



**HAL**  
open science

## Odométrie 3D vision/lidar pour les véhicules intelligents

Vincent Fremont

► **To cite this version:**

Vincent Fremont. Odométrie 3D vision/lidar pour les véhicules intelligents. Journées Nationales de la Recherche en Robotique, JNRR'09, Nov 2009, Neuvy-sur-Barangeon, France. hal-00447598

**HAL Id: hal-00447598**

**<https://hal.science/hal-00447598>**

Submitted on 15 Jan 2010

**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.

# Odométrie 3D vision/lidar pour les véhicules intelligents

Vincent Frémont

UMR CNRS 6599 Heudiasyc - Université de Technologie de Compiègne

Centre de Recherches de Royallieu

60205 Compiègne

vincent.fremont@utc.fr

*Résumé - La perception dans les véhicules intelligents demeure une fonction fondamentale pour la navigation autonome et les systèmes d'aide à la conduite. Les capteurs optiques (caméras et Lidars) fournissent des informations précieuses sur l'environnement (amers visuels, cartes de profondeurs, informations contextuelles) permettant ainsi une localisation précise en position et en orientation ainsi qu'une segmentation des objets mobiles. Néanmoins, dans le contexte qui nous intéresse, les véhicules intelligents se déplacent dans des environnements en constante évolution. En effet, si l'on se place dans un environnement routier, de nombreux objets sont en mouvement (automobiles, piétons, cyclistes, motos, etc), et la vitesse du véhicule, sa dynamique et les changements de luminosité ou les conditions météo, peuvent altérer la perception et en particulier la localisation. L'enjeu de cette problématique est d'obtenir le meilleur compromis précision/intégrité sur l'odométrie 3D en utilisant au mieux la redondance fournie par les capteurs optiques que sont les caméras et les Lidars. C'est pourquoi, à travers ce papier, nous essayons de faire le point sur les recherches existantes en odométrie optique multi-capteur et de synthétiser les verrous et difficultés existants sur cette problématique, notamment pour la navigation en milieu routier.*

*Mots-clé - Odométrie 3D, Vision stéréo, Lidar, Environnements dynamiques, véhicules intelligents*

## I. INTRODUCTION

La perception pour les véhicules intelligents est une étape indispensable dans les systèmes d'aide à la conduite telles que les situations de pré-crash, la détection d'obstacles (piétons, véhicules à l'arrêt, objets sur route, etc) ou l'estimation du temps à collision. Deux types de capteurs, largement répandus dans les applications robotiques et respectant les contraintes liées aux systèmes d'aides à la conduite, permettent de mesurer efficacement la structure tridimensionnelle d'une scène, à savoir les caméras et les télémètres laser (ou LIDAR - LIght Detection And Ranging). Les caméras, qui sont des capteurs passifs, sont souvent préférées pour la reconnaissance et l'analyse de scènes. Les LIDAR, quant à eux sont des capteurs actifs, souvent utilisés pour la détection d'obstacles [1]. Les capteurs de type Radar sont aussi largement utilisés, mais ils ne correspondent pas à la finalité qui nous intéresse ici. En effet, les capteurs optiques installés sur les véhicules présentent une certaine forme de polyvalence et peuvent ainsi servir à faire de la perception proprioceptive à partir de l'observation de l'environnement immédiat du véhicule.

De nombreux projets autour des véhicules autonomes (CyCab [2], Darpa Urban Challenge [3]) ont permis d'évaluer un certain nombre de méthodes et de faire ressortir des problèmes afin d'atteindre un état fonctionnel d'autonomie avancé. Il apparaît clairement que pour naviguer en toute sécurité dans un environnement inconnu, il convient, tout d'abord, de parfaitement se localiser. En l'absence de données GPS (par exemple dû au masquage des satellites en milieu urbain), seules les mesures proprioceptives sont

disponibles. La localisation est alors réalisée en utilisant une combinaison d'odométrie roue (disponible par exemple sur le bus CAN pour une voiture) et de capteurs inertiels (gyromètres et accéléromètres) [4]. Cette approche a deux limitations principales : les centrales inertielles sont sujettes à la dérive et l'odométrie roue reste sensible aux glissements et aux dérapages. De plus, elle ne donne qu'une information 2D dans le plan de la route.

Récemment, de nombreux travaux ont été réalisés sur ce que l'on appelle l'odométrie optique. Elle a pour objectif d'estimer l'ego-mouvement du véhicule à partir de données télémétriques 3D (LIDAR) et/ou d'images obtenues à partir d'une ou plusieurs caméras installées sur le véhicule. Ces capteurs ont l'avantage de demeurer insensibles au terrain sur lequel ils évoluent et qu'ils peuvent avoir une dérive plus faible que les centrales inertielles. Néanmoins, ils possèdent certaines limitations qui peuvent être réduites en les utilisant conjointement.

Le challenge supplémentaire dans la localisation est que les véhicules évoluent dans des environnements où d'autres d'autres objets sont en mouvement. Le problème croît alors en complexité car il faut d'abord être capable d'isoler les objets en mouvement avant de pouvoir estimer la trajectoire propre du véhicule. Il faut aussi garder à l'esprit que la dynamique propre du véhicule peut être importante (accélérations, décélérations, trajectoires circulaires, etc).

Dans cet article, nous présenterons dans un premier temps le modèle d'odométrie 3D. Nous parcourons ensuite les méthodes d'odométrie basées LIDAR. Nous continuerons ensuite sur les approches basées vision. Pour finir, nous aborderons l'approche multimodale en essayant de présenter les intérêts et les défis actuels.

## II. L'ODOMÉTRIE 3D

Le principe de l'odométrie 3D est de connaître avec précision les 6 degrés de liberté (position et orientation) d'un robot dans un repère global à chaque instant  $t$  par reconstruction incrémentale. L'information est obtenue par composition des transformations élémentaires (ego-mouvement) estimées jusqu'à l'instant  $t$ . Même dans le cadre d'une application routière, la connaissance de la pose 3D (position et orientation) est essentielle : configuration de routes non planes, échangeurs multi-niveaux, couplage avec le signal GPS exprimé en 3D, etc.

Les 6 degrés de liberté à estimer sont donc la position notée  $\mathbf{p}_t = (x_t, y_t, z_t)^T$  et l'orientation  $\mathbf{q}_t = (a_t, b_t, c_t, d_t)^T$  paramétrisée en utilisant les quaternions. Ainsi, la pose du véhicule est définie par le vecteur d'état  $\mathbf{s}_t = (x_t, y_t, z_t, a_t, b_t, c_t, d_t)^T$ . A chaque acquisition, on estime le vecteur d'odométrie  $\mathbf{u}_t = (\delta\mathbf{p}_t^T, \delta\mathbf{q}_t^T)^T$ .

La pose courante est alors définie comme :

$$\begin{aligned}\mathbf{p}_t &= \mathbf{p}_{t-1} + \mathbf{R}(\delta\mathbf{q}_t)\delta\mathbf{p}_t \\ \mathbf{q}_t &= \delta\mathbf{q}_t \star \mathbf{q}_{t-1}\end{aligned}\tag{1}$$

où  $\mathbf{R}(\delta\mathbf{q}_t)$  représente le passage quaternions vers matrice de rotation et  $\star$  le produit de deux quaternions [5]. La confiance que l'on a dans l'estimation de l'état peut être modélisée par la matrice de covariance  $\mathbf{P}_{s_t}$ . Cette matrice de covariance peut être utilisée pour la propagation des incertitudes. De plus, si un signal GPS est disponible, elle sera utile pour la mesure d'intégrité de l'estimation de la trajectoire. En effet, dans la problématique qui nous intéresse, l'objectif est d'obtenir une mesure d'odométrie 3D (sans cartographie) la plus robuste/intègre possible sous les contraintes de dynamique de scène, de temps-réel et en utilisant au maximum le couple LIDAR/vision.

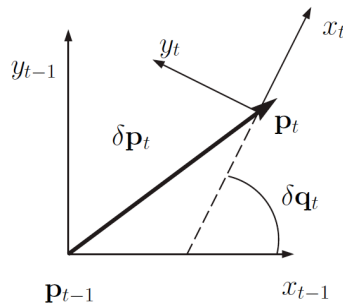


Fig. 1. Modèle odométrique

### III. ODOMÉTRIE BASÉE LIDAR

La technologie LIDAR est basée sur la mesure du temps de vol d'une onde lumineuse cohérente, proportionnelle au double de la distance d'impact. Ce type de capteur motorisé est actif et consomme de l'énergie pour fournir une mesure. Il existe différents Lidars : les Lidars 2D (à une nappe ex. Sick LMS211), les Lidars multi-nappes (ex. Ibeo Alasca XT) et les Lidars 3D (ex. Velodyne HDL-64E).

Le nombre de nappes, la résolution spatiale et temporelle et les champs de vue sont autant de paramètres qui influencent l'application et les résultats obtenus (voir Tab. I). Même si les télémètres 3D font leur apparition (Voir Fig. 2), les résolutions spatiales restent insuffisantes pour obtenir clairement la forme d'une cible, notamment en cas de fortes occultations.

	1 nappe (ex. Sick LMS211)	4 nappes (ex. Ibeo Alasca XT)	3D (ex. Velodyne HDL-64E)
<b>Fréquence de balayage</b>	75 Hz	25 Hz	5-15 Hz
<b>Champ de vue horizontal</b>	100°	240°	360°
<b>Résolution angulaire horizontale</b>	0.25° - 1.0°	0.25° - 1.0°	0.09°
<b>Champ de vue vertical</b>	-	3.2°	26.8°
<b>Résolution angulaire verticale</b>	-	0.8°	0.4°
<b>Résolution de la portée</b>	3.5cm	4cm	2cm
<b>Portée de fonctionnement</b>	0m - 80m	0.3m - 200m	50m - 120m

TABLE I  
SPÉCIFICATIONS DES LIDARS

Par rapport aux systèmes de vision, il est important de souligner que les mesures de profondeur sont directement disponibles, et ce, sans passer par les étapes d'extraction de points caractéristiques et de reconstruction 3D à partir des mesures de l'images. Il est aussi intéressant de noter que les capteurs Lidar permettent de reconnaître des amers de type arbres, poteaux, trottoir, surfaces planes, etc, par caractérisation de l'empreinte du signal produit [6].

De plus, les Lidars ne permettent pas d'utiliser une information photométrique pour la mise en correspondance temporelle des points 3D et le phénomène de la déformation du signal due au balayage est souvent occulté. Ainsi, à partir des ensembles de points obtenus à deux instants d'acquisition successifs, la plupart des approches utilise un algorithme d'association de données ou de recalage de type ICP (Iterative Closest Point) [7] et ses variantes [8]. Ce type d'algorithme prend comme entrée deux ensembles de primitives (points, segments, surfaces) 2D ou 3D, et détermine la transformation optimale (rotation et translation) qui peut exister entre eux. Le principe de l'algorithme ICP [7] est rappelé ci-après (Algo. 1) :

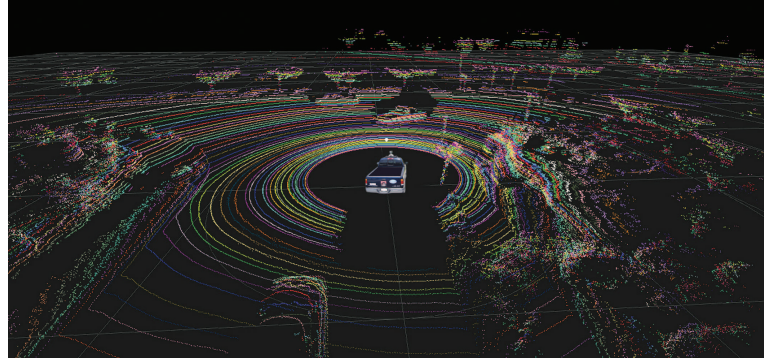
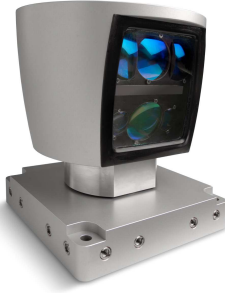


Fig. 2. Lidar 3D Velodyne HDL-64E et image des impacts (documentation Velodyne HDL-64E)

### Objectif

Estimer la transformation optimale  $\mathbf{T}$  (rotation  $\mathbf{q}$  et translation  $\mathbf{t}$ ) entre deux ensembles de points  $\mathbf{P}$  et  $\mathbf{X}$  avec respectivement  $N_P$  et  $N_X$  individus.

Entrée :  $\mathbf{P}_{(1 \times N_P)}$ ,  $\mathbf{X}_{(1 \times N_X)}$

### Algorithme

- 1) L'itération est initialisée avec  $\mathbf{P}_0 = \mathbf{P}$ ,  $\mathbf{T}_0 = [\mathbf{q}_0^T, \mathbf{t}_0^T]^T$ ,  $\mathbf{q}_0 = [1, 0, 0, 0]^T$ ,  $\mathbf{t}_0 = [0, 0, 0]^T$  et  $k = 0$ .
- 2) Calculer les plus proches points :  $\mathbf{Y}_k = \mathcal{C}(\mathbf{P}_k, \mathbf{X})$ .
- 3) Calculer la transformation  $(\mathbf{T}_k, d_k) = \mathcal{Q}(\mathbf{P}_0, \mathbf{Y}_k)$ .
- 4) Appliquer la transformation :  $\mathbf{P}_{k+1} = \mathbf{T}_k(\mathbf{P}_0)$ .
- 5) Terminer l'itération quand le changement d'erreur quadratique moyenne tombe sous un seuil prédéfini  $\tau$  :  $(d_k - d_{k+1})^2 < \tau$ .

Sortie :  $\mathbf{T}(\mathbf{q}, \mathbf{t})$

### Algorithme 1 : Iterative Closest Point, ICP

Dans les travaux de Miyasaka et al. [9], la première étape d'estimation de mouvement est réalisée par l'algorithme ICP entre les deux scans successifs d'un Lidar 4 nappes Ibeo Alasca XT. Ces paramètres de mouvement sont ensuite utilisés comme initialisation d'une étape d'association de carte locale avec le dernier balayage, comme dans les travaux de Vu et al. [10]. Ces cartes locales permettent de modéliser la probabilité d'occupation d'une région autour du véhicule, de prendre en compte la dynamique de la scène et également de pouvoir affiner l'estimation du mouvement comme dans les travaux de Laugier et al. [11], de Burllet et al. [12] et de Wang [13].

## IV. ODOMÉTRIE BASÉE VISION

Les caméras sont devenues des capteurs quasi-inévitables pour les problèmes de perception dans les véhicules intelligents. En effet, ce type de capteur peut fournir une information 3D, à bas coût et se

présente sous forme compacte avec une faible consommation d'énergie. Les technologies (CCD, CMOS), les résolutions (Haute définition, etc), les cadences (jusqu'à 100 images/seconde) ne cessent d'évoluer. De plus, les caméras permettent d'extraire de façon précise des informations géométriques et photométriques pour aboutir à des méthodes de plus haut niveau d'analyse et d'interprétation de scènes.

Preuve en est, la richesse de la littérature sur les méthodes de SLAM visuel (Localisation et cartographies simultanées par vision), bien connue dans la communauté robotique. Ces méthodes utilisent des approches de filtrage probabiliste pour mettre à jour de manière séquentielle les estimées de la position des amers (la carte) et la localisation courante de la caméra. Parmi les nombreux travaux, nous pouvons citer ceux de Davison et al. [14], Solà et al. [15], Lemaire et al. [16], Angeli et al. [17] pour les approches mono-vision, ceux de Solà et al. [15], Lemaire et al. [18] pour les méthodes basées stéréovision et enfin ceux de Scaramuzza et al. [19] pour le SLAM basé sur la vision omni-directionnelle.

Une autre alternative apparaît dans les solutions de type "Structure from Motion". Le mouvement et la structure de la scène sont estimés à partir d'une séquence d'images en appliquant d'abord une mise en correspondance robuste de points d'intérêt du type Harris [20], SIFT [21] ou CenSurE [22] afin d'estimer le mouvement relatif [23][24]. L'amélioration du suivi et de l'estimation de mouvement est réalisée par des méthodes de détections de points aberrants du type RANSAC [25]. Cette approche est souvent préférée aux autres approches robustes lorsque la contrainte de temps réel est importante. Ainsi, ces approches permettent ainsi de prendre en compte la dynamique de scène et les mauvaises associations de données temporelles [26].

Le principe de l'algorithme RANSAC [25] est le suivant (Algo. 2) :

Objectif : estimer de manière robuste, un modèle paramétrique  $\mathcal{M}$  sur un ensemble de données  $S$  qui contient des données aberrantes. Entrée :  $S$

#### Algorithme

- 1) Sélectionner de manière aléatoire un sous-ensemble de  $s$  échantillons dans  $S$  et évaluer le modèle à partir de ce sous ensemble.
- 2) Déterminer l'ensemble de données  $S_i$  qui sont inférieures à une distance  $t$  du modèle, fixée a priori. L'ensemble  $S_i$  est l'ensemble du consensus des échantillons et définit les points non aberrants de  $S$ .
- 3) Si la taille de  $S_i$  est supérieure à un seuil  $T$ , le modèle est ré-estimé à partir des données de  $S_i$  et l'algorithme se termine.
- 4) Si la taille de  $S_i$  est inférieure à  $T$ , un nouveau sous-ensemble est sélectionné et les étapes précédentes recommencent.
- 5) Après  $N$  essais, le plus grand consensus  $S_i$  est choisi et le modèle est ré-estimé à partir des données contenues dans  $S_i$ .

Sortie :  $\mathcal{M}(S_i)$

Algorithme 2 : RANdom SAMple Consensus, RANSAC

Lorsque la mise en correspondance est réalisée, une procédure itérative d'optimisation va affiner les estimations en minimisant l'erreur de reprojection : c'est l'ajustement de faisceaux [27]. Le principe de l'ajustement de faisceaux est la solution au sens du maximum de vraisemblance en supposant les bruits de mesure gaussiens : l'objectif est d'estimer les matrices de projection  $\hat{\mathbf{P}}^i$  et les points 3D  $\hat{\mathbf{X}}_j$  qui se projettent exactement sur les points d'image  $\hat{\mathbf{x}}_j^i$  tels que  $\hat{\mathbf{x}}_j^i = \hat{\mathbf{P}}^i \hat{\mathbf{X}}_j$ , et qui minimisent la distance image entre les points reprojétés et les points mesurés  $\mathbf{x}_j^i$  pour chaque vue dans lesquelles les points 3D apparaissent :

$$\min_{\hat{\mathbf{P}}^i, \hat{\mathbf{X}}_j} \sum_{ij} d(\hat{\mathbf{P}}^i \hat{\mathbf{X}}_j, \mathbf{x}_j^i)^2 \quad (2)$$

où  $d(\mathbf{x}, \mathbf{y})$  est la distance géométrique entre les deux vecteurs homogènes  $\mathbf{x}$  et  $\mathbf{y}$ .

Cette minimisation non-linéaire est réalisée de manière itérative en utilisant par exemple l'algorithme de Levenberg-Marquardt et en s'appuyant sur la structure creuse de la matrice jacobienne pour accélérer les calculs [5]. D'autres méthodes améliorent la localisation en utilisant une phase d'apprentissage des images sur une trajectoire prédéfinie comme dans les travaux de Royer et al. [28] et de Remazeilles et al [2].

Les approches précédentes sont principalement monovision. L'association de deux caméras permet d'obtenir une carte de profondeur (dense ou éparsée) et d'avoir ainsi, à chaque instant de temps, une reconstruction 3D de la scène. Les travaux sur l'estimation du mouvement d'un robot par stéréovision remontent au moins aux travaux de Moravec [29] fin des années 70. De bons résultats sur de courtes séquences en intérieur ont été obtenues au CMU [30] et à Stanford [31] dans les années 1980.

Des travaux similaires ont été menés parallèlement en France [32][33][34]. Un algorithme d'odométrie visuelle par stéréovision a été aussi utilisé sur Mars en 2004 [35]. Plus récemment, les travaux de Leibe et al. [36] permettent d'estimer le mouvement d'un véhicule et d'analyser les observations 3D afin de modéliser les trajectoires spatio-temporelles des différents objets dynamiques de la scène. D'un point de vue performances, les travaux de Howard [37] permettent d'obtenir une précision de 1 m sur un trajet de 400 m en utilisant une résolution d'images de  $1024 \times 768$  avec un traitement compris entre 10 ms et 20 ms dans un environnement statique.

L'approche de Rodriguez et al. [38] combine les avantages de plusieurs méthodes simples (stéréovision, flot optique, contraintes géométriques multi-vues) sous la forme d'un seul critère à optimiser et permet ainsi de prendre en compte les informations odométriques CAN afin d'accélérer les traitements et améliorer la robustesse vis à vis de la dynamique de scène. La première étape concerne l'extraction de points d'intérêt de type Harris et SURF. Une carte de disparité éparsée est alors calculée par corrélation croisée de type ZNCC (Zero-mean Normalized Cross Correlation). Le suivi temporel entre les instants  $t$  et  $t + 1$  est alors réalisé en utilisant l'algorithme de Lucas-Kanade [39]. Pour estimer le déplacement relatif entre les deux instants d'acquisition, une contrainte de rigidité de scène est appliquée via le tenseur quadrifocal [40] existant pour deux couples d'images. Ce tenseur permet d'encoder la position relative des caméras et de réaliser des transferts de points (warping) entre les différentes images. Basé sur ce principe, l'estimation du mouvement relatif peut être résolu comme un problème d'optimisation non linéaire. Les 6 degrés de liberté sont calculés en minimisant une fonction d'énergie non linéaire représentant l'erreur entre le mouvement apparent (flot optique) et le mouvement apparent virtuel créé par le tenseur quadrifocal. La minimisation est alors réalisée par l'algorithme de Levenberg-Marquardt sur le critère présenté ci-dessous :

$$\epsilon = \sum_{i=1}^k \mathbf{W} \left[ \|p_i - \hat{p}_i\| + \|p'_i - \hat{p}'_i\| \right] \quad (3)$$

où  $p_i$  et  $p'_i$  sont les points suivis dans les images de gauche et de droite à  $t + 1$ .  $\hat{p}_i$  et  $\hat{p}'_i$  sont les points à l'instant  $t$  transformés par le mouvement estimé.  $\mathbf{W}$  est une matrice de pondération issue d'un

M-estimateur [40] utilisée dans un schéma de minimisation itérative de type IRLS (Iteratively Reweighted Least Squares) [41]. Les poids de la matrice de pondération sont initialisés par l'odométrie roue disponible sur le bus CAN afin d'assurer une convergence plus rapide.

L'algorithme est présenté ci-après (Algo. 3) :

### Objectif

Estimer de manière robuste, le mouvement relatif  $\mathbf{u}_t = (\delta\mathbf{p}_t^T, \delta\mathbf{q}_t^T)^T$  entre deux acquisitions d'images stéréo.

Entrée : paire d'images stéréo.

### Algorithme

- 1) Extraction de points caractéristiques dans les deux vues et mise en correspondance stéréo.
- 2) Acquisition d'une nouvelle paire d'images et suivi temporel des points caractéristiques.
- 3) Initialisation de la minimisation non linéaire à partir des données de l'odométrie roue disponible sur le bus CAN.
- 4) Estimation du mouvement par schéma itératif des moindres-carrés pondérés : estimation du mouvement par minimisation non linéaire de l' Eq. 3 et calcul de la matrice de pondération à partir des résidus de la minimisation. Cette étape permet de rejeter les points aberrants (voir Fig. 3).

Sortie :  $\mathbf{u}_t = (\delta\mathbf{p}_t^T, \delta\mathbf{q}_t^T)^T$

Algorithme 3 : odométrie par stéréovision et flot optique

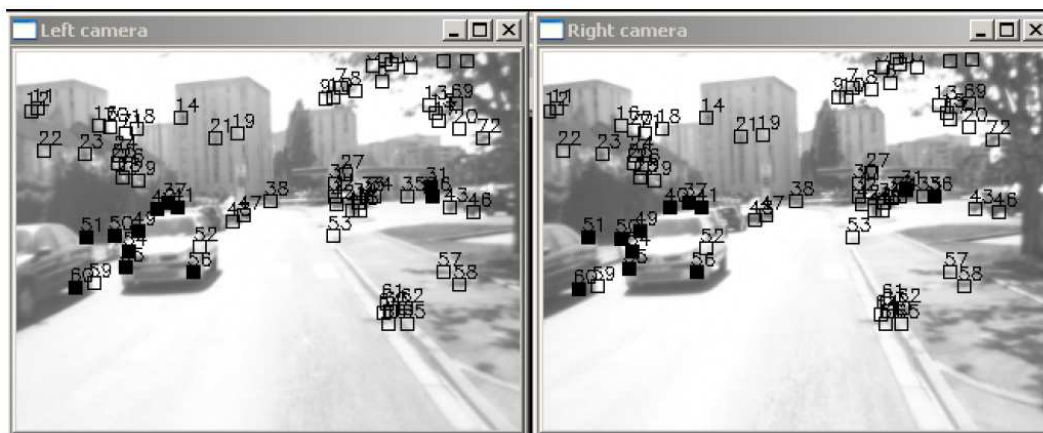


Fig. 3. Dynamique des points d'intérêt : points aberrants ■, points associés □

Les résultats obtenus sont présentés Fig. 5. Les expérimentations ont été réalisées en utilisant un banc stéréo composé de deux caméras CMOS avec des objectifs de 4.5mm et calibrées (voir Fig. 4).





Fig. 4. Véhicule CARMEN du laboratoire Heudiasyc : configuration du banc stéréo

Les images ont une résolution de  $320 \times 240$  et sont obtenues à une fréquence de 15 Hz. La vitesse du véhicule est en moyenne de 60 km/h.

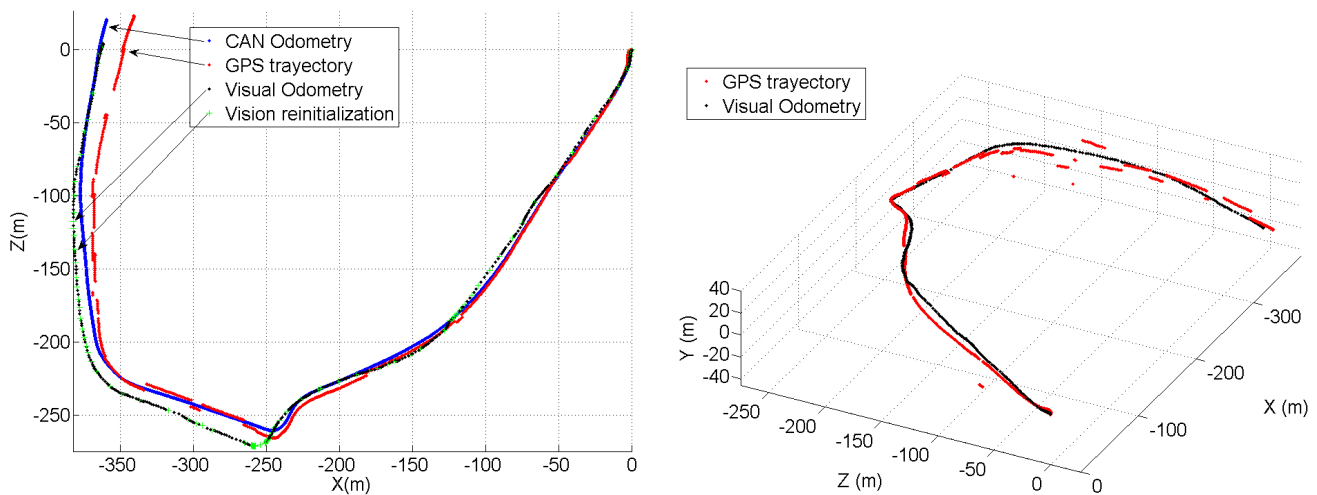


Fig. 5. Gauche : trajectoire 2D estimée avec le GPS (Septentrio PolaRx2c), les capteurs proprioceptifs (IMU-WSS) et l'odométrie visuelle. Droite : trajectoire 3D estimée

Les résultats obtenus montrent que sur une trajectoire d'environ 790m, l'odométrie dérive peu : 3.9 % dans le plan  $x - y$  et 0.25 % verticalement pour le rapport entre l'erreur en position et la distance totale parcourue. Les instants de dérive sont principalement liés au fait que le M-estimateur et l'algorithme itératif IRLS supportent mal un taux de points aberrants supérieur à 50%, ce qui peut arriver lorsque de nombreux objets sont en mouvement et qu'ils sont proches du véhicule, par exemple à l'arrivée dans un rond-point. Dans ce cas de figure, une étape de ré-initialisation est nécessaire. De nouveaux points d'intérêt sont alors détectés et l'algorithme poursuit son estimation. Comparée à l'odométrie roue, cette approche n'améliore pas la précision dans le plan de la route, mais elle fournit une estimation de mouvement sur les 6 degrés de liberté. En regardant de plus près la Fig. 5 de droite, on remarque l'intérêt de l'odométrie visuelle lors des phases de masquage des satellites sur la trajectoire GPS en rouge. Enfin, il est important

de souligner que la dérive en altitude est faible (moins de 0.3 % sur la distance parcourue), ce qui permet de fournir une alternative fiable par rapport au GPS.

## V. L'ODOMÉTRIE OPTIQUE MULTI-MODALITÉS

### A. Intérêt

Aux vues des résultats obtenus avec les caméras et les LIDAR, et en regardant leurs caractéristiques, il apparaît une forte complémentarité. L'idée est alors de les utiliser conjointement afin d'améliorer la précision et la fiabilité des résultats. Intuitivement, la combinaison d'informations de profondeur avec des images doit améliorer les performances dans la mesure où les informations d'apparence des images peuvent compléter les informations de formes issues des capteurs de profondeur, et parce que les caméras offrent une très bonne résolution angulaire.

Pour une revue exhaustive des comparaisons entre capteurs pour l'application qui nous intéresse, le lecteur pourra se référer à [42]. Nous rappelons ici les éléments importants concernant la stéréovision et le LIDAR.

	<i>LIDAR</i>	<i>Stéréovision</i>
<b>Coût</b>	important	faible
<b>Résolution</b>	Résolution spatiale limitée	Résolution spatiale importante
<b>Portée</b>	importante	limitée
<b>Conditions de luminosité</b>	insensible	fortement dépendant
<b>Importance de la texture</b>	aucune	cruciale pour avoir de bons résultats
<b>Interprétation haut niveau</b>	limitée	importante
<b>Traitements 3D</b>	aucun	extraction et triangulation

TABLE II  
AVANTAGES ET LIMITATIONS DU LIDAR ET DE LA STÉRÉOVISION

### B. Le problème du calibrage multi-capteurs

Lorsque l'on considère un problème de fusion multi-capteurs, se pose le problème du recalage d'informations, à savoir comment exprimer l'ensemble des informations dans un repère commun : c'est le problème du calibrage multi-capteurs. Dans le cas qui nous intéresse, à savoir pour des capteurs de type caméras et Lidar, l'information à fusionner est principalement géométrique. Ainsi, la consistance de l'échange d'information dépend de la transformation rigide (translation et rotation) qui les relie. L'estimation de ces paramètres doit être réalisée en utilisant les mesures provenant de ces deux capteurs. Ce problème peut être formulé comme une minimisation sur l'erreur de transformation sur un ensemble commun de mesures capteur. Cette erreur de mesure sera d'autant plus critique que les capteurs seront distants. Par exemple, dans beaucoup d'applications liées aux véhicules, les caméras sont derrière le rétroviseur et le télémètre à l'avant (voir Fig. 6).

Parmi les travaux de calibrage, nous pouvons citer ceux de Dupont et al. [43] et Brun et al. [44] qui estiment les transformations rigides entre une caméra et un télémètre à une nappe, en utilisant une mire plane. Une autre approche présentée dans [45] permet de réaliser le calibrage sans mire mais en sélectionnant des points caractéristiques, à la main dans l'image. La difficulté se trouve donc là : quel objet géométrique utiliser afin de stimuler l'ensemble des paramètres et obtenir la transformation la plus précise possible ?

Rodriguez et al. [46] propose l'utilisation d'une mire circulaire détectable et caractérisable à la fois par le télémètre et la caméra (voir Fig. 7). Les cercles sont reconstruits en 3D à la fois dans le repère lié au Lidar et celui lié à la caméra. Le calibrage est alors réalisé par recalage des points 3D sur au moins 6 poses et affiné par une minimisation non-linéaire sur la distance géométrique entre l'ensemble des estimations. Des expérimentations ont été réalisées en utilisant un Lidar 4 nappes Ibeo Alasca XT et une caméra Sony



Fig. 6. Véhicule CARMEN du laboratoire Heudiasyc : configuration de la caméra et du Lidar

DFW-VL500 de résolution  $640 \times 480$ . La mire est composée de deux cercles concentriques de rayon 20 et 25cm, et est percée en son centre afin d'être détectée par le Lidar.

L'utilisation de mire peut néanmoins poser certains problèmes quant à la portée des capteurs. En effet, pour obtenir suffisamment de précision à la fois pour le Lidar et pour la caméra, il faut disposer de cibles de grandes tailles. La manipulation peut aussi devenir critique lorsque les caméras stéréoscopiques et le Lidar n'ont pas la même portée.

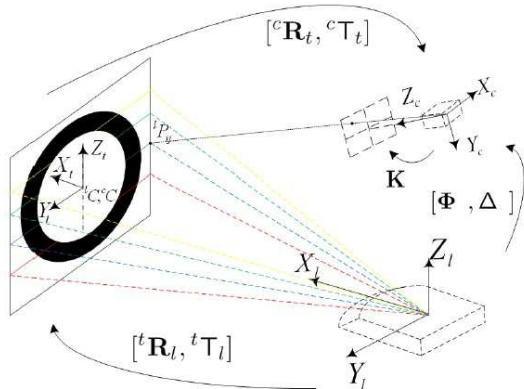


Fig. 7. Gauche : méthode de calibrage proposée par Rodriguez et al. [46]. La transformation à estimer est représentée par  $[\Phi, \Delta]$ . Droite : reprojction des impacts Lidar après calibrage.

### C. Fusion de plusieurs modalités

L'idée principale de la fusion est de déterminer la meilleure procédure pour combiner les informations provenant des différents capteurs. En reprenant la synthèse présentée dans [47], il existe différentes stratégies de fusion de données : suivant le type de fusion, la configuration des capteurs ou les caractéristiques entrée/sortie. En ce qui concerne la localisation par capteurs optiques, il est souhaitable que leurs caractéristiques assurent le compromis précision/intégrité<sup>1</sup>.

1. Précision : résultante du bruit. Intégrité : confiance et rejet de mesures aberrantes.

Si on se réfère aux travaux de Durrant-Whyte [48], il existe 3 configurations de base :

- Complémentaire ou redondante : les capteurs ne dépendent pas directement les uns des autres, mais une fois combinés ils améliorent l'image du phénomène observé.
- Compétitive : chaque capteur réalise la même fonction, le but étant alors de réduire les effets des données incertaines et erronées.
- Coopérative : la combinaison de plusieurs capteurs permet d'obtenir une information indisponible avec les capteurs seuls.

Pour l'odométrie optique 3D, les approches compétitives et redondantes sont très pertinentes dans la mesure où les caméras et le Lidar réalisent la même fonction. Ces deux capteurs sont aussi complémentaires : la stéréovision fournit de bons modèles d'apparence, utiles pour la focalisation [6] mais sensibles aux conditions d'éclairage et le Lidar améliore fortement la précision de la mesure 3D quelques soient les conditions adverses (pluie, brouillard, neige, éblouissement, etc). Cette redondance des capteurs peut aussi assurer une certaine surveillance en présence de perturbations (surexposition des caméras, mesures fantôme, etc). Deux exemples de stratégies de fusion sont présentés Fig. 8.

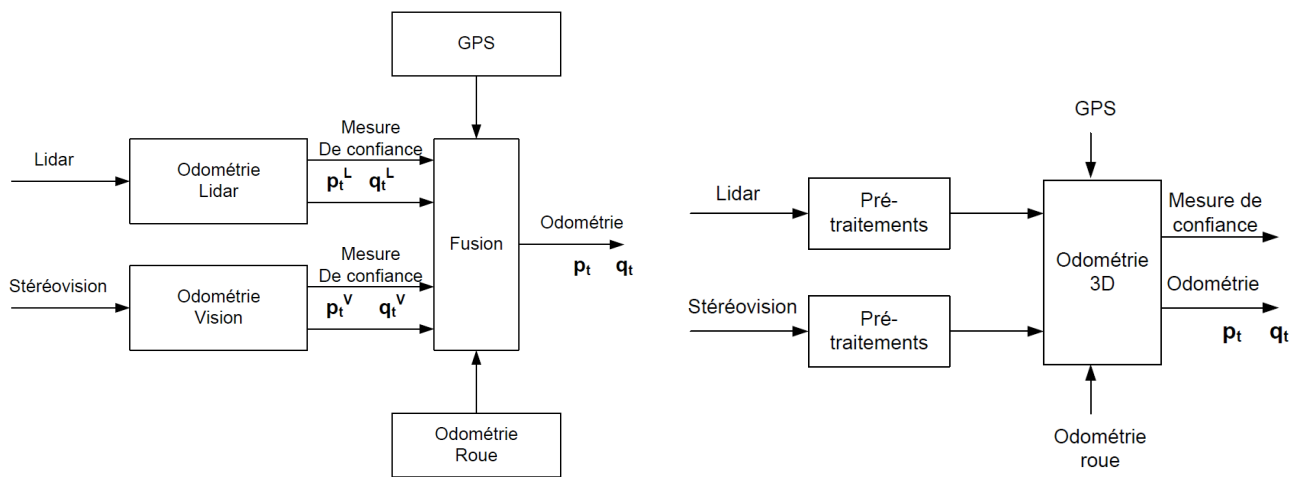


Fig. 8. Gauche : fusion lâche. Droite : fusion serrée - bas niveau.

La figure de gauche montre une fusion lâche, c'est dire que chacun des capteurs estime une odométrie et le module de fusion permet de les mettre en compétition et ainsi d'obtenir la meilleure information. Dans la figure de droite, la fusion s'opère sur les données brutes : c'est un couplage serré, pour lequel la combinaison des deux informations améliore l'odométrie. Dans les deux cas, le GPS et l'odométrie roue sont utilisés pour assurer l'intégrité de l'estimation.

Pour les capteurs optiques, la localisation est principalement réalisée par détection et suivi d'amers. Par exemple, Labayrade et al. [49] proposent d'améliorer la précision et la fiabilité des résultats de détection d'obstacles, en réalisant une fusion entre la stéréovision et un Lidar à balayage à une nappe (style Sick LMS211) qui fournit une coupe de la scène dans un plan. L'estimation des positions et des vitesses est alors effectuée par le télémètre et l'existence d'obstacles est confirmée par la stéréovision. Les travaux de Tessier et al. [6] utilisent un filtre de Kalman étendu, pour estimer l'état du véhicule et la zone d'incertitude, et estiment à chaque instant l'intégrité de la localisation par un mécanisme qui quantifie la confiance en comparant ce que perçoit le véhicule avec ce qu'il aurait dû percevoir. Mais la gestion des amers sous contrainte multimodale reste un problème central. Sans référence globale (comme des amers cartographiés [50]) l'association de deux amers disponibles dans chaque capteur en vue d'améliorer la localisation et son intégrité reste un problème ouvert. Finalement, dans les approches multimodales de perception pour

les véhicules intelligents, le challenge semble être le suivant : comment obtenir une bonne précision et intégrité à vitesse élevée et comment se localiser dans un environnement dynamique qui présente des ambiguïtés de détection ?

## VI. CONCLUSIONS ET PERSPECTIVES

Cet article a présenté un aperçu général des recherches menées sur le thème de l'odométrie optique 3D. Les défis actuels de cette thématique portent essentiellement sur la capacité à garantir une estimation précise et intègre, avec un porteur se déplaçant à grande vitesse dans une scène comportant des aspects dynamiques et/ou évolutifs dans le temps.

Dans ce cadre, la gestion des amers reste un problème ouvert, en particulier en présence d'ambiguïté de détection, de fortes occultations et de conditions adverses. Les descripteurs associés aux amers doivent être encore améliorés, notamment pour aboutir à un suivi robuste et pleinement multi-modal. Dans ce contexte, certaines méthodes d'apprentissage statistique, supervisé ou non, semblent très prometteuses.

Dans un contexte multi-modal, la représentation des données joue un rôle primordial dans la fusion d'informations, notamment pour pouvoir estimer la confiance et assurer un niveau d'intégrité du système lors de l'estimation du mouvement du véhicule. Les approches bayésiennes de fusion de données restent prédominantes, mais d'autres formalismes (comme les théories des possibilités, les approches ensemblistes et les fonctions de croyance) permettent la prise en compte simultanée de l'imprécision, de l'incertitude, et de l'incomplétude des sources d'information.

Quand on considère un problème de fusion multi-capteur, le calibrage reste aussi une étape centrale. Des efforts restent à faire à ce niveau, notamment lorsque les paramètres évoluent au cours du temps ; on parle alors du problème d'auto-calibrage.

Pour finir, on peut remarquer que les capteurs optiques sont capables de fournir maintenant une information haute résolution. Ce volume de données n'est pas encore pris en compte dans sa globalité, notamment du fait de l'augmentation de la capacité de calcul nécessaire et des contraintes de puissance liées aux systèmes embarqués. Même si les prototypes actuels des laboratoires présentent de bonnes performances (avec peu de contraintes sur la puissance de calcul), il faudra encore attendre quelques années avant le déploiement industriel de tels systèmes à bord des véhicules.

## REMERCIEMENTS

L'auteur tient à remercier Philippe Bonnifait et Sergio Rodriguez pour leurs échanges et leur contribution à la rédaction de cet article.

## REFERENCES

- [1] Fayad, F. and Cherfaoui, V., "Tracking objects using a laser scanner in driving situation based on modeling target shape", IEEE Intelligent Vehicles Symposium (IV2007), 2007
- [2] A. Diosi, A. Remazeilles, S. Segvic, F. Chaumette. Outdoor visual path following experiments. In IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'07, Pages 4265-4270, San Diego, Californie, Octobre 2007.
- [3] Junior : The Stanford Entry in the Urban Challenge.
- [4] Bonnifait, Ph. and Bouron, P. and Meizel, D. and Crubillé, P., " Dynamic Localization of Car-like vehicles using Data Fusion Of Redundant ABS sensors", The Journal Of Navigation, vol. Vol. 56, pp. 1-13, 2003.
- [5] R. Hartley and A. Zisserman, " Multiple View Geometry in Computer Vision. Second Edition", Cambridge University Press, 2003.
- [6] C. Tessier, C. Debain, R. Chapuis and F. Chausse, "Simultaneous Landmarks Detection and Data Association in Noisy Environment for Map Aided Localization", IEEE IROS 2007.
- [7] Besl, P. J. and McKay N. D., "A method for registration of 3-D Shapes", IEEE Trans. Pattern Anal. Mach. Intell., volume 14, number 2, pages. 239-256, 1992.
- [8] S. Rusinkiewicz and M. Levoy, " Efficient variants of the icp algorithm", Third Intl. Conf. on 3D Digital Imaging and Modeling, 2001.
- [9] T. Miyasaka, Y. Ohama and Y. Ninomiya, "Ego-motion Estimation and Moving Object Tracking using Multi-layer LIDAR", IEEE IV2009, 2009.
- [10] T. D. Vu, O. Aycard and N. Appenrodt, "Online Localization and Mapping with Moving Object Tracking in Dynamic Outdoor Environments", Proc. IEEE Intelligent Vehicles Symposium, pp.190-195, June 2007.

- [11] C. Laugier, D. Vasquez, M. Yguel, Th. Fraichard and O. Aycard, "Geometric and Bayesian Models for Safe Navigation in Dynamic Environments", *Intelligent Service Robotics* 1, 1 (2008).
- [12] J. Buriel, T.D. Vu and O. Aycard, "Grid-based Localization and Online Mapping with Moving Object Detection and Tracking", RR Inria, 2007.
- [13] C.C. Wang, C. Thorpe, S. Thrun, M. Hebert and H. Durrant-Whyte, "Simultaneous Localization, Mapping and Moving Object Tracking", *IJRR*, vol. 26, No. 9, pp. 889-916, 2007.
- [14] A. J. Davison, I. Reid, N. Molton and O. Stasse, "MonoSLAM : Real-Time Single Camera SLAM", *IEEE Trans. PAMI* 2007.
- [15] J. Solà, A. Monin, M. Devy and T. Vidal-Calleja, "Fusing monocular information in multi-camera SLAM". *IEEE Transactions on Robotics*, Special Issue on Visual SLAM, october 2008.
- [16] T. Lemaire, S. Lacroix. Monocular-vision based SLAM using line segments. In *IEEE International Conference on Robotics and Automation*, Roma (Italy), April 2007.
- [17] Angeli, A., Filliat, D., Doncieux, S., and Meyer, J.-A., "Visual topological SLAM and global localization", *Proceedings of the International Conference on Robotics and Automation (ICRA)*. 2009.
- [18] T. Lemaire, Cyrille Berger, I-K. Jung, S. Lacroix. Vision-based SLAM : stereo and monocular approaches. *International Journal on Computer Vision*, 74(3) :343-364, 2007.
- [19] D. Scaramuzza and R. Siegwart, "Appearance-Guided Monocular Omnidirectional visual Odometry for Outdoor Ground Vehicles", *IEEE TRO*, Vol. 24, No. 5, October 2008.
- [20] Harris, C. and Stephens, M., A Combined Corner and Edge Detector, *Proc. Fourth Alvey Vision Conference*, pp.147-151, 1988.
- [21] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91-110, 2004.
- [22] M. Agrawal, K. Konolige, M. Rufus Blas, "CenSurE : Center Surround Extremas for Realtime Feature Detection and Matching". *ECCV* (4) 2008 : 102-115.
- [23] Nister, D., "An Efficient Solution to the Five-Point Relative Pose Problem", *IEEE Conference on Computer Vision and Pattern Recognition*, Volume 2, pp. 195-202, 2003.
- [24] D. Nistér and H. Stewénus, A Minimal solution to the generalized 3-point pose problem, accepted to *Journal of Mathematical Imaging and Vision*, 2006.
- [25] Fischler, M. and Bolles, R., "Random Sample Consensus : a Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography", *Commun. Assoc. Comp. Mach.*, 24 :381- 395, 1981.
- [26] K. Konolige, M. Agrawal and J. Sola "Large Scale Visual Odometry for Rough Terrain" *International Symposium on Research in Robotics*, Japan, 2007.
- [27] A. Eudes and M. Lhuillier, "Error Propagations for Local Bundle Adjustment", *IEEE CVPR*, 2009.
- [28] E. Royer, J. Bom, M. Dhome, B. Thuilot, M. Lhuillier, F. Marmoiton, "Outdoor autonomous navigation using monocular vision", *IEEE IROS*, pp. 3395-3400, Edmonton, août 2005.
- [29] Moravec, H., "Obstacle avoidance and navigation in the real world by a seeing robot rover", PhD thesis, Stanford University, 1980.
- [30] Matthies, L. and Shafer, S., " Error modeling in stereo navigation", *IEEE Journal of Robotics and Automation*, RA-3(3), 1987.
- [31] Kriegman, D., Triendl, E. and Binford, T., "Stereo vision and navigation in buildings for mobile robots", *IEEE Trans. on Robotics and Automation*, 5(6), 1989.
- [32] Zhang, Zhengyou and Faugeras, Olivier D., "Estimation of Displacements from Two 3-D Frames Obtained From Stereo", *IEEE Trans. Pattern Anal. Mach. Intell.*, volume 14, number 12, pages. 1141-1156, 1992.
- [33] Lacroix, S., Mallet, A., Chatila, R. and Gallo, L., Rover self localization in planetary-like environments, *5th International Symposium on Artificial Intelligence, Robotics, and Automation in Space (i-SAIRAS)*, 1999.
- [34] Jung, I. and Lacroix, S., "Simultaneous localization and mapping with stereovision", *Robotics Research : the 11th International Symposium*, P. Dario and R. Chatila (eds.), Springer, 2005.
- [35] Maimone, M., Johnson, A., Willson, R. and Matthies, L., "Autonomous navigation results from the Mars Exploration Rover (MER) mission", *ISER*, 2004.
- [36] B. Leibe, N. Cornelis, K. Cornelis, L. Van Gool, "Dynamic 3D Scene Analysis from a Moving Vehicle". In *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR'07)*, Minneapolis, USA, June 2007.
- [37] Howard, A., "Real-time stereo visual odometry for autonomous ground vehicles", in *proc. of IEEE/RSJ International Conference IROS 2008*, pages. 3946-3952, septembre 2008.
- [38] S. A. Rodríguez F., V. Frémont and P. Bonnifait, "An Experiment of a 3D Real-Time Robust Visual Odometry for Intelligent Vehicles", *IEEE ITSC*, 2009.
- [39] J.-Y. Bouget, "Pyramidal implementation of the lucas kanade feature tracker description of the algorithm," *Intel Corporation Microprocessor Research Labs, Tech. Rep.*, 2002.
- [40] A.I. Comport, E. Malis and P. Rives, " Accurate Quadrifocal Tracking for Robust 3D Visual Odometry", In *Proceedings of IEEE International Conference on Robotics and Automation*, april 2007.
- [41] C. V. Stewart, "Robust parameter estimation in computer vision," *Society for Industrial and Applied Mathematics*, vol. 41, no. 3, pp. 513-537, 1999.
- [42] T. Gandhi and M. M. Trivedi, "Pedestrian Protection Systems : Issues, Survey, and Challenges", *IEEE Transactions on Intelligent Transportation Systems*, Vol. 8, No. 3, Septembre 2007.

- [43] Romain Dupont, Renaud Keriven and Philippe Fuchs, " An Improved Calibration Technique for Coupled Single-Row Telemeter and CCD Camera", In Proceedings of the International Conference on 3-D Digital Imaging and Modeling, 2005.
- [44] Xavier Brun and François Goulette, " Modeling and Calibration of Coupled Fish-Eye CCD Camera and Laser Range Scanner for Outdoor Environment Reconstruction", In Proceedings of the Sixth International Conference on 3-D Digital Imaging and Modeling, 2007.
- [45] D. Scaramuzza, A. Harati and R. Siegwart, "Extrinsic Self Calibration of a Camera and a 3D Laser Range Finder from Natural Scenes", International Conference on Intelligent Robots and Systems., 2007.
- [46] S. A. Rodriguez F., V. Frémont and P. Bonnifait. Extrinsic Calibration between a Multi-Layer Lidar and a Camera. In proceedings of IEEE Multisensor Fusion and Integration for Intelligent Systems, August 20-22, Korea University, Seoul, Korea, 2008.
- [47] H.B. Mitchell, "Multi-sensor Data Fusion An Introduction", Springer 2007.
- [48] H.F. Durrant-Whyte, "Sensor models and multi-sensor intergration", International Journal of Robotics Research, 1988.
- [49] Labayrade R., Royere C., Gruyer D., Aubert D., "Cooperative Fusion for Multi-Obstacles Detection with use of Stereovision and Laser Scanner", in Autonomous Robots, special issue on "Robotics Technologies for Intelligent Vehicles", Vol. 19, N°2, September 2005, pp. 117 - 140.
- [50] C. Cappelle, "Localisation de véhicules et détection d'obstacles, apport d'un modèle virtuel 3D urbain", Thèse de doctorat, 2008.