nuages. Après convergence, l'ensemble des *inliers* est considéré pour déduire le recalage optimal. Notons que cette technique peut également être utilisée pour recaler les objets mobiles, comme nous le présenterons par la suite.

5 Expérimentations

Nous avons conduit différentes expérimentations sur des données de synthèse et réelles. Concernant les données réelles, deux types d'acquisition ont été testés : une première avec une Kinect, une seconde tirée de la base de données KITTI [8].

5.1 Segmentation au sens du mouvement sur des données de synthèse

Nous avons construit une scène synthétique composée de n cubes de différentes tailles se déplaçant aléatoirement. Afin de vérifier la robustesse de notre approche, la position des points 3D est bruitée à l'aide d'un bruit blanc gaussien. Pour chaque expérimentation, nous calculons le taux d'erreur de classification par $\eta = \frac{\#\text{ caractéristiques mal classées}}{\#\text{ total de caractéristiques}}$. La Figure 3 présente le comportement de la méthode 3D-SSC et montre que l'approche est très robuste même avec un nombre important d'objets dans la scène et un bruit dans les données de 12% .

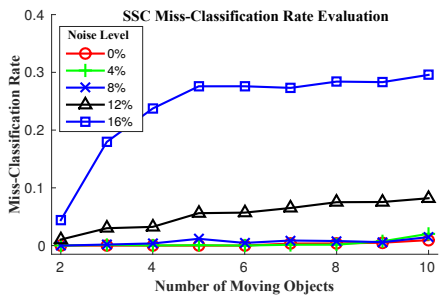


FIGURE 3 – Résultat de la méthode 3D-SSC MS sur 50 tests.

5.2 Expérimentation avec des données Kinect

Dans cette expérience, une scène 3D composée de 5 objets (un livre, une bouteille, une tasse, une lampe et une boîte) se déplaçant de façon indépendante est acquise par une Kinect. Le détail de cette expérience est présenté dans le tableau I, dans lequel les colonnes précisent la longueur de la séquence, le nombre de points caractéristiques des objets et la précision de la segmentation au sens du mouvement $(1-\eta)$. Ces résultats montrent que l'approche proposée est capable de segmenter correctement la scène dans un environnement controlé.

FIGURE 4 – Séquence Kinect.

Objects	Len.	Feat.	Acc.(%)
Tasse	15	18	100
Bouteille	32	12	100
Lampe	24	18	100
Boîte	22	18	100
Livre	20	16	93.75

TABLE 1 – Résultat de segmentation sur la séquence Kinect.

5.3 Evaluation sur la base de données KITTI

Afin d'évaluer notre méthode sur de longues séquences extérieures, nous avons utilisé la base de données KITTI [8] en considérant en quatre séquences différentes : *Highway*, *Junction*, *Station*, et *Market*. Ces séquences ont été sélectionnées pour leur longueur et leur nombre important d'objets mobiles. Les détails de ces 4 séquences sont présentés dans le Tableau 3. Dans ce tableau, la vitesse indique la vitesse relative des objets par rapport à la vitesse de déplacement de la caméra. Notons que la dynamique des objets couvre un large éventail de vitesses.

Evaluation de la méthode. Les trajectoires sont déduites de l'estimation dense du flot optique dans les images 2D et sont sous-échantillonnées par une méthode basée sur la vraisemblance des vitesses (eq. (6)). Comme nous pouvons l'observer dans la Figure 6a, un nombre important de points d'intérêt se situe sur les parties dynamiques de la scène alors que les objets en question ne correspondent qu'à une "petite" région. Cette redistribution des points nous permet de contrebalancer le déséquilibre entre le nombre des points fixes et mobiles dans la représentation éparse des données. La Figure 6b présente les trajectoires 3D des points considérés. La Figure 6c et la Figure 6d montrent respectivement la segmentation obtenue par 2D-SSC [6] et 3D-SSC. Notons que le 2D-SSC échoue à classer les panneaux routiers dans la partie fixe.

Seq.	# Objs.	Corr.	Incorr.	Dyn. Acc.(%)	Stc. Acc.(%)	Time (min.)
Highway	1	1	0	97.55	100	6.00
Junction	2	2	0	91.02	100	13.40
Station	5	5	1	91.60	92.47	3.16

TABLE 2 – Résultats quantitatifs.

Les résultats obtenus à partir du 2D-SSC et du 3D-SSC pour les quatre séquences sont présentés dans le tableau 3. La qualité de la segmentation est évaluée par la mesure de *Sensibilité* et de *Spécificité* [7]. Nous ajoutons également d'autres mesures telles que Seg. $\geq 50\%$, qui comptabilise le nombre d'objets ayant au moins la moitié de leurs points d'intérêts correctement classés, ainsi, que le rapport entre les valeurs propres $\rho = \frac{\sum_{i=1}^{n} \lambda_i}{\sum_{j=1}^{m} \lambda_j}$, où n and m correspondent respectivement au nombre de mouvements et au nombre total de trajectoires analysées. Une grande valeur de ρ révèle une meilleure représentation des sous-espaces de mouvement. Enfin, le tableau 3 reporte le temps de calcul des deux méthodes (codées sous MATLAB).

Nous pouvons observer que le 2D-SSC est très sensible aux mouvements de faibles intensités et que ses performances diminuent significativement lorsque le nombre d'objets augmente. A l'inverse, la méthode 3D-SSC reste robuste quelle que soit la vitesse de déplacement des objets mobiles. Enfin, le 3D-SSC obtient des résultats supérieurs à 2D-SSC grâce au fait que la représentation de sous-espaces 3D est plus compacte que celle du 2D-SSC comme nous pouvons l'observer dans le tableau 3 à travers la colonne décrivant le rapport des valeurs propres.

Reconstruction de la carte statique. Grâce à l'efficacité de la méthode proposée pour segmenter les objets 3D en mouvement, nous avons reconstruit les zones statiques de trois séquences : *Junction*, *Highway*, et *Train station*, (Fig. 7, 8, 9). Afin d'illustrer la qualité de la reconstruction, nous avons reconstruit la totalité de la scène et comparé avec la méthode [16] qui ne gère pas les parties mobiles présentes dans la scène. La Figure 7 présente la reconstruction de la scène (*Junction*) dans laquelle une voiture et un cycliste se déplacent. Nous remarquons que la voiture et le cycliste sont bien considérés comme des objets ayant un mouvement indépendant de la caméra.

Concernant la séquence *Highway*, une analyse qualitative est présentée en Figure 8a et Figure 8b où la zone statique reconstruite est comparée à [16]. Nous remarquons que la région encadrée en rouge (Fig. 8a) correspondant à l'ombre de l'arbre est difficilement reconnaissable avec

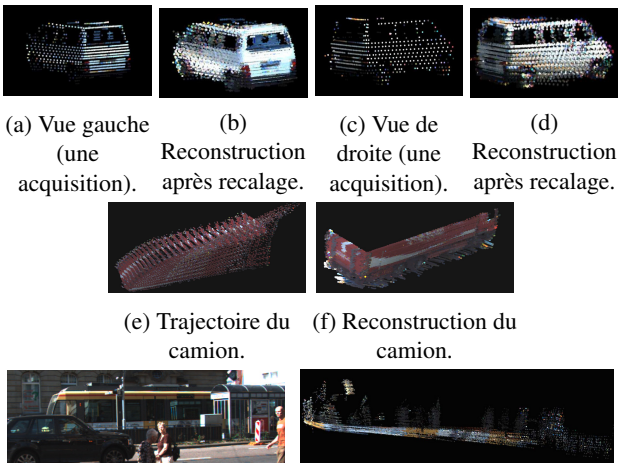
Seq.	# Images	# Objs.	# Feat.	Vit. (m/s)		Sensibilité		Spécificité		Seg. $\geq 50\%$		ρ		Temps(min.)	
				Min.	Max.	2D	3D	2D	3D	2D	3D	2D	3D	2D	3D
Highway	45	2	122	4.87	7.22	**1.0**	0.95	**1.0**	**1.0**	2	2	0.0250	0.0264	3.54	4.83
Junction	70	3	83	0.50	5.15	**1.0**	0.98	0.77	**0.99**	3	3	0.0398	0.0399	9.61	12.85
Station	9	6	77	0.35	7.12	0.62	**0.95**	0.31	**0.66**	3	**6**	0.0789	0.0979	1.39	1.68
Market	13	9	50	0.39	1.34	0.88	**1.0**	0.68	**0.98**	6	**9**	0.0666	0.1907	1.61	2.09

TABLE 3 – Comparaison entre 2D-SSC et 3D-SSC sur la base KITTI : *Seg.* $\geq 50\%$ comptabilise les objets ayant plus de 50% de points caractéristiques correctement segmentés.

la méthode [16]. Dans notre cas, l'ombre est mieux reconstruite (Fig. 8b) ce qui montre une meilleure précision dans le recalage. La séquence *Train Station* est encore plus difficile (Fig. 9). Cette séquence est composée d'une voiture se déplaçant très rapidement, de piétons avec une marche plus lente et d'un train occulté de façon intermittente par les autres objets mobiles. Les résultats présentés en Figure 9c montrent que tous les objets sont correctement détectés et reconstruits (piétons, véhicule, train). Rappelons que les objets se déplaçant à la même vitesse sont regroupés lors de la segmentation au sens du mouvement. C'est pourquoi, le véhicule et le train (objets bleus de la Figure. 9b), ainsi que les deux piétons (objets jaunes Figure. 9b) sont regroupés. Le tableau 2 résume les résultats quantitatifs où la mesure "Dyn. Acc." est calculée par $Dyn.\ Acc. = \frac{\#pointsconsiderescommedynamiques}{Total\ \#pointsdynamiques}$, et la mesure "Stc. Acc." est définie de manière similaire pour les points statiques. Une valeur importante de "Dyn. Acc." signifie un meilleur rejet des parties dynamiques de la scène alors qu'une grande valeur de "Stc. Acc." révèle une meilleure prise en compte des points fixes. Ces résultats montrent que les objets dynamiques sont correctement extraits avec une bonne précision sans pour autant altérer les zones fixes. La Figure 5 présente la qualité du recalage des différentes acquisitions au cours du temps sur les objets dynamiques de la scène. La Figure 5a-5d montre deux vues 3D de la voiture respectivement obtenues par une seule acquisition et par la séquence recalée. La deuxième ligne de cette figure présente la reconstruction d'un camion à l'aide de [16] et par notre méthode (Fig. 5f). Enfin, une reconstruction complète du train est présentée Figure 5h. Cette reconstruction est obtenue à partir du recalage des représentations partiellement occultées à chaque acquisition.

6 Conclusion et perspectives

Nous avons proposé une nouvelle méthode pour segmenter en 3D une scène dynamique en fonction du mouvement des objets qui la peuplent. Cette technique a été testée sur de nombreuses acquisitions et améliore les méthodes n'utilisant que les données 2D. L'échantillonnage des points 3D suivis en fonction de leur vitesse permet de gérer un large éventail de mouvements et un nombre important d'objets mobiles. En outre, la reconstruction 3D ainsi obtenue de la scène aboutit à une 3D photoréaliste, à la fois pour les parties statiques et dynamiques grâce au recalage des différentes acquisitions au cours du temps. Cette approche reste néanmoins sensible aux occultations. C'est pourquoi, dans la suite de ce travail, nous allons tirer parti de l'analyse de chaque trajectoire pour gérer les occultations entre chaque acquisition.



(a) Vue gauche (une acquisition). (b) Reconstruction après recalage. (c) Vue de droite (une acquisition). (d) Reconstruction après recalage.

(e) Trajectoire du camion. (f) Reconstruction du camion.

(g) Train occulté. (h) Train reconstruit

FIGURE 5 – Reconstruction des objets mobiles.

Références

[1] Ambrus, R., Bore, N., Folkesson, J., & Jensfelt, P. Metarooms : Building and maintaining long term spatial models in a dynamic world. In IROS, 2014.

[2] Alismail, H., Baker, L. D., & Browning, B. Continuous trajectory estimation for 3D SLAM from actuated lidar. In ICRA, 2014.

[3] Burgard, W., Stachniss, C., & Hähnel, D. Mobile robot map learning from range data in dynamic environments. In Auto. Nav. in Dyn. Env., 2007.

[4] Cayley, A. About the algebraic structure of the orthogonal group and the other classical groups in a field of characteristic zero or a prime characteristic. In Reine Angewandte Mathematik, 1846.

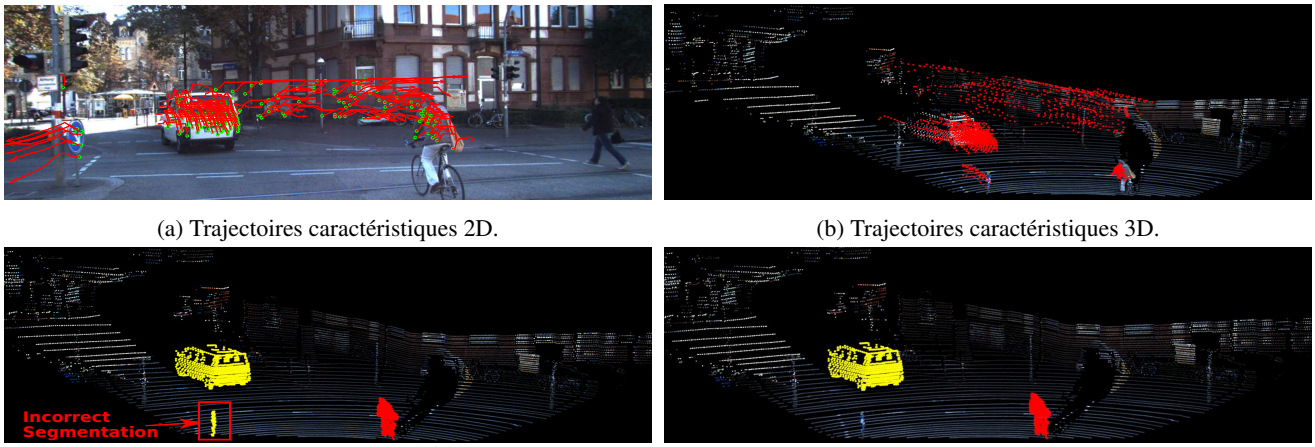
[5] Civera, J., Grasa, O., Davison, A. J., & Montiel, J. 1-Point RANSAC for extended Kalman filtering : Application to real-time structure from motion and visual odometry. In Journal of Field Robotics, 2010.

[6] Elhamifar, E., & Vidal, R. Sparse subspace clustering : Algorithm, theory, and applications. In TPAMI, 2013.

[7] Fawcett, T. An introduction to ROC analysis. In Pattern recognition letters, 2006.

[8] Geiger, A., Lenz, P., & Urtasun, R. Are we ready for autonomous driving ? the kitti vision benchmark suite. In CVPR, 2012.

[9] Jiang, C., Paudel, D. P., Fougerolle, Y., Fofi, D., & Demonceaux, C. Static-map and Dynamic Object Reconstruction in



(a) Trajectoires caractéristiques 2D.

(b) Trajectoires caractéristiques 3D.

(c) Résultat 3D basé sur une partition 2D [6].

(d) Méthode proposée.

FIGURE 6 – Comparaison entre une segmentation en 2D et en 3D

(a) Reconstruction complète de la scène à partir de [16].

(b) Zone fixe de la scène, notre approche.

FIGURE 7 – Séquence *Junction* reconstruite sur 80 acquisitions.

Outdoor Environments using 3D Motion Segmentation. In IEEE Robotics and Automation Letters, Vol. 1, Issue 1, January 2016, pp 324-331.

[10] Klein, G., & Murray, D. Improving the agility of keyframe-based SLAM. In ECCV, 2008.

[11] Koo, S., Lee, D., & Kwon, D. S. Incremental object learning and robust tracking of multiple objects from RGB-D point set data. In J. Vis. Commun., 2013.

[12] Liu, C., Yuen, J., Torralba, A., Sivic, J., & Freeman, W. T. Sift flow : Dense correspondence across different scenes. In ECCV, 2008.

[13] Michael G. & Stephen B., CVX : Matlab software for disciplined convex programming, version 2.0 beta, 2013.

[14] Ng, A. Y., Jordan, M. I., & Weiss, Y. On spectral clustering : Analysis and an algorithm. In Adv. neural inf. proc. systems, 2002.

[15] Papon, J., Kulvicius, T., Aksoy, E. E., & Worgotter, F. Point cloud video object segmentation using a persistent supervoxel world-model. In IROS, 2013.

[16] Paudel, D. P., Demonceaux, C., Habed, A., Vasseur, P., & Kweon, I. S. 2D-3D camera fusion for visual odometry in outdoor environments. In IROS, 2014.

[17] Pomerleau, F., Krusi, P., Colas, F., Furgale, P., & Siegwart, R. Long-term 3D map maintenance in dynamic environments. In ICRA, 2014.

[18] Tron, R., & Vidal, R.. A benchmark for the comparison of 3-d motion segmentation algorithms. In CVPR, 2007.

[19] Perera, S., Barnes, N., Xuming He, Izadi, S., Kohli, P., & Glocker, B. Motion Segmentation of Truncated Signed Distance Function Based Volumetric Surfaces. In WACV, 2015.

[20] Rao, S., Tron, R., Vidal, R., & Ma, Y. Motion segmentation in the presence of outlying, incomplete, or corrupted trajectories. In TPAMI, 2010.

[21] Settles, B. Active learning literature survey. Tech. Report, 2010.

[22] Sofer, Y., Hassner, T., & Sharf, A. Interactive Learning for Point Cloud Motion Segmentation. In Computer Graphics Forum, 2013.

[23] Stückler, J., & Behnke, S. Efficient Dense Rigid-Body Motion Segmentation and Estimation in RGB-D Video. In IJCV, 2015.

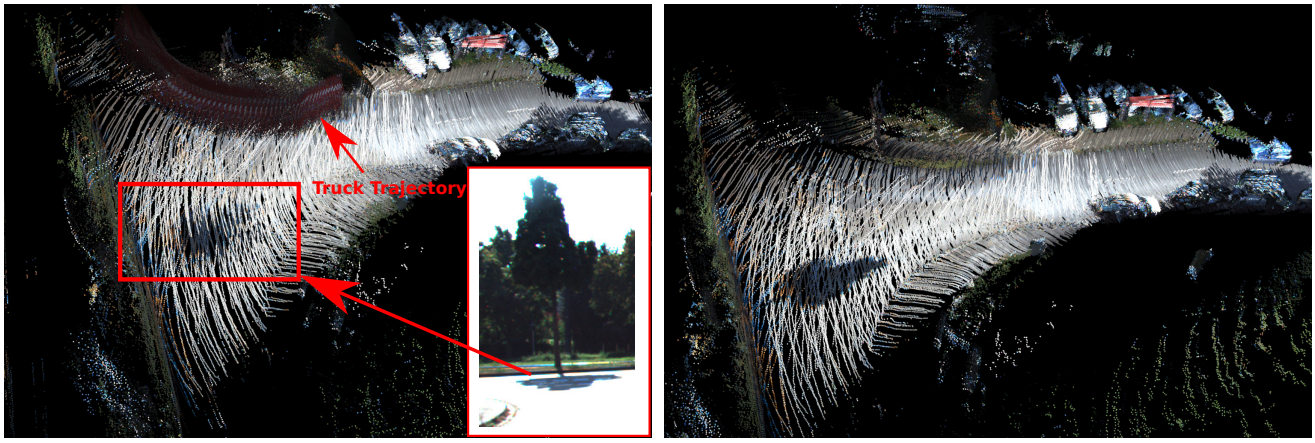
[24] Vidal, R., & Hartley, R. Motion segmentation with missing data using powerfactorization and gpca. In CVPR 2004.

[25] Vosselman, G., Gorte, B. G., Sithole, G., & Rabbani, T. Recognising structure in laser scanner point clouds. In Int. archives of photogrammetry, remote sensing and spatial information sciences, 2004.

[26] Wang, C., Thorpe, C., & Thrun, S. Online simultaneous localization and mapping with detection and tracking of moving objects : Theory and results from a ground vehicle in crowded urban areas. In ICRA, 2003.

[27] Yan, J., & Pollefeys, M. Articulated motion segmentation using ransac with priors. In Dynamical Vision, 2007.

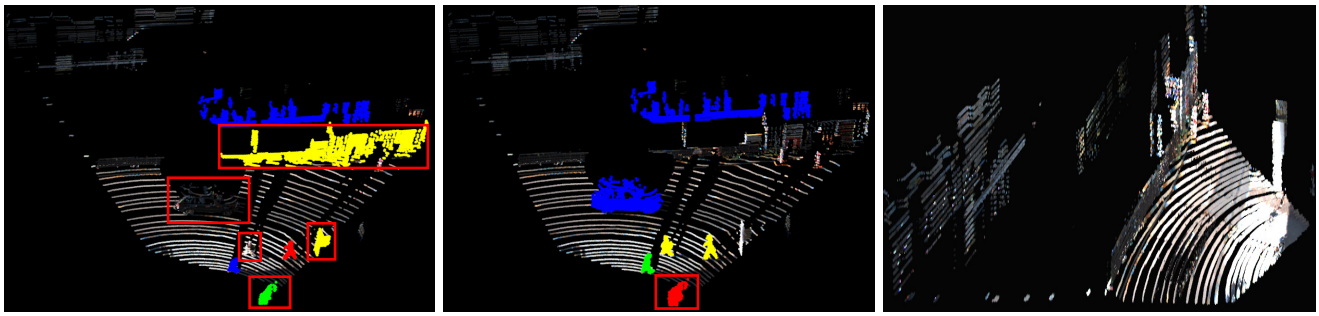
[28] Yan, J., & Pollefeys, M.. A general framework for motion segmentation : Independent, articulated, rigid, non-rigid, de-



(a) Reconstruction complète de la scène à partir de [16].

(b) Zone fixe de la scène, notre approche.

FIGURE 8 – Sequence *Highway* à partir de 45 acquisitions.



(a) Segmentation sur données 2D.

(b) Segmentation sur données 2D-3D.

(c) Zone fixe de la scène.

FIGURE 9 – Séquence *Train station* sur 9 acquisitions.

generate and non-degenerate. In Computer Vision- ECCV 2006 (pp. 94-106). Springer Berlin Heidelberg.