



N° d'ordre : 2854

**École Doctorale Mathématiques, Sciences de
l'Information et de l'Ingénieur**

UdS – INSA

THÈSE

présentée pour obtenir le grade de

Docteur de l'Université de Strasbourg

**Discipline : Sciences de l'Ingénieur
Spécialité : Robotique**

par

Marie-Anne Mittet

**Caméras 3D pour la localisation
d'un système mobile
en environnement urbain**

Soutenue publiquement le 15 juin 2015

Membres du jury

<i>Directeur de thèse :</i>	M. Pierre Grussenmeyer, Professeur, INSA de Strasbourg
<i>Rapporteur externe :</i>	M. François Goulette, Professeur, Mines ParisTech
<i>Rapporteur externe :</i>	M. Laurent Polidori, Professeur, ESGT-CNAM Le Mans
<i>Examineurs :</i>	M. Pierre Charbonnier, Directeur de Recherches, CEREMA, LRPC de Strasbourg
	Mme Tania Landes, Maître de Conférences, INSA de Strasbourg
	M. Stéphane Rialland, Société Trimble France

Remerciements

À l'issue de la rédaction de cette thèse, je suis convaincue que la thèse n'est pas un travail solitaire. En effet, je n'aurais jamais pu réaliser ce travail doctoral sans le soutien de nombreuses personnes dont la générosité, la bonne humeur et l'intérêt manifestés à l'égard de ma recherche m'ont permis de progresser et d'évoluer.

Je tiens tout d'abord à remercier Pierre Charbonnier pour m'avoir fait l'honneur de présider mon jury de thèse. Je remercie également François Goulette et Laurent Polidori, d'avoir endossé le rôle de rapporteur. Le temps consacré à la lecture de ce mémoire, et le regard critique porté sur ce travail, sont un honneur pour moi. Les remarques que m'avez faites m'ont permis d'améliorer significativement ce document.

Je tiens à adresser mes plus vifs remerciements à Pierre Grussenmeyer, mon directeur de thèse, et Tania Landes, ma co-encadrante de thèse. Ces travaux n'auraient simplement pas pu se faire sans votre aide. Pour tout le temps que vous m'avez consacré, pour l'encadrement, pour les conversations scientifiques ou non, pour les diverses relectures, tout ceci m'a été d'une grande aide durant ces 3 années.

J'adresse mes remerciements à Gregory Lepère et Stéphane Rialland pour leurs nombreuses idées et conseils avisés. Je tiens également à remercier l'ensemble des équipes de chez Trimble pour l'accueil et les conditions de travail privilégiées qui m'ont été offertes, et tout particulièrement l'équipe Hardware.

Je tiens à remercier tous mes amis qui ont toujours su trouver du temps pour organiser quelque chose ou simplement me changer les idées. Je ne peux nommer tout le monde, mais je tiens à citer : Balett, Bon'Oeil, Boré, Clé, Daï, Detapi, Doc, Domi, Ellé, José, Kcine, Lajeune, Leffé, Luidir, Locoms, Ménie, Mumu, Senpala, Stone, Zaps, Zoror'. Un grand merci à Hélène, co-doctorante, grâce à toi, mes quelques mois alsaciens sont passés bien vite. J'adresse une pensée spéciale à Yang Yifan, Geoffrey Vincent et Mikael Royère, pour leurs travaux de PFE, j'ai également une pensée pour l'ensemble de leurs collègues que j'ai pu côtoyer durant leurs stages. Je souhaite aussi remercier l'ensemble des enseignants et techniciens de la spécialité Topographie de l'INSA de Strasbourg, et notamment : Emmanuel Alby, Samuel Guillemin, Bernard Grandmougin, et Mathieu Koehl.

Mes remerciements s'adressent enfin à mes parents, qui m'ont toujours épaulée. Je souhaite également exprimer ma sincère gratitude à mes deux frères, Benjamin, et Jean-Baptiste, sans oublier Sarah et Seàn. Je souhaite également rendre hommage à mes grands-parents, enfin je remercie l'ensemble de la famille Breton.

Finalement, j'adresse mes remerciements les plus chaleureux à Antoine, pour la grande patience, les encouragements, et la confiance dont il a fait preuve tout au long de ces trois années.

Résumé

La numérisation de l'environnement urbain se développe depuis plusieurs années, grâce à une variété de systèmes d'acquisition 3D. Les systèmes statiques sont actuellement les seuls à atteindre un niveau de précision et de complétude satisfaisants, mais présentent un faible niveau de productivité lorsqu'il s'agit de numériser un quartier ou une ville entière. Afin d'automatiser les processus d'acquisition, les chercheurs et les constructeurs ont développé des systèmes mobiles de cartographie en intégrant de nombreux capteurs de positionnement et d'orientation associés aux instruments de numérisation (scanner à balayage laser, caméras numériques). Le traitement des données tridimensionnelles géoréférencées d'un relevé mobile exige de maîtriser la localisation de la plateforme d'acquisition ainsi que d'intégrer les données enregistrées par les différents capteurs composant le système.

L'objectif de cette thèse est de développer un nouveau système mobile composé de capteurs à bas coût, permettant la localisation de façon innovante en tenant compte des contraintes particulières de l'environnement urbain. Le travail s'articule ainsi autour de deux grands volets, que sont d'une part la conception d'un nouveau système mobile de localisation et d'autre part le développement d'un algorithme permettant de calculer la trajectoire du mobile à partir des données enregistrées par les capteurs composant le système.

L'étude conceptuelle et les critères d'évaluation définis en partenariat avec un industriel ont permis de proposer une solution qui repose sur l'utilisation de trois caméras 3D de type Kinect et sur l'apport complémentaire d'une caméra de type Fish-Eye. Une solution algorithmique basée sur l'odométrie visuelle permet de calculer la trajectoire du mobile en temps réel à partir des données fournies par les caméras 3D composant le système développé. L'originalité de la méthodologie mise en place réside dans l'exploitation d'orthoimages, créées à partir des nuages de points acquis en temps réel par les trois caméras. L'étude des différences entre les orthoimages successives acquises par le système mobile permet d'en déduire ses positions successives et d'en calculer sa trajectoire. Afin d'évaluer l'approche développée, différents jeux de données simulées et réelles ont été employés. Le système d'acquisition basé sur les caméras 3D est opérationnel. Placé à bord d'un charriot mobile, il a permis d'acquérir des images de profondeur et de produire des orthoimages, qui, une fois traitées, ont mené au calcul d'une trajectoire. L'expérience a été reproduite en intérieur, dans des conditions de laboratoire ainsi qu'en extérieur, dans un quartier de Strasbourg. Les atouts et les limitations identifiés à partir de ces expérimentations ont mis en exergue la complémentarité des orthoimages à champ de vision locale et des images hémisphériques acquises par une caméra Fish-Eye associée au prototype. Le couplage des données acquises par les caméras 3D et des images obtenues avec une caméra Fish-Eye constitue ainsi une perspective intéressante pour améliorer la précision de la trajectoire finale. En outre, la technologie des caméras 3D, en constante évolution, devrait permettre d'améliorer la qualité des nuages de points acquis dans le futur, notamment du point de vue de leur utilisation en extérieur et de leur portée.

Mots-clés : Localisation, caméras 3D, orthoimages, caméra Fish-Eye, odométrie visuelle, robotique, géomatique.

Abstract

The development of the digitization of urban environments is based on technological advances of 3D recording systems. Static systems are currently the only ones able to achieve a good level of accuracy and completeness, but they have a low level of productivity when it comes to acquire a neighbourhood of an entire city. In order to develop new automatic recording and mapping processes, researchers and manufacturers have developed mobile mapping systems based on the positioning and orientation of images and range based sensors such as digital cameras and laser scanners. The processing of 3D data acquired by a mobile mapping system requires the mastering of the localization of the acquisition platform and integrates the data recorded by a set of sensors.

The aim of this thesis is to develop a new kind of localization system, composed of low-cost sensors, taking into account the special constraints of urban environments. The work is organized in two steps. Firstly, the design of a new mobile localization system is presented. Secondly, the development of an algorithm able to calculate the trajectory of the mobile from the recorded data is outlined.

The conceptual study and the evaluation of some criteria have been established in collaboration with a surveying instrument manufacturer. The proposed solution is based mainly on the use of three 3D cameras (Kinect or Asus Xtion) oriented towards the ground and an additional zenithal Fish Eye camera. The localization algorithm is based on visual odometry principles in order to calculate the trajectory of the mobile platform in real time from the data provided by the 3D cameras placed on the system. One originality of the processing method lies within the exploitation of orthoimages processed from the point clouds that are acquired in real time by the three 3D cameras. From the study of the differences between successive orthoimages acquired by the mobile system, the successive positions of the mobile platform can be deduced to deliver the trajectory of the mobile system. In order to evaluate the developed approach, different sets of simulated and real data were used. Our mobile platform based on 3D cameras is operational and records depth images used to produce the orthoimages. Experiments were performed indoors in laboratory conditions as well as outdoors at nightfall in a district of Strasbourg. The assets and limitations identified from these experiments have highlighted the advantage of using orthoimages which have a local vision of the scene, and to complete the dataset with complementary hemispherical images obtained through a Fish Eye camera placed at the top of the 3D camera system. The fusion of data acquired by the 3D cameras and the Fish Eye images is thus an interesting perspective to improve the accuracy of the final processed trajectory. Furthermore, the 3D camera technology is constantly evolving and should improve the quality of point clouds acquired in the future, particularly for the use in outdoor conditions.

Keywords : Localization, 3D cameras, orthoimages, Fish Eye camera, visual odometry, robotics, geomatics.

Table des matières

Remerciements	3
Résumé	5
Abstract	7

Table des matières

1 Introduction générale

1	Contexte général	1
2	Terra Mobilita	4
3	Objectifs de la thèse	5
	3.1. Rencontre de deux communautés	6
	3.2. Plan du manuscrit	6

2 État de l'art

1	Introduction	9
2	Les types d'environnements	10
	2.1. Environnement intérieur/extérieur	10
	2.2. Environnement statique/dynamique	11
3	La localisation en environnement urbain	12
	3.1. La localisation par satellites	12
	3.2. La localisation inertielle	18
	3.3. La localisation odométrique	22
	3.4. Fusion de données	27
	3.5. La localisation à partir de capteurs de vision	31
4	Etat de l'art sur les caméras 3D	33
	4.1. Principe de la triangulation	33
	4.2. Principe du "temps de vol"	34

3 Conception d'un système de localisation

1	Introduction	39
2	Conception	39
	2.1. Analyse du besoin/contexte	40
	2.2. Concepts	42
3	Concept final	50
	3.1. Idée générale	50
	3.2. Caméras 3D	52
	3.3. Caméra omnidirectionnelle	63
	3.4. Aspect fonctionnel	74

4 Algorithme de localisation

1	Introduction	83
2	État de l'art sur l'odométrie visuelle	84
3	Pôle caméras 3D	88
	3.1. Présentation de l'approche développée	88
	3.2. Détection de points d'intérêts	93
	3.3. Mise en correspondance	103
	3.4. Élimination des correspondances aberrantes et estimation de la transformation	107
	3.5. Calcul de la transformation	112
4	Pôle caméra <i>Fish Eye</i>	114
	4.1. Détection de points d'intérêts	115
	4.2. Mise en correspondance	116
	4.3. Élimination des correspondances aberrantes	116
	4.4. Calcul de la transformation	117
	4.5. Calcul de position moyenne	119

5 Expérimentations et résultats

1	Introduction	121
2	Expérimentations	122
	2.1. Données simulées	122
	2.2. Données réelles	125
3	Résultats	130
	3.1. Pôle caméras 3D	130

3.2.	Bilan des résultats obtenus avec le pôle caméras 3D	147
3.3.	Pôle caméra <i>Fish Eye</i>	148
3.4.	Bilan des résultats obtenus avec le pôle caméra <i>Fish Eye</i>	150
3.5.	Deux pôles de vision pour une trajectoire	151
3.6.	Bilan général	152

| Conclusion générale

| Conclusion générale

1	Synthèse des résultats obtenus à l'issue de la thèse	157
1.1.	Volet 1 : Conception d'un système de localisation	158
1.2.	Volet 2 : Développement d'un algorithme de localisation	159
2	Perspectives	160
2.1.	Évolution de la solution de capteurs	161
2.2.	Évolution de l'algorithme de localisation	161
3	Contributions et encadrements	162
3.1.	Contributions	162
3.2.	Encadrements	162

| Table des figures

| Liste des tableaux

Annexes

173

A | Les systèmes de localisation par satellites

1	Le système GPS	173
2	Le système GLONASS	174
3	Le système COMPASS	174
4	Le système GALILEO	175

B | Filtre de Kalman

C | Autres concepts élaborés

D | Différentes configurations étudiées pour le placement des caméras 3D

1	2x2 caméras	183
2	3 caméras	184
3	4 caméras	185
4	4 caméras verticales	186

E | Démonstration du calcul de θ

F | Études sur les techniques de localisation par la vision

G | Étude de calibration d'une optique *Fish Eye*

H | Études annexes sur l'apport d'une optique *Fish Eye*

| Bibliographie

Introduction générale

1	Contexte général	1
2	Terra Mobilita	4
3	Objectifs de la thèse	5
3.1.	Rencontre de deux communautés	6
3.2.	Plan du manuscrit	6

1 Contexte général

La numérisation de l'environnement urbain est possible depuis plusieurs années grâce aux systèmes de numérisation statiques. Ce sont actuellement les seuls à atteindre un niveau de précision et de complétude satisfaisants. Cependant, ces systèmes présentent un niveau de productivité faible lorsqu'il s'agit de numériser un quartier voire une ville entière.

Les systèmes mobiles terrestres sont ainsi apparus comme un bon compromis [Schwarz et al., 1993]. Ils permettent de répondre au problème de productivité des systèmes statiques. Le développement des systèmes d'acquisition mobiles évolue à un rythme soutenu. Les progrès dans les domaines informatiques et électroniques permettent de faire évoluer les capteurs, mais également le traitement des données. Du fait de la polyvalence de ces systèmes, ceux-ci peuvent être déployés sur de multiples plateformes mobiles, telles que des véhicules, des 4x4, des chariots, des bateaux, etc. (Figure 1.1).

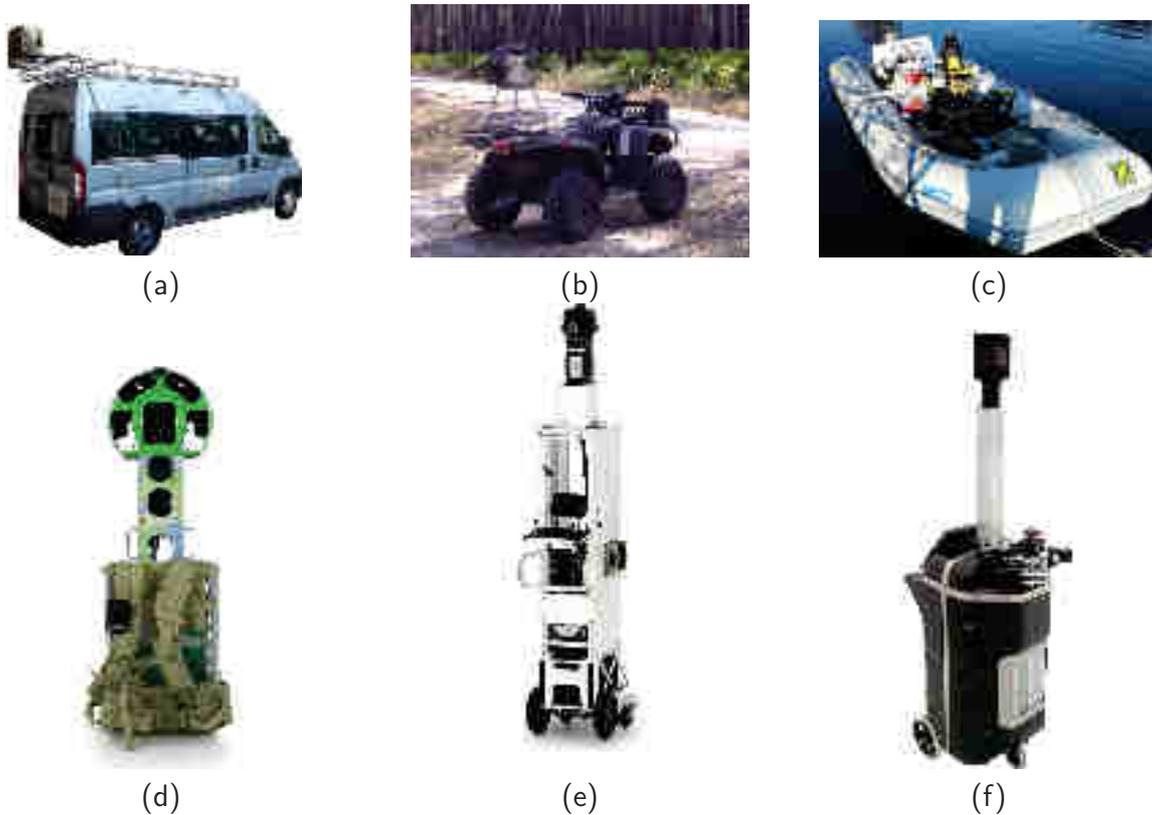


FIGURE 1.1: Exemples de systèmes de cartographie mobile : (a) L3D2 prototype développé au sein du CAOR de MINES ParisTech ; (b) Système mobile de l'université de Floride (FLREC Geomatics) ; (c) Système maritime LandScope ; (d) Google Street View Trekker ; (e) Google Street View Trolley ; (f) Système iMMS de Viametris

Un système d'acquisition mobile se décline en trois composantes principales : le système de localisation, le système de perception, et enfin la plateforme mobile (Figure 1.2).

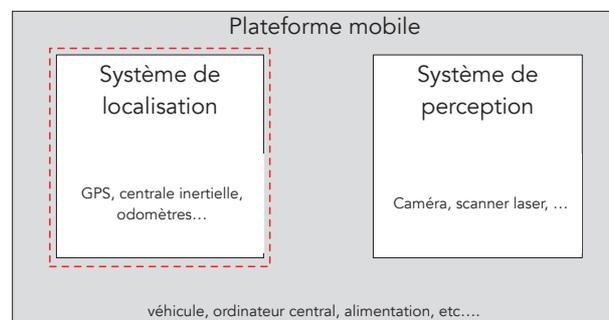


FIGURE 1.2: Composition d'un système de cartographie mobile [Yoo, 2011]

Le système de perception permet de réaliser l'acquisition des données de la scène qui l'entoure. Il est composé d'un ensemble de capteurs imageurs (caméras, scanners laser, etc...). Le système de localisation permet de calculer la position et l'orientation de la plateforme mobile à chaque instant du déplacement de celle-ci. Le travail de cette thèse concerne uniquement la partie localisation des systèmes d'acquisition mobile.

Le relevé de données tridimensionnelles géoréférencées soulève deux problématiques. La première concerne la localisation, la seconde concerne l'intégration entre les capteurs imageurs et les capteurs de localisation. La localisation implique la connaissance de la position et de l'orientation de la plateforme mobile d'acquisition à chaque instant du trajet parcouru. Évidemment, la précision de cette localisation impacte directement la qualité de l'acquisition réalisée.

Au sein de la communauté géomatique, la tâche de localisation est généralement réalisée grâce à plusieurs capteurs : le GNSS (*Global Navigation Satellite System*), la centrale inertielle, et les roues odométriques. La fusion de leurs mesures permet de fournir l'information de localisation. Le GNSS fournit une indication fondamentale et directe de positionnement absolu par rapport au repère terrestre. Cependant, en environnement urbain, nous sommes fréquemment exposés à des zones dites de "canyon urbain", il est alors habituel d'être confronté à une coupure franche de la mesure par GNSS. Une solution pour pallier ce problème consiste à équiper la plateforme d'acquisition mobile d'une centrale inertielle très haut de gamme. Cependant, les résultats obtenus ne sont pas à la hauteur des besoins en terme de précision, et le surcoût engendré limite la diffusion des systèmes d'acquisition mobiles.

La problématique de localisation est également un axe de recherche largement étudié par la communauté robotique. Elle intervient dans différentes circonstances. En effet, il est nécessaire de connaître sa position avant de planifier un déplacement, il est également important de connaître sa position par rapport aux obstacles, ou encore de connaître sa position afin de rallier un point. Les applications en robotique mobile sont nombreuses mais les plus importantes et fondatrices de la discipline, sont le déplacement et l'exploration par un robot de zones inaccessibles aux êtres humains. On peut notamment citer les robots utilisés dans le cadre de l'inspection de centrales nucléaires [Iborra et al., 2003] ou encore les rovers planétaires utilisés dans le cadre de l'exploration spatiale [Maimone et al., 2007]. On peut également citer les travaux prometteurs entrepris par l'entreprise Google, associée à Sebastien Thrun et Chris Urmson [Guizzo, 2011], concernant la navigation de véhicules autonomes en environnement urbain. Le système de pilotage automatique est déjà installé sur huit véhicules, qui ont parcouru quelques 800 000 km en Californie, sans provoquer d'accidents. À la vue des résultats obtenus, on pourrait penser que la problématique de localisation d'un mobile en environnement urbain est résolue. Cependant, il est important de noter que les capteurs utilisés par Thrun et Urmson sont principalement des capteurs laser haut de gamme, dont la commercialisation à grande échelle n'est pas envisageable.

Aussi bien en robotique qu'en géomatique, on remarque que les solutions trouvées pour répondre à la problématique de localisation sont des solutions onéreuses et non déployables à grande échelle. Il est alors nécessaire de développer des méthodes de localisation basées sur des capteurs à bas coût. Les systèmes de vision constituent une bonne option en raison de leur prix, mais également de leur facilité d'approvisionnement. Nous nous intéressons plus particulièrement à un nouveau type de capteur apparu il y a une dizaine d'années : les caméras 3D. Elles permettent l'acquisition de données tridimensionnelles ainsi que l'acquisition de données colorimétriques. La rapidité d'acquisition d'un nuage de points et la mobilité de ces capteurs en font des outils très prometteurs. De plus, leur prix par rapport aux systèmes d'acquisition disponibles actuellement les rend attrayantes auprès du grand public, mais également auprès de la communauté scientifique. On observe un engouement pour ces équipements, depuis la commercialisation de la Kinect par Microsoft en 2010 et de la Xtion Pro par Asus en 2011, toutes deux basées sur la technologie développée par PrimeSense [Arieli et al., 2010]. Compte tenu de l'ensemble de ces avantages, les caméras 3D seront un élément clé du système développé au cours de cette thèse.

2 Terra Mobilita

Notre travail s'inscrit dans le cadre du projet Terra Mobilita par l'intermédiaire de la société Trimble qui a financé ces travaux. Le but de ce projet est de mettre au point de nouveaux processus automatiques de création et de mise à jour de modèles tridimensionnels de l'espace public, afin d'en obtenir différentes informations. L'obtention de ces modèles est réalisée grâce aux acquisitions obtenues à partir de systèmes mobiles de cartographie. Ceux-ci sont équipés de matériels permettant des acquisitions lasergrammétriques ou photogrammétriques. L'exploitation des modèles tridimensionnels obtenus permet alors de nombreuses applications, parmi lesquelles nous pouvons citer :

- Étude de l'encombrement urbain (stationnement, poubelles, etc.)
- Diagnostic de dégradation et d'entretien des voiries,
- Plans de Déplacement Urbain (PDU) pour les circulations douces (piétons, vélos, rollers, poussettes, fauteuils roulants, etc.),
- Diagnostics d'accessibilité de la voirie pour les personnes à mobilité réduite (PMR).

L'objectif fixé par Terra Mobilita est de fournir des solutions pour l'acquisition, le traitement de données 3D et la production de cartes, plus automatisées et moins coûteuses que les méthodes actuelles. De plus, il est attendu la possibilité de mise à jour des modèles tridimensionnels obtenus de façon fréquente (mensuelle par exemple).

Le projet Terra Mobilita est réparti en différents sous-projets. Ceux-ci sont classés en trois catégories, que nous présentons dans la Figure 1.3. Les travaux réalisés dans le cadre de cette thèse interviennent dans le sous-projet 2 (cadre vert). Celui-ci a pour tâche la mise au point de techniques de numérisation 3D de qualité topographique.

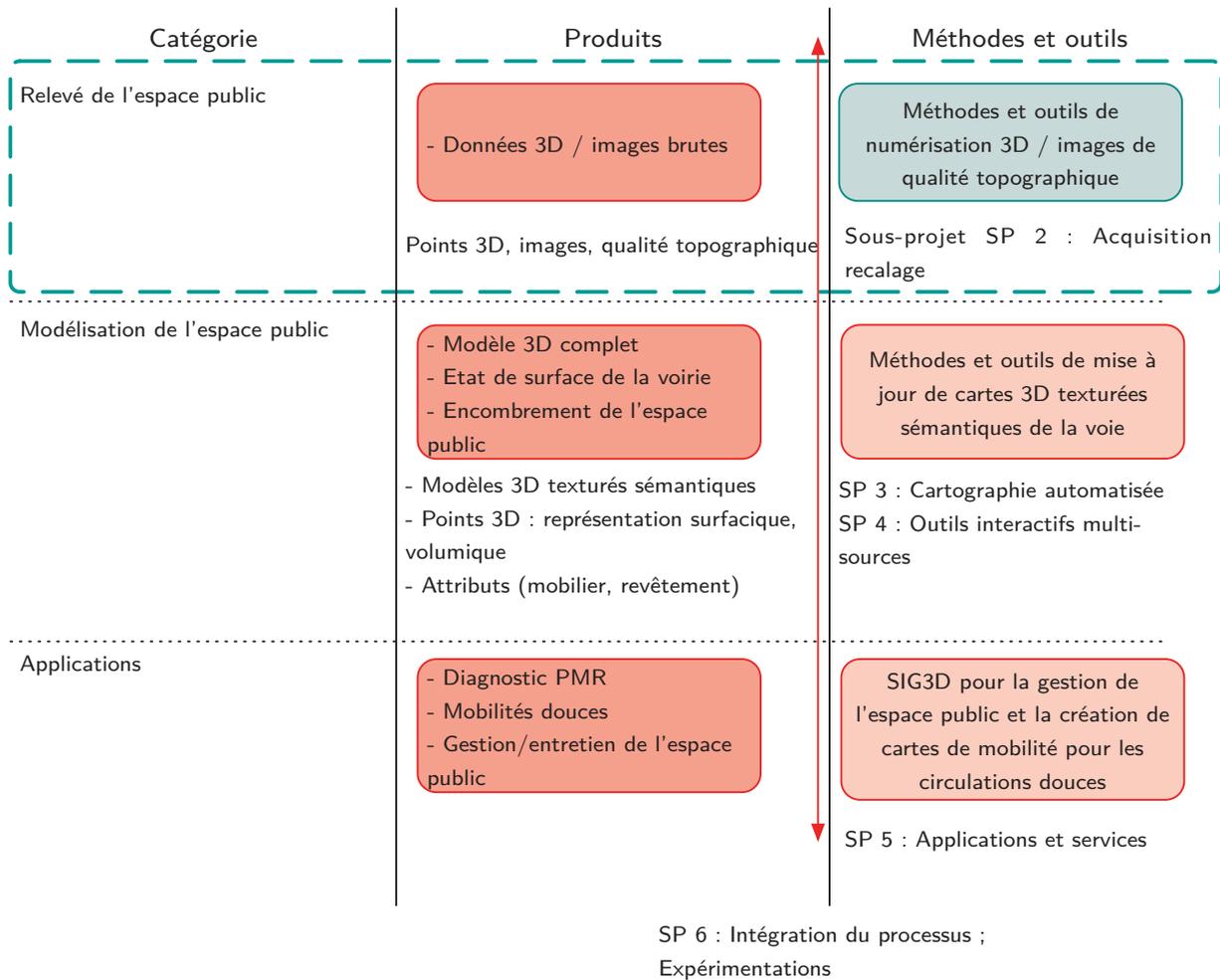


FIGURE 1.3 : Diagramme de décomposition fonctionnelle du projet Terra Mobilita. (Source : Site Terra Mobilita)

3 Objectifs de la thèse

L'objectif de ce travail de thèse est donc de définir un nouveau système de capteurs à bas coût ayant la capacité de localisation en environnement urbain. Le premier problème posé par le sujet de cette thèse consiste en l'élaboration d'un prototype matériel, fonctionnel, qui soit en capacité de fournir l'ensemble des données nécessaires au calcul de la trajectoire du mobile. Le second problème posé par le sujet consiste à définir une méthodologie de localisation permettant le positionnement précis (centimétrique) et en temps réel, d'un mobile, à l'aide des données fournies par le prototype élaboré.

3.1. Rencontre de deux communautés

Comme nous avons pu le voir précédemment, l'opération de localisation tient une place importante dans différentes tâches, aussi bien en robotique qu'en géomatique. Ainsi, l'objectif des deux communautés pour répondre au problème de localisation est identique : le développement d'un système de localisation qui soit suffisamment précis et non limité par les capteurs utilisés. Dans ce contexte notre approche a consisté à tirer profit des connaissances des deux communautés, afin d'apporter une solution satisfaisante au problème de localisation en environnement urbain.

L'utilisation de la vision a toujours suscité un vif intérêt dans la communauté robotique. Ainsi, l'utilisation des capteurs de vision pour répondre à la problématique de localisation est historique et classique.

Les capteurs généralement utilisés par la communauté géomatique sont employés à une autre fin. En effet, les capteurs de vision ont tout d'abord été utilisés dans un objectif de restitution et non de localisation. Les capteurs exploités pour la localisation sont de type GNSS, auxquels il est courant de coupler une centrale inertielle et des odomètres, et ceci afin d'assurer une continuité dans la localisation ainsi qu'un complément d'information (Figure 1.4).

Cependant, les techniques utilisées en robotique et en géomatique montrent des limitations, aussi bien en termes de diffusion qu'en termes de précision. Depuis quelques années, les deux communautés ont tendance à se rapprocher afin de tirer profit des avantages de l'une et de l'autre. Aussi, il est de plus en plus courant de trouver des applications en robotique exploitant les approches développées par la communauté géomatique et inversement. Les deux communautés améliorent ainsi les performances de localisation quels que soient l'environnement et les capteurs utilisés [Kim & Sukkarieh, 2005], [Dusha & Mejias, 2010].

Dans le cadre de cette thèse, nous nous proposons de réaliser un nouveau système de localisation destiné à être utilisé en environnement urbain. Celui-ci permettra de déterminer les positions et orientations relatives à partir d'une position de départ, tout en répondant aux exigences de précision nécessaires à la réalisation future de modèles tridimensionnels.

3.2. Plan du manuscrit

Avant de présenter le système proposé pour répondre aux différentes problématiques posées par cette thèse, le chapitre deux revient tout d'abord sur les différents types d'environnements. En effet, leurs spécificités vont avoir un impact direct sur les techniques de localisation qu'il est possible de déployer en leurs seins. Nous aborderons ensuite la localisation en environnement urbain, cadre d'évolution induit par Terra Mobilita. Finalement, nous présenterons les caméras 3D qui sont au cœur du prototype développé au cours de ces travaux.

Le chapitre trois portera sur l'élaboration et la conception de notre système. Le choix des capteurs, leur nombre, ainsi que leurs positions et orientations sur le mobile seront discutés et justifiés.

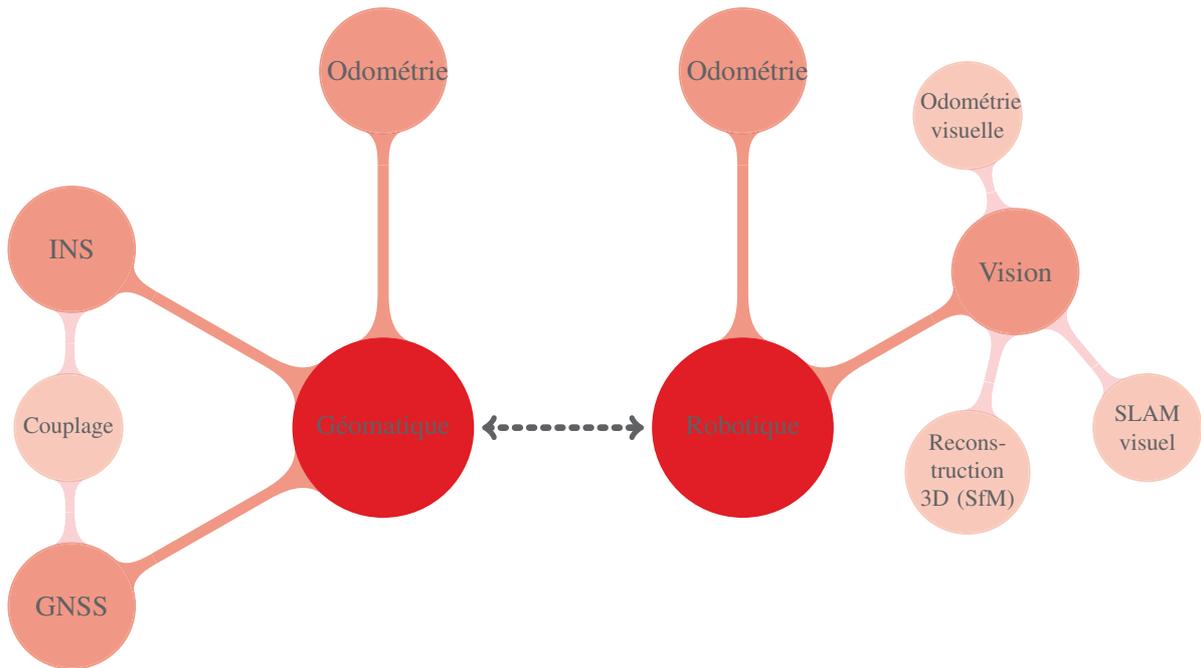


FIGURE 1.4 : Approches développées pour la tâche de localisation, positionnées selon leur appartenance à l'une ou l'autre des communautés

Sur la base du système conçu pour répondre à la problématique posée par le sujet de thèse, nous avons élaboré un algorithme de localisation qui a pour originalité d'être basé sur l'exploitation d'orthoimages obtenues à partir des nuages de points fournis par les caméras 3D. La formation de ces images ainsi que l'intérêt de leur utilisation seront présentés dans le chapitre quatre. La méthodologie développée est issue de l'odométrie visuelle, c'est pourquoi un état de l'art y sera consacré dans cette partie. Une évaluation de détecteurs de points d'intérêts a été menée, et sera également présentée.

Une des difficultés dans la réalisation d'un système de localisation est l'évaluation de la précision de la localisation obtenue à partir des données fournies par celui-ci. Il est nécessaire de mettre en place un système de suivi secondaire aussi précis, voire plus précis que l'approche que nous souhaitons mettre en place. Différentes expérimentations ont été menées. Dans le chapitre cinq nous présentons les plus significatives d'entre elles pour la thèse, et les conditions dans lesquelles elles ont été réalisées.

Enfin, le chapitre six présentera les résultats obtenus, ainsi qu'une discussion critique de ceux-ci. Finalement des perspectives d'amélioration du système seront proposées.

Chapitre

2

État de l'art

1	Introduction	9
2	Les types d'environnements	10
	2.1. Environnement intérieur/extérieur	10
	2.2. Environnement statique/dynamique	11
3	La localisation en environnement urbain	12
	3.1. La localisation par satellites	12
	3.2. La localisation inertielle	18
	3.3. La localisation odométrique	22
	3.4. Fusion de données	27
	3.5. La localisation à partir de capteurs de vision	31
4	Etat de l'art sur les caméras 3D	33
	4.1. Principe de la triangulation	33
	4.2. Principe du "temps de vol"	34

1 Introduction

Pour localiser un véhicule en environnement urbain, le capteur par excellence est le GNSS. Celui-ci fournit la position du récepteur que l'on souhaite localiser dans un référentiel national géographique de précision. Ce capteur est aujourd'hui très répandu et est employé dans de nombreux domaines. La configuration de l'environnement dans lequel on souhaite réaliser la tâche de localisation va avoir une influence sur la précision de la position fournie par le capteur GNSS.

Afin de répondre aux problèmes de localisation en environnement urbain, il est fréquent d'ajouter au GNSS un ensemble de capteurs (centrale inertielle, odomètres...). La fusion des données fournies par l'ensemble de ces capteurs permet de pallier les dérives qui peuvent intervenir lors de la tâche de localisation en environnement urbain. Cependant, le coût de ces systèmes limite largement leur diffusion. Il convient alors de se tourner vers d'autres types de capteurs et d'approches alternatives.

Ce chapitre a pour but de faire un état de l'art des différentes approches de localisation. Dans un premier temps, nous définirons les types d'environnement dans lesquels la tâche de localisation peut être réalisée. Une fois cette notion définie, nous aborderons les principes du positionnement par satellite et les limites posées par l'environnement dans lequel nous évoluons. Nous présenterons également les différents capteurs aidant à la tâche de localisation lorsque le GNSS est inopérant. La fusion de l'ensemble des informations fournies par ces capteurs sera également présentée. Enfin, nous aborderons la localisation basée sur les capteurs de vision. Finalement, nous ferons une présentation des caméras 3D, ces capteurs constituant le pilier du système développé au cours de cette thèse.

2 Les types d'environnements

De manière générale, la localisation d'un objet consiste à déterminer sa position géographique. Localiser un mobile consiste donc à le situer par rapport à son environnement. Les environnements peuvent être divers et variés, posant des problématiques différentes. En robotique mobile, on peut distinguer les environnements de plusieurs façons : intérieur/extérieur ou statique/dynamique, par exemple.

2.1. Environnement intérieur/extérieur

La localisation en intérieur s'intéresse principalement à la localisation de robots à l'intérieur de bâtiments. Ces environnements sont structurés, souvent de faible taille et n'entraînent généralement pas de problèmes de franchissement d'obstacles. Les sonars, télémètres lasers et caméras sont les capteurs les plus utilisés dans ce type d'environnement. Les odomètres et les centrales inertielles sont également particulièrement indiqués. Ces capteurs sont connus pour souffrir d'une dérive importante. Cependant, les surfaces parcourues en intérieur sont généralement planes, les distances de déplacement faibles et les vitesses de déplacement réduites. Ainsi la dérive de ces capteurs se trouve limitée, grâce à ces conditions favorables. Naturellement, l'utilisation du GNSS n'est pas envisageable dans ce genre d'environnement.

Les environnements d'intérieur sont souvent considérés comme plans et horizontaux. Cette hypothèse de sol plan est communément admise et elle autorise, lorsque la configuration cinématique du mobile le permet, de réduire le nombre de paramètres de déplacement à estimer [Scaramuzza et al., 2009]. Par ailleurs, la structuration des environnements d'intérieur est à l'origine de nombreuses techniques de localisation basées sur la détection des contours à partir des données obtenues par un sonar [Leonard & Durrant-Whyte, 1991].

Le développement des robots mobiles d'extérieur a connu une accélération à la fin des années 1980. L'exploration spatiale a largement contribué à cet essor, notamment avec les financements de la NASA. Nous pouvons également citer les travaux entrepris par le laboratoire américain DARPA, ayant obtenu des subventions pour plusieurs projets. Nous trouvons ainsi des projets tels que le projet "Urban Challenge", apparu en 2007, dont l'objectif est de permettre à un véhicule de suivre

une route de manière autonome ou encore le projet "Unmanned Ground Vehicle", qui consiste à explorer un terrain non structuré inconnu. En conséquence, les autres laboratoires de robotique se sont rapidement intéressés à la problématique de localisation en environnement extérieur. Les environnements extérieurs sont des environnements complexes, souvent non structurés et de grande taille. La localisation par GNSS est une technique très répandue dans ce cadre. La fusion de données satellitaires avec des centrales inertielle et des odomètres a favorisé le développement de mobiles d'extérieur ayant la capacité de parcourir des centaines de kilomètres sur des autoroutes ou encore de réaliser des tâches de navigation dans des environnements ouverts non structurés, tels que le milieu agricole. La localisation en environnement extérieur est ainsi répandue dans de nombreux domaines d'application : l'exploration spatiale, l'exploration sous-marine, ou encore la cartographie. Malgré leur diversité, ces exemples possèdent tous un point commun : l'utilisation de la vision. En effet, les informations fournies par les caméras sont riches, et ont contribué à rendre la localisation plus robuste. Nous reviendrons sur ces aspects par la suite.

2.2. Environnement statique/dynamique

Dans le cadre de notre étude, il est également important de classifier l'environnement comme étant statique ou dynamique. On dit d'un environnement qu'il est dynamique lorsque son état change au cours du temps, ainsi une rue avec des voitures est considérée comme un environnement dynamique. Ce genre d'environnement peut présenter une difficulté dans la tâche de localisation. En effet, si un mobile cherche à se localiser sur une carte dont il a connaissance, la présence d'obstacles dynamiques peut induire une erreur dans la localisation. De même, un mobile équipé d'un capteur de vision permettant de se localiser grâce à des amers, éprouvera des difficultés de localisation, s'il est confronté à des éléments mouvants. [Hahnel et al., 2003] et [Wolf & Sukhatme, 2005] ont mené des travaux spécifiques sur la localisation en environnements dynamiques.

A l'opposé de l'environnement dynamique, nous trouvons l'environnement statique, par exemple un bâtiment vide. Ce contexte de travail est plus simple à considérer, notamment en ce qui concerne les objets mobiles, qui peuvent être problématiques dans la tâche de localisation. Même si ce type d'environnement est moins réaliste, il constitue un contexte théorique intéressant et a fait l'objet de nombreuses études [Thrun, 2002].

La problématique posée par le sujet de cette thèse définit clairement l'environnement dans lequel nous évoluons : l'environnement urbain. Il s'agit d'un environnement extérieur et dynamique. Ces spécificités ont donc dû être prises en compte dans l'approche que nous avons développée.

3 La localisation en environnement urbain

3.1. La localisation par satellites



Le positionnement à partir de satellites offre la possibilité à l'utilisateur de connaître sa position à n'importe quel moment, quelle que soit sa position sur Terre, et de façon absolue. Depuis quelques années, il a fait son apparition dans notre vie quotidienne : véhicule, téléphonie mobile, sport, etc. Les systèmes de localisation et de navigation par satellites sont fondés sur la détermination de la position d'un récepteur à partir de la méthode de trilatération. Cette méthode nécessite la connaissance de la position d'au moins quatre satellites et la mesure de la pseudo-distance entre le récepteur (position inconnue) et les différents satellites (position connue).

3.1.1. Le calcul de la position

Le positionnement par satellites repose principalement sur deux méthodes. La première, appelée positionnement absolu, utilise le principe de trilatération à partir des informations émises par les satellites visibles. La seconde, appelée positionnement relatif par satellites, utilise une station de base dont la position est parfaitement connue. Cette station transmet aux récepteurs les corrections des erreurs de trilatération, permettant ainsi un positionnement précis.

Le positionnement absolu obéit au principe de la trilatération [Duquenne et al., 2005]. La trilatération consiste à utiliser uniquement les distances afin de déterminer la position relative d'un point, contrairement à la triangulation qui utilise les angles pour positionner un point. Les ondes radio émises par chaque satellite sont précisément datées. Les signaux émis sont codés selon des fréquences pseudo-aléatoires. Le récepteur génère, en même temps que le satellite, une réplique de la séquence du code du satellite. Le décalage que doit subir la réplique afin de coïncider avec le code reçu est égal au temps nécessaire pris par le signal pour parcourir la distance satellite-récepteur. À partir de cette différence de temps, il est alors possible de déterminer la pseudo-distance entre le satellite et le récepteur.

La Figure 2.1 illustre le principe de la trilatération. Trois émetteurs sont représentés sous la forme de satellites. Leurs positions respectives sont connues du récepteur représenté par la voiture sur notre illustration. Le récepteur connaît les distances le séparant des émetteurs (pd_1, pd_2, pd_3), ainsi il est en mesure de savoir qu'il se situe quelque part sur les cercles gris, vert et rouge. Le récepteur peut calculer les coordonnées $x_{rec}, y_{rec}, z_{rec}$ de sa position à partir des pseudos-distances selon les équations en (2.1).

$$\begin{aligned}
 pd_1 &= \sqrt{(x_{rec} - x_{sat1})^2 + (y_{rec} - y_{sat1})^2 + (z_{rec} - z_{sat1})^2} \\
 pd_2 &= \sqrt{(x_{rec} - x_{sat2})^2 + (y_{rec} - y_{sat2})^2 + (z_{rec} - z_{sat2})^2} \\
 pd_3 &= \sqrt{(x_{rec} - x_{sat3})^2 + (y_{rec} - y_{sat3})^2 + (z_{rec} - z_{sat3})^2} \\
 pd_4 &= \sqrt{(x_{rec} - x_{sat4})^2 + (y_{rec} - y_{sat4})^2 + (z_{rec} - z_{sat4})^2}
 \end{aligned} \tag{2.1}$$

En théorie, afin de déterminer une position, il est nécessaire de connaître trois distances. Le récepteur doit donc pouvoir recevoir les signaux de trois satellites. Dans la pratique, les erreurs de synchronisation d'horloge entre les satellites et le récepteur ne permettent pas d'obtenir un positionnement suffisamment précis. La connaissance d'une quatrième pseudo-distance permet de résoudre ce problème et de fournir les coordonnées tridimensionnelles du récepteur, ainsi qu'une estimation de l'erreur d'horloge du récepteur.

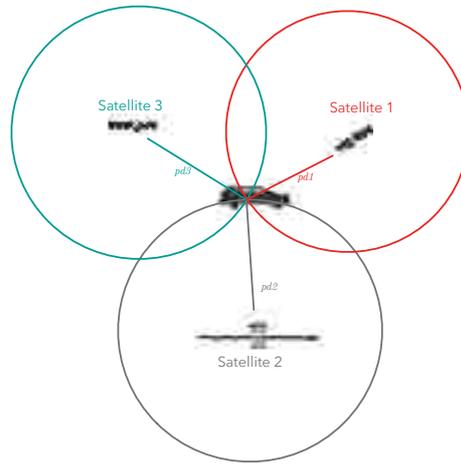


FIGURE 2.1 : Principe de la trilatération où pd_1 , pd_2 et pd_3 représentent les distances séparant le récepteur (la voiture) des émetteurs (satellites)

La précision du calcul de position dépend de la prise en compte et du traitement des erreurs suivantes ([Arpin, 2003], [Olynik, 2002]) :

- l'erreur d'éphéméride E_{orb} causée par les écarts entre les paramètres précis des satellites et les paramètres transmis au récepteur,
- l'erreur d'horloge du récepteur ϵ_{rec} et des satellites ϵ_{sat} due à une mauvaise synchronisation et à l'inexactitude de la mesure de réception du signal par le récepteur,
- les erreurs atmosphériques E_{atm} , représentant des retards dus à la traversée de l'ionosphère et de la troposphère,
- le bruit du récepteur B_{rec} , dépendant de la qualité du matériel électronique utilisé lors de la fabrication de celui-ci,
- l'erreur représentant les multi-trajets, E_{multi} , causée par les réflexions ou les diffractions.

Ainsi, la pseudo-distance ρ_i mesurée est différente de la mesure géométrique pd_i (satellite-récepteur). ρ_i est alors obtenue à partir de pd_i , à laquelle il est nécessaire d'ajouter l'ensemble des erreurs listées auparavant :

$$\rho_i = pd_i + c \cdot (\epsilon_{sat} - \epsilon_{rec}) + E_{orb} + E_{atm} + B_{rec} + E_{multi} \quad (2.2)$$

avec c la vitesse de propagation des ondes satellites.

Les erreurs atmosphériques peuvent être corrigées en utilisant des modèles basés sur les données météorologiques ou à l'aide de l'exploitation de deux fréquences [Ioannides & Strangeways, 2002]. Les erreurs d'éphémérides peuvent être corrigées en faisant des mesures précises des positions des satellites. Celles-ci sont ensuite actualisées par les stations de contrôle. Les erreurs liées aux horloges sont faibles, grâce à l'utilisation de plusieurs horloges par satellites. Les erreurs subsistantes sont principalement celles causées par les multi-trajets et par l'environnement dans lequel le récepteur évolue. Nous reviendrons sur ces limitations par la suite.

Dans le cas du positionnement relatif par satellites, le récepteur est équipé d'un boîtier supplémentaire. Celui-ci permet de recevoir des corrections fournies par des stations DGNSS (GNSS différentiel). Le principe de fonctionnement reste le même que pour le positionnement absolu. Les mesures effectuées en un lieu, dont la position est bien connue, sont comparées aux valeurs théoriques. La différence obtenue permet de fournir des corrections transmises au récepteur, qui les intègre dans le calcul de sa position. Ainsi, le positionnement est considérablement amélioré. Cependant un tel système de positionnement présente un inconvénient majeur : sa portée, qui n'excède pas une centaine de kilomètres autour des stations d'émissions.

3.1.2. Les limitations de la localisation par satellites en environnement urbain

Nous avons pu observer que la précision du positionnement par satellite va directement dépendre de la qualité et de la précision des signaux émis par les satellites. La mauvaise qualité des signaux peut avoir des origines diverses, telles que l'environnement dans lequel se situe le récepteur. Par exemple, le milieu forestier est problématique, les feuilles atténuant le signal GNSS. De même, un émetteur TV est générateur de bruit et rend ainsi la mesure difficile. Les environnements urbains ne sont pas moins problématiques : la typologie urbaine (densité, hauteur de bâtiments, canyons urbains) entraîne parfois un manque de visibilité, et les matériaux ainsi que le mobilier urbain les composant sont propices aux phénomènes de multi-trajets.

a. Masquage des satellites

L'un des problèmes majeurs de l'utilisation du GNSS en milieu urbain est le problème des masques. Celui-ci est dû aux différents obstacles (bâtiments, arbres, etc...) que peuvent rencontrer les signaux émis par les satellites. On désigne par "masque" un obstacle physique qui se trouve placé entre le récepteur et les satellites (Figure 2.2).

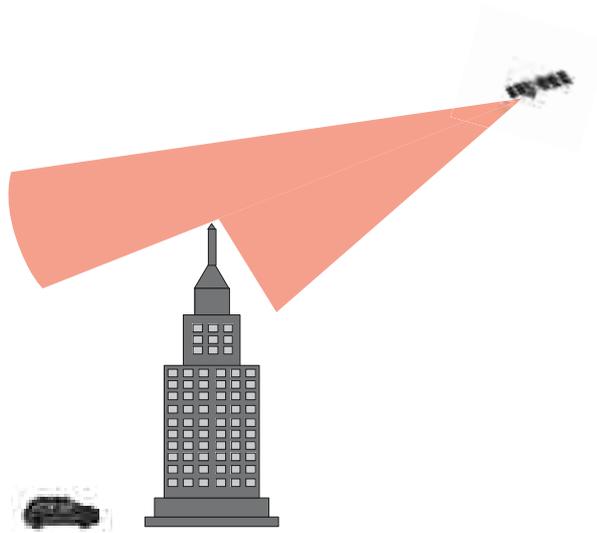


FIGURE 2.2 : Phénomène du masquage des satellites pour un récepteur (véhicule).

Dans [Malicorne et al., 2001], une classification du milieu urbain est proposée selon 4 catégories établies en fonction de la hauteur des bâtiments : milieux urbains, milieux résidentiels, milieux industriels, et milieux urbains denses (Tableau 2.1) .

	Hauteur moyenne des bâtiments (m)	Écart-type de la hauteur (m)	Largeur des rues (m)
Urbain	25	6	20-25
Résidentiel	6	3	7
Industriel	8	2	19
Urbain dense	30	10	45-60

TABLE 2.1 : Catégorie des milieux urbains [Malicorne et al., 2001]

Sur la base de cette classification, [Nahimana, 2009] propose également une classification des milieux urbains en considérant l'angle d'élévation des constructions. Cet angle est obtenu à partir de la hauteur des bâtiments et de la largeur des rues (Figure 2.3 et Tableau 2.2).

Angle d'élévation des masques	Milieu
$< 10^\circ$	Dégagé (campagne)
$10^\circ - 40^\circ$	Visibilité réduite (zone industrielle)
$40^\circ - 60^\circ$	Visibilité très réduite
$> 60^\circ$	Mauvaise visibilité (centre ville)

TABLE 2.2: Visibilité en milieux urbains

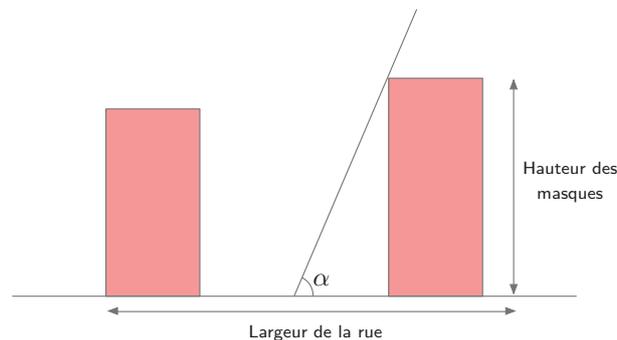


FIGURE 2.3: Mesure de l'angle d'élévation des masques [Nahimana, 2009]

Les zones urbaines les plus contraintes sont nommées "canyons urbains". Cette configuration se retrouve classiquement dans le cas de petites rues étroites, ainsi que dans les zones urbaines du type du quartier de "La Défense" à Paris ; les rues y sont larges mais les immeubles hauts, leur hauteur limitant très fortement la portion de ciel visible (Figure 2.4).

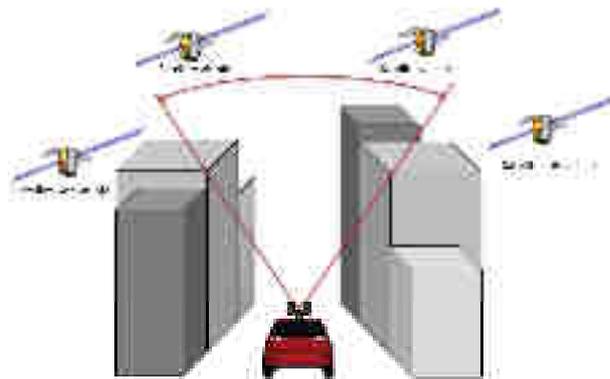


FIGURE 2.4: Phénomène de canyon urbain

Le masquage peut également être partiel, dans ce cas, la puissance du signal émis est atténuée. Le récepteur a alors plus de difficultés à détecter le signal, et la qualité du positionnement sera directement impactée.

b. Le multi-trajet

Une autre problématique rencontrée en milieu urbain est celle du multi-trajet. On parle de multi-trajet lorsqu'il existe, à proximité de l'antenne d'un récepteur, des éléments susceptibles de réfléchir et de diffracter les signaux émis par les satellites. Le multi-trajet a pour effet de rallonger le chemin parcouru par le signal et engendre donc une erreur de positionnement. Selon certaines études, on observe que ces réflexions peuvent engendrer des erreurs de positionnement de plusieurs centimètres. La Figure 2.5 illustre ce phénomène. En vert, nous pouvons observer un signal qui suit un trajet direct entre le satellite et le récepteur, aussi appelé LOS (*Line of Sight*), alors que le signal en rouge représente un signal ayant subi une réflexion sur un bâtiment.

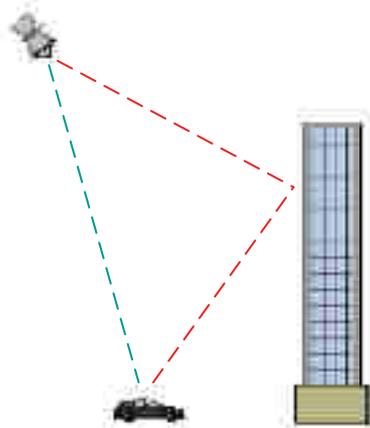


FIGURE 2.5: Phénomène de multi-trajet(en rouge)

3.1.3. Conclusions sur le positionnement par satellites

La localisation par satellites est en plein essor. L'apparition de nombreux systèmes de localisation par satellites tel que GALILEO va permettre une couverture du globe terrestre de plus en plus complète (Annexe A). L'avantage principal de cette technique est sa capacité à fournir une information de positionnement directe et absolue (par rapport au repère terrestre). Cependant, en environnement urbain, les conditions de visibilité des satellites sont souvent dégradées du fait de la présence de bâtiments et de tunnels. Ainsi les phénomènes de masques ou de multi-trajets peuvent limiter la disponibilité des signaux et donc la précision du positionnement. De plus, l'information fournie par le GNSS n'est qu'une information de position. Pour obtenir d'autres informations telles que la vitesse ou l'angle de lacet, il est nécessaire d'exploiter plusieurs mesures successives. Pour pallier ces inconvénients, il est essentiel d'utiliser d'autres types de capteurs, tels que les centrales inertielles ou les odomètres. Ceux-ci sont présentés ci-après.

3.2. La localisation inertielle



Comme nous avons pu le voir auparavant, la localisation par satellites en environnement urbain peut être confrontée à différents problèmes. Une des solutions consiste à utiliser d'autres capteurs afin de prendre le relai du GNSS lorsqu'il est inopérant. Parmi ces capteurs, on trouve les centrales inertielle.

La localisation inertielle repose sur les mesures fournies par des accéléromètres et des gyroscopes. Les capteurs sont contenus dans ce que l'on appelle une IMU (*Inertial Measurement Unit*) ou centrale inertielle. L'IMU est un dispositif composé de trois accéléromètres et de trois gyroscopes [Kelly, 1994], [Stovall, 1997], [Titterton & Weston, 2004]. Les données fournies par ces capteurs sont employées pour estimer la position et l'orientation d'un objet relativement à un point de départ, à une orientation et à une vitesse connus. Afin de réaliser un système de localisation ou de navigation inertielle nommé INS (*Inertial Navigation System*), il est nécessaire d'associer à l'IMU un calculateur de type intégrateur.

La navigation inertielle repose sur les lois de la mécanique classique, telles qu'elles ont été décrites par Isaac Newton. Le mouvement d'un corps est uniformément rectiligne en ligne droite, à moins qu'une force ne lui soit appliquée, qui va produire une accélération proportionnelle au poids de ce corps. Ainsi, si l'accélération est mesurable, il est possible de calculer la variation de la vitesse v et la position p en effectuant des intégrations mathématiques successives de l'accélération a . Nous avons alors :

$$\text{vitesse } v = \int a \, dt \quad (2.3)$$

$$\text{position } p = \int v \, dt \quad (2.4)$$

$$\text{orientation } \Omega = \int \omega \, dt \quad (2.5)$$

où Ω est l'orientation et ω la vitesse angulaire du corps en mouvement.

Pour obtenir une information de position en trois dimensions, la centrale inertielle doit pouvoir mesurer les accélérations linéaires selon chaque axe. Il est également nécessaire d'utiliser trois gyroscopes pour mesurer les accélérations angulaires afin d'obtenir l'orientation dans l'espace.

3.2.1. Principe de fonctionnement

Parmi les INS, nous pouvons citer deux architectures principales : l'architecture à plateforme stabilisée (*Gimballed INS*) ou l'architecture à composants liés (*Strapdown*). La principale différence entre ces deux catégories repose sur le repère de référence dans lequel les gyroscopes et les accéléromètres de l'INS fonctionnent [Kelly, 2013].

a. Les systèmes à plateformes stabilisées

Ces systèmes utilisent des gyroscopes mécaniques. Leur attitude est conservée fixe dans le temps, en d'autres termes les capteurs inertiels sont montés sur une plateforme qui est isolée de tout mouvement de rotation externe. Pour conserver une attitude fixe dans le temps, il existe deux solutions : la première repose sur l'effet gyroscopique d'une masse en mouvement, celle-ci va permettre le maintien d'une direction fixe, dans la mesure où le système mécanique des cardans lui laisse la liberté de mouvement. La seconde solution consiste à compenser les mouvements de rotation, détectés par les gyroscopes, par une rotation inverse effectuée par un moteur. Les accéléromètres montés sur la plateforme sont utilisés pour mesurer la force spécifique sur chaque axe de navigation. Ces mesures sont alors exploitées pour déterminer la position de l'objet en mouvement. Ces systèmes mécaniques sont complexes, coûteux et encombrants, cependant, ils sont très précis.

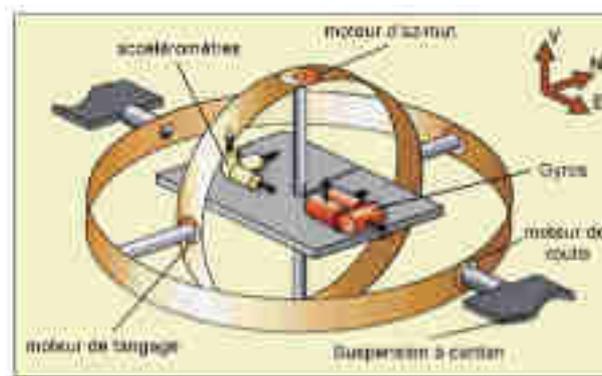


FIGURE 2.6 : Système à plateformes stabilisées [King, 1998]

b. Systèmes à plateformes liées

Les progrès en électronique et en informatique ont permis l'apparition d'un nouveau type d'architecture : les systèmes à plateformes liées. Les gyroscopes et les accéléromètres sont fixés de façon rigide à l'objet en mouvement. Ainsi, les mesures sont effectuées dans le repère mobile (objet) plutôt que dans le repère global (inertiel) (Figure 2.7). Dans ces systèmes, les capteurs mesurent la dynamique du véhicule de façon à ce que les mesures et l'état de l'objet en déplacement soient connus en continu. La charge de calcul est donc supérieure comparativement aux systèmes à plateformes stabilisées. Grâce à leur structure simplifiée et miniaturisée, leur diffusion est grande, cependant cette simplification est également synonyme d'une moins bonne précision d'estimation du mouvement.

3.2.2. Les capteurs

Les systèmes de localisation inertielle sont composés de deux types de capteurs : les gyroscopes et les accéléromètres, dont nous allons présenter le principe de fonctionnement.

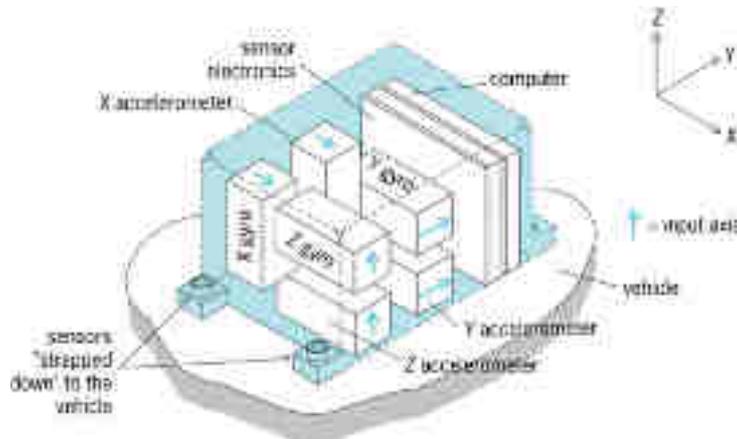


FIGURE 2.7: Système à plateformes liées

a. Le gyroscope

Les détecteurs de vitesse angulaire ou gyroscopes sont les capteurs utilisés pour calculer l'attitude d'un système de navigation inertielle. Ils mesurent la vitesse de rotation suivant l'axe sur lequel ils sont montés. On trouve différents types de gyroscopes.

Le *gyroscope mécanique* est le plus basique des gyroscopes. Il est composé d'un rotor en rotation et exploite le principe de conservation du mouvement angulaire. Cette loi veut qu'en l'absence de couple appliqué à un solide en rotation, celui-ci conserve son axe de rotation invariable. Cette propriété donne lieu à une seconde propriété, nommée *precession*, qui représente la façon dont le gyroscope réagit lorsqu'un couple est appliqué sur son axe de rotation.

L'équation fondamentale décrivant le comportement du gyroscope est :

$$\vec{N} = \frac{d\vec{L}^i}{dt} = \frac{d(I^i\vec{\omega})}{dt} \quad (2.6)$$

où les vecteurs \vec{N} et \vec{L}^i sont respectivement le moment appliqué au gyroscope et le moment cinétique exprimé dans le repère inertielle, I^i son moment d'inertie et $\vec{\omega}$ la vitesse angulaire.

Les *gyroscopes optiques* peuvent être classés selon deux catégories : les gyroscopes à fibre optique (*FOG*) et les gyroscopes à laser (*RLG*). Ces deux types de gyroscopes reposent sur le même principe de mesure basé sur l'effet de Sagnac. Pour mesurer la rotation, deux faisceaux lumineux sont émis dans des directions opposées. Si le capteur subit une rotation, alors le faisceau voyageant dans le sens de la rotation aura un parcours plus long que celui du faisceau voyageant dans le sens de la contre rotation.

Dans le cas des gyroscopes de type *RLG*, les rayons sont dirigés autour d'un chemin fermé utilisant des miroirs. La précision de ce type de gyroscope va dépendre en grande partie de la longueur du parcours de la lumière, or celle-ci est contrainte par la taille du dispositif. Les

gyroscopes à laser sont les plus simples à réaliser, et leur performance est plus que suffisante dans la plupart des cas.

Dans le cas des gyroscopes *FOG*, une diode laser émet un signal optique modulé, celui-ci est divisé dans des directions opposées dans une fibre optique. En sortie de fibre, les deux signaux sont recombinaés. Il est ainsi possible d'en extraire leur différence de phase, qui est proportionnelle à la vitesse angulaire.

Les gyroscopes MEMS : grâce aux progrès techniques de ces dernières années, la technologie MEMS a considérablement évolué. Ainsi nous avons pu observer l'apparition de systèmes mécaniques microscopiques permettant de détecter des mouvements. Le principe de fonctionnement de base de ces capteurs est d'induire une vitesse à une masse dans une direction, et de détecter ensuite les mouvements dans une direction orthogonale ; ceux-ci sont causés par la force de Coriolis lorsque le dispositif subit une rotation. Bien évidemment on ne peut pas déplacer une masse sur une distance indéfinie dans un capteur. Les MEMS sont donc composés d'éléments virants qui mesurent l'effet de Coriolis. Les MEMS présentent de nombreux avantages : leur faible coût, une taille réduite, cependant ils restent moins précis que les gyroscopes optiques [Woodman, 2007].

b. Les accéléromètres

L'ensemble des accéléromètres reposent sur un même principe : il s'agit d'un dispositif capable de mesurer la force appliquée à une masse fixée au boîtier de l'appareil par un dispositif élastique (ressort, électromagnétisme, etc.). Soit une masse m ponctuelle, positionnée en P , on peut alors lui appliquer la seconde loi de Newton (éq. 2.7).

$$m \cdot \vec{a}_P = \sum \vec{F} \quad (2.7)$$

avec :

- \vec{a}_P l'accélération de masse m en P ,
- $\sum \vec{F}$ désigne les forces extérieures exercées sur l'objet (gravité, force de réaction, perturbations,...).

Les accéléromètres sont montés orthogonalement les uns par rapport aux autres dans l'IMU. De façon identique aux gyroscopes MEMS, nous trouvons des accéléromètres MEMS.

3.2.3. Avantages et inconvénients de la localisation inertielle

L'utilisation combinée de ces différents capteurs présente l'avantage d'obtenir une information de localisation complète. En effet, partant d'une position de départ, il est possible de connaître la position, la vitesse, le cap, la rotation, ainsi que l'accélération d'un objet en mouvement. Un autre avantage est l'indépendance des centrales inertielles, puisqu'il n'est pas nécessaire d'avoir recours à un dispositif extérieur pour obtenir la localisation d'un mobile. Enfin, leur cadence de fonctionnement ainsi que leur compacité sont des atouts.

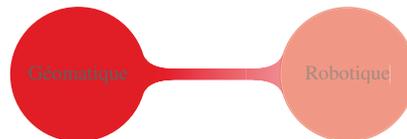
Malgré ces nombreux avantages, les centrales inertielles présentent des inconvénients majeurs. La limitation principale est leur forte dérive au cours du temps. Les capteurs composant les centrales inertielles sont victimes du même type d'erreurs. Nous trouvons dans la littérature des études

sur la modélisation de ces erreurs [Nassar, 2003]. Bien souvent, la précision des données fournies par ces capteurs va dépendre de la qualité de fabrication du capteur, et va donc implicitement dépendre du coût de celui-ci. Dans [Cramer, 1997], l'auteur propose une classification des centrales inertielles en fonction de leur précision et de leur prix (Tableau 2.3). Ce tableau révèle que la niveau de précision susceptible d'être fourni par la centrale inertielle augmente avec son coût.

Précision	Haute	Moyenne	Basse
Position			
1 h	0.3 - 0.5 km	1-3 km	200-300 km
1 min	0.3 - 0.5 m	0.5 - 3 m	30-50 m
1 s	0.01 - 0.02 m	0.03 - 0.1 m	0.3-0.5 m
Attitude			
1 h	3-8 mdeg	0.01-0.05 deg	1-3 deg
1 min	0.3-0.5 mdeg	4-5 mdeg	0.2-0.3 deg
1 s	< 0.3 mdeg	0.3-0.5 mdeg	0.01-0.03 deg
Coût	750 000 \$	100 000 \$	10 000 \$

TABLE 2.3: Classification de la précision des centrales inertielles [Cramer, 1997]

3.3. La localisation odométrique



L'odométrie est l'opération qui consiste à évaluer le déplacement d'un mobile à partir de l'exploitation des données renvoyées par des capteurs placés sur les roues de ce mobile. Ce capteur est appelé odomètre et mesure le nombre de tours de roue effectués à travers des codeurs incrémentaux, pendant une période de temps. La précision de la mesure va dépendre de la résolution du codeur. En plus de la distance parcourue par le mobile, il est aussi possible d'obtenir l'orientation du mobile en étudiant la différence de distance parcourue par la roue droite et la roue gauche.

Le calcul de la position et de l'orientation relatives $(x, y, \theta)^T$ du mobile est réalisé par intégrations successives de ses déplacements élémentaires. La connaissance du diamètre des roues, l'entre-axe des roues et la cinématique du véhicule est requise. Dans cette section, nous présentons deux modèles cinématiques élémentaires employés en localisation par odométrie : le modèle de type char, et le modèle de type tricyle. Avant toute chose, il semble utile de rappeler la notion de contrainte non-holonome, celle-ci intervenant dans la définition des modèles cinématiques.

🔍 **Contrainte non-holonome** *Définition*

La locomotion à l'aide de roues exploite la friction au point de contact entre roue et sol. Pour cela, la nature du contact a une forte influence sur les propriétés du mouvement relatif de la roue par rapport au sol. Dans des conditions idéales de déplacement, on dira qu'il y a roulement sans glissement de la roue sur le sol. Pour vérifier cette hypothèse, il est nécessaire de remplir les conditions suivantes :

- le contact entre la roue et le sol doit être ponctuel,
- les roues doivent être indéformables, et de rayon constant r

On parle de contrainte non-holonome lorsque le système ne peut pas effectuer certains mouvements instantanément. Par exemple, dans le cas de la roue, il ne peut y avoir de translation instantanée parallèle à l'axe de la roue. Ainsi on peut décrire le mouvement d'un mobile de type tricyle comme un ensemble de rotations.

3.3.1. Le modèle cinématique de type char ou unicycle

On désigne par unicycle, un mobile équipé de deux roues indépendantes (Figure 2.8). Un odomètre est installé sur chacune des roues. On considère que le mobile se déplace sans glissement.

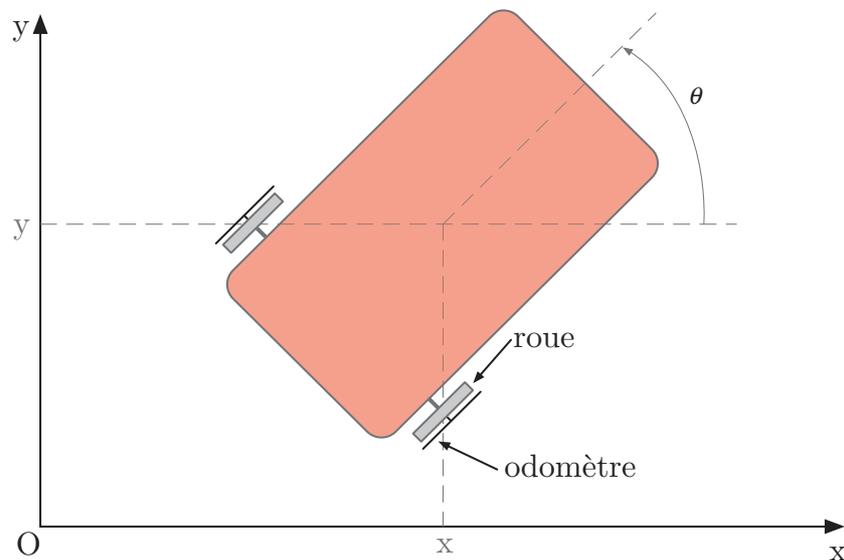


FIGURE 2.8: Modèle cinématique de type unicycle

Pour déterminer la vitesse longitudinale v_{long} , la vitesse latérale v_{lat} et la vitesse de rotation $\dot{\theta}$ du véhicule en fonction des vitesses angulaires ω_g et ω_d , on pose :

$$\begin{aligned} v_{long} &= \frac{r(\omega_d + \omega_g)}{2} \\ v_{lat} &= 0 \\ \dot{\theta} &= \frac{r(\omega_d - \omega_g)}{2l} \end{aligned} \quad (2.8)$$

où r est le diamètre des roues (identique à droite et à gauche) et l la distance entre les roues. Considérant que le déplacement se fait sans glissement, la vitesse latérale est considérée comme nulle. La vitesse longitudinale et la vitesse latérale du véhicule peuvent aussi être exprimées en fonction de x et y [Siegwart et al., 2011] de la façon suivante :

$$\begin{aligned} v_{lat} &= -\dot{x}\sin(\theta) + \dot{y}\cos(\theta) \\ v_{long} &= \dot{x}\cos(\theta) + \dot{y}\sin(\theta) \end{aligned} \quad (2.9)$$

où θ représente l'orientation du véhicule.

Ainsi à partir des équations (2.8), (2.9) et de l'intégration des vitesses, le déplacement du mobile entre les instants k et $k - 1$ est décrit tel que :

$$\begin{aligned} x_k &= x_{k-1} + \frac{r(\omega_d + \omega_g)}{2} \cos(\theta_k) \\ y_k &= y_{k-1} + \frac{r(\omega_d + \omega_g)}{2} \sin(\theta_k) \\ \theta_k &= \theta_{k-1} + \frac{r(\omega_d - \omega_g)}{2l} \end{aligned} \quad (2.10)$$

3.3.2. Le modèle cinématique de type tricycle

On désigne par tricycle, un mobile avec deux roues arrières et une roue avant directrice (Figure 2.9). Les odomètres peuvent être montés sur les roues arrières ou sur la roue avant. L'état du mobile est alors décrit par sa position et son orientation $(x, y, \theta)^T$ ainsi que par l'angle de braquage (α_b) de la roue.

On peut alors décrire ce modèle dynamique selon deux formulations [Siegwart et al., 2011] (respectivement les équations (2.11) et (2.12)).

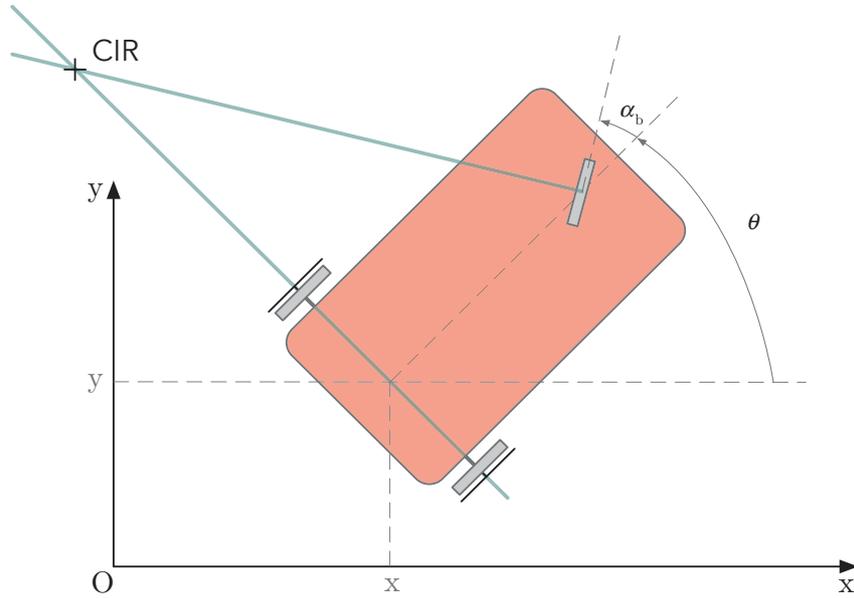


FIGURE 2.9: Modèle cinématique de type tricycle (où CIR signifie Centre Instantané de Rotation)

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\alpha}_b \end{pmatrix} = \begin{pmatrix} \cos(\theta) \\ \sin(\theta) \\ \tan(\frac{\alpha_b}{l}) \\ 0 \end{pmatrix} v_{long} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} v_r \quad (2.11)$$

$$\begin{pmatrix} \dot{x} \\ \dot{y} \\ \dot{\theta} \\ \dot{\alpha}_b \end{pmatrix} = \begin{pmatrix} \cos(\alpha_b)\cos(\theta) \\ \sin(\alpha_b)\cos(\theta) \\ \sin(\frac{\alpha_b}{l}) \\ 0 \end{pmatrix} v_{long} + \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} v_r \quad (2.12)$$

où $\dot{\alpha}_b$ est la vitesse de braquage de la roue directrice et v_{long} la vitesse longitudinale des roues arrières ou de la roue avant selon la position des odomètres.

Considérant l'équation 2.11, et en admettant que l'angle de braquage soit connu (au moyen d'un gyroscope par exemple), l'intégration de \dot{x} , \dot{y} et $\dot{\theta}$ dans le temps permet alors de déterminer le déplacement du véhicule entre les instants $k - 1$ et k tel que :

$$\begin{aligned}x_k &= x_{k-1} + v_{long} \cos(\theta_k) \\y_k &= y_{k-1} + v_{long} \sin(\theta_k) \\ \theta_k &= \theta_{k-1} + \tan\left(\frac{\alpha_b}{l}\right)\end{aligned}\quad (2.13)$$

Le modèle est réaliste sous la contrainte qu'à chaque instant, $|\theta| < \theta_{max} \leq \frac{\pi}{2}$. Cela induit que l'angle de braquage α_b doit être borné. Cette contrainte peut être exprimée en fonction du rayon de courbure instantané maximal R_{max} et de la vitesse longitudinale :

$$|\theta| \leq |v_{long}| \cdot R_{max} \quad \text{où} \quad R_{max} = \frac{l}{\tan(\alpha_{bmax})}\quad (2.14)$$

Le cas du modèle de type voiture est très proche de celui du tricycle. La principale différence se situe au niveau du train avant, qui comporte deux roues libres au lieu d'une sur le tricycle.



Centre instantané de rotation *Définition*

Le centre instantané de rotation (CIR) est un terme utilisé en mécanique classique et plus particulièrement en cinématique pour désigner le point autour duquel tourne un solide à un instant donné par rapport à un repère de référence.

Dans le cas du tricycle, le CIR se situe à la rencontre des axes des roues fixes et de la roue orientable. Dans le cas d'un mobile de type voiture, on considère une roue virtuelle qui transformerait ce mobile en tricycle, en plaçant la roue orientable du tricycle au centre de l'axe des roues avant.

3.3.3. Avantages et inconvénients de l'odométrie

L'odométrie est une technique très répandue. Sa facilité de mise en œuvre, son faible coût, et sa fréquence d'acquisition en font une solution intéressante. Cependant, celle-ci est victime d'une dérive qui peut devenir importante en fonction des défauts du terrain. Dans [Borenstein et al., 1996], l'auteur classe ces erreurs selon deux catégories :

- les erreurs systématiques : engendrées soit par les erreurs mécaniques du système (diamètres des roues différents, mauvais alignement des roues, etc...) qui peuvent être corrigées par calibration, soit par les erreurs de mesures (résolution du capteur, fréquence d'échantillonnage erronée, etc...). Ces dernières ne peuvent pas être calibrées, et doivent être prises en compte dans le modèle cinématique du véhicule.
- les erreurs non systématiques liées aux contacts roue/terrain. Elles sont aléatoires et dépendent directement de la surface sur laquelle le mobile évolue.

3.4. Fusion de données

Comme nous avons pu le voir, il existe de nombreux capteurs capables de fournir une indication sur la localisation d'un véhicule lorsque celui-ci effectue un déplacement. Il s'avère que leurs avantages et leurs inconvénients sont complémentaires. La localisation par centrales inertielles et la localisation grâce aux odomètres, aussi appelée localisation à l'estime, est disponible à tout moment du déplacement, ce qui n'est pas le cas de la localisation par satellites. Cependant, elle reste très dépendante du terrain, et subit une dérive importante à long terme. En conséquence, l'utilisation combinée de ces différents types de capteurs dans un même véhicule paraît intéressante. Une fusion de ces capteurs devrait permettre d'améliorer la précision et la robustesse de la localisation.

Parmi les méthodes de fusion de données pour la localisation, l'une des plus répandues est le filtre de Kalman [Kalman, 1960]. Celui-ci est une approche puissante permettant de coupler différents capteurs de localisation. Le filtre de Kalman permet de tirer profit des avantages et des caractéristiques des différents systèmes de capteur afin de produire un système de localisation présentant des performances supérieures à l'un ou l'autre des capteurs utilisés [Grewal et al., 2007].

Avant de présenter les différents types de fusion existants, nous présentons rapidement le filtre de Kalman, celui-ci étant la clé de la fusion de données.

3.4.1. Introduction au filtre de Kalman

Le filtre de Kalman est une méthode statistique d'assimilation de données, dont le principe est de corriger la trajectoire du modèle en combinant les observations avec l'information fournie par le modèle de façon à minimiser l'état vrai et l'état filtré. Le filtre de Kalman fonctionne de façon récursive, en d'autres termes, pour estimer l'état du système à un moment donné, il est nécessaire de connaître l'état antérieur du système et les observations [Kalman, 1960]. L'utilisation du filtre de Kalman est recommandée lorsque l'imprécision sur les données traitées peut être interprétée comme un bruit aléatoire dont la distribution est connue *a priori* et Gaussienne. Selon la théorie des systèmes linéaires, la dynamique d'un système linéaire peut être représentée par un modèle d'état, où un ensemble d'équations différentielles du premier ordre exprime l'écart par rapport à une trajectoire de référence. En disposant d'une information statistique sur la nature du bruit, il est possible de construire une estimation optimale de l'état du système (Annexe B). Le filtre de Kalman est utilisé dans de nombreuses applications, parmi lesquelles on peut citer la perception, la localisation, l'association de données, la calibration.

3.4.2. Les différents types de fusion

Nous présentons ici différents types de fusion. Dans un premier temps, nous abordons l'intégration GNSS/INS. Cette approche est connue depuis longtemps [Cox, 1978], et est très détaillée dans la littérature. Nous présentons également deux types de fusion différents, à savoir la fusion GNSS/Odométrie, et la fusion Odométrie/INS.

a. GNSS et INS

Le couplage GNSS/INS a été fortement développé par la communauté scientifique. Le positionnement approximatif fourni par le GNSS, et la grande précision de l'INS permettent de reconstituer le parcours d'un mobile avec une bonne précision dans un repère global. Il est possible de coupler ces deux capteurs de différentes façons. La complexité du couplage du système va dépendre de la précision à atteindre sur le positionnement. De manière générale, on trouve trois types d'intégrations : le mode non couplé, le mode faiblement couplé ou couplage lâche, et le mode fortement couplé ou couplage serré.

Mode non couplé

Ce mode consiste à utiliser séparément les informations de localisation fournies par les deux systèmes (Figure 2.10). Le gain en exactitude n'est obtenu qu'en moyennant les deux positions fournies par chacun des systèmes. L'avantage de ce type de couplage est qu'il est simple et peu cher à mettre en place. Cependant, l'un des principaux défauts est qu'il ne permet pas d'éliminer les erreurs de positionnement du GNSS, ni les dérives des centrales inertielles.

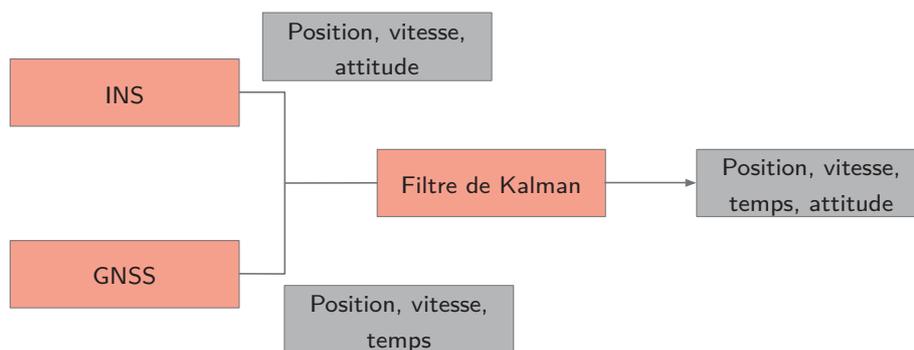


FIGURE 2.10: Mode non couplé

Mode faiblement couplé

Contrairement au mode non couplé, la sortie du couplage lâche est utilisée comme support d'évaluation de la trajectoire réalisée par l'INS. Les données GNSS et la trajectoire de l'INS sont calculées à l'avance. Cependant, pour l'évaluation de la trajectoire de l'INS (Figure 2.11), la sortie du filtre de Kalman de la période précédente est introduite en tant qu'information additionnelle. Ce procédé facilite le calcul de la trajectoire de l'INS ainsi que les dérives et corrections de la dérive. Les erreurs de désalignements sont considérablement réduites grâce à lui. La boucle rétroactive est particulièrement importante pour les INS de moyenne et basse précision ; elle permet d'en borner les erreurs.

L'avantage du mode faiblement couplé est qu'il ne nécessite pas de développement majeur. De plus, la dimension du vecteur d'état du filtre de Kalman reste faible, le rendant ainsi rapide en terme de temps de calcul. Le principal désavantage reste que les erreurs de positionnement du GNSS ne sont pas détectées, comme pour le mode non couplé.

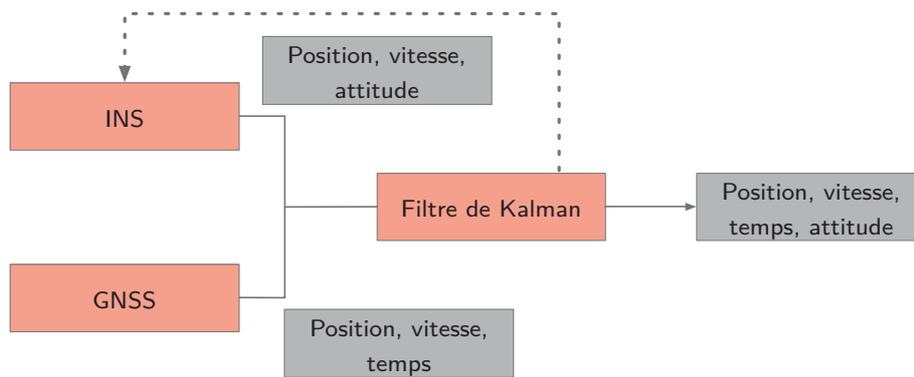


FIGURE 2.11 : Mode faiblement couplé

Mode fortement couplé

Dans un système fortement couplé, les filtres de Kalman pour la navigation et pour le GNSS sont combinés en un seul filtre (Figure 2.12). Le vecteur d'état du filtre de Kalman inclut désormais le vecteur d'erreur de l'INS ainsi que des nouveaux états représentant le biais de l'horloge du GNSS et sa dérive. Les éléments du vecteur d'état qui représentent les erreurs de l'INS sont utilisés pour calibrer l'INS et corriger les estimations de positions et de vitesse, ainsi que l'attitude du mobile. Les estimations du biais de l'horloge et de la dérive sont utilisées pour corriger les mesures du GNSS. Les sorties du filtre de Kalman et de l'INS peuvent également être utilisées pour aider les boucles de repérage des satellites du récepteur GNSS. L'architecture fortement couplée utilise ainsi de manière plus efficace les mesures disponibles et les informations *a priori* pour déterminer et corriger les erreurs du système. Ainsi lorsque peu (ou pas) de satellites sont visibles, nous sommes en mesure de prédire le déplacement du récepteur et donc de restreindre les fenêtres de recherche lorsque les signaux GNSS sont à nouveau disponibles. Ceci permet de réduire le temps de réacquisition des signaux, d'améliorer le suivi des signaux de faible puissance et rend possible la détection des multi-trajets.

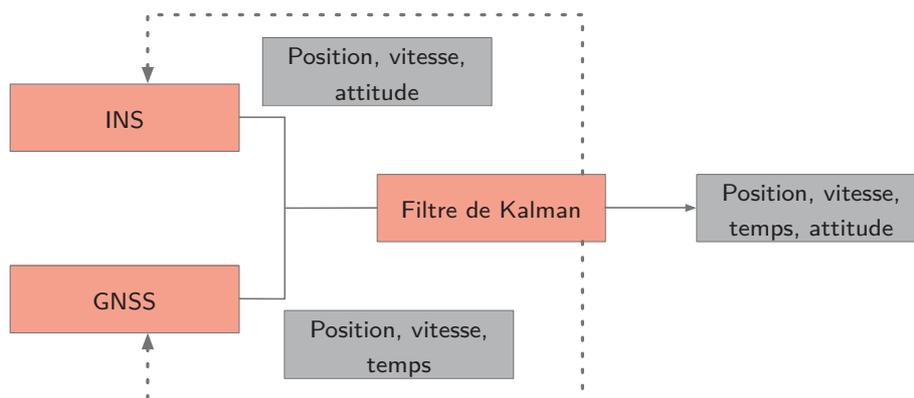


FIGURE 2.12 : Mode fortement couplé

b. GNSS et odométrie

Nous avons pu observer que l'odométrie permet de fournir une bonne exactitude de positionnement à court terme, mais elle subit une dérive importante dès lors que la distance parcourue augmente. Il est alors nécessaire de recalibrer la position du mobile, ce qui est possible grâce au GNSS [Ohno et al., 2003].

L'utilisation couplée de ces deux systèmes permet de gagner en précision, car les odomètres vont permettre de réaliser une prédiction sur la mesure entre deux mesures GNSS. Cette prédiction sera également la source principale de localisation lorsque la mesure GNSS ne sera pas disponible. La difficulté principale de ce couplage est la différence de fréquence de fonctionnement des deux systèmes. En effet, l'odométrie mesure à une fréquence supérieure à 10 Hz tandis que les récepteurs GNSS sont souvent cadencés à 1 Hz . Il est donc nécessaire de faire appel à des systèmes de fusion, tels que le filtre de Kalman, afin de réaliser un recalage soit temporel, soit fonction de la distance parcourue. Nous pouvons également appliquer les couplages lâches et forts pour la fusion GNSS/Odométrie comme présentés précédemment pour la fusion GNSS/INS.

En environnement urbain, le déplacement est réalisé sur un terrain peu accidenté, les odomètres conservent donc une précision de localisation suffisante, ainsi ce couplage prend tout son sens pour notre étude.

c. Odométrie et INS

Comme nous avons pu le voir auparavant, l'odométrie et les centrales inertielles sont deux systèmes qui présentent une dérive dépendant de la distance parcourue. Leur fréquence de fonctionnement est élevée, ainsi, le couplage de ces deux capteurs est plus aisé. Le but de ce couplage est d'obtenir une dérive globale inférieure à celle de chacun des systèmes pris indépendamment l'un de l'autre. Dans [Borenstein & Feng, 1996] un gyroscope est associé à une roue odométrique, l'approche développée consistant à faire un choix entre la vitesse de lacet du gyroscope et la vitesse de lacet déduite des roues. L'erreur d'orientation du robot est alors bien plus faible qu'en utilisant l'un ou l'autre de ces capteurs. Dans [Liu et al., 2005], l'utilisation d'un IMU et de roues odométriques permet d'améliorer les performances du capteur inertiel. Grâce aux informations odométriques et à la modélisation des contraintes du véhicule, les erreurs de l'IMU sont estimées et exploitées pour corriger les mesures fournies par celle-ci. Là encore, les résultats sont convaincants, et les erreurs de mesure de la centrale inertielle grandement réduites. Ce type de couplage permet d'évaluer les erreurs de chacun des capteurs; en effet la redondance des mesures permet de détecter les erreurs brusques. De plus, c'est un couplage facile à mettre en place, compte tenu des fréquences d'acquisition proches pour ces deux capteurs.

Il ressort de cette étude sur les couplages envisageables, que l'utilisation combinée de plusieurs approches permet de tirer profit des avantages de chacune et d'atteindre une précision de localisation suffisante pour positionner un mobile en environnement urbain. Malgré tout, cette précision va fortement dépendre de la qualité des capteurs utilisés et donc de leur coût, ce dernier devenant le facteur limitant. Il est alors nécessaire de se tourner vers d'autres capteurs. Les capteurs de vision font de bons candidats. De nombreuses applications ont été développées par la communauté robotique autour de ceux-ci [Ortin & Montiel, 2001], [Ho & Newman, 2007]. Ces capteurs

présentent l'avantage d'être abordables en terme de coût, mais également facilement disponibles. De plus, ils sont simples à intégrer aussi bien d'un point de vue mécanique, que d'un point de vue algorithmique. Dans le chapitre suivant, nous revenons sur les approches de localisation basées sur l'utilisation de capteur de vision.

3.5. La localisation à partir de capteurs de vision



La localisation d'un robot à partir de capteurs de vision peut être considérée comme l'estimation de la trajectoire d'une caméra par rapport à un repère initial. Les techniques de localisation par la vision peuvent être réparties selon trois familles d'approches : les méthodes de reconstruction 3D à partir du mouvement (*Structure from Motion*), la cartographie et la localisation simultanée (*SLAM ou Simultaneous Localization And Mapping*), et l'odométrie visuelle. Initialement, l'odométrie visuelle ne s'attache qu'à déterminer la trajectoire de la caméra. En ce qui concerne la reconstruction 3D, celle-ci porte sur la restitution de la géométrie de l'environnement, tandis que le SLAM est axé à la fois sur la détermination de la trajectoire, et sur la reconstruction 3D. Ces approches sont très liées et se rejoignent sur de nombreux aspects.

3.5.1. Localisation et cartographie simultanée

Le processus de localisation et de cartographie simultanée est un axe de recherche relativement ancien et particulièrement étudié dans le domaine de la robotique [Durrant-Whyte & Bailey, 2006] [Thrun, 2002]. Le principe consiste à construire incrémentalement une carte de l'environnement à partir d'une position de départ. Cette carte est ensuite utilisée pour la localisation, et chaque nouvelle mesure permet de la mettre à jour. Pour répondre à ce problème, différents outils statistiques sont utilisés tel que le filtre à particules [Eade & Drummond, 2006] ou encore le filtre de Kalman étendu [Jazwinski, 1970]. La principale limitation de ces outils est leur efficacité calculatoire limitée. En effet, dans le cas d'une approche basée sur le filtre de Kalman, le vecteur d'état de celui-ci va contenir à la fois la position de la caméra, et la position de l'ensemble des amers qui constituent la carte. L'étape de mise à jour consiste à calculer la position courante de la caméra et à raffiner la carte. Cette étape a une complexité de calcul en N^2 , où N est la taille du vecteur d'état. Ainsi, dès que la carte contient trop d'amers, un traitement en temps réel n'est plus envisageable. Cependant, il existe des solutions pour remédier à ce problème : [Civera et al., 2009] suppriment de la carte les points qui ne sont plus observés, [Leonard & Newman, 2003] proposent quant à eux de gérer plusieurs sous-cartes. Les premières approches de SLAM proposées restent des approches incrémentales et souffrent donc d'une dérive au cours du temps. Une solution pour pallier ce problème consiste à détecter les lieux où est déjà passé le robot, et à

procéder au calcul de fermeture de boucles [Chapoulie et al., 2004]. Ce procédé est également mis en oeuvre de façon usuelle dans le domaine de la topographie. Ainsi, il est possible d'appliquer des contraintes d'optimisation à la trajectoire [Grisetti et al., 2007].

3.5.2. La reconstruction 3D ou SfM (*Structure from Motion*)

La communauté de la géomatique s'est également intéressée au problème de localisation d'une caméra en mouvement. Les outils utilisés ici diffèrent de ceux utilisés par la communauté robotique. Par exemple le filtre de Kalman, classiquement employé pour les approches SLAM, est remplacé par une technique d'ajustement de faisceaux largement répandue en photogrammétrie [Triggs et al., 2000]. Le premier défi de cette communauté a tout d'abord été de réaliser une reconstruction 3D à partir d'une collection d'images. Cette reconstruction est effectuée en post traitement. Le principal inconvénient des techniques SfM est que la reconstruction à très grande échelle reste difficile à appliquer, bien que certains algorithmes fournissent des reconstructions de zones de tailles relativement grandes (façades de bâtiments) [Furukawa & Ponce, 2010].

3.5.3. L'odométrie visuelle

L'odométrie est l'opération qui consiste à estimer la trajectoire d'un mobile à partir d'un ensemble de relevés de positions relatives. Comme présenté auparavant, l'odométrie que nous nommerons "classique" consiste à placer des capteurs sur les roues d'un mobile. Les données collectées par ces capteurs permettent d'interpoler le déplacement de celui-ci. La dérive de l'odométrie, lors de déplacements longs ou dans des milieux accidentés, a poussé les chercheurs à trouver d'autres techniques de localisation. Les capteurs de vision pour l'estimation de trajectoire se sont révélés très efficaces. Ainsi est apparue l'odométrie visuelle. Elle est un sujet de recherche qui a fait son apparition dans les années 1980 [Moravec, 1980]. L'idée générale de cette approche est de détecter les similitudes entre deux images prises successivement par une caméra. L'étude de ces similitudes va permettre de déduire le déplacement de la caméra et donc la trajectoire du mobile sur lequel elle est installée. Nous revenons plus en détail sur la technique d'odométrie visuelle dans le chapitre décrivant l'approche développée (Chapitre 4 - section 2).

Pour conclure sur ces trois approches de localisation à partir de capteurs de vision, nous pouvons constater que dans le cas des techniques de SLAM et de SfM, l'objectif est identique : localiser en temps réel une caméra dans un milieu inconnu et créer une carte de l'environnement. Dans le cas de l'odométrie visuelle, la création de carte n'est pas essentielle, mais reste cependant tout à fait possible [Scaramuzza et al., 2009]. Ces trois approches ont donc des finalités communes et à ce jour il est de plus en plus difficile de les distinguer les unes des autres.

Concernant notre projet, une approche basée sur l'odométrie visuelle a été développée. Notre choix s'est orienté vers cette approche pour sa rapidité de mise en place, ainsi que son indépendance vis-à-vis de l'environnement dans lequel elle est appliquée.

Lors de notre étude nous avons vu que l'ensemble des approches de localisation présentées permet également l'acquisition de données tridimensionnelles. En comparaison aux techniques de

mesure classique (LIDAR, acoustiques, radar, etc.), ces approches présentent l'avantage de permettre l'acquisition de données 3D de façon simultanée sans nécessiter d'actionneur mécanique. Cependant, elles présentent tout de même certains inconvénients. Premièrement, nous pouvons citer le temps de traitement nécessaire à la construction de la troisième dimension. Deuxièmement, elles nécessitent souvent des conditions d'éclairage adéquates et des surfaces suffisamment texturées dans le cas de mises en correspondance automatiques.

4 Etat de l'art sur les caméras 3D

Un nouveau type de capteur est apparu il y a une dizaine d'années : les caméras 3D. À la frontière entre les systèmes de vision et les systèmes de mesure 3D, elles présentent de nombreux avantages. Issues des techniques de mesure 3D optiques, elles en reprennent les principes de fonctionnement. Les caméras 3D étant au coeur du prototype développé au cours de cette thèse, il est essentiel de les présenter.

Les techniques de mesure 3D optiques peuvent être classées selon trois catégories : l'interférométrie, la triangulation, et le "temps de vol" [Remondino & Stoppa, 2013]. Dans le cadre de notre approche, nous n'aborderons en détail que la triangulation et le "temps de vol". En effet, l'interférométrie ne permet pas une mesure au-delà de quelques centimètres [Remondino & Stoppa, 2013].

4.1. Principe de la triangulation

La triangulation est une technique permettant de déterminer la distance du capteur à un point. Nous distinguons la triangulation active de la triangulation passive. Nous parlons de système de mesure passif, lorsque la mesure de distance repose uniquement sur l'information reçue. Ce type de système n'émet aucun signal. Un système de mesure actif émet quant à lui un signal, puis réceptionne le signal renvoyé par la scène. Nous ne présenterons ici que la triangulation active, qui concerne plus directement les caméras 3D.

La triangulation active utilise une source de lumière structurée (laser par exemple), dont la direction est parfaitement déterminée, et un capteur sur lequel se forme l'image de ce seul point source. La partie émission (laser) et la partie réception du signal sont séparées d'une distance connue.

La connaissance de la distance entre la source d'émission et le récepteur, ainsi que la position du signal reçu sur le récepteur, permettent de déterminer la distance D séparant l'objet du système, par résolution du triangle ABC (Figure 2.13).

En effet, nous avons :

$$D = \frac{base}{\frac{1}{\tan\alpha} + \frac{1}{\tan\beta}} \quad (2.15)$$

Avec :

- D : distance entre l'objet et la caméra (m),
- $base$: distance entre la source d'émission et le récepteur (m),
- α : angle de réception (radians),
- β : angle d'émission (radians).

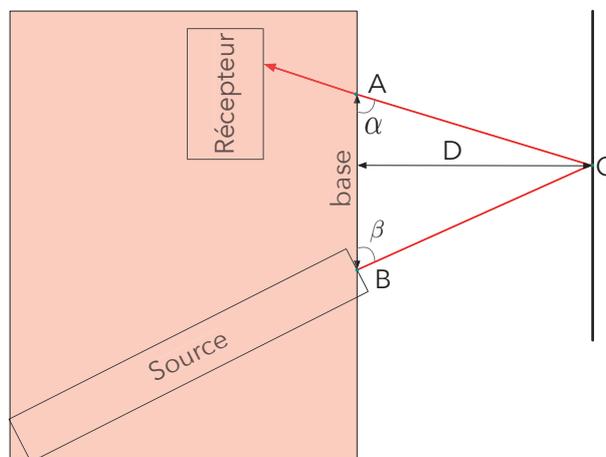


FIGURE 2.13: Principe de la triangulation active

Cette technique présente un avantage : la résolution élevée du récepteur permet une mesure précise de D . Cependant, elle présente aussi des limitations. En effet, sa portée est limitée par la position et la taille du récepteur. De plus, pour être mesuré, un point doit être à la fois visible par la source et par le récepteur, ainsi nous pouvons être confrontés à des effets d'ombres.

L'utilisation d'un laser simple présente l'inconvénient de nécessiter un déplacement mécanique et un nombre d'acquisitions important. L'utilisation d'un motif lumineux permet de contourner ce problème. L'étude de la déformation du motif va permettre d'évaluer un ensemble de distance, en vue d'obtenir un nuage de points.

4.2. Principe du "temps de vol"

Les capteurs dits à "temps de vol" (*ToF*) se répartissent eux aussi selon deux catégories. D'une part, il existe le temps de vol direct, qui mesure la distance aux objets grâce au temps de parcours d'une onde lumineuse émise par la caméra. D'autre part, nous trouvons le temps de vol indirect, ou décalage de phase, qui fournit la distance par l'étude du décalage de phase entre le signal émis par la caméra et le signal renvoyé par l'objet.

4.2.1. Temps de vol direct

Cette catégorie de capteurs n'utilise qu'une impulsion lumineuse pour chaque trame d'acquisition. L'impulsion laser passe à travers une optique de transmission, elle est ensuite réfléchiée par un objet, et est finalement captée par une matrice de photo-diode (Figure 2.14). Les pixels de cette matrice sont indépendants et capables de déterminer le temps de parcours de l'impulsion laser. Ce temps de parcours est dépendant de la distance à laquelle se trouve l'objet. La distance (D) peut alors être déterminée par l'équation (2.16)

$$D = \frac{1}{2}cT \quad (2.16)$$

où c représente la vitesse de la lumière et T le temps nécessaire à l'impulsion laser pour aller de l'émetteur à l'objet, et retourner à la matrice de photo-diodes.

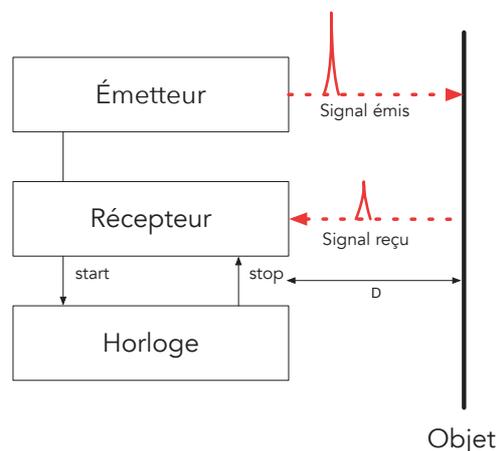


FIGURE 2.14 : Principe de mesure du temps de vol direct

La puissance du signal reçu est plus faible que celle du signal émis. Plusieurs facteurs causent ce phénomène, nous pouvons notamment citer l'albédo de l'objet sur lequel le signal se réfléchit, mais également l'angle d'incidence du signal, ou encore la divergence du faisceau laser. Cette technique présente un avantage : la mesure de distance est possible à grande portée (plusieurs centaines de mètres).

4.2.2. Temps de vol indirect

Cette seconde catégorie de capteurs exploite le décalage de phase pour évaluer la distance à l'objet. Un signal modulé en amplitude est émis de façon continue par la source laser. Ce signal est réfléchi par les différentes surfaces/objets de la scène ; le récepteur reçoit donc un signal de même fréquence, mais déphasé.

Connaissant le décalage de phase $\Delta\varphi$, il est alors possible de déduire la distance parcourue par le signal (2.17) [Kolb et al., 2008].

$$\Delta\varphi = 2\pi f \Delta t = 2\pi f \frac{2D}{c} \quad (2.17)$$

Soit :

$$D = \frac{\Delta\varphi c}{4\pi f} \quad (2.18)$$

Avec :

- $\Delta\varphi$: le décalage de phase,
- f : la fréquence du signal,
- Δt : décalage temporel entre le signal émis et le signal reçu,
- D : la distance parcourue (m),
- c : la vitesse de la lumière

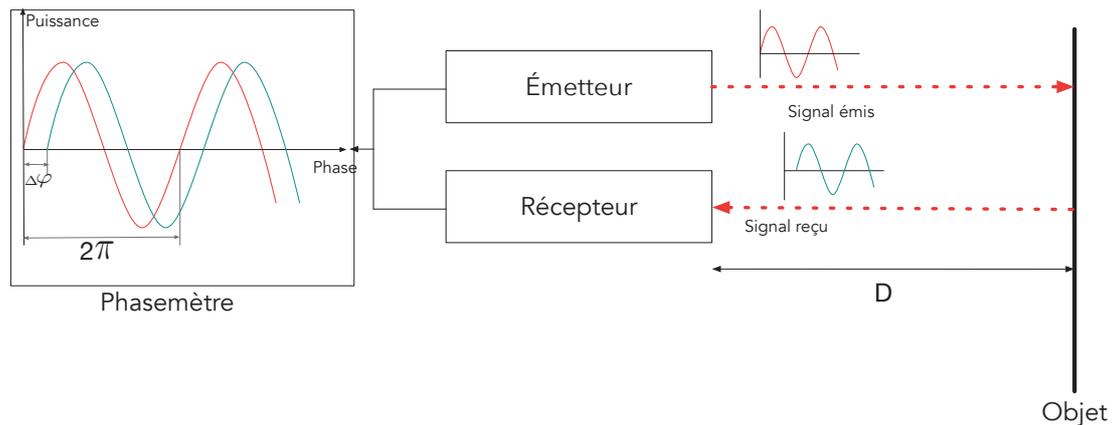


FIGURE 2.15 : Principe de mesure du temps de vol indirect

Cette méthode de mesure présente également l'avantage d'être très rapide et très précise. Toutefois, il existe une distance d'ambiguïté à prendre en compte. En effet, pour une fréquence de modulation de 20 MHz , la longueur d'onde sera de 15 m , ce qui nous donne une distance d'ambiguïté de 7.5 m^1 . Ainsi, plus la fréquence du signal est petite, plus la distance maximale sans ambiguïtés de mesure est grande. Cependant cela impacte la qualité des mesures.

Les scanners laser ou encore les stations totales sont un bon exemple d'appareils reposant sur ces principes [Landes & Grussenmeyer, 2011]. Les scanners laser permettent l'acquisition de millions de distances par seconde grâce à l'utilisation de plusieurs moteurs rotatifs. À partir de

1. Distance d'ambiguïté $D = \frac{\Delta\varphi c}{4\pi f} = \frac{2\pi \cdot 3 \cdot 10^8}{4\pi \cdot 20 \cdot 10^6} = 7.5 \text{ m}$

ces distances et des angles horizontaux et verticaux observés, il est alors possible d'obtenir un nuage de points. La plage de mesure des scanners laser peut varier de quelques décimètres à quelques kilomètres en fonction des appareils. Ils présentent cependant quelques inconvénients : leurs dimensions, leurs poids, mais surtout leurs prix.

Dans ce contexte, les caméras 3D ont fait leur apparition. Ces capteurs permettent l'acquisition de nuage de points de façon continue et ceci sans recourir à un quelconque système mécanique. Elles ont la capacité de produire des données tridimensionnelles à des vitesses élevées, et ainsi produire des nuages de points en temps réel. Cependant, les caméras 3D présentent également des inconvénients. Les capteurs les composant sont souvent de faible résolution (quelques dizaines de milliers de pixels), tandis que la précision de la mesure est de l'ordre de quelques centimètres.

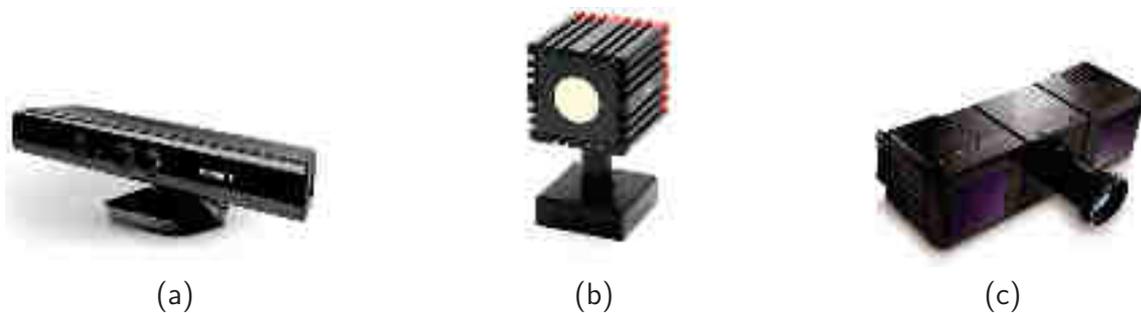


FIGURE 2.16 : Exemples de caméras 3D : (a) Kinect de Microsoft ; (b) SR-4000 de MESA Imaging ; (c) PMDCamCube3.0 de PMD Technologies GmbH

Les domaines d'application des caméras 3D sont nombreux, et nous les retrouvons dans divers domaines : la robotique [Weingarten et al., 2004], la reconnaissance de geste [Soutschek et al., 2008], ou encore la reconstruction 3D [Ellekilde et al., 2007].

Depuis les années 2010, nous avons pu observer un engouement pour les caméras 3D, surtout depuis la commercialisation de la Kinect par Microsoft en 2010, et de la Xtion Pro par Asus en 2011. Toutes deux basées sur le principe de triangulation mis au point par PrimeSense [Arieli et al., 2010], elles présentant l'avantage de permettre l'acquisition simultanée de la géométrie et de la texture à une fréquence élevée. De plus, leur prix est relativement faible (150 €) par rapport aux autres systèmes de mesure présentés auparavant.

Conclusion du chapitre 2

Dans ce chapitre, nous avons dressé un état des lieux du problème et des solutions de localisation. Cet axe de recherche a été très développé depuis de nombreuses années, et ceci dans différentes communautés scientifiques. Ainsi, un nombre important d'approches ont pu être développées.

Il ressort de l'étude de ces différentes approches que :

- L'utilisation du positionnement par GNSS est souvent indiquée pour la tâche de localisation. Malgré tout, en milieu urbain, la configuration de l'environnement rend souvent la localisation par satellite difficile ou de qualité moindre.
- L'ajout de capteurs (centrale inertielle, odomètres) au GNSS permet de corriger ce problème dans une certaine mesure.
- La fusion de ces différents capteurs permet de corriger les inconvénients des uns et des autres et mène à l'obtention d'une information de localisation précise.
- Cependant, la précision, et donc la valeur ajoutée de ces capteurs, va principalement dépendre de l'investissement réalisé.
- Dans ce contexte, il a alors été nécessaire de mettre en place de nouvelles solutions basées sur des capteurs différents. Ainsi, les capteurs de vision ont fait leur apparition comme capteurs de localisation en robotique. De nombreuses techniques et méthodes ont été développées autour de ceux-ci.
- Ces approches présentent différents inconvénients, toutefois depuis quelques années, l'apparition des caméras 3D laissent entrevoir des solutions innovantes. Ce sont des capteurs puissants et complets. De plus, leur prix par rapport aux systèmes d'acquisitions habituels les rend attrayantes.

Conception d'un système de localisation

1	Introduction	39
2	Conception	39
	2.1. Analyse du besoin/contexte	40
	2.2. Concepts	42
3	Concept final	50
	3.1. Idée générale	50
	3.2. Caméras 3D	52
	3.3. Caméra omnidirectionnelle	63
	3.4. Aspect fonctionnel	74

1 Introduction

La problématique de la localisation est un axe de recherche développé par de nombreuses communautés. Nous avons vu, dans le chapitre précédent, qu'un nombre important d'approches ont été mises en place pour répondre à cette problématique.

De cet examen ressortent plusieurs conclusions, qui vont permettre d'alimenter la phase de création d'un prototype permettant la localisation d'un mobile en environnement urbain.

Dans ce chapitre, nous abordons tout d'abord la démarche de conception que nous avons mise en œuvre. Puis, nous présentons les premiers concepts que nous avons pu imaginer dans le cadre de cette réflexion, définis et imaginés pour répondre à la première problématique posée par la thèse. Ensuite, nous présentons le concept dit "final". Deux aspects de ce concept seront abordés. D'une part, nous présenterons l'élaboration et la réalisation d'un prototype composé de capteurs. Ce prototype devra permettre le développement d'une méthodologie de localisation en environnement urbain. D'autre part, nous aborderons l'aspect fonctionnel de ce prototype devant aboutir à un capteur capable d'acquérir les données nécessaires à cette tâche de localisation.

2 Conception

Afin de concevoir une solution de capteurs répondant à notre problématique, nous avons suivi un processus de conception qui nécessite, avant d'aboutir au concept final développé dans le cadre

de cette thèse, de franchir différentes étapes (Figure 3.1). Chacune de ces étapes sera présentée dans les sous-parties suivantes.

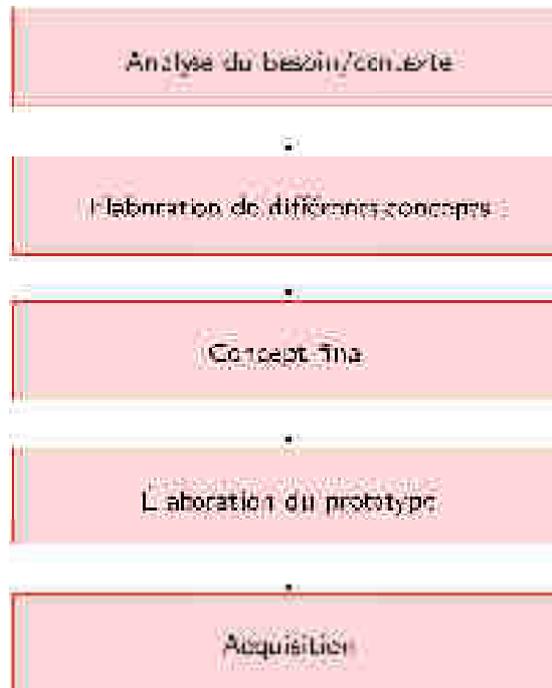


FIGURE 3.1 : Différentes étapes de notre démarche de conception

2.1. Analyse du besoin/contexte

Avant toute chose, il est nécessaire de réaliser une analyse du contexte. En effet, il est indispensable de rechercher et de caractériser les conditions dans lesquelles nous évoluerons. Cette première phase d'analyse a été réalisée dans un état d'esprit proche de l'analyse fonctionnelle. Ce type de démarche consiste "à rechercher et à caractériser les fonctions offertes par un produit pour satisfaire les besoins de son utilisateur" [Tassinari, 2003]. Ce type d'étude intervient principalement dans le cadre de la création (conception) ou de l'amélioration (reconception) de produit. Sans en appliquer toutes les étapes, nous nous sommes inspirés de cette démarche afin de délimiter parfaitement les besoins auxquels la solution envisagée doit répondre.

La première étape consiste donc à évaluer, quantifier et qualifier précisément le contexte de l'étude, mais également le résultat que nous souhaitons obtenir.

Nous rappelons que ce travail de thèse intervient dans le cadre du projet Terra Mobilita, construit autour de deux axes principaux :

- Développer de nouvelles méthodes et outils de création et mise à jour de plans de voirie.
- Développer des applications innovantes d'aide à la mobilité, sur la base des plans établis.

Nos travaux interviennent plus particulièrement dans le cadre du sous-projet : " Techniques avancées d'amélioration de la localisation ". Les travaux de ce sous-projet visent à établir la position et l'orientation d'un mobile : "que ce soit un véhicule, un porteur de type chariot, un vélo, ou encore l'utilisateur lui-même".

Notre contribution consiste donc en des travaux de recherche portant sur la création d'un système basé sur des capteurs de type caméra 3D, pour la localisation d'un système mobile, compte tenu des contraintes particulières de l'environnement urbain.

Nous pouvons d'ores et déjà lister des éléments clés à prendre en compte :

- **Environnement :** comme nous avons pu le voir dans le Chapitre 2, l'environnement urbain dans lequel nous évoluons a une grande influence sur la solution développée et va influencer les orientations algorithmiques de la méthodologie à développer. De plus, la conception mécanique de la solution de capteurs envisagée va également être dépendante de cet environnement.
- **Accessibilité :** à l'heure actuelle, les collectivités territoriales sont dans l'obligation de produire des diagnostics d'accessibilité des espaces publics pour les personnes à mobilité réduite. Notre étude visant particulièrement les circulations dites douces, il est nécessaire d'élaborer un prototype pouvant prendre place sur un mobile circulant sur des trottoirs. Il est donc essentiel de prendre en compte cette spécificité dans l'élaboration de notre solution. Afin d'assurer la navigabilité de notre système en environnement urbain, nous avons considéré que celui-ci serait installé sur un mobile de largeur inférieure ou égale à 80 cm, largeur représentative du gabarit d'un fauteuil roulant.
- **Coût :** dans le Chapitre 2, nous avons pu voir que les systèmes d'acquisition mobile sont composés de capteurs de localisation dont la qualité dépend du prix. Cela a largement limité leur diffusion. Il s'agit dans nos travaux de développer un nouveau système de localisation ayant un coût limité, permettant ainsi sa large diffusion.
- **Intégration avec d'autres capteurs :** Notre solution de localisation sera intégrée sur une plateforme mobile d'acquisition de données tridimensionnelles, donc constituée de capteurs d'acquisition. La solution envisagée ne doit engendrer qu'une occultation minimale lors de l'acquisition des données. Ainsi, le concept retenu doit être suffisamment compact.
- **Productivité et précision :** comme nous avons pu le voir auparavant, les systèmes d'acquisitions fixes présentent un niveau de précision qui comble largement les besoins du projet Terra Mobilita. Cependant, leur niveau de productivité n'en font pas une solution viable. L'apparition des systèmes mobiles permet de répondre au problème de productivité. La complexité de ces systèmes réside dans leur capacité à effectuer une double mesure : le relevé laser en lui-même, ainsi que la position et l'orientation de la plateforme mobile. Ainsi, la précision de la localisation durant l'acquisition va directement impacter la qualité du livrable (plan, carte, modèle, etc.). Dans ce contexte, il est nécessaire d'imaginer une solution qui permette de conserver la productivité des systèmes mobiles, tout en visant

une précision de localisation suffisante pour la production de données de qualité.

- **Fiabilité/robustesse** : compte tenu de la difficulté à mettre en place les campagnes d'acquisition (autorisations légales, etc.), il est nécessaire de réaliser un système de localisation qui soit fiable, celui-ci doit donc être en mesure de fonctionner sans défaillance durant l'acquisition.
- **Complexité** : la complexité intervient à plusieurs niveaux. D'une part, il est nécessaire d'imaginer une solution qui ne soit pas trop complexe en terme de réalisation et de mise en œuvre. D'autre part, l'information de localisation doit être fournie en temps réel. Ainsi, la complexité algorithmique de la méthodologie développée, à partir de la solution de capteurs imaginée, ne doit pas affecter le bon déroulement du calcul de l'estimation de la trajectoire.
- **Innovation** : l'objectif de nos travaux est de développer une solution de capteurs, composée de caméra(s) 3D, permettant la localisation de façon totalement innovante compte tenu des contraintes particulières de l'environnement urbain.
- **Aspect temporel / temps réel** : le temps réel est défini comme "un mode de traitement qui permet l'admission des données à un instant quelconque et l'obtention immédiate des résultats". Il est nécessaire d'imaginer un concept capable de fournir une information de localisation en temps réel. Cependant, il est envisageable d'imaginer un post-traitement afin d'accroître la précision de la localisation *a posteriori*.

Une fois cette analyse terminée, notre démarche a consisté à réaliser des acquisitions d'images en environnement urbain, de façon à pouvoir analyser le type de scènes auxquelles nous serons confrontés. Les zones d'acquisitions ont été choisies de façon à être cohérentes avec les zones d'expérimentations définies par le projet Terra Mobilita (Figure 3.2).

Les différentes acquisitions de test réalisées en environnement urbain nous ont permis de formuler un ensemble d'observations :

- Le mobilier urbain est une information essentielle de la scène : l'environnement dans lequel nous allons évoluer est régulièrement équipé de ce type d'objet. De plus, ceux-ci présentent l'avantage d'être des objets fixes.
- Les bâtiments sont également porteurs d'informations. Les fenêtres et les façades sont de bons candidats pour la détection de points d'intérêts. Ils présentent aussi l'avantage d'être fixes, à noter également qu'ils peuvent être visibles sur de longues distances.

Sur la base de ces constats, la prochaine étape de notre démarche a donc été d'élaborer un ensemble de concepts répondant aux différentes contraintes du projet.

2.2. Concepts

Dans cette section, seuls les trois concepts les plus aboutis et développés au cours de la première année de la thèse et établis dans le cadre de cette démarche, seront détaillés.



FIGURE 3.2 : Exemple de scène enregistrée durant l'acquisition terrain (Rue Mauconseil, Fontenay-sous-Bois)

2.2.1. Concept 1

Ce premier concept repose sur l'utilisation conjointe d'un scanner laser, et d'une caméra 3D (Figure 3.3). Nous souhaitons étudier la possibilité de mettre en relation les nuages de points fournis par ces deux capteurs.

a. Aspect fonctionnel

Le système imaginé fonctionnerait en mode "*stop & go*". Il serait prévu de réaliser une première acquisition à partir du scanner laser. Parallèlement à cela, la caméra 3D serait en mesure de fournir un nuage de points en temps réel. À partir de l'ensemble de ces informations, il serait alors possible de réaliser une mise en correspondance entre ces deux nuages de points. Connaissant les caractéristiques du scanner et la vitesse de déplacement de l'utilisateur, nous serions en mesure de déterminer à quel moment il serait nécessaire de réaliser une nouvelle acquisition avec le scanner.

Ainsi, si nous considérons la Figure 3.4, nous réaliserions, à la position P_0 , une première acquisition, à l'aide du scanner laser (Nuage 1 (scanner)). Le mobile se déplacerait alors à la position P_1 . Grâce au nuage obtenu par l'intermédiaire de la caméra 3D, nous serions en mesure de déterminer la transformation entre P_0 et P_1 . Il en va de même pour la transformation entre P_0 et P_2 .

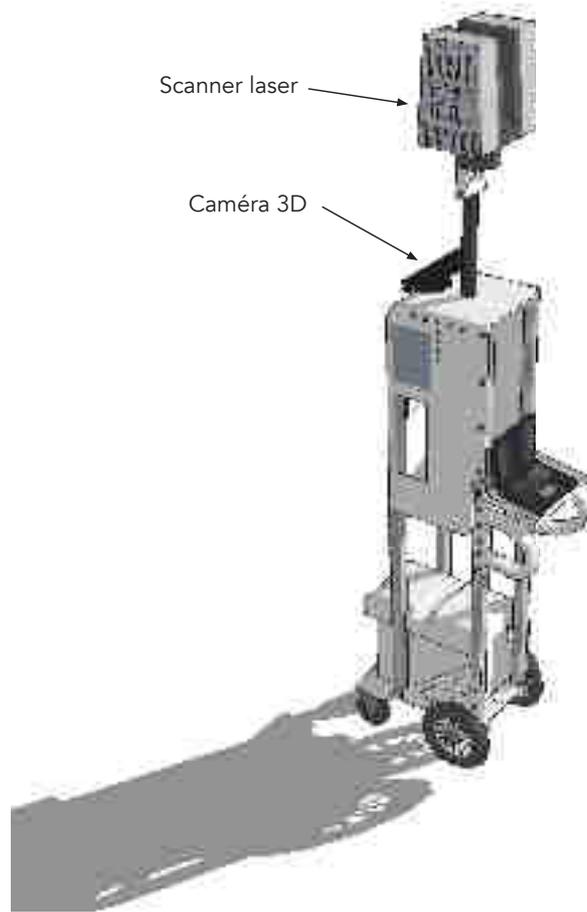


FIGURE 3.3: Modélisation du premier concept établi

Lors du déplacement nous amenant en P_3 , nous considérerions que la distance parcourue nécessite à nouveau une acquisition du scanner (Nuage 2 (scanner)). Le recouvrement entre les deux acquisitions du scanner nous permettrait alors de déterminer la transformation entre P_0 et P_3 . Lors du déplacement suivant, nous serions en mesure de déterminer la transformation entre P_3 et P_4 grâce au nuage de la caméra 3D.

Il serait également prévu d'équiper le mobile d'un système de radio-identification (RFID). Ce système permet de mémoriser, et de récupérer des informations en utilisant des marqueurs appelés "radio-étiquettes". L'utilisation de ce système imposerait un équipement de la scène au préalable. Cependant, cela ouvre plusieurs possibilités : recalage de la trajectoire, détection des fermetures de boucles etc.

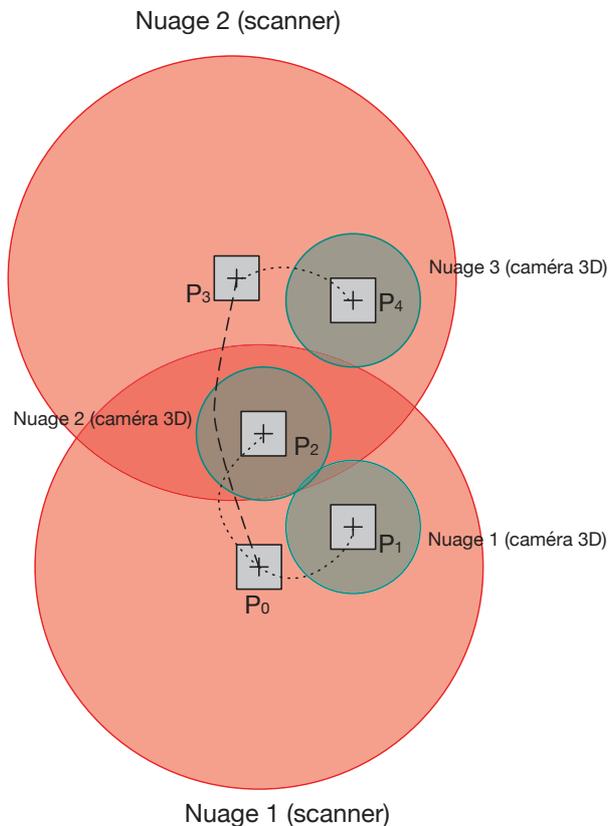


FIGURE 3.4: Fonctionnement du premier concept

b. Bilan

Sur la base des critères établis dans le paragraphe précédent (paragraphe 2.1.), nous pouvons évaluer ce premier concept. Nous avons reporté sur le graphique en radar présenté sur la Figure 3.5 les différents critères évalués.

Concernant le critère de productivité, l'utilisation et l'instrumentalisation de la scène par des puces RFID (*Radio Frequency IDentification*) rendraient le système lourd à mettre en œuvre. Cependant, celles-ci rendraient la solution fiable et précise. En effet, il serait possible de recalculer la trajectoire estimée durant le parcours du mobile. Concernant le coût, celui-ci serait limité, dans la mesure où le scanner laser du système d'acquisition serait utilisé dans la tâche de localisation. Ainsi, il serait nécessaire de n'investir que dans une unique caméra 3D. De plus, son intégration plus serait alors plus aisée. Du point de vue de l'innovation, cette solution serait satisfaisante, sans pour autant être une innovation de rupture.

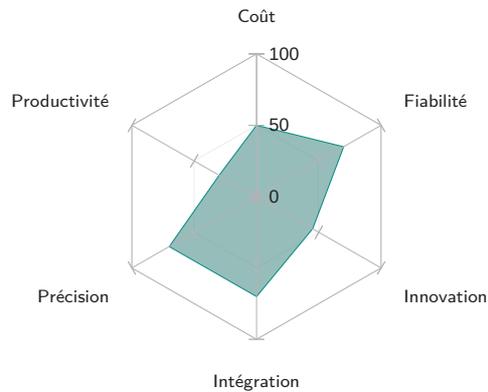


FIGURE 3.5 : Évaluation du premier concept

2.2.2. Concept 2

Ce concept repose, comme le précédent, sur l'utilisation conjointe de deux capteurs. Il s'agit ici de mettre en corrélation les données obtenues grâce à une caméra RVB et les données fournies par une caméra 3D (Figure 3.6 (a)).

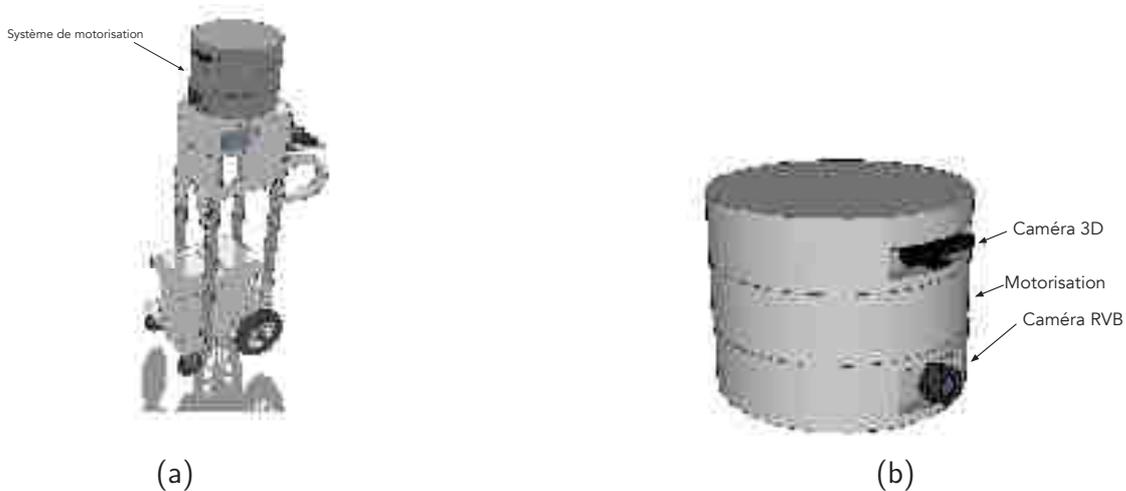


FIGURE 3.6 : (a) Modélisation du second concept - (b) Motorisation des caméras

La caméra RVB permet ici de détecter des points d'intérêts dans la scène, tandis que la caméra 3D permet d'en évaluer la distance. Il a été choisi d'utiliser une caméra RVB externe à la caméra 3D dans la mesure où les caméras 3D ne sont que rarement équipées de ce capteur de couleur. De plus, lorsqu'il est présent, le capteur RVB est souvent de qualité moindre. Les caméras 3D ayant un champ de vision restreint, il est prévu de placer la caméra 3D sur un système motorisé afin de l'orienter vers la partie de la scène porteuse d'information (Figure 3.6(b)).

a. Aspect fonctionnel

Les acquisitions réalisées lors de la première phase de l'étude nous ont permis d'observer que les rues parcourues étaient équipées de mobilier urbain, mais également "d'objets types" :

- fenêtres de bâtiments,
- arbres,
- passages piéton,
- panneaux de signalisation,
- ...



FIGURE 3.7 : Exemple d'objets pouvant être détectés (surlignés en rouge)

Dans le cadre de ce concept, l'approche choisie viserait à détecter les objets mis évidence dans la Figure 3.7, par reconnaissance d'objets sur la base de processus d'apprentissages. L'intérêt de l'utilisation de ces objets résiderait dans l'assurance qu'ils sont fixes dans la scène. Une fois cette détection d'objets réalisée, il serait possible d'orienter la caméra 3D en fonction de la répartition des objets dans la scène. Connaissant parfaitement le lien entre les deux caméras, et après calibration, nous serions en mesure de retrouver les objets détectés dans les images RVB et dans les nuages de points. Il s'agirait alors de mettre en correspondance les objets d'une image à l'autre et d'exploiter la transformation entre leurs nuages de points, afin d'évaluer le déplacement du mobile.

b. Bilan

Sur la base des éléments clés présentés précédemment, nous pouvons juger ce second concept (Figure 3.8).

Le coût de ce concept peut être élevé, d'une part du fait de l'achat de la caméra RVB, mais surtout pour la réalisation mécanique permettant de motoriser la caméra 3D. Du point de vue de l'intégration, celle-ci peut être rendue difficile. En effet, la caméra RVB et la caméra 3D motorisée peuvent devenir d'importantes sources d'occultations pour un système d'acquisition. La fiabilité de ce système est moyenne dans la mesure où les deux caméras exploitent des objets identiques.

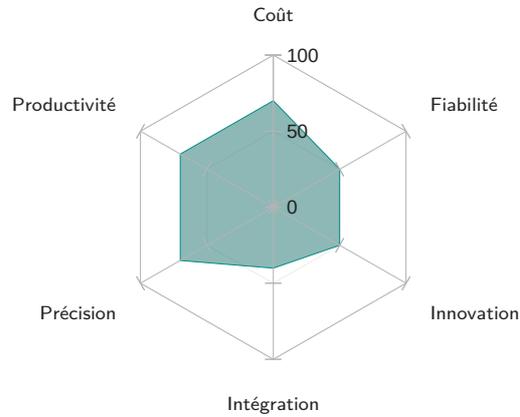


FIGURE 3.8: Évaluation du second concept

De plus, il n'existe pas de moyen de recalage, contrairement au concept précédent. Cependant, ce concept permet d'éliminer la contrainte d'instrumentalisation et de gagner ainsi en productivité.

2.2.3. Concept 3

Ce concept se distingue des autres par son approche basée sur les traitements à effectuer sur les données fournies par une caméra 3D, plutôt que de fournir une solution de capteur. Il repose sur l'exploitation des différences entre images de profondeur.

Pour ce faire, nous nous sommes servis d'un jeu de données mises à disposition par le groupe Computer Vision de la Technische Universität München (Figure 3.9). Le jeu de données contient les images RVB et les cartes de profondeur fournies par une Kinect sur une scène acquise en intérieur. La trajectoire de la caméra a également été enregistrée grâce à un système de capture de mouvement.



FIGURE 3.9: Image extraite du jeu de données utilisé dans le cadre du concept 3 (Technische Universität München)

Dans le cadre de ce concept, notre démarche consisterait à étudier les différences entre images de profondeur. Ainsi, si l'on réalise la différence entre deux images de profondeur successives, nous obtenons le résultat présenté Figure 3.10.



FIGURE 3.10 : Différence entre deux cartes de profondeur successives sur la scène présentée Figure 3.9

Estimant que le mobile se déplace à une vitesse de 1 m/s , nous pouvons borner les différences que nous souhaitons observer :



FIGURE 3.11 : Différence bornée entre deux cartes de profondeur successives

Ceci nous permet de mieux apprécier les informations que l'on souhaite extraire. Il est possible d'appliquer un ensemble de traitements sur ces images de différences. Par exemple, le calcul du gradient nous permet de faire apparaître différentes informations :

- contours des objets immobiles dans la scène,
- direction du déplacement.

Dans le cadre de ce concept, l'approche choisie consisterait à exploiter l'ensemble de ces informations pour la localisation. Nous pouvons, par exemple, imaginer une approche SLAM reposant sur l'utilisation des objets fixes détectés dans la scène.

a. Bilan

Bien que la définition de ce concept soit moins poussée, il est possible d'évaluer les performances de celui-ci, par rapport aux critères définis auparavant.

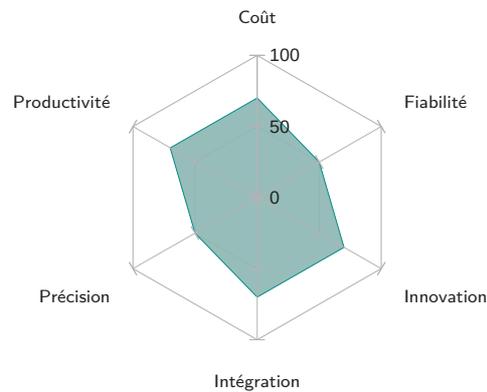


FIGURE 3.12: Évaluation du troisième concept

L'utilisation d'une unique caméra 3D limiterait considérablement le coût du concept, et son intégration au sein d'un système d'acquisition serait plus aisée. La productivité atteinte grâce à ce système serait bonne, cependant la précision et la fiabilité de l'approche dépendront directement du type de caméra 3D utilisé.

2.2.4. Les autres concepts

Durant cette période d'élaboration de concepts, plusieurs autres systèmes ont vu le jour. Ceux-ci sont souvent des déclinaisons basées sur ces trois premiers concepts, ou des améliorations d'approches existantes. Pour le lecteur intéressé, le détail de ces concepts est fourni en annexe (Annexe C).

3 Concept final

3.1. Idée générale

L'élaboration des différents concepts mentionnés précédemment a donné lieu à une réflexion collective permettant d'aboutir à une solution finale. Comme nous l'avons vu précédemment, les acquisitions d'images tests nous ont permis de tirer quelques conclusions sur l'environnement dans lequel le système va évoluer.

a. Aspect fonctionnel

Sur la base de cette réflexion, une idée principale a vu le jour : la nécessité d'utiliser une ou plusieurs caméras 3D et éventuellement, dans un second temps, de compléter le système avec une caméra omnidirectionnelle. Les caméras 3D seront orientées de façon à observer le mobilier urbain, et la caméra omnidirectionnelle pourra être positionnée de façon à observer les bâtiments (Figure 3.13). L'utilisation couplée de ces deux systèmes permettra d'exploiter au mieux les informations présentes dans l'environnement. Il est prévu d'exploiter les avantages de chacune de ces solutions de vision pour obtenir la trajectoire la plus précise possible.

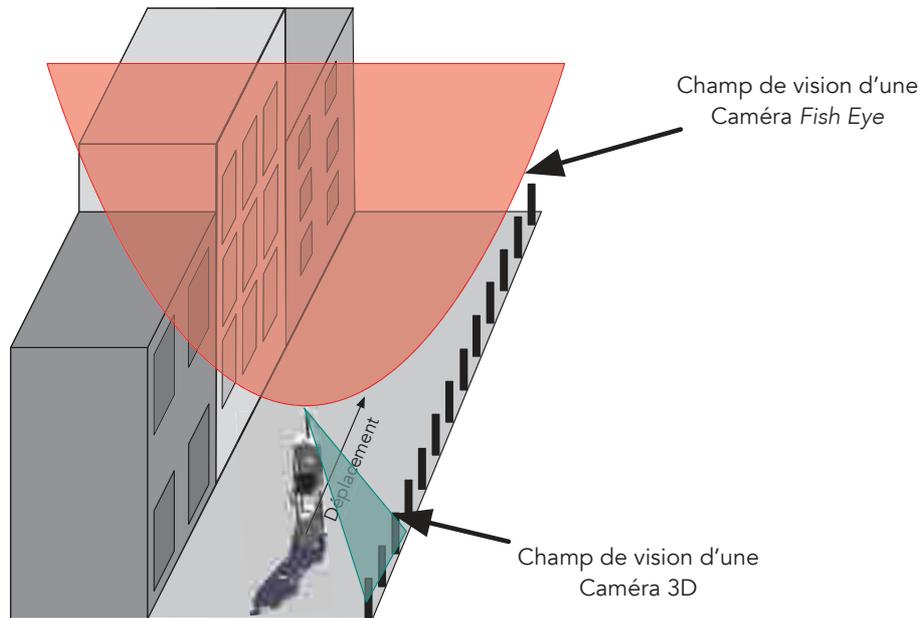


FIGURE 3.13: Présentation du concept final

Les caméras 3D permettent l'acquisition de nuage de points en temps réel. Cependant, la quantité de données fournies par celles-ci n'est pas toujours compatible avec la contrainte de temps réel fixée. C'est pourquoi, il est apparu nécessaire de simplifier les données produites par la caméra 3D, avant de les traiter. Pour ce faire, nous avons décidé de générer une orthoimage à partir des nuages de points fournis par les caméras 3D.

En parallèle de cette première étape de conception, différents travaux ont été menés autour de l'odométrie visuelle, et notamment durant un projet de fin d'étude encadré au cours de la thèse [Yang, 2012]. Sur la base des résultats obtenus, notre approche a consisté à appliquer une odométrie visuelle sur les orthoimages générées.

De façon similaire, les images obtenues grâce à la caméra omnidirectionnelle permettent également la mise en place d'un algorithme d'odométrie visuelle.

Enfin, il est prévu de réaliser une fusion des trajectoires obtenues respectivement pour chacun des systèmes.

b. Bilan

De façon identique aux autres concepts, nous réalisons un premier bilan sur ce concept final (Figure 3.14).

Le choix des caméras a un impact direct sur le coût du prototype à réaliser. Nous verrons par la suite que cela a eu une influence sur le modèle de caméra utilisé. En terme de productivité, ce concept présente de bons résultats dans la mesure où il n'est pas nécessaire d'équiper la scène

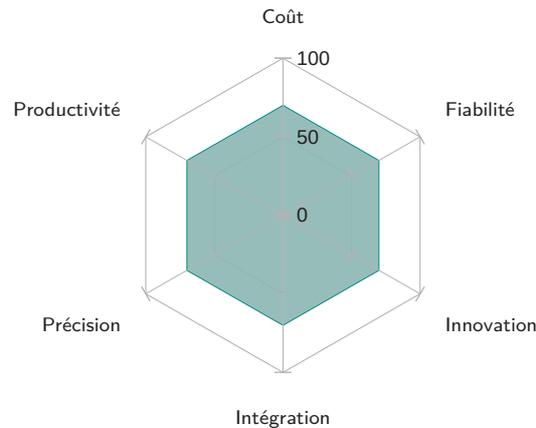


FIGURE 3.14 : Évaluation du concept final

pour aider à la localisation. Concernant la fiabilité, l'utilisation de deux systèmes de vision va permettre de réaliser un contrôle entre eux afin d'éviter toute erreur durant l'estimation de la trajectoire. L'intégration de cette solution de capteur au sein du système d'acquisition va directement dépendre du nombre de caméras et de leur position sur le mobile. Cette approche nous semble novatrice et en rupture avec les systèmes de localisation existants.

Dans les sous-parties suivantes, nous allons présenter ces deux systèmes de vision. La solution à développer doit exploiter les données de ces deux systèmes de vision, en vue d'estimer la trajectoire d'un mobile en environnement urbain de la façon la plus précise possible.

3.2. Caméras 3D

Nous présentons ici le pôle caméras 3D. Aussi, nous proposons une étude sur le choix de la ou des caméra(s) utilisée(s) lors de la réalisation du prototype. Il sera également question du nombre de caméras à utiliser et de leur position sur le mobile. Enfin, nous aborderons la calibration de la "tête" de caméras 3D, ainsi que l'exploitation des données fournies par celles-ci.

3.2.1. Choix du modèle de caméra 3D

Les modèles de caméras 3D existants sont de plus en plus nombreux [Dario Piatti, 2010]. Cependant, dans la mesure où nous réalisons un prototype sur lequel s'appuiera notre approche de localisation, notre choix s'est porté sur la caméra Xtion Pro de Asus. Outre son prix, celle-ci répond à nos besoins en terme de compacité et d'approvisionnement. Cette caméra a vu le jour à la suite de la commercialisation de la Kinect de Microsoft. Ces deux caméras sont basées sur le même principe de fonctionnement, issu des travaux menés par Prime Sense [Arieli et al., 2010], société rachetée en novembre 2013 par Apple.

D'abord conçue dans un but de reconnaissance de forme pour l'industrie du jeu vidéo, ce modèle de caméra a connu un réel engouement auprès des développeurs. Ainsi, PrimeSense a mis à disposition, au début de l'année 2011, des pilotes officiels permettant une utilisation plus ouverte de ce capteur. À partir de ceux-ci, il est possible d'extraire les données brutes fournies par la caméra. Ce modèle a également connu un large succès au sein de la communauté scientifique, dans de multiples domaines, tels que la robotique [Newcombe et al., 2011] et la reconnaissance d'objet [Janoch et al., 2013].

Le système Asus Xtion Pro fonctionne sur le principe de triangulation (Chapitre 2 - paragraphe 4.1.). Ce système est donc équipé de différents éléments (Figure 3.15) :

- un système d'émission infrarouge,
- un système de réception captant l'infrarouge,
- une caméra RGB d'une résolution de 640x480.



FIGURE 3.15 : Différents éléments de la caméra Xtion Pro de Asus

Le système de réception de la caméra est équipé d'un capteur CMOS comme pour une caméra classique. La détermination de la distance se fait pour l'ensemble des pixels du capteur. Cette détermination est réalisée à l'aide d'une émission de lumière structurée. Plusieurs approches de lumières structurées sont déjà utilisées, comme la méthode exploitant des franges de lumière ou encore les approches basées sur l'utilisation d'un laser [David, 2015].

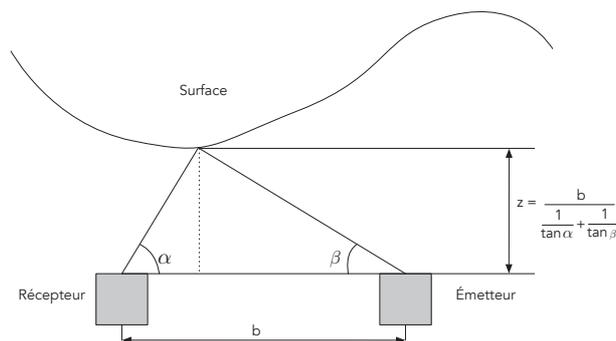


FIGURE 3.16 : Principe de fonctionnement du système Asus Xtion Pro

Cependant, dans notre cas, la détermination de la profondeur est plus complexe car la projection de l'image est codée. Ainsi, par corrélation entre l'image de référence et l'image déformée

par la scène, connaissant les angles α et β (Figure 3.16), il est possible de déterminer z pour l'ensemble des pixels simultanément. Nous obtenons alors une image de profondeur, où chaque pixel contient l'information de distance.

Connaissant les caractéristiques du système de réception (optique : focale $f = 12$ mm et capteur : taille d'un pixel : $21 \mu m$), il est par conséquent possible de calculer les points 3D de la scène (Figure 3.17).

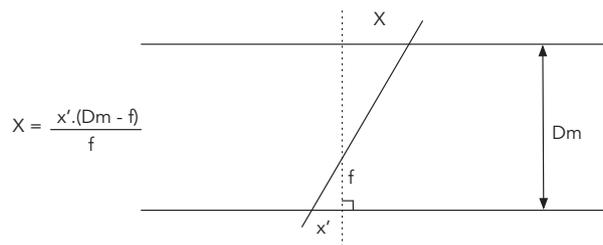


FIGURE 3.17 : Calcul des coordonnées 3D des points acquis à partir de la carte de profondeur

où X représente les coordonnées 3D, x' les coordonnées pixel, et D_m la distance du point.

La caméra Xtion Pro (Asus) permettant d'obtenir une image de profondeur d'une taille de 640×480 pixels, il est possible d'obtenir un nuage de points composé de 307200 points (Figure 3.18).

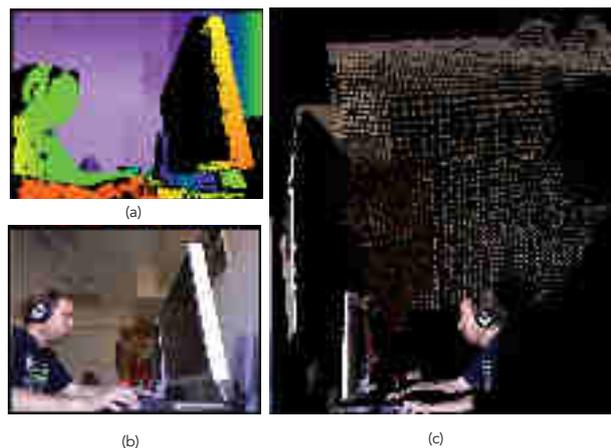


FIGURE 3.18 : Exemple de données obtenues à partir de la Xtion Pro : (a) Carte de profondeur - (b) Image RGB - (c) Nuage de points (Source : <https://graphics.stanford.edu/~mdfisher/Kinect.html>)

3.2.2. Nombre et position des caméras

Connaissant le modèle de caméras utilisé, il s'agit désormais de déterminer le nombre de caméras et leur position sur le mobile. Nous souhaitons utiliser la/les caméra(s) 3D afin d'évaluer le déplacement du mobile, en exploitant le mobilier urbain présent dans la scène. En effet, ces objets sont des points de repères robustes dans l'environnement dans lequel nous évoluons. Nous positionnons donc la/les caméra(s) en hauteur, en orientant leur champ de vision vers le sol, de façon à pouvoir étudier les objets qui "sortent" du sol.

Afin de choisir le nombre et la position des caméras à utiliser, nous avons réalisé une étude de positionnement. Avant toute chose, il est nécessaire de tenir compte de plusieurs critères :

- La solution de capteur doit rester compacte en vue de son intégration sur une plateforme d'acquisition.
- Il est nécessaire d'éviter le recouvrement entre les champs des caméras, d'une part pour éviter un dysfonctionnement des caméras, et d'autre part pour couvrir au sol la plus grande zone possible.

Sur la base de ces critères, plusieurs configurations ont été testées. Afin de mieux évaluer celles-ci, plusieurs modélisations ont été réalisées. Elles permettent de visualiser l'empreinte au sol obtenue, mais également l'intégration des caméras 3D sur le mobile. Nous présentons ici la configuration retenue, mais on retrouvera l'ensemble des configurations modélisées en Annexe D.

La solution finale retenue comprend trois caméras. Nous nous sommes limités à ces trois caméras afin d'offrir une compacité suffisante. De plus, toujours dans un souci de compacité, les caméras ont été placées de façon verticale. Ainsi, leur intégration au sein d'une plateforme d'acquisition sera plus aisée, tout en garantissant une empreinte au sol pertinente pour l'approche développée (Figure 3.19).

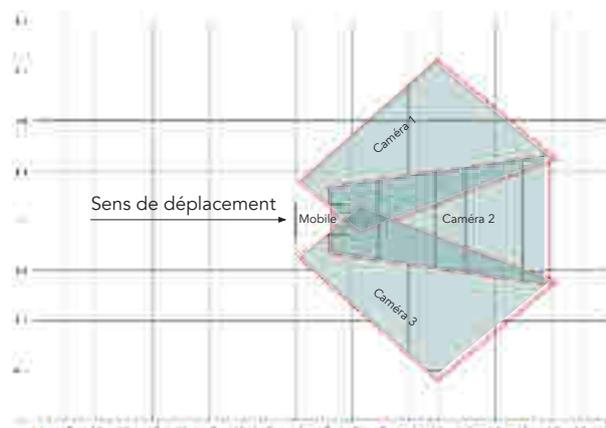


FIGURE 3.19 : Empreinte au sol obtenue à partir de la configuration choisie (vue du dessus)

Afin de conserver la continuité dans le champ de vision des caméras, il a été choisi d'appliquer

un angle de 40° entre la caméra centrale et les deux caméras latérales (Figure 3.20). Enfin, chacune des caméras est inclinée avec un angle de 30° par rapport à la verticale. Ceci permet de "pencher le regard" des caméras vers les objets de la scène.

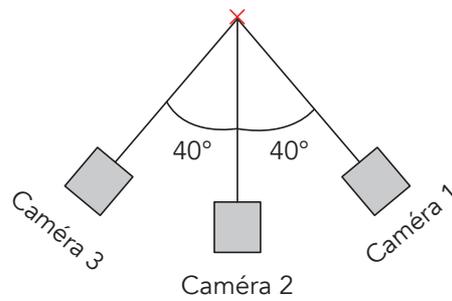


FIGURE 3.20 : Position des caméras (vue de dessus)

Sur la Figure 3.21, nous pouvons observer les champs de vision des trois caméras positionnées de façon verticale sur le mobile. Cette modélisation nous permet de vérifier la bonne intégration des caméras sur le mobile. Nous pouvons également observer les possibles occlusions qui pourraient affecter les nuages fournis par les caméras, mais également le système d'acquisition (scanner laser sur la Figure 3.21) présent sur la plateforme d'acquisition.

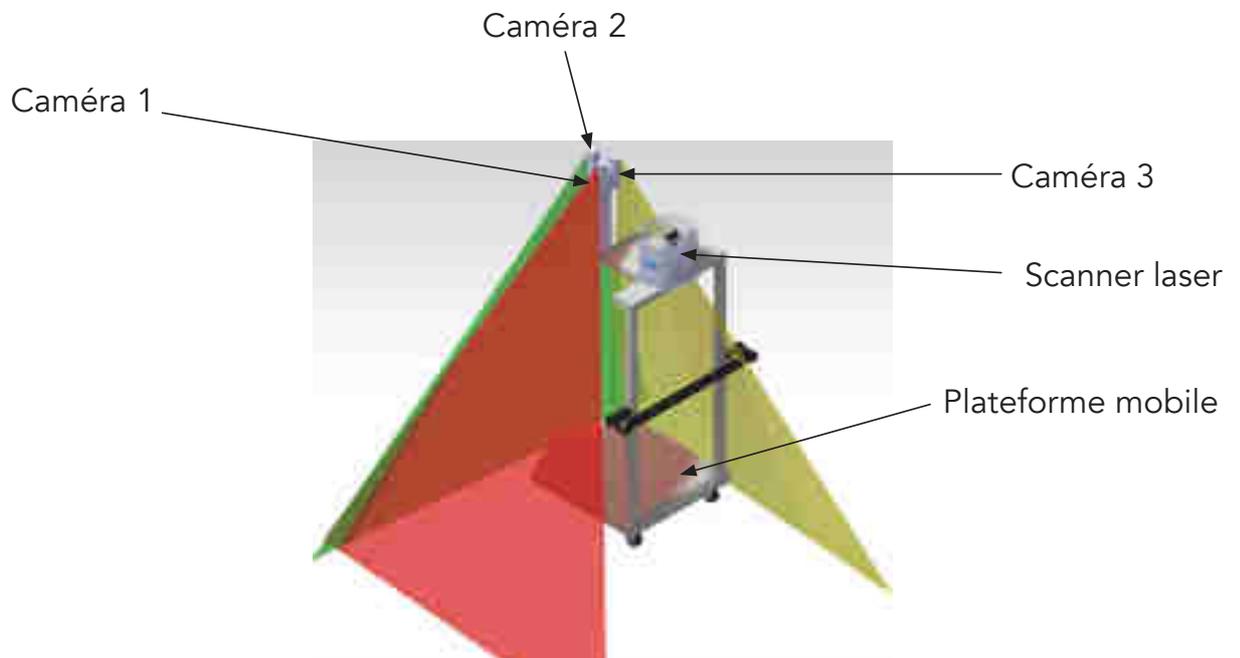


FIGURE 3.21 : Champs de vision des trois caméras à partir de la configuration choisie

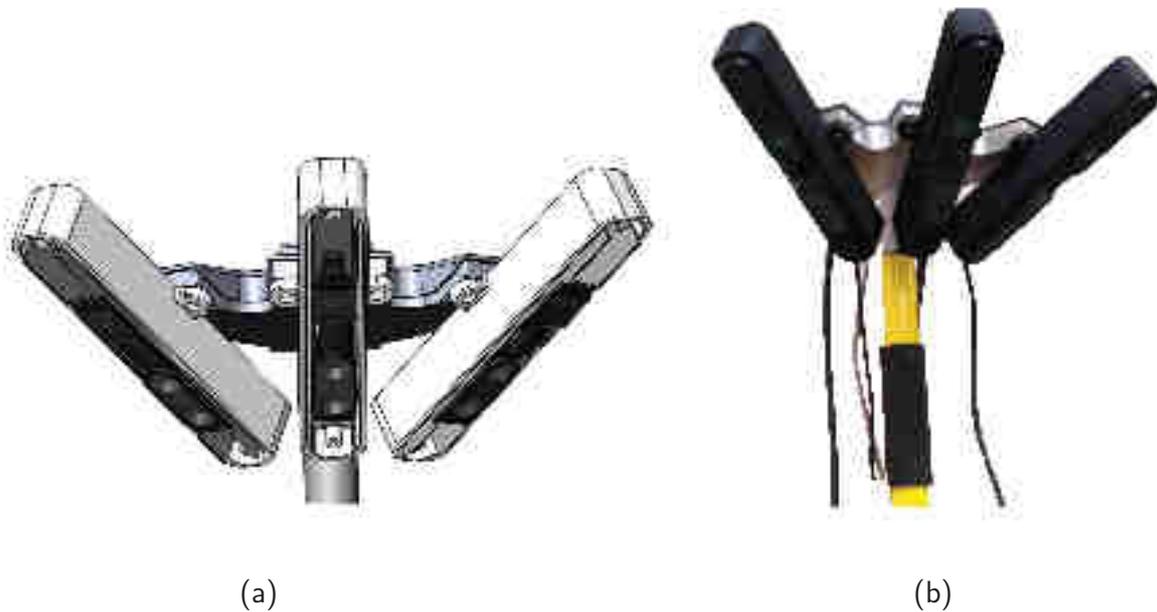


FIGURE 3.22 : (a) Modélisation de la configuration finale du système de caméras 3D - (b) Prototype réalisé

Sur la base de ces informations, nous avons réalisé la modélisation sur plan grâce à un logiciel de conception assistée par ordinateur (Figure 3.22 (a)). La réalisation du prototype (Figure 3.22 (b)) a été confiée à une entreprise spécialisée, et l'assemblage a eu lieu dans les ateliers de Trimble.

3.2.3. Calibration

La position et l'orientation des caméras ont été définies pour répondre à des besoins d'intégration, mais également à des contraintes de champ de vision. Sur la base de la configuration choisie, une modélisation a pu être réalisée. Grâce à cette modélisation, les pièces mécaniques ont pu être fabriquées puis assemblées. Cependant, l'ensemble de ces étapes introduit des écarts entre le plan théorique et la réalisation. Ceci se traduit par des erreurs de positionnement et/ou d'orientation des capteurs. Ces erreurs doivent être évaluées.

Afin de positionner les caméras les unes par rapport aux autres, nous avons choisi d'utiliser une méthode de topographie classique. Ainsi, nous avons eu recours à un ensemble de cibles. Celles-ci sont facilement identifiables, et leurs coordonnées dans le référentiel capteur peuvent être obtenues directement grâce aux caméras 3D. En effet, à partir des nuages de points et des images infrarouge fournies par les caméras, il est possible de retrouver la position de chacune des cibles. Pour mener à bien cette calibration, nous avons réalisé une mire (Figure 3.23).

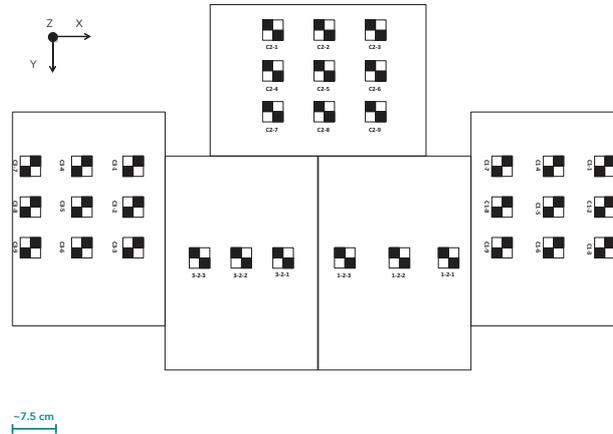


FIGURE 3.23 : Mire de calibration pour le contrôle du positionnement des caméras 3D

Notre objectif est d'exprimer les coordonnées de chacune des caméras dans le repère de la mire, afin d'obtenir les transformations entre chaque caméra. Pour cela, nous levons les coordonnées des cibles visibles dans le repère capteur. Nous serons ensuite en mesure de fusionner les trois nuages de points issus de chaque caméra en un seul (Figure 3.24).

Par construction, nous connaissons les coordonnées de chacune des cibles dans le repère de la mire. Ainsi, à partir de l'ensemble des coordonnées levées et de l'ensemble des coordonnées réelles des cibles, nous pouvons déterminer les coordonnées des caméras dans le repère de la mire (Tableau 3.1). La transformation rigide permettant de passer d'un repère à l'autre peut s'écrire sous la forme :

$$T = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \quad (3.1)$$

où $R \in SO(3) \subset \mathbb{R}^{3 \times 3}$ est une matrice de rotation, et $t \in \mathbb{R}^3$ est un vecteur de translation de dimensions 3×1 . Il s'agit alors de retrouver les paramètres Ω , Φ , κ , qui représentent les angles de rotation en degrés et T_x , T_y , T_z , les translations. Afin de les estimer, nous utilisons l'approche proposée par [Dewitt, 1996].

Sur la Figure 3.24, nous pouvons observer les nuages issus de chacune des caméras. Nous pouvons observer sur chacun des nuages quelques zones "blanches", celles-ci correspondant à des ombres dues aux objets présents dans la scène.

	Camera 1	Camera 2	Camera 3
Ω	16.7186 ° (± 0.2353)	39.3136 ° (± 0.1405)	13.7818 ° (± 0.2319)
Φ	36.1954 ° (± 0.2413)	-2.2857 ° (± 0.1826)	-37.2252 ° (± 0.3102)
κ	-29.5936 ° (± 0.2054)	-93.7636 ° (± 0.1634)	-158.8719 ° (± 0.2216)
T_x	0.656 m (± 0.003)	0.580 m (± 0.003)	0.517 (± 0.004)
T_y	0.792 m (± 0.003)	0.768 m (± 0.002)	0.837 m (± 0.003)
T_z	-0.618 m (± 0.003)	-0.615 m (± 0.002)	-0.605 m (0.003)

TABLE 3.1 : Coordonnées de chacune des caméras dans le repère de la mire. Ω , Φ , κ représentent les angles de rotation en degrés et T_x , T_y , T_z les translations.

3.2.4. Approche algorithmique

Dans cette sous-partie, nous présentons l'approche algorithmique développée autour des caméras 3D afin d'aboutir à la trajectoire du mobile sur lequel elles sont placées. Il s'agit ici de définir les lignes directrices de la méthodologie mise en place au cours de cette thèse. Les détails de celle-ci seront présentés dans le Chapitre 4.

Les caméras 3D présentes au sein de notre solution de capteurs sont orientées de façon à observer la scène urbaine dans laquelle le mobile évolue. Chacune de ces caméras est en mesure de fournir une image de profondeur permettant la génération du nuage de points associé. Celui-ci peut contenir jusqu'à 300 000 points du fait des spécificités techniques liées au modèle de caméra choisi. La calibration de la tête de caméra permet de fusionner les trois nuages pour n'en former qu'un seul. Ainsi, notre système fournit des nuages de points composés d'environ 900 000 points. Cependant, l'ensemble de ces données peut devenir volumineux et engendrer un temps de traitement important.

Nous rappelons ici la contrainte temporelle à laquelle notre solution doit répondre. Cette contrainte nous a amenés à réfléchir à une simplification des données à traiter, tout en conservant leur intégrité afin de profiter des différents avantages fournis par les caméras 3D. Dans ce contexte, il nous est apparu opportun d'orthoprojeter le nuage de point global obtenu après fusion. Il s'agit de projeter verticalement chaque point du nuage sur le plan représentant le sol.

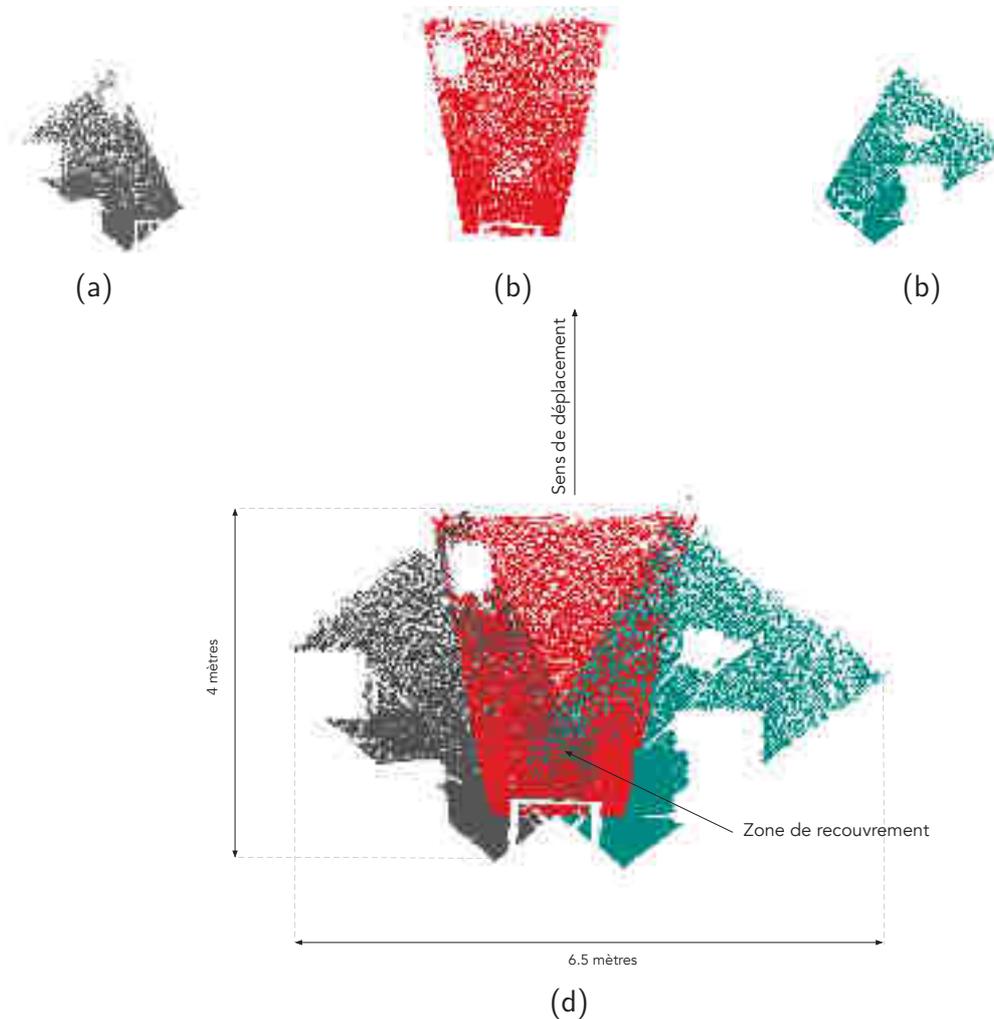


FIGURE 3.24: (a/b/c) Nuages de point issus de chacune des caméras - (d) Nuage fusionné

Afin de générer une orthoimage, il est nécessaire de déterminer la surface couverte au sol par un point. Compte tenu de la configuration géométrique de notre système (Figure 3.26), plus un point est éloigné, plus la surface qu'il va "couvrir" au sol va être importante.

Il est alors possible de déterminer Δy de la façon suivante :

$$\Delta y = \frac{\sqrt{h^2 + y^2} \tan \Delta \theta}{\cos \theta} \quad (3.2)$$

Nous pouvons ainsi imaginer exploiter différents attributs pour coloriser les orthoimages. Par exemple, les pixels de l'orthoimage peuvent être colorisés suivant l'altitude des points projetés. Cependant, nous pouvons également imaginer une colorisation fonction des informations RVB ou infrarouge du point.

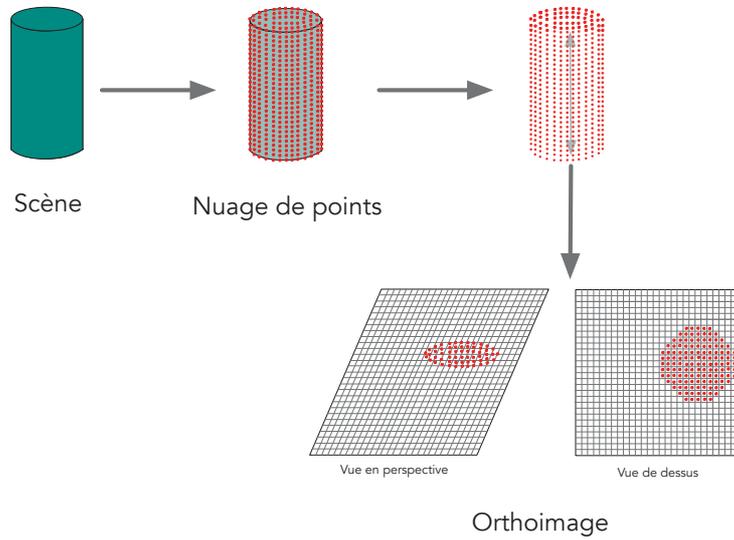


FIGURE 3.25 : Principe de l'orthoprojection développée dans notre approche

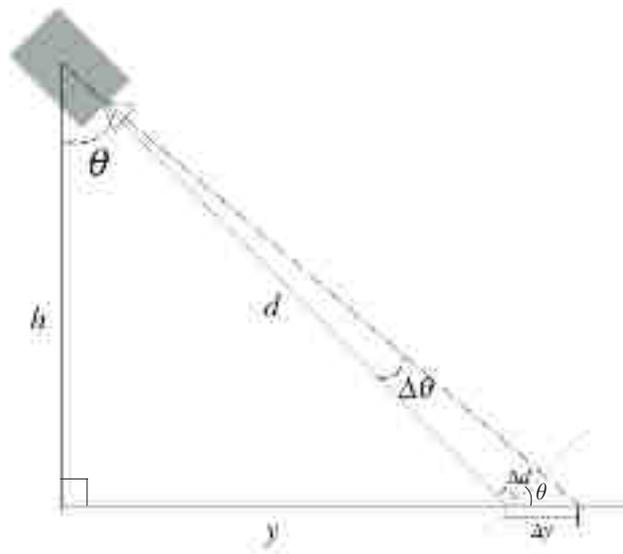


FIGURE 3.26 : Géométrie d'acquisition du système développé

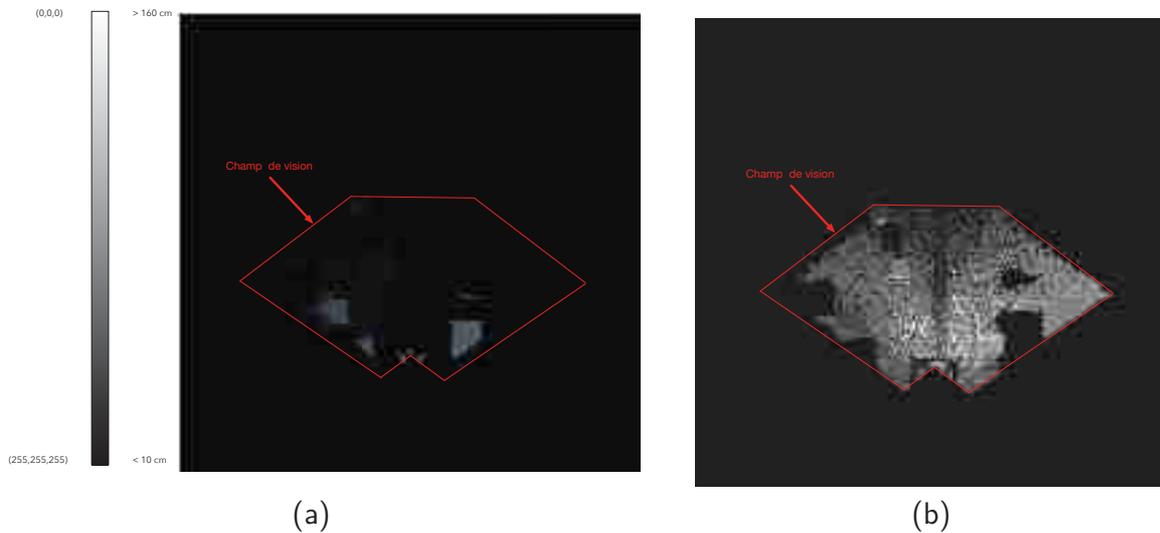


FIGURE 3.27: (a) Image d'orthoprojection colorisée suivant l'altitude des points - (b) Image d'orthoprojection colorisée en fonction des données infrarouge fournies par les caméras 3D (à partir du nuage présenté Figure 3.24)

Sur la Figure 3.27, nous pouvons observer deux exemples d'orthoimages. La colorisation en fonction de l'altitude des points est réalisée grâce à une table de correspondance (*Look Up Table*). Ainsi, nous colorisons un point ayant une altitude supérieure à 1.6 m en blanc. A contrario les points dont l'altitude est inférieure à 0.1 m sont colorisés en noir (Figure 3.27 (a)). Nous pouvons également observer une orthoimage obtenue à partir des données infrarouges des caméras 3D (Figure 3.27 (b)).

La représentation sous forme d'orthoimages présente plusieurs avantages :

- L'utilisation de l'altitude des points pour coloriser les orthoimages permet de conserver l'information tridimensionnelle ;
- Le traitement des données 2D est moins coûteux d'un point de vue traitement informatique ;
- Il est possible d'exploiter des algorithmes de traitement d'image 2D éprouvés.

Sur la base de ces orthoimages, notre méthodologie consistera à mettre en place une approche basée sur l'odométrie visuelle. Cette dernière consiste à évaluer le déplacement d'un mobile à partir de l'étude des images fournies par une caméra placée sur celui-ci. Dans notre cas, l'odométrie visuelle prendra, en entrée, les orthoimages obtenues à partir des nuages de points fournis par les 3 caméras 3D.

Le cahier des charges fixé par l'entreprise finançant ces travaux était très clair : il s'agissait d'évaluer la capacité des caméras 3D à permettre la localisation d'un système mobile en environnement urbain. Dans ce contexte, les développements réalisés ont visé à exploiter autant que possible ce capteur. Cependant, l'environnement dans lequel notre système va évoluer laisse à

penser qu'il sera nécessaire d'employer un second système de vision. La sous-partie suivante est consacrée à la présentation du pôle caméra omnidirectionnelle, deuxième système de vision qui pourrait intervenir dans le concept final.

3.3. Caméra omnidirectionnelle

Nos observations de l'environnement urbain nous ont permis de mettre en évidence la nécessité d'observer les bâtiments nous environnant. Il nous est alors apparu important d'introduire un capteur de vision permettant d'avoir une vue la plus globale possible, autorisant une vision à 360° horizontalement et 180° verticalement (Figure 3.28).

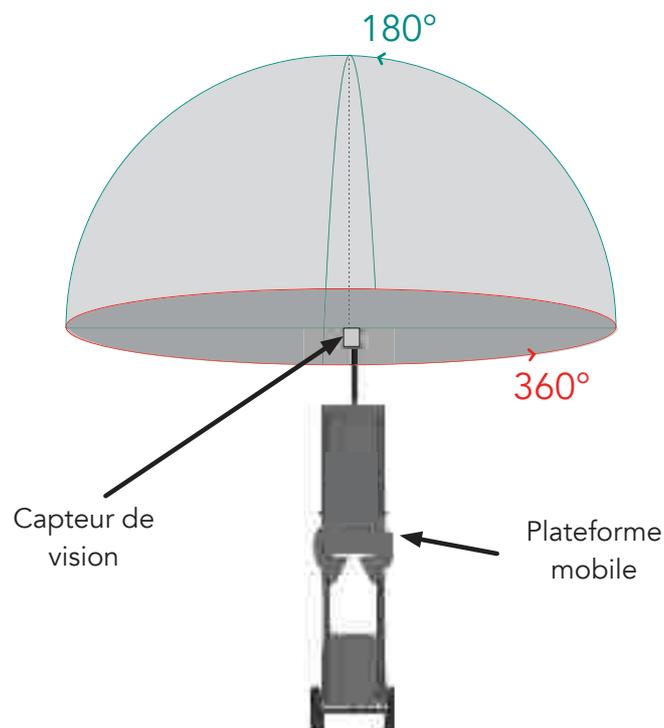


FIGURE 3.28: Champ de vision de la caméra omnidirectionnelle souhaité

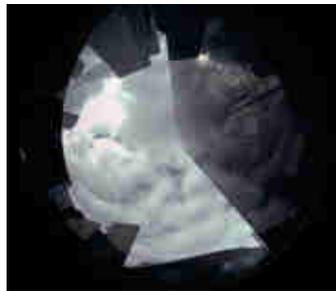
Comme nous allons le voir, différents types de capteur permettent d'obtenir une vision omnidirectionnelle. Nous reviendrons plus en détails sur le capteur choisi, et sur la calibration de celui-ci, étape nécessaire pour l'exploitation de ses données.

3.3.1. La vision omnidirectionnelle

Il existe plusieurs façons d'accroître le champ de vision d'une caméra. Nous trouvons tout d'abord l'utilisation de lentilles spécifiques permettant d'élargir le champ visuel d'une caméra conventionnelle. Les objectifs hypergones, plus communément appelés *Fish Eye* emploient des lentilles permettant d'obtenir un champ visuel plus large. Leurs distances focales courtes per-



(a)



(b)



(b)

FIGURE 3.29: (a/b) Exemples d'images obtenues à partir d'un objectif *Fish Eye* - (c) Lentille *Fish Eye* (Fujinon)

mettent l'observation d'une demi-sphère. La Figure 3.29 présente un exemple d'image acquise par une caméra équipée de ce type d'objectif.

L'utilisation combinée d'une caméra et d'un objectif *Fish Eye* présente l'avantage d'être facile à mettre en œuvre. Cependant, de fortes distorsions sont visibles sur l'image.

La seconde solution consiste à générer une mosaïque à partir d'une série d'images issues d'une ou de plusieurs caméras conventionnelles. Nous pouvons notamment citer le système proposé par [Cutler et al., 2002]. Les images acquises par leurs caméras sont traitées en temps réel en vue de générer une image panoramique.

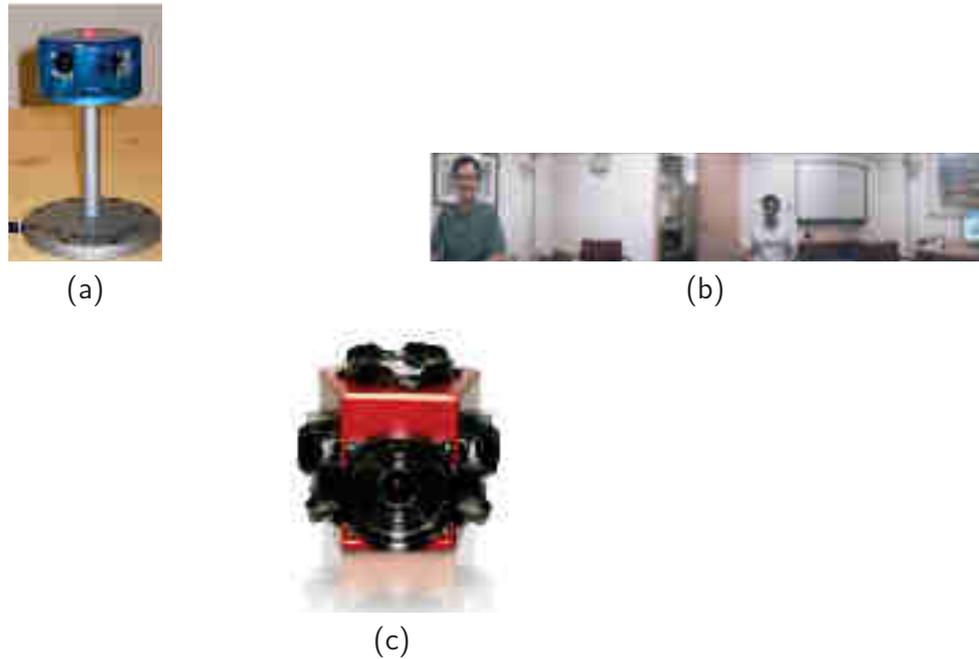


FIGURE 3.30 : (a) Capteur panoramique RingCam [Cutler et al., 2002] - (b) Image panoramique produite par la caméra RingCam - (c) Caméra LadyBug (PointGrey)

Nous pouvons également citer des solutions commerciales, comme par exemple celle proposée par Point Grey (Figure 3.30(c)). Plus simple à mettre en place que les systèmes "artisanaux", elle présente l'inconvénient d'être onéreuse comparativement à la solution que nous souhaitons développer.

Enfin, la troisième solution consiste à coupler des miroirs à des lentilles conventionnelles (Figure 3.31). Il s'agit alors d'observer le projection de l'espace sur le miroir par l'intermédiaire de l'imageur. Nous parlons alors de systèmes catadioptriques. Ces solutions présentent l'avantage d'être très répandues, cependant, elles ont l'inconvénient de ne pas offrir un dôme complet de vision. Enfin, elles sont moins compactes, et donc plus difficiles à intégrer dans une solution de capteur.

3.3.2. Choix du modèle de caméra omnidirectionnelle

Les systèmes de type objectifs *Fish Eye* répondent à nos besoins en terme de champ de vision, tout en restant abordables. Nous avons donc choisi d'orienter notre choix vers ce type de solution.

Ainsi, nous nous sommes tournés vers le modèle *Gige Vision MvBlueCougar_X125a*, proposé par la société Matrix Vision. Cette caméra est équipée d'un capteur CCD ayant une résolution de 2448×2050 pixels. Nous avons équipé celle-ci d'un objectif Fujinon *FE 185 CO57HA-1*, permettant d'obtenir un angle de vue de $185^\circ \times 185^\circ$.

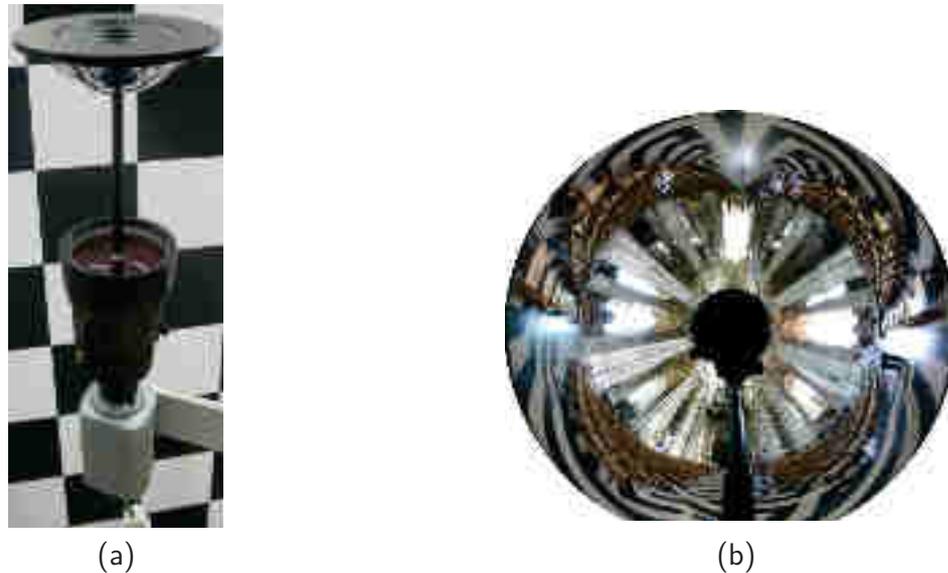


FIGURE 3.31 : (a) Caméra catadioptrique - (b) Image omnidirectionnelle à l'intérieur de la cathédrale d'Amiens obtenue avec un objectif catadioptrique.



FIGURE 3.32 : (a) Caméra *MvBlueCougar_125a* de Matrix Vision équipée de l'objectif Fujinon *FE 185 CO57HA-1* - (b) Exemple d'image capturée avec cette configuration

3.3.3. Calibration

La calibration de l'optique installée sur la caméra est une étape essentielle. Celle-ci revient à modéliser le processus de formation des images, c'est-à-dire trouver la relation entre les coordonnées spatiales d'un point de l'espace avec le point associé dans l'image prise par la caméra. Cette étape permet alors l'utilisation d'images dans de nombreux travaux de mesure, tels que la modélisation en 3D, mais également la détermination de l'orientation externe d'une caméra

(position et orientation).

a. Le modèle de sténopé

On appelle modèle de caméra, l'ensemble des lois géométriques définissant la façon dont se projette, lors du processus de saisie d'une image, un point de l'espace à trois dimensions sur un plan à deux dimensions. Le modèle le plus simple et le plus souvent utilisé est le modèle de sténopé. Géométriquement, le modèle de sténopé repose sur une projection centrale ou perspective [Faugeras, 1993], [Hartley & Zisserman, 2003]. Celui-ci suppose que le système optique de la caméra respecte les conditions de Gauss, c'est-à-dire que l'angle entre les rayons incidents et l'axe optique est faible, et que les rayons lumineux passent par le centre optique. Pour chaque point P de la scène, nous considérons ainsi la droite passant par p et par le centre optique C comme étant le trajet de la lumière perçue par la caméra en provenance de P . L'intersection de cette droite avec le plan image est la projection p du point P (Figure 3.33).

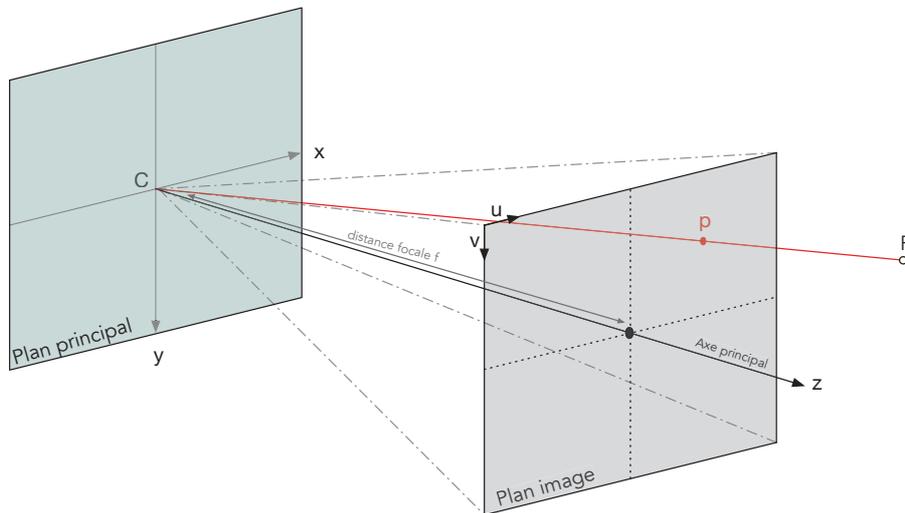


FIGURE 3.33 : Le modèle de sténopé [Hartley & Zisserman, 2003]

Le point P a pour coordonnées (X_C, Y_C, Z_C) dans le repère \mathcal{R}_C de la caméra. Nous pouvons alors exprimer les coordonnées de sa projection p de coordonnées (x, y) dans le plan image :

$$x = f \cdot \frac{X_C}{Z_C}, y = f \cdot \frac{Y_C}{Z_C} \quad (3.3)$$

Le passage en coordonnées homogènes permet de réécrire la projection perspective sous forme linéaire, on aura alors :

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} \quad (3.4)$$

Il est ensuite nécessaire de passer des coordonnées métriques (x, y) du point p dans le plan image à ses coordonnées pixel (u, v) . Les coordonnées pixel représentent les coordonnées d'un pixel sur une image, elle même étant une discrétisation du plan image (Figure 3.34).

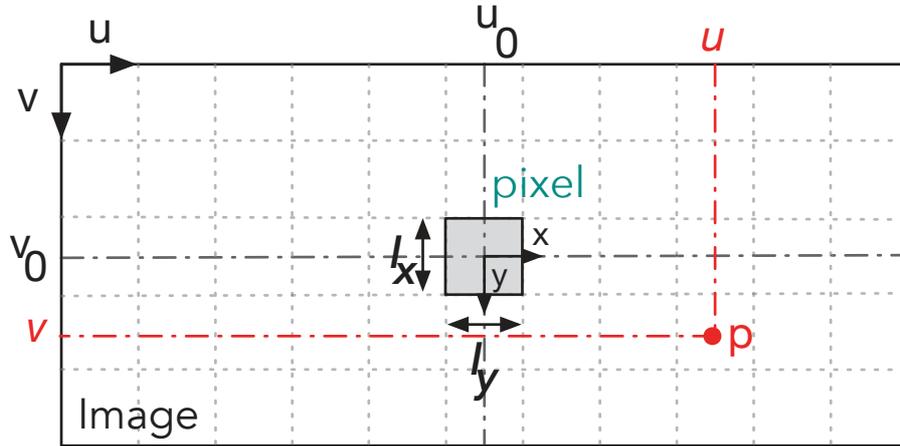


FIGURE 3.34 : Coordonnées pixel (u, v) et métriques (x, y) d'un pixel p

Les coordonnées pixel dépendent des dimensions d'un pixel (l_x, l_y) ainsi que des coordonnées pixel du point principal dans l'image (u_0, v_0) :

$$\begin{aligned} u &= u_0 + \frac{x}{l_x} \\ v &= v_0 + \frac{y}{l_y} \end{aligned} \quad (3.5)$$

Cette relation peut alors s'écrire dans une représentation homogène sous la forme :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{l_x} & 0 & u_0 \\ 0 & \frac{1}{l_y} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (3.6)$$

Nous pouvons alors écrire :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{l_x} & 0 & u_0 \\ 0 & \frac{1}{l_y} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X_C \\ Y_C \\ Z_C \\ 1 \end{pmatrix} \quad (3.7)$$

Le modèle de sténopé de la caméra dépend donc de quatre paramètres, dit paramètres intrinsèques : les coordonnées du point principal (u_0, v_0) dans l'image, et les deux paramètres de changement d'échelle $(\frac{f}{l_x}, \frac{f}{l_y})$.

L'hypothèse faite sur le respect des conditions de Gauss n'est jamais vérifiée. En effet, les conditions de Gauss ne sont pas respectées : il est nécessaire de tenir compte des erreurs de fabrication du dispositif.

Ces défauts vont engendrer des distorsions de deux types :

- Les distorsions tangentielles : dues à de légers défauts d'alignement entre les différentes parties optiques de la caméra (lentilles, diaphragme, capteur).
- Les distorsions radiales : dues à la position du diaphragme dans le système optique et aux distorsions des optiques ; plus le diaphragme est loin du centre optique, plus la distorsion radiale est forte.

De nombreux modèles ont été décrits afin de corriger ces distorsions. Le modèle de Brown [Brown, 1971] est très répandu, il permet de corriger les coordonnées-images à l'aide de coefficients déterminés par calibration.

b. Le modèle de caméra *Fish Eye*

Le modèle de projection à perspective centrale (ou modèle de sténopé) ne permettant pas de projeter l'ensemble du champ de vision sur un plan, il a fallu définir de nouveaux types de projections. Ces nouvelles projections sont appelées "fonctions de représentation". Les équations de projections les plus utilisées dans les objectifs *Fish Eye* actuels sont les suivantes :

- Projection équidistante :

$$r = f \times \theta \quad (3.8)$$

- Projection orthographique :

$$r = f \times \sin(\theta) \quad (3.9)$$

- Projection à angle équisolide :

$$r = k_1 \times f \times \sin\left(\frac{\theta}{k_2}\right) \quad (3.10)$$

- Projection stéréographique :

$$r = k_1 \times f \times \tan\left(\frac{\theta}{k_2}\right) \quad (3.11)$$

avec :

- r : distance radiale : $\sqrt{u^2 + v^2}$,
- f : distance focale de l'objectif,
- θ : angle d'incidence,
- k_1, k_2 : coefficients constants fixés par le constructeur de la lentille.

c. Calibration d'une optique *Fish Eye*

Il existe différentes approches pour réaliser la calibration d'un objectif *Fish Eye* [Puig et al., 2012]. Nous allons en détailler deux.

La **première approche** fait intervenir l'équation de projection. Dès lors que celle-ci est connue, il est possible de modéliser les distorsions en prenant en compte le modèle de Brown [Brown, 1971]. Dans le cadre d'un stage de fin d'étude d'études à l'INSA de Strasbourg, réalisé en soutien de la thèse, [Vincent, 2013] a exploité ce modèle afin de développer une application de calibration pour objectif *Fish Eye* "imcalib". Un résumé de ces travaux est présenté en annexe (Annexe G). Cette application présente l'avantage de ne nécessiter qu'un seul cliché pris avec la caméra à calibrer et avec une mire de calibration développée à cette fin. Cependant, le mode opératoire pour réaliser cet unique cliché est difficile à mettre en œuvre. En effet, il est nécessaire que la mire photographiée soit strictement parallèle au capteur et que le rayon émergent du point principal soit aligné avec le centre de la mire. De plus, il est nécessaire de connaître l'équation de projection pour utiliser cette application.



FIGURE 3.35 : Dispositif de prise de cliché en vue de la calibration [Vincent, 2013]

Nous trouvons également dans la littérature plusieurs approches visant à déterminer cette équation de projection. Nous pouvons citer les travaux de [Bakstein & Pajdla, 2002], dont l'approche permet de déterminer la distance radiale par rapport à l'angle d'incidence et de déterminer ensuite par l'intermédiaire de l'algorithme de Levenberg-Marquart [Moré, 1978], l'équation de projection de leur caméra. Cette approche est concluante, cependant le mode opératoire peut paraître difficile à mettre en œuvre. En effet, il est nécessaire de tracer sur un cylindre des cercles correspondants à des angles d'incidence espacés de 5° à 10° . Il faut ensuite positionner la caméra de façon à ce que l'axe du cylindre et le rayon lumineux passant par le point principal de la caméra soient confondus. De plus, le lieu géométrique où se rejoignent l'ensemble des rayons lumineux doit être le plan passant par le cercle correspondant à l'angle d'incidence de 180° .



FIGURE 3.36 : Mode opératoire pour la détermination de l'équation de projection [Bakstein & Pajdla, 2002]

La **seconde approche** de calibration qu'il nous paraît intéressant d'évoquer, a été étudiée pour répondre aux difficultés à évaluer l'équation de projection. Cette approche repose sur la représentation sur une sphère virtuelle et sur l'utilisation d'une fonction polynomiale afin de calibrer l'objectif. Ainsi, [Geyer & Daniilidis, 2000] ont exprimé la projection des différents cônes de révolution d'une caméra catadioptrique sur une sphère virtuelle. Ce modèle a été transposé dans [Ying & Hu, 2004] aux caméras *Fish Eye*. Enfin, nous trouvons dans les travaux de [Scaramuzza, 2008] un modèle unifié reposant sur l'utilisation d'une fonction polynomiale de Taylor. Ce modèle présente l'intérêt de permettre l'utilisation d'algorithmes valables pour des caméras "standard", à condition d'utiliser des coordonnées homogènes. Dans [Scaramuzza, 2007], l'auteur propose une *toolbox* "Ocamcalib" permettant d'exploiter cette approche. À partir de celle-ci, il est possible de déterminer les coordonnées pixel du point principal, les paramètres c, d, e qui sont les paramètres qui décrivent les potentiels défauts d'alignement entre le miroir et la caméra, et les paramètres a_0, a_1, a_2, \dots , qui correspondent aux coefficients du polynôme de Taylor décrivant la fonction de projection associée à la lentille : $f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \dots + a_n \cdot x^n$.

Dans ce contexte, notre approche a tout d'abord consisté à exploiter la *toolbox* fournie par [Scaramuzza, 2007]. Les résultats obtenus sont satisfaisants. En effet, en tenant compte des fortes distorsions de la lentille, l'écart type de 1.36 pixels permet l'utilisation des paramètres de calibration dans les algorithmes de localisation que nous souhaitons développer. Cependant l'acquisition des clichés nécessaires à la calibration est réalisée à 10 cm de la mire, est-ce cohérent avec l'approche que nous souhaitons mener ?

Nous avons donc contacté Fujinon à plusieurs reprises afin d'obtenir l'équation de projection de notre lentille. À défaut d'obtenir l'équation de projection, l'entreprise nous a fourni le tableau de valeurs de la Figure 3.37.

FO DISTORTION

$$FO \text{ Distortion} [\%] = \frac{f0 \text{ ideal image height [mm]} - \text{image height on Sensor [mm]}}{\text{ideal image height [mm]}} \times 100$$

Image height		Half angle of view	FO Distortion
(%)	[mm]	[degrees]	(%)
100	2.85	82.51	0.00
90	2.97	82.55	0.00
80	2.28	73.15	0.36
70	2.00	63.08	0.43
60	1.71	54.85	0.39
50	1.43	45.75	0.30
40	1.14	36.03	0.20
30	0.86	27.50	0.12
20	0.57	18.34	0.05
10	0.29	9.18	0.01

FIGURE 3.37 : Tableau de valeurs fourni par Fujinon

Ce tableau est intéressant, car il indique la distance radiale en fonction de l'angle d'incidence. Dans un premier temps, l'utilisation de celui-ci a permis d'éliminer certaines fonctions de représentations. Dans le cadre des travaux menés par [Royère, 2014], durant un stage de fin d'étude à l'INSA de Strasbourg, réalisé en soutien de la thèse, une approche de calibration similaire à celle présentée dans [Bakstein & Pajdla, 2002] a pu être développée. Cette méthodologie nous permet d'aboutir à un résultat dont le résidu maximal est inférieur à 1,5 pixels. Le lecteur intéressé pourra retrouver en annexe (Annexe H) un résumé des développements nécessaires pour aboutir à ce résultat.

Ainsi, nous avons pu mettre en place la calibration de l'objectif en utilisant deux approches différentes. La première approche mise en place peut être vue comme une calibration "analytique", elle se base sur les travaux de [Scaramuzza, 2007], qui approxime le comportement de la lentille par une fonction polynomiale. Elle présente l'avantage de permettre l'utilisation d'algorithmes valables pour tout type de projection. La seconde approche mise en place, peut quant à elle, être vue comme une calibration "géométrique". Elle se base sur l'équation de projection de la lentille et se veut plus précise. Cependant, elle présente l'inconvénient principal de ne pas être directement utilisable dans les calculs de transformation, du fait de la non-linéarité de l'équation de projection estimée.

Nous nous sommes donc tournés vers le premier type de calibration. Il est nécessaire de réaliser un contrôle dans des conditions proches de notre cas d'étude. Les travaux menés par [Royère,

2014] ont permis la mise au point d'une application permettant de contrôler la calibration réalisée.

Le programme de contrôle développé par [Royère, 2014] permet le calcul de la matrice caméra, et la déduction des paramètres d'orientation externe (centre de l'objectif et orientation de la caméra). Celui-ci nous permet de comparer les coordonnées mesurées sur l'image et les coordonnées issues de la reprojection des points d'intérêts grâce à la matrice des paramètres intrinsèques de la caméra.

Une campagne de mesure a permis la détermination d'un canevas de points fixes sur les façades des bâtiments de l'INSA de Strasbourg. Nous nous sommes donc servis de ces points fixes pour contrôler la calibration.

Plusieurs prises de vue des bâtiments de l'INSA ont été réalisées. Celles-ci ont été faites dans différentes positions, ainsi nous avons orienté l'objectif vers le haut afin de retrouver les conditions finales d'utilisation. L'objectif a également été orienté de façon plus classique pour permettre d'évaluer la calibration dans un cadre plus général. La position approximative de l'objectif a également été mesurée à l'aide d'une station totale, ceci dans un souci de contrôle.

Nous pouvons trouver dans le Tableau 3.2 les résultats obtenus pour trois clichés.

Photos			
	(1)	(2)	(3)
Nombre de points mesurés	27	16	14
Position de l'objectif en m (station totale)	$\begin{pmatrix} 1000072.60 \\ 112368.60 \\ 140.65 \end{pmatrix}$	$\begin{pmatrix} 1000072.70 \\ 112368.50 \\ 140.60 \end{pmatrix}$	$\begin{pmatrix} 1000072.70 \\ 112368.50 \\ 140.60 \end{pmatrix}$
Position de l'objectif en m (Calculée)	$\begin{pmatrix} 1000072.75 \\ 112368.53 \\ 139.66 \end{pmatrix}$	$\begin{pmatrix} 1000075.21 \\ 112362.42 \\ 141.59 \end{pmatrix}$	$\begin{pmatrix} 1000071.94 \\ 112368.06 \\ 140.55 \end{pmatrix}$
Écart de positionnement en m	$\begin{pmatrix} 0.15 \\ 0.07 \\ 0.99 \end{pmatrix}$	$\begin{pmatrix} 2.49 \\ 6.08 \\ 1.01 \end{pmatrix}$	$\begin{pmatrix} 0.76 \\ 0.56 \\ 0.05 \end{pmatrix}$
Moyenne des résidus	3.7 <i>pixels</i>	1.5 <i>pixels</i>	2.2 <i>pixels</i>
Résidu max	10.3 <i>pixels</i>	3.5 <i>pixels</i>	6.4 <i>pixels</i>

TABLE 3.2: Tableau comparatif des écarts et résidus de 3 clichés réalisés autour des bâtiments de l'INSA de Strasbourg

À la lecture des résultats, nous pouvons affirmer que la calibration mise en place à partir de l'approche proposée par [Scaramuzza, 2008] est validée. En effet, même si nous pouvons noter des imprécisions sur le calcul de la position de l'objectif, les résidus sur la reprojection des points d'appuis sont corrects. Dans le Tableau 3.2, nous pouvons remarquer que la précision est correcte lorsque la répartition des points est uniforme (photo (1)). Il est important de noter que cette précision chute lorsque les points d'appuis sont coplanaires ou mal répartis (photo (2)). Cependant, compte tenu des distorsions mises en jeu, nous pouvons considérer ce résultat comme tout à fait acceptable pour l'approche que nous souhaitons développer.

Comme nous avons pu le décrire dans le chapitre précédent, il existe de nombreuses approches sur la base d'images. Les différentes études menées autour du pôle de caméras 3D nous ont amenés à privilégier l'odométrie visuelle. Naturellement, nous nous tournons de façon identique vers cette approche pour le pôle caméra *Fish Eye*.

3.4. Aspect fonctionnel

Après avoir établi le concept final du point de vue des capteurs à mettre en oeuvre et d'un point de vue mécanique, cette section a pour but de présenter l'aspect fonctionnel relatif à notre solution. Ainsi, nous présentons les outils informatiques nécessaires au bon fonctionnement de l'ensemble des capteurs. De plus, nous présentons les processus mis en place pour assurer la synchronisation des différents flux de données. Enfin, nous présentons l'infrastructure matérielle de notre système.

3.4.1. Solutions de développement

Afin d'implémenter l'algorithme de localisation développé autour du concept élaboré, nous avons choisi les solutions suivantes :

- développement des différents algorithmes en C++,
- utilisation des bibliothèques OpenCV et PCL,
- exploitation du "middleware"¹ ROS.

a. Le langage C++

Le langage C++ est un langage libre de droits développé durant les années 1980 par Bjarne Stroustrup. Celui-ci est une évolution orientée objet du langage C de Brian Kernighan et Denis Ritchi. Cependant, il repose sur le paradigme de Programmation Orientée Objet (POO). Il est aujourd'hui un des langages de programmation les plus utilisés en matière de développement informatique. Sa relative simplicité en fait un outil attractif, ainsi que les nombreux outils compatibles avec celui-ci. Ces arguments nous ont poussés à l'utiliser.

1. Un middleware ou intergiciel est un réseau d'échanges entre différentes applications informatiques. Dans le cas de ROS, il s'agit de faire communiquer différents sous-ensembles d'un robot.

b. OpenCV



FIGURE 3.38 : Logo d'OpenCV

OpenCV (*Open Computer Vision*) est une bibliothèque graphique libre, écrite en C++, initialement développée par Intel. Celle-ci est spécialisée dans le traitement d'images en temps réel. Depuis 2008, la société Willow Garage en assure le support. Cette librairie met à disposition de nombreuses fonctionnalités dont la plupart des opérations classiques en traitement bas niveau des images :

- lecture, écriture affichage d'une image ;
- lissage, filtrage ;
- détection de contours ;
- détection de points d'intérêts ;
- segmentation ;
- etc.

OpenCV comprend également différents modules de plus haut niveau tel que les entrées/sorties et l'interface utilisateur permettant de réaliser des interfaces graphiques très simples. Nous y trouvons également un module de calibration de caméras.

Au cours de nos développements, OpenCV nous a permis de générer les orthoimages nécessaires à notre approche, mais également de travailler sur la détection de points d'intérêts et la mise en correspondance entre images.

c. PCL

FIGURE 3.39 : Logo de PCL (*Point Cloud Library*)

PCL (*Point Cloud Library*) est une bibliothèque libre écrite en C++, spécialisée dans le traitement de nuage de points et dans la géométrie 3D. Cette librairie est développée par la société Willow Garage depuis mars 2010. Elle met à disposition de nombreux outils permettant de gérer

les données 3D et notamment les nuages de points. Elle offre un ensemble d'algorithmes, notamment pour la reconstruction de surface, pour la segmentation ou encore pour la détection de points caractéristiques 3D.

PCL nous a permis d'exploiter les nuages de points fournis par les caméras 3D de notre système et ainsi d'en extraire les informations nécessaires à la formation des orthoimages, base de notre approche.

d. ROS



FIGURE 3.40 : Logo de ROS (*Robot Operating System*)

ROS (*Robot Operating System*) est un ensemble d'outils informatiques libres permettant de développer des logiciels pour la robotique. Développé par la société Willow Garage depuis 2007, nous retrouvons au sein de ROS les bibliothèques OpenCV et PCL. Les langages de programmation disponibles au sein de ROS sont le C++ et le Python, mais il est également possible de travailler avec Lisp ou encore R.

ROS fournit un ensemble d'outils qui permettent de développer et déployer différentes applications robotiques. Ainsi, on retrouvera sous forme de paquets un nombre important de briques algorithmiques, mais également des paquets propres à la gestion des capteurs.

ROS est un environnement reposant sur la notion de nœud. Chaque nœud possède un rôle qui lui est propre. La communication entre les nœuds est réalisée grâce à l'utilisation de messages transitant par l'intermédiaire d'un "démon".

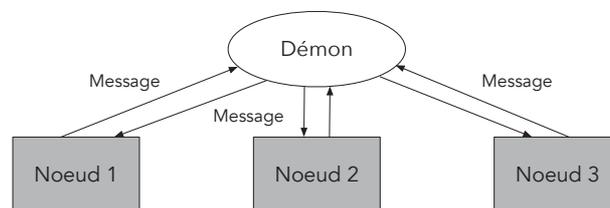


FIGURE 3.41 : Principe de communication dans ROS (www.ros.org)

Ces messages sont échangés grâce au mécanisme "*Publish/Subscribe*" à travers un bus de communication appelé ROS-Topic.

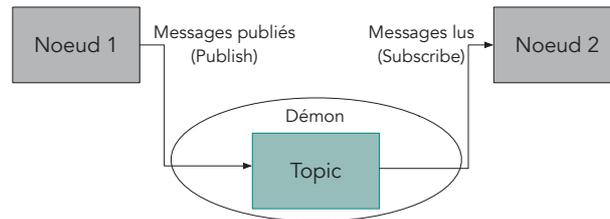


FIGURE 3.42 : Principe de communication dans ROS (www.ros.org)

Ainsi, le noeud 1 publie des messages sur ce que l'on appelle un "topic". Le noeud 2 est abonné à ce "topic" afin de pouvoir lire les messages publiés sur celui-ci.

La réception d'un message au sein d'un noeud permet le déclenchement d'algorithme, eux mêmes générateurs de messages.

Une des forces de ROS est sa capacité à faire abstraction du matériel. En effet, les messages qui transitent au sein de ROS sont des structures de données comportant des données typées. Par exemple, le noeud permettant de gérer les capteurs de type laser vont générer des messages de type `sensor_msgs/LaserScan` quel que soit le modèle de capteur utilisé. Cette spécificité présente un avantage certain, car elle nous permet de rendre nos développements compatibles avec tous les modèles de caméras 3D ou *fish-eye*.

Enfin ROS permet de travailler sur la synchronisation des messages reçus. Les noeuds permettant de gérer les capteurs offrent la possibilité d'enregistrer les données fournies par ceux-ci. Ainsi, sans modifier la fréquence d'acquisition des capteurs, il est possible de déclencher des actions lorsque plusieurs capteurs publient des messages dans un intervalle de temps défini.

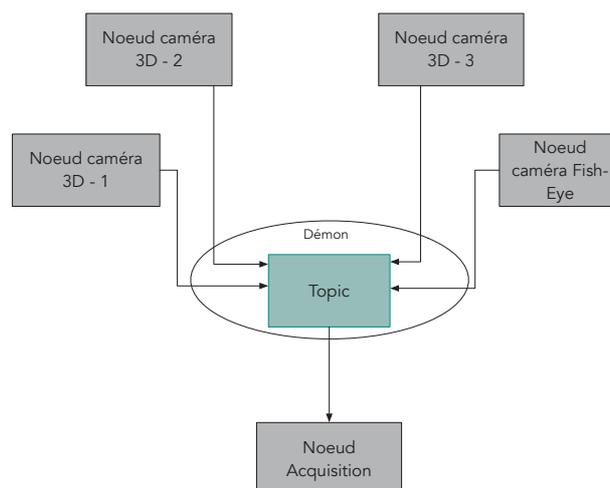


FIGURE 3.43 : Communication mise en place au sein de notre prototype

Le nœud permettant de gérer l'acquisition des données fournies par notre système est donc à l'écoute des 4 nœuds gérant chacune des caméras.

Lors du démarrage de l'application, chacun des nœuds des caméras va émettre des messages contenant les données issues de celles-ci. Chacun de ces messages émis est horodaté sur une base commune. Ainsi, lorsque chaque nœud émet un message dans un intervalle de temps inférieur à 1 *ms*, nous considérons que l'ensemble des messages est synchronisé. Nous enregistrons ensuite chacun des 4 messages. Lors de nos acquisitions, cette méthodologie a été mise en place. Ainsi, à chaque instant de l'acquisition, trois nuages de points et une image, issue de la caméra *fish-eye*, sont enregistrés.

3.4.2. Infrastructure matérielle

Il s'agit ici de décrire l'aspect matériel de notre prototype. Comme nous l'avons présenté auparavant, les deux pôles de vision le composant sont constitués de deux types de caméras. Nous trouvons en premier lieu les caméras 3D, qui sont de type Asus Xtion Pro. Au sein du second pôle, nous trouvons une caméra *Gige Vision MvBlueCougar_X125a* proposée par la société Matrix Vision. Cette caméra est équipée d'un capteur CCD ayant une résolution de 2448 x 2050 pixels. Nous avons équipé celle-ci d'un objectif Fujinon *FE 185 CO57HA-1*.

Comme nous l'avons présenté dans le paragraphe 3.2.2. de ce chapitre, les caméras 3D sont assemblées grâce à un montage mécanique. Le concept final prévoit également le montage de la caméra *Fish Eye* de façon à ce que celle-ci soit orientée vers le ciel (Figure 3.44). Cet ensemble est installé sur un chariot équipé de 4 roues, qui rentre dans la catégorie des tricycles (Chapitre 2 - paragraphe 3.3.2.), permettant de réaliser des acquisitions en environnement urbain (Figure 3.45).

Durant les acquisitions de données, le chariot est équipé d'un ordinateur portable. Celui-ci sert à l'enregistrement des données fournies par chacun des capteurs du prototype. Chaque caméra 3D est reliée à cet ordinateur par câble USB. Ainsi, leur alimentation en courant est assurée par l'ordinateur. Pour pouvoir exploiter ces trois caméras de façon synchronisée, il a fallu les connecter à trois bus USB indépendants. Cette contrainte nécessaire au bon fonctionnement du pilote des caméras permet également de conserver un débit suffisant pour l'enregistrement des données. La caméra *Fish Eye* est reliée à l'ordinateur par réseau Ethernet. L'alimentation de cette caméra est réalisée grâce à une batterie externe ayant une autonomie importante. L'autonomie du système ne dépend alors que de l'autonomie de l'ordinateur portable utilisé pour l'acquisition.



FIGURE 3.44 : Assemblage de l'ensemble des pôles de vision

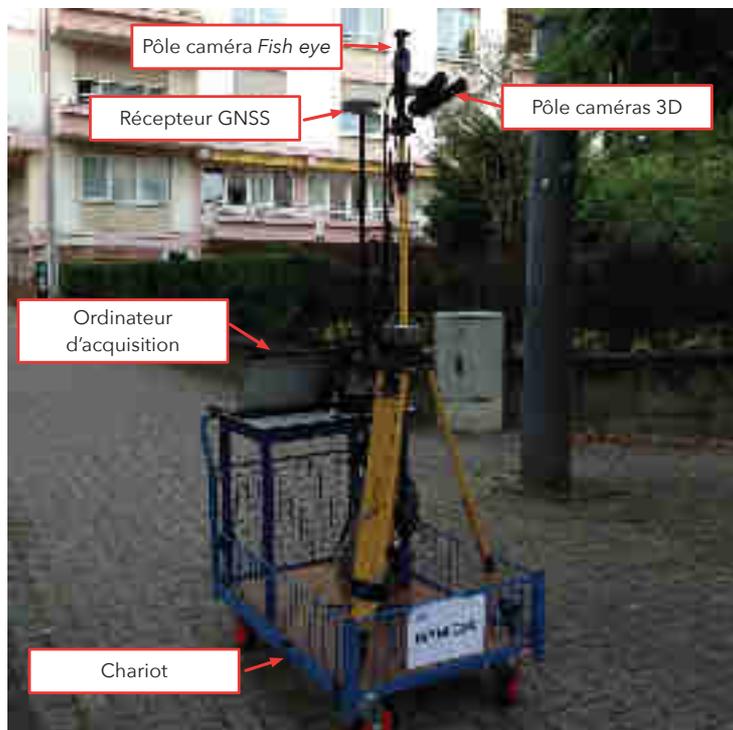


FIGURE 3.45 : Prototype en condition d'acquisition

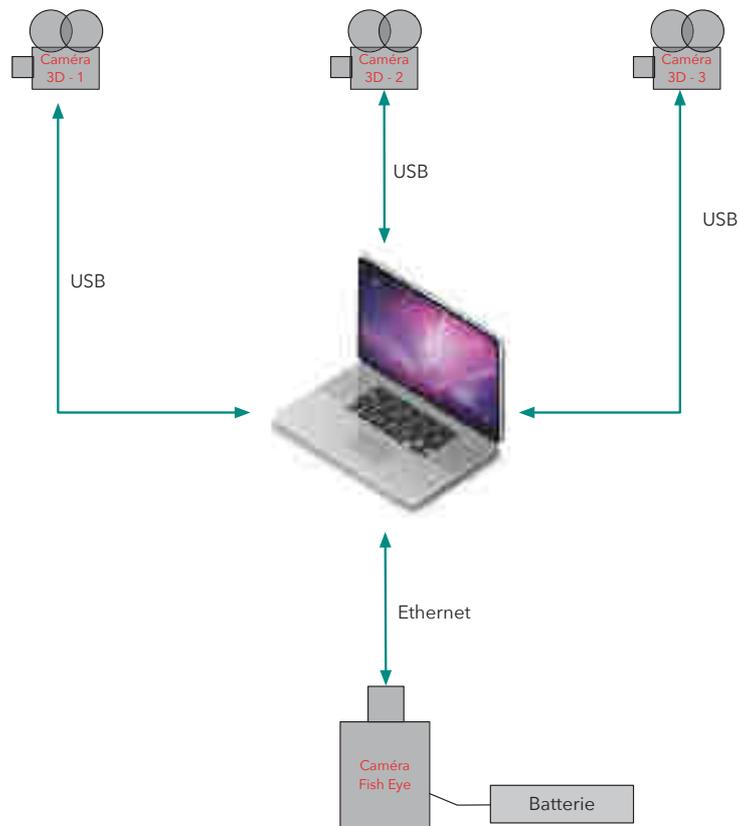


FIGURE 3.46 : Schéma de l'architecture matérielle du prototype

Conclusions du chapitre 3

Dans ce chapitre, nous avons décrit l'ensemble des concepts ayant mené à la réalisation du prototype final élaboré au cours de cette thèse. Nous avons également présenté les outils informatiques utilisés en vue d'exploiter ce prototype, de même que l'infrastructure matérielle du prototype. Le concept final développé s'articule donc autour de deux pôles :

- Le pôle principal est composé de caméras 3D. Celles-ci sont installées sur le mobile en déplacement de façon à observer le mobilier urbain. Nous obtenons ainsi une vue locale et précise de la scène au sol. Dans ce contexte, la première étape de la méthodologie que nous développons consiste à générer des orthoimages représentant le sol et le mobilier urbain. À partir de celles-ci, une approche d'odométrie visuelle s'est imposée.
- Le pôle secondaire, qui vient compléter le premier, est une caméra de type *Fish Eye*. Contrairement aux caméras 3D, la caméra *Fish Eye* est orientée vers le ciel afin d'obtenir une vue globale de la scène. La projection particulière de ce type de caméra nous a poussés à mettre en place une étape de calibration propre à notre modèle de caméra. La méthode de calibration choisie nous permet d'exploiter des approches développées pour des objectifs standards. Sur la base des images obtenues par cette caméra, nous avons décidé de mettre en place une approche d'odométrie visuelle.

À partir des deux systèmes de vision intégrés dans notre prototype, l'objectif final de notre méthodologie de localisation sera de produire une trajectoire finale, qui profitera des atouts de chacun de ces pôles de caméras. Nous reviendrons plus en détails sur les différentes étapes composant cette méthodologie dans le chapitre suivant.

Algorithme de localisation

1	Introduction	83
2	État de l'art sur l'odométrie visuelle	84
3	Pôle caméras 3D	88
3.1.	Présentation de l'approche développée	88
3.2.	Détection de points d'intérêts	93
3.3.	Mise en correspondance	103
3.4.	Élimination des correspondances aberrantes et estimation de la transformation	107
3.5.	Calcul de la transformation	112
4	Pôle caméra <i>Fish Eye</i>	114
4.1.	Détection de points d'intérêts	115
4.2.	Mise en correspondance	116
4.3.	Élimination des correspondances aberrantes	116
4.4.	Calcul de la transformation	117
4.5.	Calcul de position moyenne	119

1 Introduction

Dans le chapitre précédent, nous avons présenté le prototype proposé et développé dans le cadre de cette thèse. À présent, le second volet de la thèse peut être abordé. Il s'agit de proposer et développer un algorithme de localisation permettant, à une plateforme mobile, de se localiser en environnement urbain à partir des données fournies par le prototype développé.

Ce quatrième chapitre débutera par un état de l'art sur l'odométrie visuelle, les approches développées pour le pôle caméras 3D et pour le pôle caméra *Fish Eye* reposant sur cette technique. Nous aborderons chacune des étapes qui composent la méthodologie mise en place plus particulièrement pour le pôle caméras 3D et présenterons les travaux entrepris pour le pôle caméra *Fish Eye*.

2 État de l'art sur l'odométrie visuelle

Quel que soit le système de vision utilisé, les images acquises par celui-ci sont porteuses d'informations (information de colorimétrie, de distance dans le cas d'images de profondeurs, de luminescence, ou encore de géométrie). Les travaux traitant de la problématique de localisation et exploitant les systèmes de vision sont nombreux. Comme nous avons pu le voir dans le Chapitre 2 (section 3.5.), nous trouvons tout d'abord différentes techniques développées par la communauté des spécialistes en vision qui travaillent sur la problématique de "Structure from Motion". Celle-ci consiste à reconstruire, à partir d'une séquence d'images, la géométrie de l'environnement et le déplacement effectué par une caméra. Les premiers travaux dans ce domaine ont été menés par [Longuet-Higgins, 1987] et [Harris & Pike, 1987]. Nous pouvons considérer l'odométrie visuelle comme un cas particulier du problème de "Structure from Motion". En effet, l'odométrie visuelle est l'opération qui consiste à estimer le déplacement d'un mobile (véhicule, humain ou robot) en utilisant uniquement les données obtenues grâce à un système de vision installé dessus. Les recherches menées dans ce domaine ont débuté dans les années 1980, et le terme odométrie visuelle a fait son apparition en 2004 [Nistér et al., 2004]. Les domaines d'applications sont nombreux, et nous pouvons citer : la robotique [Nistér et al., 2004], la réalité augmentée [Oskiper et al., 2011], ou encore l'automobile [Scaramuzza et al., 2009]. L'odométrie visuelle présente de nombreux avantages. Nous pouvons citer son indépendance vis à vis du terrain, en opposition aux roues odométriques victimes de dérive en cas de terrain accidenté, mais surtout son efficacité dans les espaces confinés ou non couverts par le système GNSS (canyons urbains, milieu marin....).

La problématique d'estimation de position d'un véhicule à partir d'images est apparue au début des années 1980 et a été décrite par [Moravec, 1980]. À ce jour, le cadre posé par ces travaux reste la référence dans le domaine de l'odométrie visuelle. Ces travaux ont été menés dans le cadre de l'exploration spatiale au moyen d'un rover. Depuis lors, ils ont été soutenus par le programme d'exploration martienne de la NASA [Maimone et al., 2007].

Nous pouvons classer les systèmes d'odométrie visuelle en deux catégories. D'une part, on trouve les systèmes composés de deux (ou plusieurs) caméras. Nous parlons alors d'odométrie visuelle stéréoscopique. Cette terminologie souligne le fait que les positions 3D des points d'intérêts, détectés dans les images successives, sont calculées par triangulation pour chaque position du mobile. L'alternative aux systèmes stéréoscopiques est l'utilisation d'une unique caméra. Dans ce cas, on parlera d'odométrie visuelle monoculaire. Le désavantage de cette approche est que la trajectoire du mobile est obtenue modulo un facteur d'échelle. Ce facteur peut être déterminé à partir de mesures directes (par exemple, par la mesure d'éléments de la scène), ou grâce à l'intégration de données provenant d'autres capteurs installés sur le mobile (IMU, laser, odomètres classiques....).

Les premiers systèmes d'odométrie visuelle étaient principalement des systèmes stéréoscopiques. À partir des travaux menés par [Moravec, 1980], [Matthies & Shafer, 1987] ont proposé la

définition d'un modèle permettant de représenter l'erreur de positionnement des points d'intérêts, support de l'étape de triangulation. Cette approche a permis d'obtenir des résultats intéressants, puisque pour une trajectoire linéaire de 5.5 *m*, l'erreur de positionnement était de l'ordre de 2 % soit 11 *cm*.

[Olson et al., 2000] et [Olson et al., 2003] ont fait évoluer ces premiers travaux en introduisant un autre capteur au système, à savoir le gyroscope. Celui-ci fournit une information sur l'orientation du mobile en déplacement et permet de réduire significativement l'erreur de positionnement, soit une erreur de l'ordre de 1.2% (24 *cm*) pour un déplacement linéaire de 20 *m*.

En 2004, l'approche mono-caméra proposée par [Nistér et al., 2004] a été novatrice. Outre le fait que le terme "odométrie visuelle" soit apparue à la suite de cet article, l'implémentation proposée est la première fonctionnant en temps réel. Une des nouveautés importantes de cette approche concerne l'utilisation et la mise en place d'un algorithme de transformation à 5 points [Nistér, 2004]. Ainsi, les manipulations réalisées dans différentes conditions de terrain ont permis d'obtenir des résultats de positionnement atteignant une erreur de l'ordre de 5 % pour des déplacements de plusieurs centaines de mètres (≈ 250 *m*).

Dans [Tardif et al., 2008], une acquisition sur une distance de plusieurs kilomètres a été réalisée. Contrairement aux travaux précédents, les estimations de rotation et de translation sont réalisées de façon indépendante. La rotation est estimée à partir de points situés à "l'infini", tandis que la translation est obtenue à partir d'une carte 3D. Les résultats obtenus grâce à cette approche sont de l'ordre de 2.5 % d'erreur de positionnement pour un déplacement de 2.5 *km*. Nous pouvons également citer différents travaux exploitant la configuration cinématique des mobiles sur lesquels les systèmes de vision sont installés. Ainsi [Scaramuzza, 2011] propose une approche ne nécessitant qu'un appariement pour appliquer le paradigme de RANSAC ; ceci permet alors la mise en place d'un algorithme de positionnement ayant une cadence de 400 *Hz*.

Malgré le nombre important d'approches développées ces dernières années dans le cadre de l'odométrie visuelle, les principales étapes définies par [Nistér et al., 2004] restent valables. Il nous semble donc opportun de les résumer ici.

Comme nous l'avons précisé précédemment, l'odométrie visuelle consiste à évaluer le déplacement d'un mobile. Le mouvement de celui-ci est évalué à partir d'images obtenues grâce à un système de caméra placé dessus. Ce système acquiert des images à intervalles de temps réguliers. Nous notons l'ensemble de ces images à chaque temps k : $I_{0:n} = I_0, \dots, I_n$. Nous considérons que le repère du mobile en déplacement est assimilable au repère du système de caméras. Deux positions successives du système de caméras aux instants $k - 1$ et k sont liées par la transformation rigide $T_{k,k-1} \in \mathbb{R}$, définie telle que :

$$T_{k,k-1} = \begin{bmatrix} R_{k,k-1} & t_{k,k-1} \\ 0 & 1 \end{bmatrix} \quad (4.1)$$

où, $R_{k,k-1}$ est la matrice de rotation, appartenant au groupe des rotations $SO(3)$, et $t_{k,k-1} \in \mathbb{R}^{3 \times 1}$ est le vecteur de translation. L'ensemble $T_{1:n} = T_{1,0}, \dots, T_{n,n-1}$ contient les transformations

successives décrivant le déplacement du mobile. Enfin, on trouve l'ensemble $C_{0:n} = C_0, \dots, C_n$ qui contient les positions successives du système de caméra. La position C_n peut alors être déterminée en concaténant les T_k transformations. Ainsi on obtient : $C_n = C_{n-1}.T_n$ avec C_0 la position de départ, qui peut être fixée.

L'objectif principal de l'odométrie visuelle revient donc à déterminer les transformations T_k à partir de l'étude des images I_k et I_{k-1} et ensuite à les concaténer, afin d'obtenir l'ensemble des positions $C_{0:n}$ permettant de reconstruire la trajectoire du mobile (Figure 4.1). L'odométrie visuelle est une approche incrémentale, la position courante de la caméra étant construite à partir des transformations précédentes. Il est également possible d'exploiter les n dernières transformations pour apporter plus de précision au calcul de la position en cours (Figure 4.1).

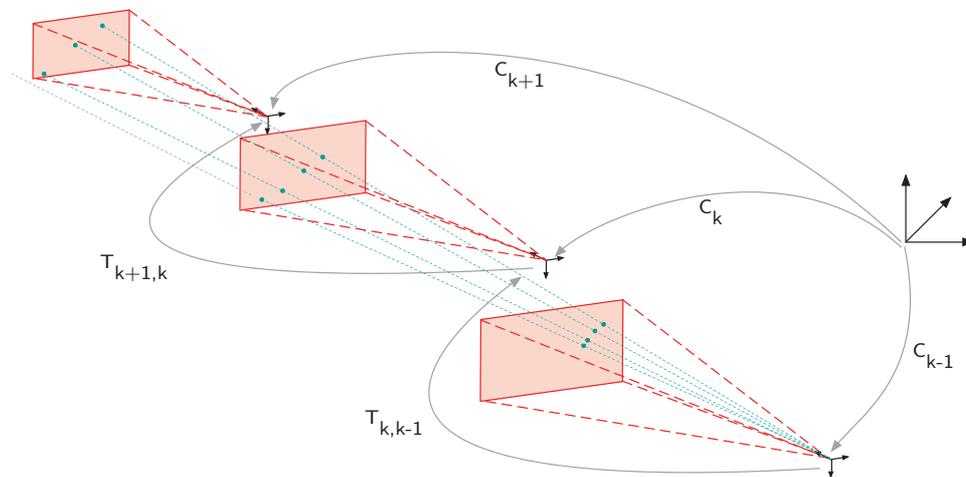


FIGURE 4.1 : Principe de l'odométrie visuelle

Les cinq étapes de l'odométrie visuelle sont décrites dans la Figure 4.2. La première étape consiste à détecter des points d'intérêts dans les images fournies par le système de caméras monté sur le mobile. Ensuite, les points détectés sont mis en correspondance avec ceux détectés dans l'image précédente. Une fois la mise en correspondance réalisée et l'élimination des correspondances aberrantes finalisée, la dernière étape consiste à estimer le mouvement effectué. Chacune de ces étapes sera décrite par la suite.

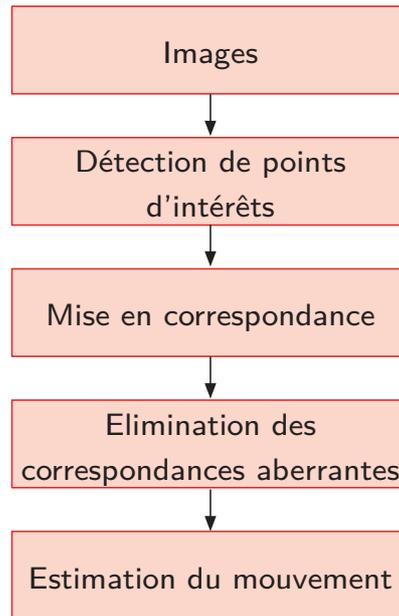


FIGURE 4.2 : Étapes de l'odométrie visuelle ([Scaramuzza & Fraundorfer, 2011])

Les différentes approches de l'odométrie visuelle ont montré leurs performances dans de nombreuses applications. Cependant, elles présentent encore aujourd'hui certaines limites. La première limite que nous pouvons citer concerne la géolocalisation. En effet, dans une approche classique d'odométrie visuelle, aucune information sur l'environnement dans lequel évolue la caméra n'est disponible. La seule information fournie par l'odométrie visuelle est le déplacement relatif de la caméra au fil du temps. L'absence de géolocalisation n'est toutefois pas nécessairement bloquante dans certaines applications (trajectographie, suivi de convois, etc.).

La seconde limite que nous pouvons mettre en avant concerne la dérive que l'on peut observer au cours du temps. En effet, une des étapes principales de l'odométrie visuelle consiste à détecter des points d'intérêts dans les images successives. Ces mêmes points d'intérêts sont ensuite utilisés pour calculer la trajectoire de la caméra. Or, la détection de ces points n'est pas "parfaite". Ainsi les erreurs faites lors de la détection ont un impact direct sur la localisation à l'instant k , et sur les localisations futures. L'odométrie visuelle étant un processus incrémental, la position de la caméra à l'instant k dépend des données collectées à l'instant $k - 1$. Or, si les données sont entachées d'erreurs, le calcul du déplacement relatif entre les instants k et $k - 1$ sera impacté par cette transmission d'erreurs. De plus, à l'erreur réalisée sur ce déplacement s'ajoute l'erreur globale à l'instant $k - 1$. Au cours du processus d'odométrie visuelle, les erreurs s'ajoutent donc à chaque nouvelle itération de l'algorithme et produisent une accumulation d'erreurs.

En contrepartie de ces limitations, l'odométrie visuelle présente de nombreux avantages. D'une part, nous avons pu voir qu'elle a atteint un certain degré de maturité dans de nombreux domaines (spatial, terrestre, ou encore sous-marin) et qu'elle peut être exploitée comme une brique fondamentale d'une approche plus complète de SLAM comme nous avons pu le voir dans le chapitre 2 (section 3.5.). D'autre part, un avantage considérable de l'odométrie visuelle est sa facilité de mise en œuvre aussi bien d'un point de vue matériel, que d'un point de vue logiciel. De plus, en comparaison aux systèmes de type LIDAR, elle propose une solution plus abordable et plus facile à mettre en œuvre. Enfin, la miniaturisation des systèmes de vision de ces dernières années offre la possibilité de développer des systèmes de plus en plus petits, capables d'estimer le déplacement d'un mobile.

Dans le cadre d'un projet de fin d'études encadré au cours de la thèse, nous avons pu évaluer les possibilités offertes par l'odométrie visuelle [Yang, 2012]. Les résultats obtenus dans ce projet ont confirmé les avantages énumérés ci-dessus et nous ont définitivement convaincus de suivre cette approche de localisation basée sur l'odométrie visuelle. Dans les sections suivantes, nous présentons la méthodologie mise en place autour du pôle caméras 3D avant d'aborder les travaux engagés autour du pôle caméra *Fish Eye*.

3 Pôle caméras 3D

3.1. Présentation de l'approche développée

3.1.1. Rappel sur la transformation de pose

Soit un repère orthonormé \mathcal{R}^* , appartenant à l'espace euclidien, nommé repère de référence, et un repère orthonormé \mathcal{R} nommé repère courant. Soit la matrice homogène $T \in SE(3) \subset \mathbb{R}^{4 \times 4}$, appartenant au groupe Spécial Euclidien, de dimension 4, telle que :

$$T = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \quad (4.2)$$

où $R \in SO(3) \subset \mathbb{R}^{3 \times 3}$ est une matrice de rotation, et $t \in \mathbb{R}^3$ est un vecteur de translation de dimensions 3×1 . La matrice T définit le déplacement 3D rigide entre les repères \mathcal{R}^* et \mathcal{R} , également nommée "transformation de pose" entre les deux repères (Figure 4.3).

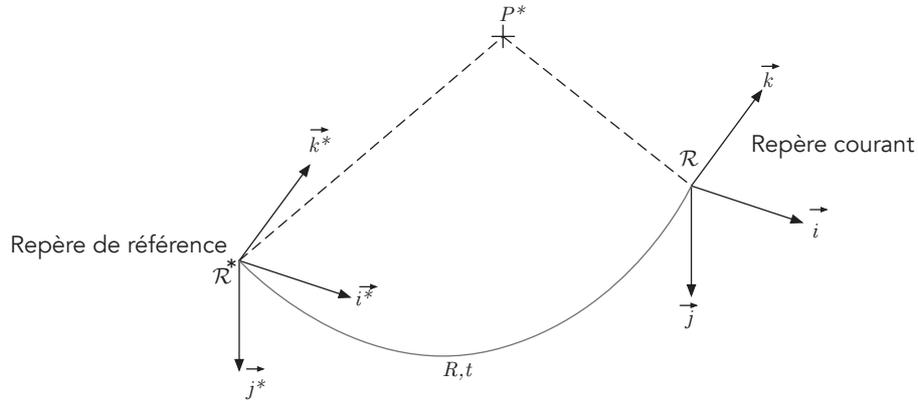
Soit $P^* = [x, y, z]^T \in \mathbb{R}^3$, un point 3D de l'espace euclidien défini dans le repère \mathcal{R}^* . Le point P^* peut être exprimé dans le repère \mathcal{R} grâce à la transformation T vue dans l'équation 4.3 (équation 4.3).

$$P = R.P^* + t \quad (4.3)$$

3.1.2. Principe général et formulation

Afin de déterminer la position du mobile dans le repère global \mathcal{R}_g , nous avons tout d'abord besoin de définir différents repères (Figure 4.4) :

- Le repère système \mathcal{R}_s : il représente le repère du système de caméra. Le premier axe est dirigé dans le sens du déplacement, le second est dirigé orthogonalement au premier vers la

FIGURE 4.3: Transformation entre les repères \mathcal{R}^* et \mathcal{R}

droite, enfin le dernier est défini de façon à définir un repère orthogonal dans le sens direct. Dans le cas de notre étude, nous considérons ce repère \mathcal{R}_s comme le repère principal, soit le repère du mobile.

- Le repère global \mathcal{R}_g : il s'agit de notre repère de travail. La position du système de caméras est exprimée dans ce repère orthogonal.

À l'instant k du déplacement du mobile, la position et l'orientation du système (de caméras) dans \mathcal{R}_g sont définies par le vecteur de dimension trois $P(k) = [x_s(k), y_s(k), \theta_s(k)]$ où (x_s, y_s, z_s) sont les coordonnées dans O_s (Figure 4.4) et θ_s l'angle $(O_g \widehat{X}_g, O_s \widehat{X}_s)$.

À l'instant $k + 1$, la position et l'orientation du système dans \mathcal{R}_g sont alors définies par le vecteur $P(k + 1) = [x_s(k + 1), y_s(k + 1), \theta_s(k + 1)]$. Nous cherchons à déterminer la matrice de transformation $T_{k,k+1}$ définissant le déplacement 2D rigide entre les repères $\mathcal{R}_s(k)$ et $\mathcal{R}_s(k + 1)$.

Nous considérons que le mobile se déplace dans le plan principal que représente le sol. Pour des raisons de simplifications, l'approche développée au cours de cette thèse ne considère pas de changement d'altitude durant le déplacement.

Nous pouvons alors définir la transformation $T_{k,k+1}$ de dimension 3x3 telle que :

$$T_{k,k+1} = \begin{pmatrix} R & t \\ 0 & 1 \end{pmatrix} \quad (4.4)$$

où $R \in SO(2) \subset \mathbb{R}^{2 \times 2}$ est une matrice de rotation et t un vecteur de translation de dimension 2. Nous pouvons alors écrire :

$$P_s(k) = R.P_s(k - 1) + t \quad (4.5)$$

À l'instant $k + 1$, nous avons alors :

$$P_s(k + 1) = R.P_s(k) + t \quad (4.6)$$

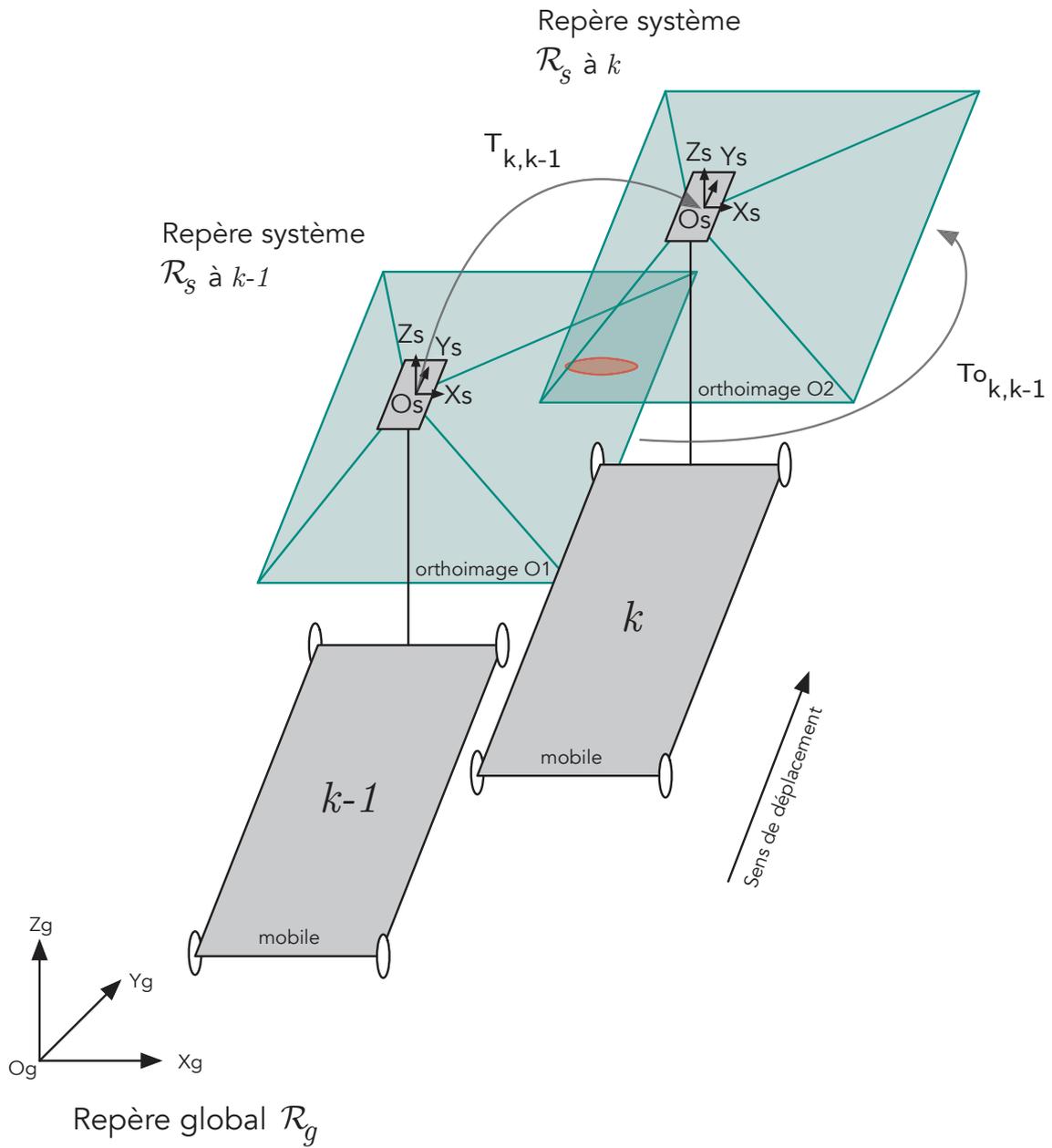


FIGURE 4.4: Schéma représentant le principe de notre approche de localisation

Ainsi, la position $P_s(n)$ peut être déterminée en concaténant les positions successives du système. Nous pouvons écrire :

$$P_s(n) = R.P_s(n-1) + t \quad (4.7)$$

La position $P(0)$ peut être fixée par l'utilisateur.

L'estimation de la transformation $T_{O_k, k+1}$ entre les orthoimages O_1 et O_2 permet de déterminer la transformation $T_{k, k+1}$ entre les positions $P_s(k)$ et $P_s(k+1)$ du système. En déterminant l'ensemble des transformations entre les orthoimages successives nous serons en mesure de déterminer les positions successives de notre système.

Dans ce contexte, s'inspirant de l'odométrie visuelle pour exploiter les orthoimages, notre méthode de localisation se décompose en quatre étapes principales :

- **Détection de points d'intérêts** : Il s'agit de détecter automatiquement les points d'intérêts présents sur une orthoimage.
- **Mise en correspondance** : Une fois les points d'intérêts détectés dans l'orthoimage courante, il est nécessaire de "lier" ces points à ceux détectés dans l'orthoimage précédente.
- **Élimination des correspondances aberrantes** : Bien que très efficace, la mise en correspondance entraîne parfois des valeurs aberrantes, qu'il est indispensable de supprimer. Cette étape est fondamentale dans notre approche, car elle conditionne la qualité de la transformation calculée par la suite.
- **Calcul de la transformation** : Une fois l'ensemble des mises en correspondances filtrées, nous sommes en mesure d'estimer une transformation entre les orthoimages successives.

L'ensemble de ces étapes est présenté en détail dans la suite du manuscrit et repris en détail sur la Figure 4.5.

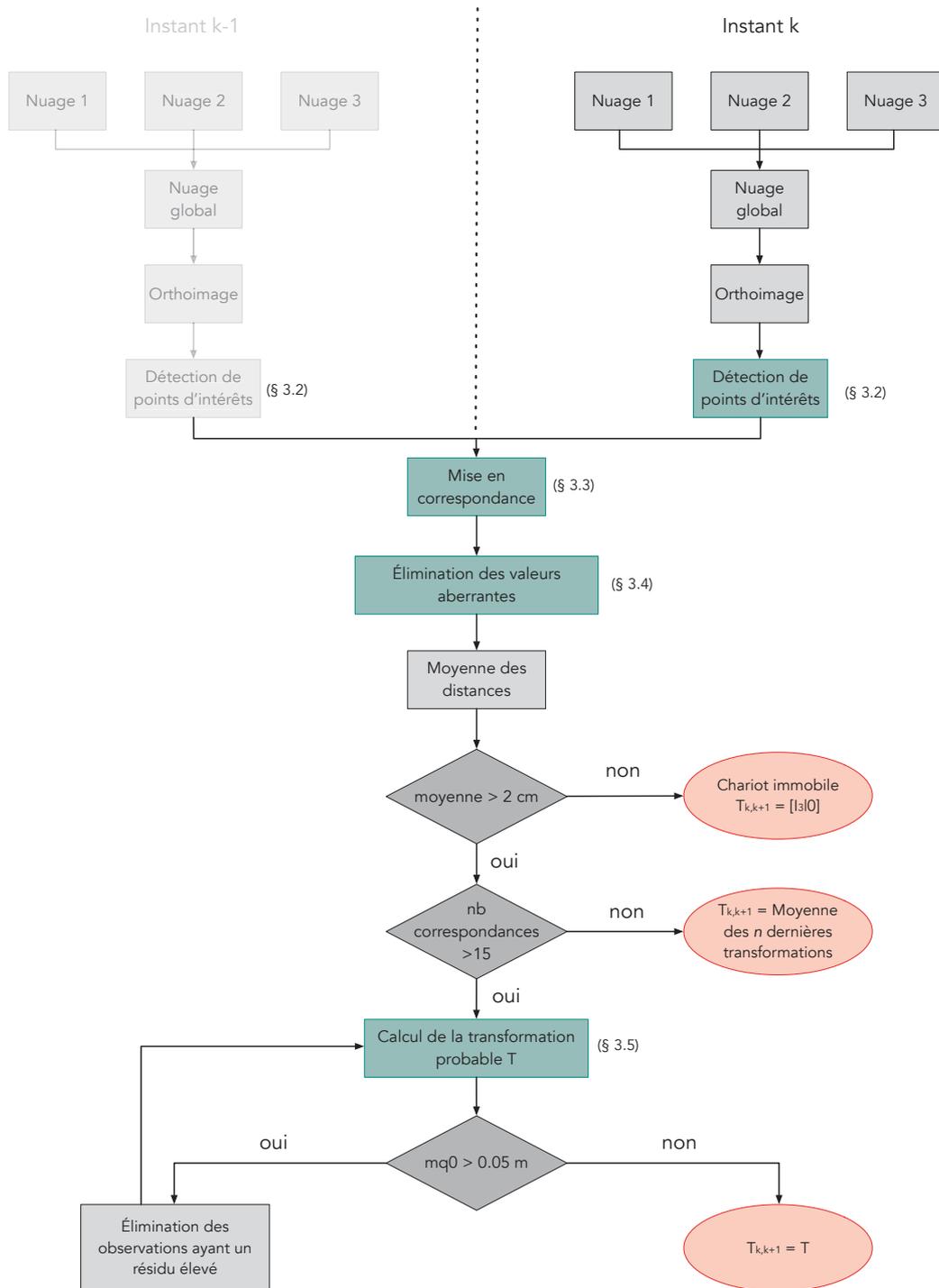


FIGURE 4.5 : Méthodologie complète de localisation développée (pour chaque étape principale, le paragraphe associé est indiqué entre parenthèses)

3.2. Détection de points d'intérêts

Dans cette étape, il s'agit de détecter les points d'intérêts présents dans les orthoimages obtenues à partir des nuages de points fournis par les 3 caméras 3D. Cette détection est réalisée pour chacune des nouvelles orthoimages générées.

Les points d'intérêts d'une image sont définis comme étant des points possédant des caractéristiques qui permettent de les distinguer des autres. L'étape de détection de points d'intérêts est souvent une étape préliminaire à de nombreux processus de photogrammétrie ou de vision par ordinateur. Ces points d'intérêts, dans une image, correspondent à des doubles discontinuités de la fonction d'intensités. Suivant le contenu de l'image, celles-ci peuvent apparaître lors de discontinuités de la fonction de réflectance ou lors de discontinuités de profondeur. Ainsi, seront détectés comme points d'intérêts : les coins, les jonctions en T, ou encore les pixels de fortes variations de texture. Dans la tâche d'odométrie visuelle, le choix d'un bon détecteur de points est primordial. En effet, la précision finale de la trajectoire calculée va directement dépendre de la "qualité" des points utilisés pour le calcul. Ainsi, il est nécessaire d'énumérer les critères qui nous permettront de comparer et d'évaluer les détecteurs : précision, répétabilité, rapidité de détection, robustesse, invariance.

Une recherche bibliographique nous a permis de constater que les détecteurs de points d'intérêts dans le cadre de l'odométrie visuelle peuvent être classés selon deux catégories :

- les détecteurs de coins, tels que [Moravec, 1980], [Harris & Pike, 1987], ou encore [Shi & Tomasi, 1994] (Figure 4.6(a)),
- les détecteurs basés sur les différences d'intensités également nommés "blob detector" en anglais, tels que [Lowe, 2004],[Bay et al., 2006],[Rublee et al., 2011] et [Mikolajczyk & Schmid, 2004] (Figure 4.6(b)).

Évidemment, chaque détecteur présente des avantages et des inconvénients. Les détecteurs de coins sont réputés pour être rapides mais moins distinctifs, alors que les détecteurs de "blob" sont plus distinctifs, mais plus lents dans le processus de détection. Les détecteurs de coins localisent plus précisément les points d'intérêts, mais supportent mal les changements d'échelle. Cela signifie que les coins ne seront pas aussi bien redétectés que les blobs lors de grands changements d'échelle ou de points de vue. Malgré tout, l'utilisation des détecteurs de blob n'est pas toujours la meilleure solution. Par exemple SIFT [Lowe, 2004] ne prend pas en compte les coins, or en environnement urbain c'est un élément que l'on retrouve très régulièrement (fenêtre, portes, trottoirs, etc.). Il est donc nécessaire de réaliser une étude afin de choisir le détecteur le plus approprié à notre besoin spécifique, en limitant toutefois nos investigations aux détecteurs les plus reconnus et utilisés.

Il existe de nombreux détecteurs de points d'intérêts. Nous n'en décrivons que quelques uns, à savoir les détecteurs de blobs : SIFT, et SURF et les détecteurs de coins : ORB, et SHI . En effet, SIFT a prouvé son efficacité dans de nombreuses applications, mais souffre d'un temps d'exécution trop long, aussi il n'est pas adapté à des applications en temps réel. Face à ce problème, de nombreuses améliorations ont été proposées telles que SURF ou encore ORB, lui-même basé sur les approches de [Rosten & Drummond, 2006] et [Calonder et al., 2010]. Nous avons également

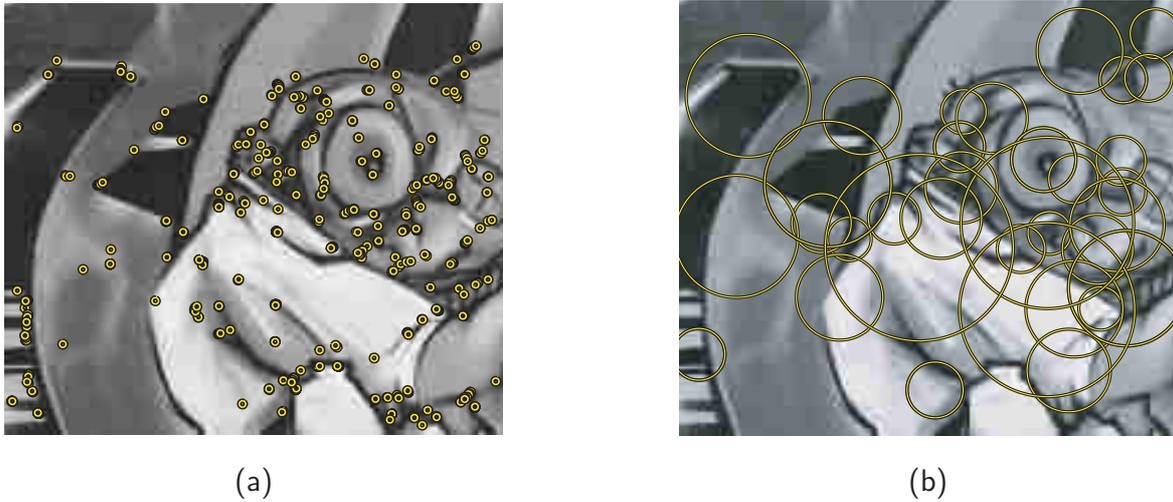


FIGURE 4.6 : Exemple de détection de points d'intérêt pour chacune des catégories de détecteurs mentionnées : (a) Détecteurs de coins (Harris [Harris & Pike, 1987]) - (b) Détecteurs basés sur les différences d'intensité (Hessian Laplace [Mikolajczyk & Schmid, 2004])

choisi de présenter l'approche proposée par [Shi & Tomasi, 1994], particulièrement appropriée à la tâche d'odométrie visuelle, et qui a été développée à partir du détecteur de [Harris & Stephens, 1988].

3.2.1. Détecteur de Harris

Le détecteur de Harris fait partie des détecteurs basés sur le signal. Ce détecteur étudie l'information locale de niveau de gris afin de détecter les points d'intérêts. En s'inspirant du détecteur de [Moravec, 1980], [Harris & Stephens, 1988] ont proposé un nouveau détecteur plus stable, et plus performant. Le détecteur de Harris est basé sur le calcul de la matrice d'auto-corrélation décrivant les gradients de couleurs autour du point détecté. Un lissage gaussien (d'écart type σ) est effectué avant de calculer les dérivées du signal. La matrice d'auto-corrélation M est calculée pour chaque pixel (x, y) de l'image selon l'équation 4.8.

$$M = e^{-\frac{x^2+y^2}{2\sigma^2}} \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \quad (4.8)$$

où I_x est la dérivée de la fonction d'intensité du premier ordre selon x , et I_y est la dérivée de la fonction d'intensité du premier ordre selon y . L'étude des valeurs propres (λ_1, λ_2) de la matrice M permet d'évaluer la présence ou non de points d'intérêts. Ces deux valeurs propres sont proportionnelles aux courbures principales de la fonction d'auto-corrélation [Harris & Stephens, 1988]. Il y a alors trois cas à considérer :

- si les deux courbures (λ_1, λ_2) sont faibles, alors la fonction d'auto-corrélation est approximativement constante, ce qui indique une région homogène de l'image,

- si l'une des courbures est forte, alors que la seconde est faible, nous sommes en présence d'un contour,
- enfin si les deux courbures sont fortes, nous sommes en présence d'un point d'intérêt.

[Harris & Stephens, 1988] ont proposé d'éviter le calcul des valeurs propres en étudiant le déterminant et la trace de la matrice M . Ainsi la fonction de réponse C , pour chaque pixel de l'image, est définie telle que :

$$\begin{aligned} C(x, y) &= \lambda_1 \lambda_2 - k(\lambda_1 + \lambda_2)^2 \\ &= \det(M(x, y)) - k \cdot \text{trace}^2(M(x, y)) \end{aligned} \quad (4.9)$$

où λ_1 et λ_2 sont les valeurs propres de la matrice M et $k \in [0.04, 0.15]$. Les coins sont détectés en exhibant les maxima locaux de C supérieurs à un seuil donné.

Le détecteur de Harris est probablement le détecteur de points d'intérêt le plus connu et le plus répandu en traitement d'images. De nombreuses implémentations de ce détecteur ont été proposées, car cinq paramètres peuvent être choisis lors de son utilisation : le filtre dérivatif, le filtre Gaussien, le paramètre k , le voisinage de l'extraction des maxima locaux, et le seuillage final. Son succès provient non seulement des résultats qu'il fournit, mais également de sa facilité d'implémentation.

3.2.2. Détecteur de Shi & Tomasi (GFTT)

Parmi les différentes améliorations du détecteur de Harris, nous trouvons l'approche proposée par [Shi & Tomasi, 1994], également nommée "*Good Features To Track*". Cette variante propose d'étudier les matrices de corrélation en utilisant une pondération uniforme, et de calculer $C(x, y)$ directement en utilisant $\min(\lambda_1, \lambda_2)$. Leur étude, portant sur des transformations affines des images, les a amenés à recommander l'usage des valeurs propres λ_1 et λ_2 les plus faibles comme fonction de coût. En effet, la mesure de $\min(\lambda_1, \lambda_2)$ remplace l'utilisation de $C(x, y)$. La Figure 4.7 présente un exemple de détection de points d'intérêts avec le détecteur de Shi & Tomasi.

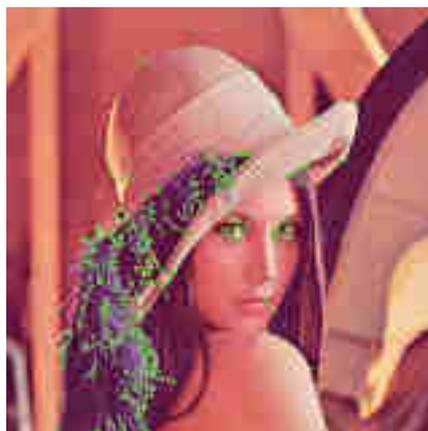


FIGURE 4.7 : Exemple de détection de points d'intérêt (en vert) avec le détecteur de Shi & Tomasi

3.2.3. Détecteur SIFT (*Scale Invariant Feature Transform*)

Le détecteur SIFT, décrit dans [Lowe, 2004], à l'image de Harris est également un détecteur très répandu que l'on retrouve dans de nombreuses applications ([Se et al., 2001],[Skrypnik & Lowe, 2004]). Commençons par définir la notion d'espace échelle utilisée par le détecteur SIFT. Une définition de l'espace échelle pour les signaux 1D est la suivante :

🔍 Espace échelle *Définition*

Soit $f(x)$ une fonction et $G_\sigma(x)$ la gaussienne d'écart type σ . On appelle espace échelle, le lieu des réalisations de la transformation S définie par :

$$(Sf)(x, \sigma) = f * G_\sigma(x), (x, \sigma) \in \mathcal{R} \times \mathcal{R}^+$$

On appellera S l'opérateur de changement d'échelle et on notera $E = (x, \sigma)$ l'espace échelle.

Cette transformation conduit à la représentation d'une fonction sous la forme d'une surface décrite dans E . On peut donc étudier cette surface en utilisant des arguments de géométrie différentielle. La représentation dans l'espace échelle d'une image I est définie par une fonction $E(x, y, \sigma)$, obtenue par la convolution de I avec une gaussienne $G(x, y, \sigma)$ d'écart-type variable :

$$E(x, y, \sigma) = G(x, y, \sigma) * I(x, y) \quad (4.10)$$

où $*$ est l'opérateur de convolution en x et y et

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2+y^2)/2\sigma^2} \quad (4.11)$$

Les points d'intérêts sont détectés au travers d'un calcul qui comprend un changement d'échelle. Les points invariants à ce changement d'échelle sont considérés comme des points d'intérêt. L'approche consiste donc à détecter les points qui sont stables dans l'espace échelle. Les extrema locaux de l'opérateur DoG (*difference of gaussian*) sont calculés dans l'espace échelle. Étant données deux échelles séparées par une constante multiplicative k , on calcule les extrema locaux par l'équation 4.12 :

$$\begin{aligned} DoG(x, y, \sigma) &= (G(x, y, k\sigma) - G(x, y, \sigma))I(x, y) \\ &= E(x, y, k\sigma) - E(x, y, \sigma) \end{aligned} \quad (4.12)$$

L'opérateur DoG est rapide et fournit une bonne approximation de l'opérateur laplacien. Pour détecter les points d'intérêts, chaque point d'une image $DoG(x, y, \sigma)$ est comparé à ses 8 voisins de la même échelle, ainsi qu'à ses 16 voisins des deux échelles immédiatement inférieure et supérieure. On détecte ainsi les minima et les maxima locaux. SIFT n'est pas un détecteur strictement invariant affine. En effet, il n'est invariant qu'au changement d'échelle. Cependant, le descripteur qui lui est associé est assez robuste et fournit de bons résultats, même lors de changements de point de vue importants. Nous pouvons observer sur la Figure 4.8 un exemple de détection de points d'intérêt avec le détecteur SIFT.

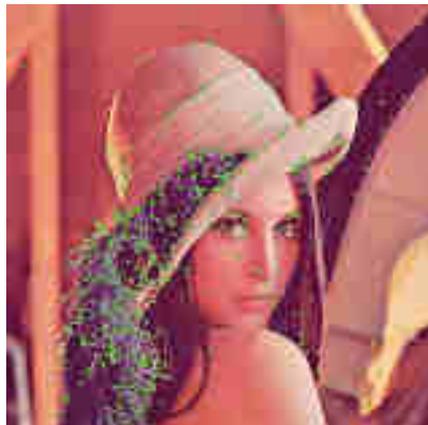


FIGURE 4.8: Exemple de détection de points d'intérêt (en vert) avec le détecteur SIFT

3.2.4. Détecteur SURF (*Speeded Up Robust Features*)

Afin de résoudre le problème de sensibilité à certaines transformations de SIFT, [Bay et al., 2006] ont proposé une méthode de détection très influencée par SIFT. Ils ont ajouté une étape de recalage de la zone d'analyse grâce à la construction d'un histogramme de gradient orienté. Ainsi, la principale amélioration de SURF par rapport à SIFT est une invariance aux changements d'échelle et aux rotations. Un exemple de détection de points d'intérêt avec le détecteur SURF est présenté sur la Figure 4.9.

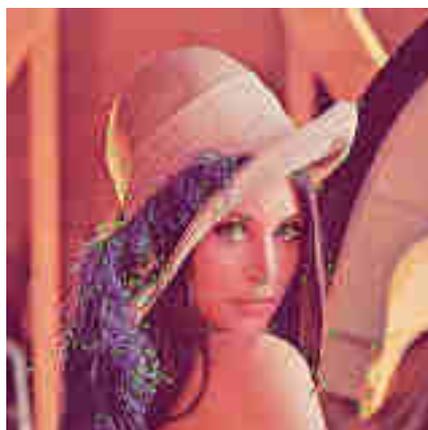


FIGURE 4.9: Exemple de détection de points d'intérêt (en vert) avec le détecteur SURF

3.2.5. Détecteur ORB (*Oriented FAST and Rotated BRIEF*)

ORB a été proposé par les équipes d'OpenCV [Rublee et al., 2011] comme une alternative à SIFT et SURF (Figure 4.10). Cette approche est présentée comme une alternative plus efficace en termes de temps de calcul, de performance, mais surtout en terme de droit d'utilisation. En effet, les algorithmes SIFT et SURF sont protégés par différents brevets, rendant leur utilisation

tion payante dans certaines situations, notamment dans le cadre du développement d'applications commerciales, ce qui n'est pas le cas de ORB.

ORB est essentiellement basé sur le détecteur FAST [Rosten & Drummond, 2006] et sur le descripteur BRIEF [Calonder et al., 2010]. Cependant, un ensemble de modifications a été apporté à ces deux approches afin d'en améliorer leurs performances. Dans un premier temps, FAST est utilisé pour détecter les points d'intérêt, puis un détecteur de Harris [Harris & Pike, 1987] leur est appliqué afin de n'en conserver que les plus "forts". Le problème principal de FAST dans sa version originale est sa variance à la rotation. C'est pourquoi, les auteurs du détecteur ORB [Rublee et al., 2011] ont suggéré d'appliquer un détecteur de Harris dans un second temps.

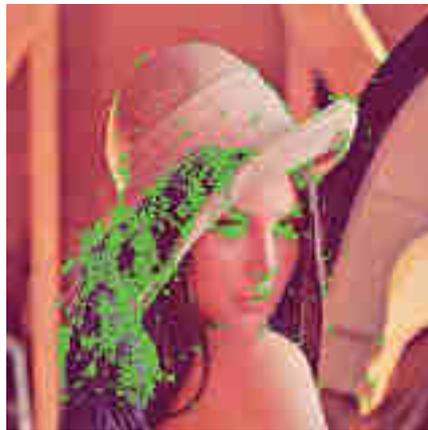


FIGURE 4.10: Exemple de détection de points d'intérêt (en vert) avec le détecteur ORB

3.2.6. Choix du détecteur

Les données à traiter dans le cadre de notre approche sont très particulières, puisqu'il s'agit d'images produites par orthoprojection de nuages de points, donc d'images relativement uniformes et comportant assez peu de variations en termes de comptes numériques. Il nous est apparu nécessaire de trouver le détecteur de points d'intérêt en mesure de nous fournir suffisamment de points de bonne qualité. Afin d'évaluer la performance de différents détecteurs de points d'intérêt, il est nécessaire de tenir compte de différents critères, à savoir : la répétabilité, le temps de détection, et le nombre de points détectés. Nous avons restreint nos travaux à ces critères, dans la mesure où ceux-ci nous ont semblé pertinents pour l'évaluation que nous souhaitons réaliser.

a. Répétabilité des détecteurs

Dans [Mikolajczyk & Schmid, 2005] et [Miksik & Mikolajczyk, 2012], les auteurs proposent d'analyser la répétabilité des détecteurs. On définit la répétabilité d'un détecteur comme sa capacité à repérer un point d'une image confrontée à différentes transformations. Ces transformations peuvent être le changement d'échelle, la rotation, la translation, le changement de luminosité, etc. On considérera que la répétabilité d'un détecteur est forte lorsque, pour deux conditions diffé-

rentes d'observation d'une même scène, un fort pourcentage des points détectés dans la première image est aussi visible dans la seconde.

Afin de déterminer la répétabilité d'un détecteur, il est nécessaire de connaître la transformation appliquée d'une image à l'autre. Connaissant cette transformation, l'inverse de celle-ci est appliquée aux points détectés dans la seconde image. Si la région autour des points reprojétés a suffisamment de recouvrement avec la région autour des points correspondants de la première image, nous considérons que la correspondance est satisfaisante. Ainsi, si deux images contiennent respectivement n_1 et n_2 points détectés, le critère de répétabilité est défini tel que présenté dans l'équation 4.13.

$$\text{répétabilité} = \frac{\text{nb_bonnes_correspondances}}{\min(n_1, n_2)} \quad (4.13)$$

Afin de se placer dans des conditions les plus proches possibles de notre situation, nous avons utilisé des orthoimages pour évaluer les détecteurs. Ces orthoimages ont été calculées sur des données simulées produites via l'application Blensor [Gschwandtner et al., 2011]. Blensor nous a permis de modéliser une scène en 3 dimensions (Figure 4.11 (a)) et de simuler le nuage de points obtenu par une caméra 3D dans cet environnement. Nous reviendrons plus en détail sur la présentation de cette application de simulation dans le chapitre concernant les expérimentations. L'utilisation de cette application nous permet de maîtriser parfaitement la transformation entre les orthoimages dont une est présentée dans la Figure 4.11 (b). La création de ces orthoimages a été présentée dans le Chapitre 3 (paragraphe 3.2.4.). Comme nous l'avons représenté sur la palette de niveau de gris, l'orthoprojection des points dont l'altitude est inférieure à 10 cm est représentée en noir, et l'orthoprojection des points dont l'altitude est supérieure à 160 cm est représentée en blanc (Figure 4.11 (b)).

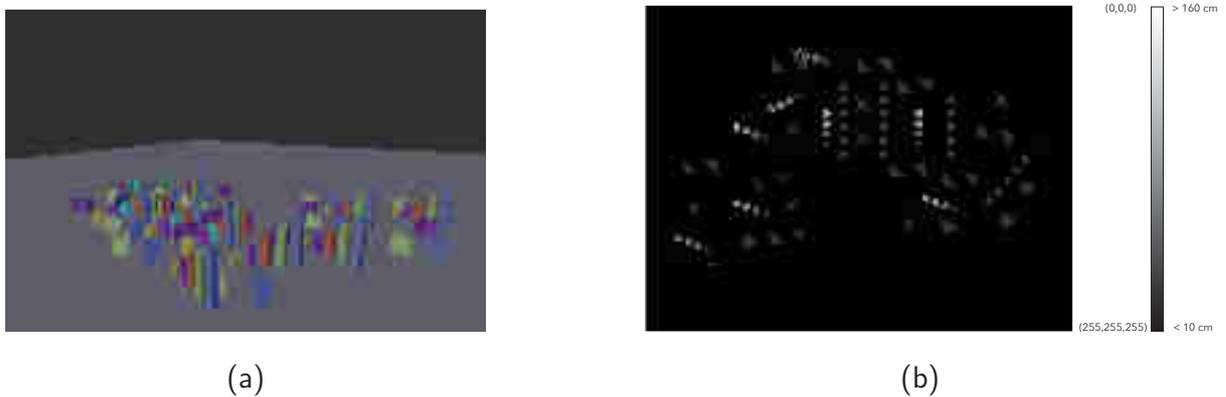


FIGURE 4.11 : Données utilisées pour l'évaluation des détecteurs : (a) Scène 3D modélisée pour la simulation de nuage de points - (b) Orthoimage obtenue à partir de cette scène

La répétabilité des différents détecteurs a été évaluée pour deux types de transformations, à savoir la translation et la rotation. En effet, le changement d'échelle n'est pas une situation à

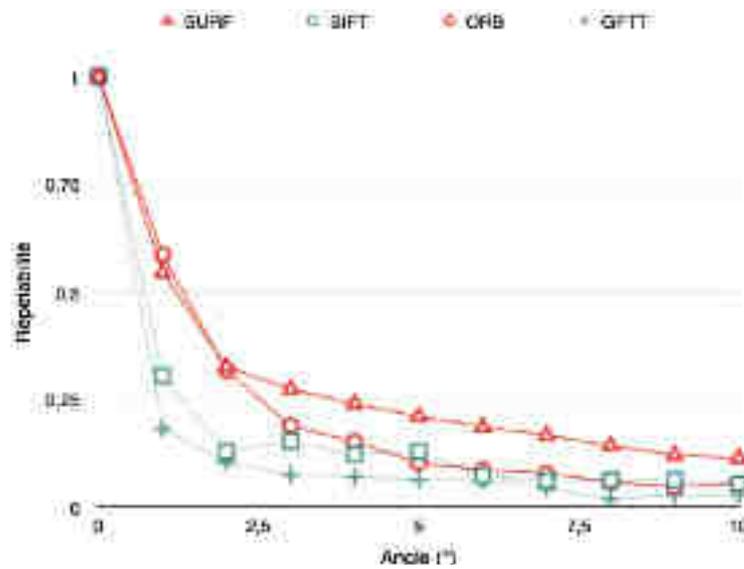


FIGURE 4.12 : Répétabilité des détecteurs après rotations

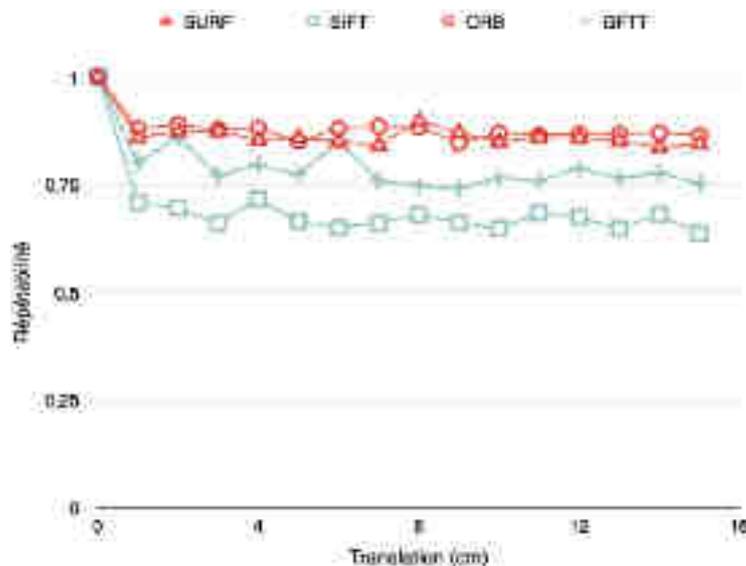


FIGURE 4.13 : Répétabilité des détecteurs après translations

laquelle nous serons confrontés au cours du déplacement du mobile.

Sur la Figure 4.12, nous pouvons observer que pour de faibles angles de rotation, SURF et ORB présentent une meilleure répétabilité que les autres détecteurs. Cependant, pour des angles de rotation supérieurs à $2,5^\circ$, SURF est le détecteur présentant la meilleure répétabilité.

Concernant la répétabilité dans le cas d'une translation, nous pouvons observer que SURF et ORB présentent à nouveau des résultats supérieurs aux autres détecteurs (Figure 4.13).

b. Temps de détection

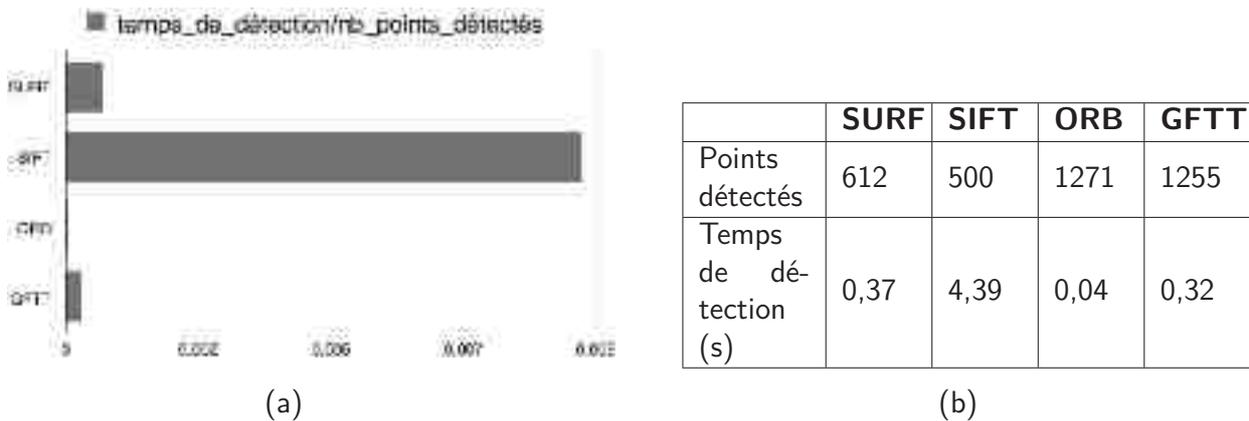


FIGURE 4.14: Évaluation du temps de détection pour chaque détecteur de notre étude : (a) Ratio $\text{temps_de_détection}/\text{nb_points_détectés}$ (b) Nombre de points détectés sur l'orthoimage présentée Figure 4.11(b) et temps de détection

L'aspect temporel joue un rôle important dans notre approche. Ainsi, il est nécessaire de prendre en considération le temps de détection dans le choix du détecteur approprié. Nous avons donc également considéré comme critère, le ratio $\text{temps_de_détection}/\text{nb_points_détectés}$ (Figure 4.14). Nous pouvons observer que SIFT présente le plus mauvais ratio, à l'inverse de ORB et GFTT.

c. Nombre de points détectés

Évidemment, le nombre de points d'intérêts détectés joue également un rôle important dans notre approche, puisqu'il faudra disposer d'un nombre de points suffisant si l'on souhaite obtenir un nombre de correspondances suffisant avant de calculer la transformation entre deux orthoimages. Nous avons donc appliqué la détection de points d'intérêts sur une orthoimage unique et étudié le nombre de points issus de chacun des détecteurs.

Il ressort de cette étape que ORB et GFTT sont les deux détecteurs fournissant le plus de points (Tableau 4.14(b) et Figure 4.15). SIFT au contraire est le détecteur fournissant le moins de points.

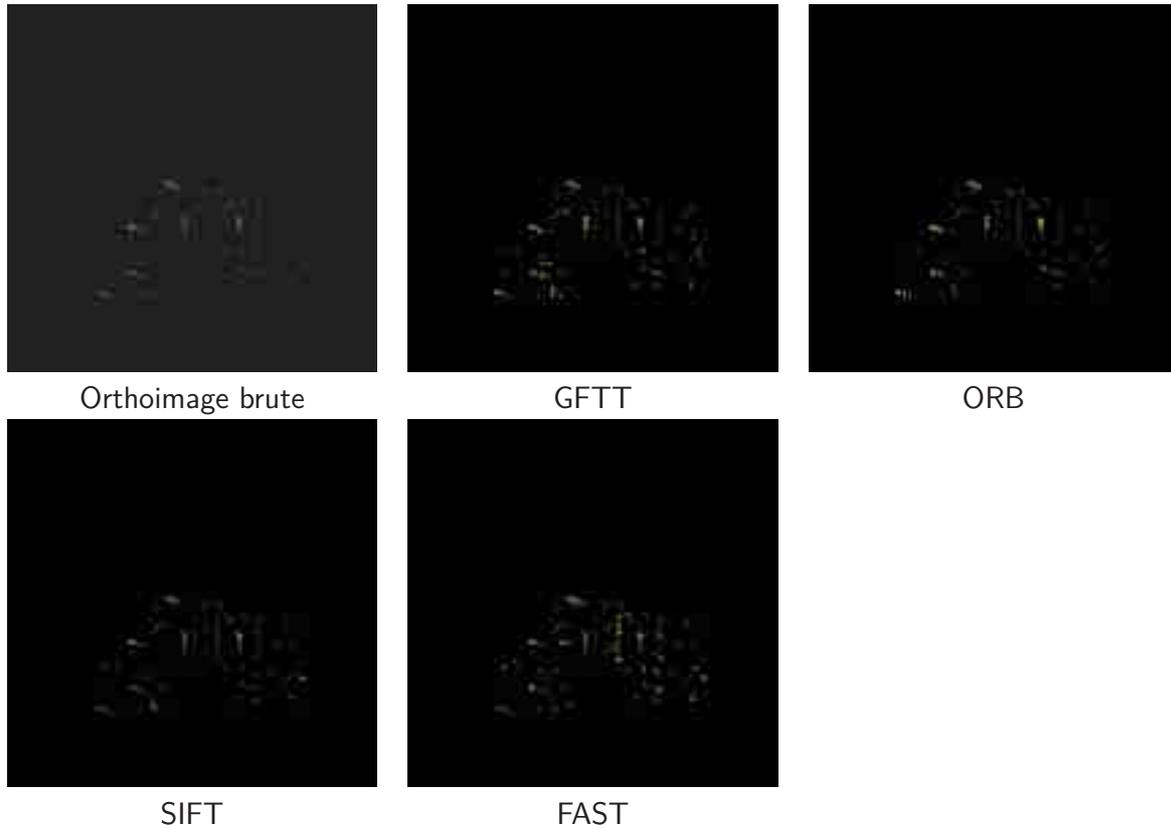


FIGURE 4.15: Application des détecteurs de points d'intérêt évalués sur l'orthoimage présentée sur la Figure 4.11(b) (points détectés en vert)

À l'issue de cette étude sur la répétabilité, le temps de détection, et le nombre de points détectés, nous sommes en mesure de choisir le détecteur à utiliser dans le cadre de notre approche. Concernant la répétabilité, ORB présente de bons résultats, aussi bien pour les rotations que pour les translations. Pour ce qui est du temps de détection, ORB présente à nouveau des résultats supérieurs, de même en ce qui concerne le nombre de points détectés. Pour l'ensemble de ces raisons, ORB est donc le détecteur de points d'intérêt retenu.

3.3. Mise en correspondance

Estimer le mouvement entre deux images revient à faire correspondre les éléments de la première avec ceux de la seconde. De manière générale, la mise en correspondance consiste à retrouver, dans les images, les points d'intérêts homologues détectés préalablement.

La littérature met en évidence deux familles d'algorithmes pour la mise en correspondance : l'appariement (*matching*) et le suivi de points (*tracking*). L'appariement est une approche reposant sur la mesure des similitudes entre les points détectés d'une image à l'autre. Le tracking quant à lui est une approche reposant sur le suivi de points entre images. Nous présentons ici un état de l'art des deux approches.

3.3.1. Appariement (ou *matching* en anglais)

Nous ne pouvons aborder le thème de l'appariement de points sans aborder les descripteurs de points d'intérêts. Un descripteur permet, comme son nom l'indique, de caractériser un point détecté dans une image. Cette caractérisation est en fait une description au voisinage du point d'intérêt. Il peut s'agir tout simplement des niveaux de gris des pixels voisins. Mais dans certains cas, il peut s'agir d'une description plus complexe. [Lowe, 2004] propose, avec son détecteur SIFT, un descripteur du même nom, calculé pour chaque point d'intérêt, et composé d'un vecteur de 128 éléments obtenus à partir d'histogrammes des orientations des vecteurs de gradient au voisinage des points. Nous trouvons également dans [Bay et al., 2006] un descripteur proposé avec le détecteur SURF.

La technique la plus simple pour réaliser l'appariement entre les points d'une image est de comparer, pour chaque point, son descripteur avec ceux de tous les points détectés dans la seconde image. Les descripteurs sont comparés grâce à un calcul de distance (Manhattan ou euclidienne par exemple) entre eux. Il s'agit ensuite d'un problème de recherche des plus proches voisins. Ce problème est étudié depuis de nombreuses années et il existe de nombreuses solutions de type Quadtree [Samet & Webber, 1985], Octree [Eberhardt et al., 2010], kd-tree ([Friedman et al., 1977], [Muja & Lowe, 2009]) ou encore ANN [Arya et al., 1998].

La technique d'appariement présente quelques désavantages, comme le nombre limité de points pouvant être traités. En effet, au-delà d'un certain nombre de points, l'appariement peut devenir très coûteux en temps. De plus, il sera souvent nécessaire de réaliser un appariement "croisé". Cela consiste, si l'on considère les images A et B, à réaliser l'appariement de A vers B, puis de B vers A. L'intersection des résultats obtenus dans les premières étapes permettra ainsi d'éliminer une partie des erreurs d'appariement.

3.3.2. Suivi de points (ou *tracking* en anglais)

La mise en correspondance dans le cadre du suivi de points consiste à rechercher les points détectés en effectuant une recherche locale de similitude. Cette approche nécessite un recouvrement suffisant entre les deux images que l'on cherche à mettre en correspondance. Pour réaliser ce suivi, une des approches consiste à poser l'hypothèse qu'un pixel conserve son intensité lumineuse au cours de son déplacement. [Horn, B. K., & Schunck, 1981] et [Lucas & Kanade, 1981] dans leurs travaux fondateurs, utilisent une version approchée de la condition de conservation de l'intensité lumineuse.

🔍 Condition de conservation de l'intensité lumineuse *Définition*

Cette condition peut s'écrire sous la forme :

$$I(x + \delta x, y + \delta y, t + \delta t) - I(x, y, t) = 0 \quad \forall (x, y) \in \Omega \text{ et } t \in \mathbb{R} \quad (4.14)$$

avec (x, y) les coordonnées du point P considéré, et $\delta x, \delta y$ les coordonnées du vecteur décrivant le déplacement de P au cours de l'intervalle de temps δt .

L'approche proposée par [Lucas & Kanade, 1981] a été développée par [Tomasi & Kanade, 1991] et on la retrouve dans la littérature sous le nom de KLT pour Kanade-Lucas-Tomasi. On trouvera une description complète dans [Baker & Matthews, 2003]. Nous nous contenterons ici d'en faire une brève description.

Soient I et J deux images, et $P = (x, y)^T$ un point de l'image I . On souhaite trouver le point P' de J tel que $I(P)$ et $J(P')$ soient "similaires". L'objectif de la méthode développée par [Lucas & Kanade, 1981] consiste à trouver une transformation W telle que :

$$P' = W(P, T) \quad (4.15)$$

où T désigne les paramètres de la transformation reliant les images I et J . Pour trouver la meilleure transformation, on va donc minimiser l'écart $\epsilon(T)$, qui évalue la similarité entre les deux régions des deux images autour des points P et P' :

$$\epsilon(T) = \sum_{x \in \Omega} [J(W(x, T)) - I(x)] \quad (4.16)$$

avec Ω un voisinage du point P , appelé aussi fenêtre d'intégration. Le déplacement qui minimise l'erreur ϵ est déterminé de façon itérative.

Une limite de cette approche est le fait qu'elle n'est valable que dans le cadre de petits déplacements. Si l'on considère le problème à plusieurs résolutions et en considérant que sur les niveaux les plus grossiers, l'échelle des déplacements est plus petite, alors on peut traiter le problème à des échelles plus grossières et appliquer KLT à des grilles de plus en plus fines. C'est la solution proposée par [Bouguet, 2001] dans son approche multi-résolutions.

3.3.3. Choix de l'approche de mise en correspondance

La mise en correspondance joue un rôle clé dans l'approche d'odométrie visuelle, puisque la précision et la justesse des mises en correspondance va avoir un impact sur le résultat du calcul de la transformation entre deux positions successives de la caméra. Ainsi, il est nécessaire de choisir l'approche la plus efficace au regard des données que nous exploitons. Pour ce faire, nous avons comparé les deux approches présentées précédemment avec un jeu de données simulées (Figure 4.16).

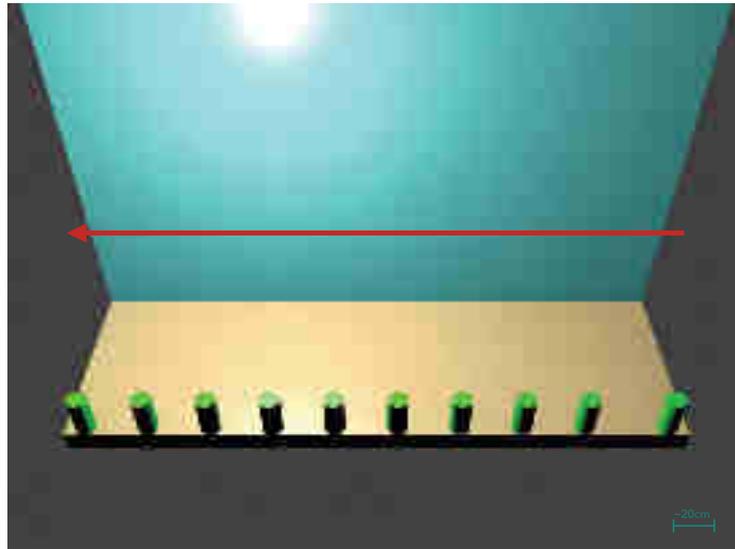


FIGURE 4.16 : Scène simulée utilisée pour évaluer les différentes approches de mise en correspondance (flèche rouge : sens du déplacement)

La scène utilisée pour réaliser cette comparaison a été créée dans Blensor [Gschwandtner et al., 2011]. Il s'agit ensuite d'appliquer la mise en correspondance sur les orthoimages calculées à partir des nuages obtenus par simulation. Nous cherchons donc à mettre en correspondance des points d'intérêts détectés dans deux orthoimages successives. Le déplacement réalisé entre ces deux orthoimages est une translation pure, représentée en rouge sur la Figure 4.16.

L'étude menée précédemment sur la détection de points d'intérêt nous a permis de déterminer que ORB [Rublee et al., 2011] était l'approche la plus adaptée à nos données.

Pour choisir l'approche de mise en correspondance, nous allons appliquer tout d'abord la solution proposée par Lucas & Kanade et implémentée par [Tomasi & Kanade, 1991].

La Figure 4.17 présente le résultat de la mise en correspondance basée sur le suivi de points. Les mises en correspondance sont affichées en vert. Nous pouvons observer que la plupart des mises en correspondances semblent cohérentes (même direction des vecteurs, même norme).

Pour tester la solution de mise en correspondance par appariement, nous avons choisi d'appliquer celle présente dans OpenCV de [Muja & Lowe, 2009]. Elle repose sur l'utilisation des kd-trees aléatoires multiples, permettant d'améliorer le temps de recherche des correspondances. Nous pouvons observer que les résultats de la mise en correspondance par appariement sont bien plus bruités que ceux obtenus par suivi de points (Figure 4.18). L'uniformité des orthoimages peut expliquer ce phénomène. En effet, les descripteurs des points d'intérêts détectés seront peu distinctifs entre eux.

À l'issue de plusieurs tests de ce type, et au vu des résultats qu'ils ont produits, nous avons choisi d'utiliser le suivi de points dans le reste de l'approche concernant le pôle caméras 3D.

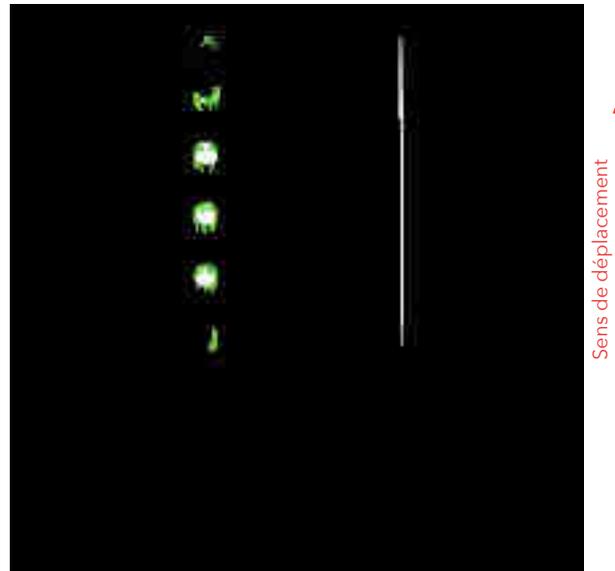


FIGURE 4.17 : Résultat de la mise en correspondance grâce au suivi de points

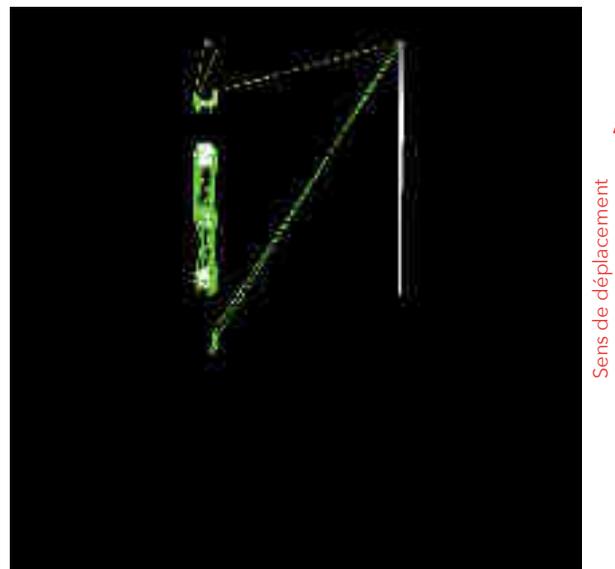


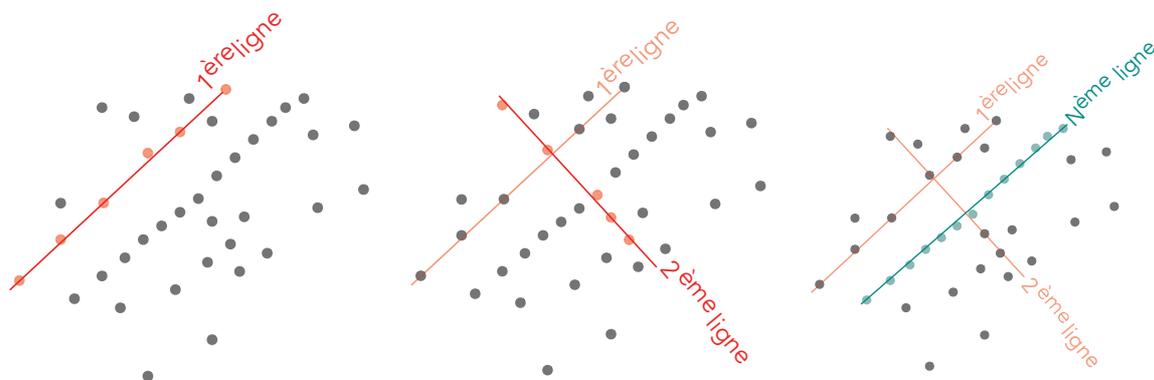
FIGURE 4.18 : Résultat de la mise en correspondance grâce à l'appariement

3.4. Élimination des correspondances aberrantes et estimation de la transformation

Malgré des techniques de mise en correspondance performantes, il subsiste cependant de mauvaises associations. Il existe différentes causes à ces erreurs : le bruit, les occlusions, le bruit de mouvement, des changements d'illumination ou encore de trop brusques changements de points de vue. Le paradigme de RANSAC est une méthode fréquemment utilisée pour résoudre le problème d'élimination des données aberrantes.

3.4.1. RANSAC (RANdOm SAMple Consensus)

Le principe du paradigme de RANSAC consiste à estimer un modèle [Fischler & Bolles, 1981]. À chaque itération de l'algorithme, un sous-ensemble est tiré de façon aléatoire parmi les données à traiter. À partir de ce sous-ensemble, une hypothèse de modèle est établie. Pour chaque valeur du jeu de données restantes, on contrôle si elle vérifie cette hypothèse. Si celle-ci est vérifiée, on considère la valeur testée comme valide, sinon elle est classifiée comme aberrante. Lorsque l'ensemble des valeurs ont été testées, un nouveau sous-échantillon de tests est tiré au hasard. Une fois le nombre d'itérations maximum atteint, l'hypothèse qui est vérifiée par le plus grand nombre de valeurs est considérée comme la solution répondant au problème.



Étape 1 : une ligne est choisie au hasard. On compte le nombre de points entrant dans la tolérance fixée.

Étape 2 : tant qu'aucune ligne choisie au hasard n'est "meilleure", la première est retenue.

Étape 3 : après N itérations, la dernière "meilleure" est retenue.

FIGURE 4.19 : Principe de l'algorithme RANSAC. Illustration des étapes du paradigme de RANSAC pour la détection de ligne 2D.

Dans le cas de l'odométrie visuelle, les données en entrée de l'algorithme de RANSAC sont les couples de points mis en correspondance à l'étape précédente et le modèle à établir est la transformation (R, t) entre deux positions successives de la caméra.

Le nombre d'itérations de l'algorithme RANSAC nécessaires pour obtenir un résultat correct peut être calculé de la façon suivante :

$$N = \frac{\log(1-p)}{\log(1-(1-\eta)^s)} \quad (4.17)$$

avec N le nombre d'itérations, s le nombre d'inconnues, η le pourcentage de valeurs aberrantes, et p la probabilité de succès de l'algorithme. Comme nous avons pu le voir auparavant, RANSAC repose sur un tirage aléatoire et fournit des solutions différentes à chaque exécution ; la solution permettant de le rendre plus stable consiste à augmenter le nombre d'itérations. Nous pouvons observer dans l'équation 4.17 que le nombre d'itérations nécessaires pour converger vers une solution valable augmente en fonction du nombre d'inconnues s . Dans [Longuet-Higgins, 1987], l'auteur propose un algorithme à huit points pour calculer la transformation entre deux positions d'une caméra ; [Nistér et al., 2004], quant à eux, ont proposé une solution à cinq points. On peut également citer les travaux de [Kneip et al., 2011], qui exploitent les données fournies par un IMU afin d'évaluer la matrice de rotation, et donc de réduire le nombre de paramètres à estimer. Une attention particulière est portée aux travaux de [Scaramuzza et al., 2009], dont l'approche présente différents avantages. L'approche proposée considère le cas d'un problème plan ; dans ce cas, la complexité du déplacement peut être décrite avec deux points [Ortin & Montiel, 2001]. [Scaramuzza et al., 2009] exploite la contrainte non-holonomique du mobile utilisé, pour réduire encore le nombre de degrés de liberté (2) et pour aboutir à un algorithme ne nécessitant qu'une unique mise en correspondance de points. Une évaluation de ces différentes approches est proposée dans [Scaramuzza, 2011].

3.4.2. Méthodologie mise en place pour l'élimination des valeurs aberrantes

La méthodologie choisie en vue d'éliminer les mises en correspondances aberrantes s'appuie sur les analyses précédentes.

Le mobile sur lequel est placé notre système répond à la contrainte non-holonomique exploitée par [Scaramuzza et al., 2009]. L'exploitation de cette contrainte permet de limiter le nombre d'inconnues composant la transformation à déterminer, c'est pourquoi cette approche est intéressante. Dans le Chapitre 2, nous avons introduit la notion de Centre Instantané de Rotation (CIR) et de contrainte non-holonome (paragraphe 3.3.). Ainsi, lorsqu'un véhicule de type tri-cycle se déplace, nous pouvons décrire l'ensemble de ses déplacements comme des rotations. La trajectoire décrite par la roue libre est un cercle dont le centre n'est autre que le CIR (Figure 4.20).

Nous considérons notre système comme positionné en O_s sur la plateforme mobile (Figure 4.20). L'axe Z_s est orthogonal au plan dans lequel le déplacement a lieu, et l'axe X_s est perpendiculaire à l'axe des roues arrières du mobile. Après un déplacement, nous pouvons définir l'origine O'_x et les axes X'_s et Z'_s du système. Le mouvement du système entre ces deux positions peut être décrit localement comme un déplacement circulaire. Ainsi, ce déplacement peut être décrit grâce à deux paramètres, l'angle de rotation θ et le rayon de courbure ρ (celui-ci tend vers l'infini dans le cas d'une translation pure). Dans ce contexte, nous pouvons définir chaque mise en correspondance comme un "vecteur de déplacement" également défini par les paramètres θ et

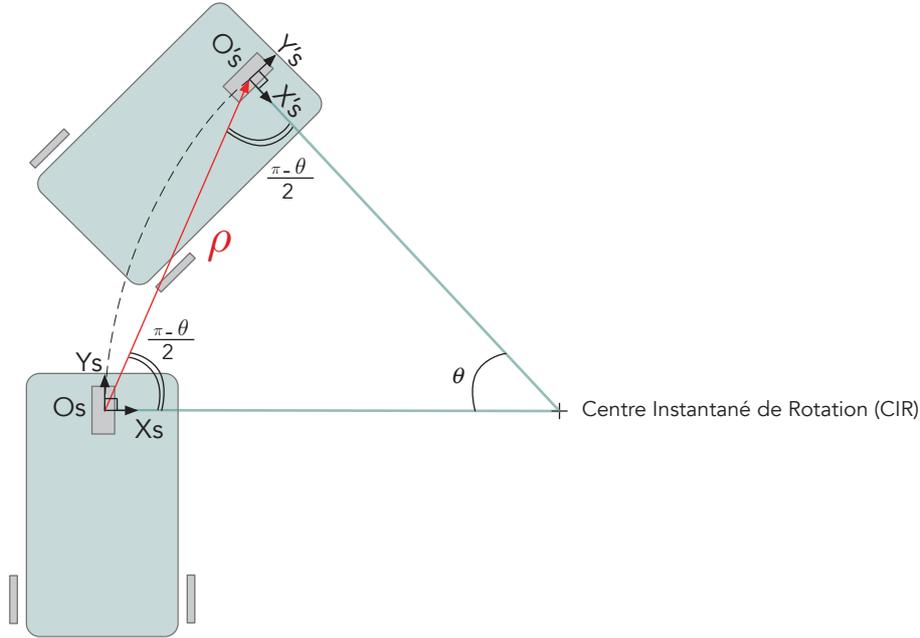


FIGURE 4.20 : Mouvement du système rapporté au CIR pendant un déplacement

ρ . Comme nous l'avons montré auparavant, la transformation nécessaire pour passer du repère \mathcal{R}_s au repère \mathcal{R}'_s se compose d'une rotation $R_{s'}^s$ et d'une translation $t_{s'}^s$ (équations 4.18 et 4.19).

$$R_{s'}^s = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.18)$$

$$t_{s'}^s = \rho \begin{pmatrix} \sin\frac{\theta}{2} \\ \cos\frac{\theta}{2} \\ 0 \end{pmatrix} \quad (4.19)$$

Nous noterons p et p' ($p = [x, y, 1]^T$ et $p' = [x', y', 1]^T$), en coordonnées homogènes, des points présents dans deux orthoimages successives et mis en correspondance. Sachant que p peut être exprimé dans \mathcal{R}_s et p' dans \mathcal{R}'_s , nous pouvons écrire :

$$p' = R_{s'}^s \cdot p + t_{s'}^s \quad (4.20)$$

Ainsi, on obtient :

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta & \rho \sin\frac{\theta}{2} \\ -\sin\theta & \cos\theta & \rho \cos\frac{\theta}{2} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (4.21)$$

Si on pose $\frac{\theta}{2} = \Phi$, on aura alors :

$$\rho = \frac{x' - x \cos 2\Phi - y \sin 2\Phi}{\sin \Phi} \quad (4.22)$$

$$\rho = \frac{y' + x \sin 2\Phi - y \cos 2\Phi}{\cos \Phi} \quad (4.23)$$

Il est alors possible d'en déduire θ (Démonstration en annexe E) de la façon suivante :

$$\theta = 2 \tan^{-1} \left(\frac{x' - x}{y' + y} \right) \quad (4.24)$$

À ce stade, pour chaque mise en correspondance, nous sommes en mesure de calculer la valeur θ . L'analyse des valeurs de θ pour chaque vecteur de déplacement va permettre d'éliminer les valeurs aberrantes. La solution développée consiste à construire, à partir de l'ensemble des valeurs de θ , un histogramme de fréquences d'occurrences. Chacune des classes de cet histogramme représente le nombre de vecteurs ayant la même valeur de θ (Figure 4.21).

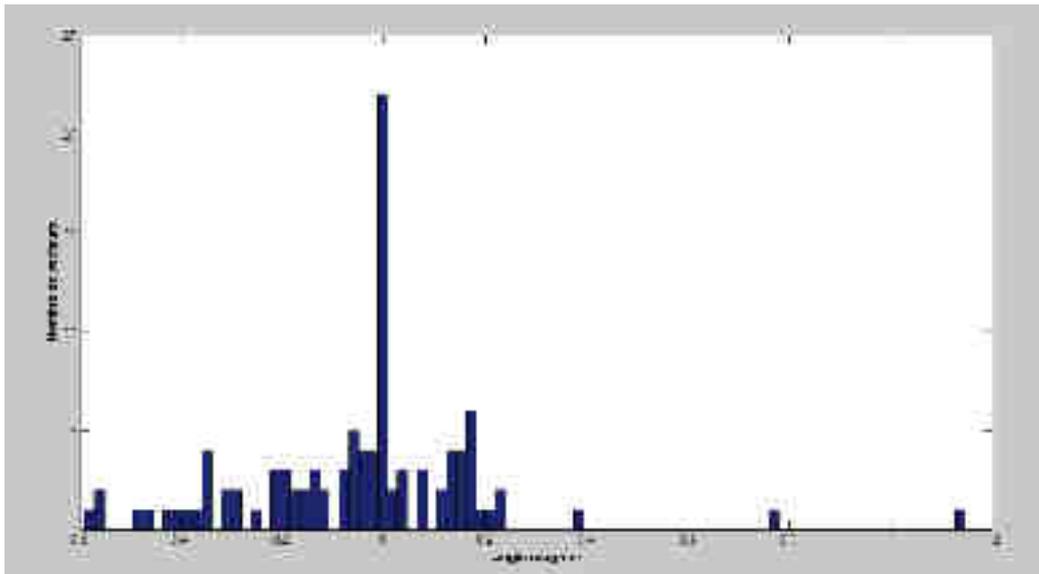


FIGURE 4.21 : Histogramme des valeurs de θ des mises en correspondance présentées dans la Figure 4.17

Sur la base de cet histogramme, nous pouvons sélectionner les vecteurs dont la valeur de θ est égale à $\text{argmax}\{H\} \pm \epsilon$. La valeur de ϵ a été fixée à 0.1° , de façon à éliminer le plus de valeurs aberrantes possibles.

L'utilisation de la contrainte non-holonomique et de l'histogramme des valeurs de θ présente plusieurs avantages. Premièrement, celle-ci n'est pas itérative, contrairement au paradigme de

RANSAC "classique". Ceci permet de réduire la complexité algorithmique du développement. Deuxièmement, comme nous l'avons précisé au paragraphe 3.4.1., la formation des orthoimages implique que les points d'intérêts les plus lointains sont les moins précis. Ainsi, dans le cas du paradigme de RANSAC, il est possible que les données sélectionnées lors du tirage soient de qualité moindre. La transformation finale estimée pourrait alors s'en trouver affectée.

Bien que l'approche de [Scaramuzza et al., 2009] soit pertinente et permette l'élimination d'un grand nombre de correspondances fausses, nous avons malgré tout mis en place plusieurs filtres supplémentaires permettant d'améliorer la qualité des données exploitées avant le calcul de transformation entre deux orthoimages successives.

Le premier test mis en place cherche à éliminer les correspondances générées par la présence de bruit (Figure 4.5). En effet, le bruit de mesure observé entre deux acquisitions peut entraîner un "faux déplacement". Les caméras utilisées pour la réalisation de notre prototype fonctionnent par intervalles de mesures [Bernard, 2011]. Ainsi, entre deux acquisitions successives, le nuage de points fourni par le système de caméra va varier. Cette variation, même minime, va avoir un impact sur les orthoimages générées à partir de ces nuages. Ce phénomène va alors engendrer des mises en correspondance fictives.

Il est nécessaire de détecter ces situations. Pour ce faire, nous avons mis en place un test, qui évalue la distance séparant deux points homologues pour chacune des mises en correspondance. Si la moyenne arithmétique des distances est inférieure à 2 cm, alors nous considérons qu'il n'y a pas de déplacement. La transformation résultante, entre les instants k et $k + 1$ prend alors la forme suivante :

$$T_{k,k+1} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.25)$$

Si ce premier test met en évidence un déplacement, nous effectuons un nouveau test. Celui-ci concerne cette fois le nombre de mises en correspondances conservées à l'issue des précédentes étapes. En effet, si le nombre d'observations n'est pas suffisant, nous ne pouvons assurer une estimation de la transformation représentative du mouvement de la plateforme mobile. Ainsi, nous considérons qu'en-dessous de 15 mises en correspondance évaluées comme étant valables, nous ne pouvons pas considérer les observations comme suffisantes pour estimer la transformation entre deux orthoimages successives. En opérant ainsi, nous émettons l'hypothèse que le déplacement du mobile est réalisé à vitesse constante et nous utilisons les transformations précédentes pour calculer la transformation en cours. La transformation $T_{k,k+1}$ est donc calculée à partir des n dernières transformations. Cette valeur n sera déterminée en tenant compte de la vitesse de déplacement du mobile.

3.5. Calcul de la transformation

L'étape finale de notre approche consiste à calculer la transformation entre deux jeux de points mis en correspondance (voir organigramme de la Figure 4.5, page 94). Nous cherchons donc, lors de cette étape, à déterminer les paramètres de transformation les plus probables, en respectant le principe des moindres carrés, permettant de passer du repère $\mathcal{R}_s(k)$, représentant la position du mobile à l'instant k , au repère $\mathcal{R}_s(k+1)$, représentant la position du mobile à l'instant $k+1$.

Si l'on considère l'ensemble n des points $(x_1, y_1)_i$, avec $i = 1 \dots n$, appartenant à l'orthoimage 1, et les n points $(x_2, y_2)_i$ leurs homologues dans l'orthoimage 2, nous cherchons à retrouver les paramètres de la transformation rigide à 2 dimensions entre deux images ramenées dans le même plan. Ainsi, les paramètres de transformation à déterminer sont une rotation et deux translations. De plus, nous ne considérons aucune déformation entre les orthoimages, le coefficient d'échelle est donc égal à un. La transformation de Helmert est une similitude permettant de passer d'un système de coordonnées à un autre, en minimisant l'écart quadratique moyen entre les positions des points connus dans le système cible et leurs transformées depuis le système source, à condition de disposer d'observations excédantes [Ghilani, 2010]. Parmi les différentes transformations linéaires, sa simplicité, et ses caractéristiques en font une approche fréquemment utilisée. Celle-ci permet de déterminer une rotation et deux paramètres de translation, ainsi qu'un facteur d'échelle. Elle répond donc à nos besoins en termes de paramètres de transformation à évaluer. La transformation de Helmert 2D à déterminer est définie telle que présentée dans l'équation 4.26.

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = k \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (4.26)$$

où θ représente l'angle de rotation, t_x et t_y sont les paramètres de translation, et k le facteur d'échelle (égal à 1 dans notre cas).

L'équation 4.26 n'est pas linéaire en θ . La linéarisation permet de s'affranchir des itérations, et ainsi de diminuer le temps nécessaire à la convergence vers la solution répondant au principe des moindres carrés. Cette linéarisation permet également de fixer des valeurs approchées des inconnues à 0, ce qui réduit la complexité algorithmique de notre système. Afin de linéariser l'équation 4.26, on pose $a = k\cos\theta$ et $b = k\sin\theta$ [Strang & Borre, 1997]. L'équation 4.26 prend alors la forme suivante :

$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} a & b \\ -b & a \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} t_x \\ t_y \end{pmatrix} \quad (4.27)$$

Afin d'améliorer les précision de la translation estimée, précision qui dépend de la position d'origine des deux systèmes, un changement de base est effectué. Les coordonnées des points de chacune des orthoimages sont donc réduites aux barycentres respectifs de chaque ensemble de points.

Si l'on considère n le nombre d'observations, et q le nombre d'inconnues à déterminer, on cherche alors à minimiser la formule suivante :

$$V_{(n,1)} = A_{(n,q)} \cdot d_x - K_{(n,1)} \quad (4.28)$$

où $V_{(n,1)}$ est le vecteur des écarts apparents, $A_{(n,q)}$ la Jacobienne, d_x le vecteur des inconnues et $K_{(n,1)}$ le vecteur des $-k_i$ (écarts entre observation brute et observation compensée approchée).

Le vecteur des inconnues est défini tel que :

$$d_x = N_{(q,q)}^{-1} \cdot n_k \quad (4.29)$$

où N est la matrice normale et n_k la matrice des coefficients en k . Ces deux matrices sont obtenues à partir des équations 4.30 et 4.31.

$$N_{(q,q)} = A_{(n,q)}^t \cdot P \cdot A_{(n,q)} \quad (4.30)$$

$$n_k = A_{(n,q)}^t \cdot P \cdot K_{(n,1)} \quad (4.31)$$

où P représente la matrice des poids des observations, définie telle que :

$$P_{(n,n)} = \begin{pmatrix} p_1 & 0 & \dots & 0 \\ 0 & p_2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & 0 & \dots & p_n \end{pmatrix} \quad (4.32)$$

À ce stade de notre étude, nous avons attribué un poids identique à l'ensemble des observations. Une évolution aisée à réaliser serait de faire varier le poids des observations en fonction de leurs positions dans les orthoimages. En effet, l'approche utilisée pour former ces images et le fonctionnement des caméras 3D engendrent une précision des points détectés dépendante de leur position, ainsi les points les plus éloignés du mobile seront moins précis que les plus proches.

L'approche choisie pour le calcul de la transformation permet également de calculer l'erreur moyenne quadratique de l'unité de poids (emq_0), qui représente la précision d'une observation dont le poids vaut 1 (équation 4.33)¹. Le principe des moindres carrés est respecté lorsque l' emq_0 atteint un minimum. Lorsqu'elle est élevée, cela signifie que les corrections ($V_{(n,1)}$) des observations brutes sont élevées. Cela peut provenir du fait que les observations sont mauvaises ou que les correspondances ne sont pas satisfaisantes. Cette grandeur statistique devient ainsi un critère de qualité intéressant à prendre en compte dans la détection de correspondances aberrantes.

1. Si l'on suppose que les observations en entrée au calcul ne sont pas entachées d'erreurs systématiques, l'erreur moyenne quadratique est un estimateur de l'écart-type. L'équation 4.33 présente ainsi la variance de l'unité de poids. [Landes, 2011]

$$emq_0^2 = \frac{V^t P V}{n - q} \quad (4.33)$$

Cet indicateur rentre directement en jeu dans notre approche. En effet, lorsque la valeur de emq_0^2 est trop élevée (valeur seuil fixée empiriquement à 0.05 m^2), nous contrôlons les écarts venant corriger chaque observation.

Malgré les différents filtrages réalisés auparavant, il peut subsister encore de mauvaises mises en correspondance. Typiquement, et nous le verrons dans la suite du manuscrit, il se peut que certaines paires de points aient une valeur de θ cohérente avec le déplacement, mais que les points mis en correspondance n'aient pas à l'être. Afin de déterminer les observations à éliminer, un seuil est déterminé à partir de la moyenne arithmétique des résidus. Ainsi, lorsqu'une observation présente un écart supérieur à ce seuil, celle-ci est éliminée du jeu de données, et la transformation est à nouveau estimée.

A l'issue de l'ensemble de ces étapes, nous sommes en mesure de fournir la matrice de transformation traduisant le déplacement de notre système dans le repère global. Les résultats obtenus pour le calcul de trajectoire sont présentés dans le chapitre 5 de ce manuscrit.

4 Pôle caméra *Fish Eye*

Comme mentionné au paragraphe 3.3. du chapitre 3, nous envisageons d'équiper notre système d'une caméra de type *Fish Eye*. Dans le cadre des travaux menés par [Royère, 2014], nous avons pu étudier la capacité de localisation d'un mobile en environnement urbain à partir d'une caméra équipée d'une optique *Fish Eye*.

L'approche développée autour de la caméra *Fish Eye* est également inspirée de l'odométrie visuelle. Comme nous avons pu le voir précédemment dans ce document (Chapitre 3 - paragraphe 3.3.3.), le modèle de calibration choisi pour l'objectif utilisé nous permet d'appliquer des approches normalement appliquées à des objectifs classiques. Il s'agit ici d'étudier les différences entre les images successives fournies par la caméra *Fish Eye* pour évaluer le déplacement de notre système.

L'approche développée est composée de cinq étapes principales :

- **Détection de points d'intérêts** : il s'agit ici de détecter les points d'intérêts dans les images fournies par la caméra *Fish Eye*.
- **Mise en correspondance** : les points détectés entre les images successives sont mis en correspondance.
- **Élimination des correspondances aberrantes** : parmi les points détectés et appariés grâce aux étapes précédentes, il subsiste des erreurs de détection ou d'appariement, qu'il est nécessaire d'éliminer avant le calcul de transformation.
- **Calcul de la transformation** : à partir des appariements filtrés auparavant, il est alors possible de calculer la transformation effectuée entre deux positions de la caméra.

- **Extrapolation** : la position et l'orientation de la caméra *Fish Eye*, mais également ses caractéristiques techniques, permettent d'avoir une vision à "long terme" de la scène. Dans ce contexte, il est possible de travailler avec plusieurs images successives.

Nous revenons de façon brève sur chacune de ces étapes, et sur les choix réalisés pour les mettre en œuvre. Le lecteur intéressé pourra retrouver l'ensemble des travaux développés en annexe (Annexe F). Il est à noter également que les développements n'ont pas été réalisés dans les mêmes conditions que pour le pôle caméras 3D, ce qui explique certains choix notamment concernant l'étude sur les points d'intérêts. L'outil Matlab a été utilisé pour l'ensemble des programmes réalisés en rapport avec la caméra *Fish Eye*.

4.1. Détection de points d'intérêts

Un ensemble de détecteurs de points d'intérêts a à nouveau été évalué. Les détecteurs SURF [Bay et al., 2006], FAST [Muja & Lowe, 2009], Shi & Tomasi [Shi & Tomasi, 1994], MSER [Matas et al., 2004] et Harris [Harris & Stephens, 1988] ont été considérés. Cette étude s'est limitée à ces différents détecteurs, car leurs implémentations sont disponibles au sein de l'outil Matlab.

Détecteur	FAST	MSER	SURF	Shi_Tomasi	Harris
Points détectés	516	334	998	2847	814
Durée détection	0,06s	1,39s	1,74s	2,17s	1,96s

TABLE 4.1 : Tableau de comparaison des méthodes de détection de points d'intérêt sur une image *Fish Eye* acquise en extérieur

Afin de choisir le détecteur le plus adapté à notre approche, nous avons comparé la rapidité et le nombre de points détectés par chacune de ces implémentations. Compte tenu du fait que des mises en correspondances sont susceptibles d'être aberrantes, il est nécessaire de détecter suffisamment de points d'intérêts dans une image. Cependant, ce nombre doit rester limité pour ne pas accentuer le temps de calcul, ni compromettre la précision de la trajectographie. L'aspect temporel joue un rôle important dans notre approche, il est donc nécessaire de prendre en compte ce critère pour choisir le détecteur de points d'intérêts adéquat.

Nous avons évalué chacun de ces critères sur une même image acquise avec notre caméra *Fish Eye* (Figure 4.22).

Le tableau des résultats permet de tirer plusieurs conclusions (Tableau 4.1). Le nombre de points détectés par le détecteur MSER apparaît comme étant trop faible pour l'étude que nous souhaitons mener. A contrario, le détecteur Shi & Tomasi fournit un nombre de points trop important. La comparaison entre les détecteurs Harris et SURF peut être faite sur la répartition des points détectés. Nous pouvons observer sur la Figure 4.22 que les points détectés sont répartis de façon plus homogène avec le détecteur SURF. Ceci aura une influence sur la précision de la transformation estimée, il est donc important d'en tenir compte. Dans ce contexte, notre choix

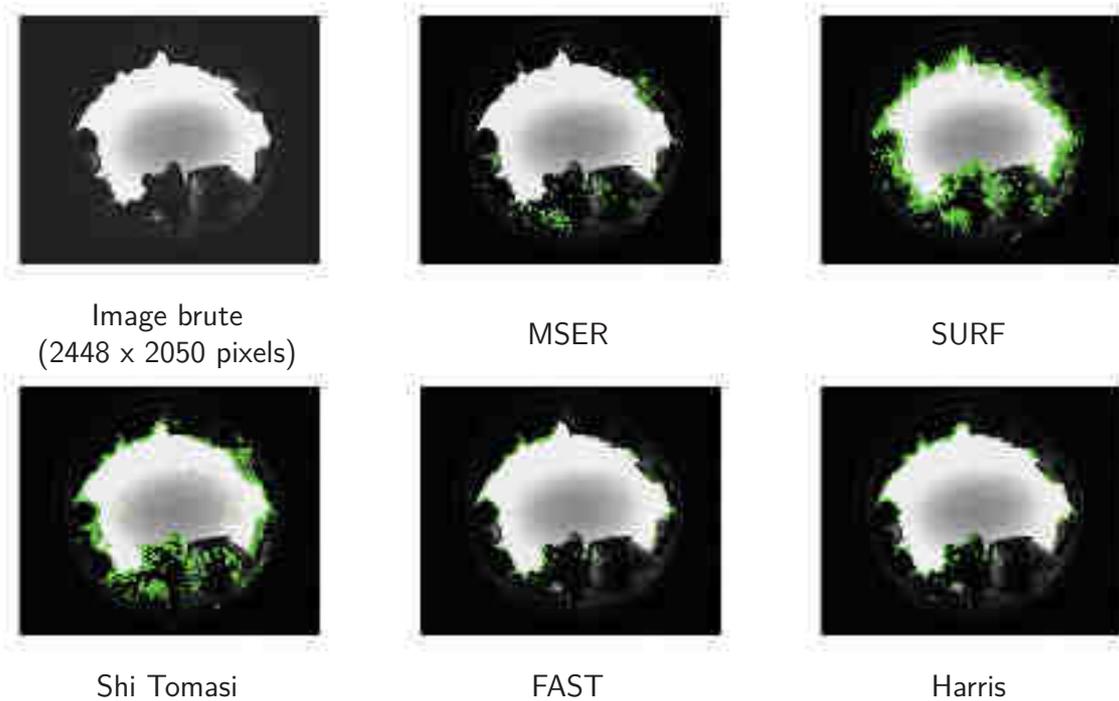


FIGURE 4.22: Application des détecteurs de point d'intérêts évalués sur une image *Fish Eye* acquise avec la caméra de notre système

s'est tourné vers SURF pour la détection de points d'intérêts au sein des images acquises par la caméra *Fish Eye*.

4.2. Mise en correspondance

Comme déjà vu, la mise en correspondance repose sur deux approches distinctes : l'appariement, et le suivi de point. L'appariement de points repose sur l'utilisation de descripteurs de points d'intérêts. Le détecteur de points d'intérêts SURF est proposé avec un descripteur du même nom [Bay et al., 2006]. La mise en correspondance mise en place, dans le cadre de la méthodologie développée autour de la caméra *Fish Eye*, repose sur l'utilisation d'un appariement grâce au descripteur fourni par SURF.

4.3. Élimination des correspondances aberrantes

Après ces deux premières étapes, des points ont été détectés dans les images fournies par la caméra, et ceux-ci ont été mis en correspondance au travers des images. Cependant, il subsiste de nombreuses erreurs de détection ou d'appariement, qu'il est nécessaire d'éliminer afin de garantir la précision de la transformation finale.

Pour cette étape, nous avons choisi d'exploiter les résultats obtenus précédemment avec le pôle caméras 3D. La solution mise en place est donc basée sur les travaux réalisés par [Scaramuzza

et al., 2009] et sur l'exploitation de la contrainte non-holonomique du mobile sur lequel le système est installé. L'intérêt de cette contrainte est qu'elle limite le nombre d'inconnues composant la transformation à déterminer. À nouveau, nous avons exploité l'histogramme des valeurs de θ pour guider l'élimination des correspondances aberrantes.

4.4. Calcul de la transformation

À partir de l'ensemble des mises en correspondance restantes, nous sommes en mesure de calculer la transformation entre les images fournies par la caméra *Fish Eye*. Comme nous travaillons ici sur des images acquises par une caméra numérique, nous nous trouvons confrontés à des problématiques connues en photogrammétrie. La géométrie épipolaire dense est souvent employée en reconstruction 3D. On la retrouve également sous le nom de "corrélation épipolaire dense". Comme précisé dans [Hullo, 2009], la contrainte épipolaire est une technique dite "passive", car elle permet la reconstruction en trois dimensions sans aucune autre information sur l'objet que les photographies. Ainsi, lorsque l'on prend en photo une même scène, sous un angle différent, le point objet X se projette sur l'image gauche en x_G (Figure 4.23). Nous cherchons à déterminer les coordonnées x_D du point dans la seconde image. L'utilisation de la contrainte épipolaire et la connaissance des épipoles e_G et e_D respectivement dans les images 1 et 2, nous permet de déterminer un rayon de recherche de la position du point dans la seconde image, à savoir la ligne correspondant à l'intersection entre le plan de l'image 2 et le plan passant par les deux épipoles e_G , e_D et le point image x_G .

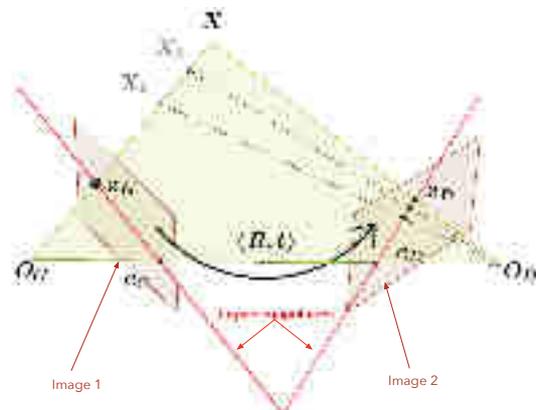


FIGURE 4.23: Principe de la géométrie épipolaire

où O_G et O_D sont les centres de projection des caméras, (R, t) représente la transformation intervenue entre les deux prises de vues.

Nous pouvons trouver différentes approches menant à la détermination des paramètres de la géométrie épipolaire. En photogrammétrie, les épipoles sont obtenues grâce au calcul de l'orientation relative, à l'aide des équations de colinéarité [McGlone, 2013]. En vision par ordinateur, le calcul de la géométrie épipolaire repose sur le calcul de la matrice fondamentale [Hartley &

Zisserman, 2003].

Dans le cadre des travaux menés au cours de cette thèse, l'utilisation de la géométrie épipolaire n'a pas pour finalité la réalisation d'une reconstruction tridimensionnelle. L'emploi de celle-ci a pour objectif de trouver la transformation entre deux images. Notre approche utilise la matrice essentielle afin de déterminer les paramètres de rotation et de translation entre deux images. Cette même matrice est obtenue grâce aux paramètres d'orientation interne de la caméra, et grâce à la matrice fondamentale.

La matrice fondamentale est la matrice carrée F de dimension 3 qui vérifie l'équation 4.34 où x_G et x_D sont les coordonnées homogènes dans les images 1 et 2 (ou gauche et droite).

$$x_D^T \cdot F \cdot x_G = 0 \quad (4.34)$$

Dans [Hartley & Zisserman, 2003], il est précisé que la matrice fondamentale est la représentation algébrique de la géométrie épipolaire. Celle-ci est simple à calculer et ses propriétés sont nombreuses. Par exemple, les deux épipoles e_G et e_D sont calculés par résolution des équations $F \cdot e_D = 0$ et $F^T \cdot e_G = 0$. À partir de F , il est alors possible de déterminer la matrice essentielle E , qui permettra à son tour de déterminer les paramètres de rotation et de translation. La matrice E est une spécialisation de la matrice F ; elle tient compte des coordonnées normalisées, et des paramètres de calibration de la caméra.

Les paramètres internes des caméras rassemblés dans les matrices de calibration K_G et K_D permettent le calcul de la matrice essentielle E (équation 4.35).

$$E = K_D^T \cdot F \cdot K_G \quad (4.35)$$

La matrice fondamentale, ou dans le cas des coordonnées normalisées, la matrice essentielle, peut être calculée de différentes manières. Parmi les principaux algorithmes permettant de déterminer ces matrices, on trouve l'algorithme à 8 points proposé par [Longuet-Higgins, 1987]. Cette approche présente l'avantage d'être simple à implémenter. [Nistér, 2004] propose un algorithme à 5 points qui présente l'avantage de nécessiter moins de points que l'approche proposée par [Longuet-Higgins, 1987]. Cependant, elle est plus complexe à mettre en place.

A l'aide de la matrice essentielle, il est possible de déterminer les paramètres de transformation R et t par l'équation 4.36.

$$E = [t]_x R \quad (4.36)$$

avec :

— $[t]_x = \begin{pmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{pmatrix}$ matrice carrée antisymétrique de dimension 3 représentant les variables du vecteur de translation t .

— R la matrice rotation.

À partir de cette relation, il est possible de construire la matrice de rotation permettant de passer de l'image 1 à l'image 2. Comme pour le pôle caméras 3D, un ensemble de contrôles est mis en place afin de rendre le calcul plus robuste. On peut citer notamment des contrôles sur la distance entre images successives mis en œuvre pour éliminer de "faux" déplacements.

4.5. Calcul de position moyenne

Le champ de vision important de la caméra *Fish Eye* lui permet d'observer la scène durant une longue période, comparée à celle durant laquelle une caméra voit le même objet dans son champ de vision. De plus, l'orientation de l'objectif sur le système amplifie ce phénomène. Il est alors possible d'introduire ici deux paramètres : un premier permettant de fixer un intervalle sur le nombre d'images à utiliser (une image sur deux, une image sur trois, etc...) et un second permettant de fixer la taille de la fenêtre de calcul à considérer dans l'estimation de la position en cours (Figure 4.24). Ainsi, lors du calcul de la position du mobile, une moyenne des positions estimées, à partir des n dernières transformations calculées est opérée. Cette moyenne est pondérée en fonction de la distance séparant les deux images utilisées.

La transformation est déterminée à un facteur d'échelle près, fixé arbitrairement au départ. Cependant, il est possible de déterminer sa valeur réelle : soit en utilisant un moyen de mesure extérieur (utilisation de points connus en 3 dimensions), soit en exploitant les informations fournies par le pôle caméras 3D.

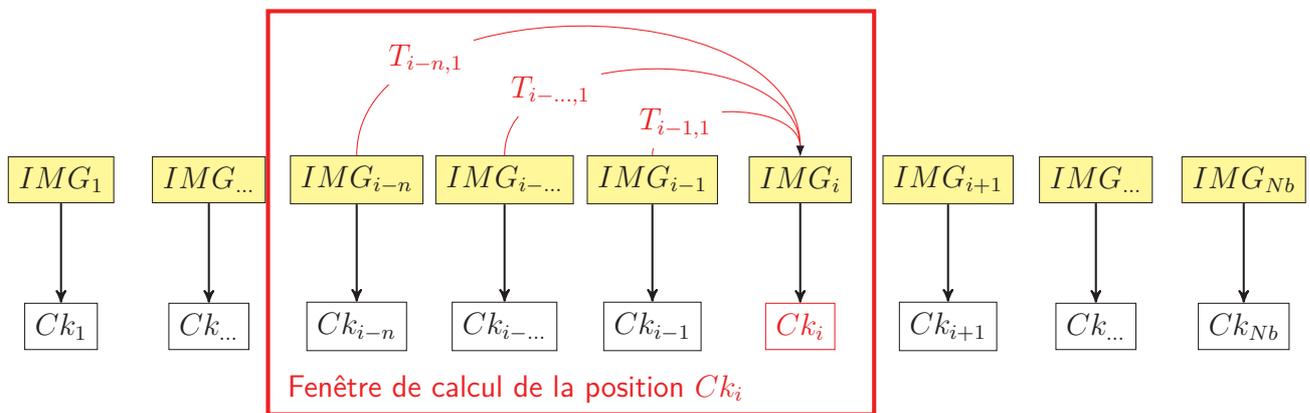


FIGURE 4.24 : Procédure de calcul de position moyenne, utilisant les " n " précédentes images pour le calcul de la matrice position Ck_i . On notera Ck_{Nb} la dernière position à calculer.

Conclusions du chapitre 4

Dans ce chapitre, nous avons tout d'abord détaillé la technique d'odométrie consistant à évaluer le déplacement d'un mobile à partir d'images enregistrées par un système de vision. Dans notre cas, il s'agit d'évaluer le déplacement d'un mobile à partir du traitement des orthoimages fournies par le système de caméras 3D, et à partir d'images fournies par la caméra *Fish Eye*. Nous avons ensuite présenté les différentes méthodologies mises en place pour permettre le calcul de la localisation de notre système.

D'une part, nous avons présenté l'approche développée autour du pôle caméras 3D. Celle-ci repose sur quatre étapes principales : la détection de points d'intérêts dans les orthoimages fournies par notre système, la mise en correspondance de ces points, l'élimination des correspondances aberrantes et enfin l'estimation de la transformation entre deux orthoimages successives.

Bien qu'elle n'intervienne que dans un second temps et qu'elle n'a été abordée qu'en fin de travail de thèse, la méthodologie développée autour de la caméra *Fish Eye* a également été présentée dans cette partie, car le développement réalisé est assez proche de celui mené autour des caméras 3D. Les quatre premières étapes mises en place sont effectivement identiques à celles vues auparavant. Une cinquième étape vient compléter celles-ci. Il s'agit d'une étape de calcul de déplacement moyen rendu possible compte tenu des caractéristiques de la caméra utilisée.

Ces deux chaînes de traitement adaptées à deux pôles de capteurs de vision ont pour objectif d'aboutir au calcul de la localisation de notre système en environnement urbain. Le chapitre suivant expose les résultats obtenus à partir du prototype développé.

Expérimentations et résultats

1	Introduction	121
2	Expérimentations	122
	2.1. Données simulées	122
	2.2. Données réelles	125
3	Résultats	130
	3.1. Pôle caméras 3D	130
	3.2. Bilan des résultats obtenus avec le pôle caméras 3D	147
	3.3. Pôle caméra <i>Fish Eye</i>	148
	3.4. Bilan des résultats obtenus avec le pôle caméra <i>Fish Eye</i>	150
	3.5. Deux pôles de vision pour une trajectoire	151
	3.6. Bilan général	152

1 Introduction

Dans le Chapitre 4, nous avons présenté les deux méthodologies développées autour de chacun des pôles de notre système. Chacune de ces chaînes algorithmiques est en mesure de fournir une information de localisation.

Ce cinquième chapitre débute par une présentation des différentes expérimentations réalisées pour évaluer les approches développées. Nous présenterons deux types de données, d'une part des données formées à partir de simulations, et d'autre part des données réelles, obtenues lors de différentes campagnes d'acquisition.

Nous présentons ensuite les résultats fournis par la chaîne algorithmique de localisation, basée sur les orthoimages générées grâce aux caméras 3D. Compte tenu des résultats obtenus à partir de cette approche, un premier bilan sera effectué. Celui-ci confirmant la nécessité d'utiliser un second pôle de vision, les résultats de localisation obtenu avec ce nouveau pôle seront présentés. Nous comparerons ensuite les deux pôles de vision présents au sein de notre système. Finalement, nous réaliserons un bilan global de notre solution par rapport au cahier des charges initialement défini.

2 Expérimentations

Afin d'évaluer les méthodologies développées au cours de la thèse, il a été nécessaire de réaliser deux types d'expérimentations. Les premières expérimentations portent sur des données obtenues à partir de simulations. Elles ont pour avantage de permettre de tester l'approche de localisation avant même de posséder un prototype fonctionnel. Les expérimentations suivantes portent sur des données obtenues à partir d'acquisitions réelles. Ces acquisitions ont été réalisées en intérieur (en condition de laboratoire) et en extérieur (en environnement urbain). Dans cette partie, nous présentons l'ensemble de ces données.

2.1. Données simulées

Les données simulées ne concernent que le pôle caméras 3D. Lors de l'élaboration du concept mis en place dans le cadre de cette thèse et avant même la réalisation mécanique de celui-ci, il est apparu nécessaire de valider l'approche de localisation en cours de développement pour le pôle caméras 3D exclusivement. L'utilisation de simulations nous a aidés dans cette tâche.

Les données simulées ont été obtenues grâce au plugin Blensor [Gschwandtner et al., 2011]. Ce dernier propose à la fois un environnement de simulation et une interface de modélisation 3D. Ainsi, il est possible de simuler différents types de capteurs télémétriques (caméra temps de vol, Kinect/Asus Xtion Pro, Lidar, etc...) ayant des propriétés spécifiques (Figure 5.1) dans des environnements modélisés par l'utilisateur (Figure 5.2). Il nous a alors été possible de réaliser des simulations modélisant un modèle de caméra identique à celui envisagé pour notre prototype.



FIGURE 5.1: Capteurs disponibles dans Blensor et paramètres

Cette approche est rendue possible grâce à l'intégration de l'outil de simulation Blensor à Blender [Blender, 1995], un logiciel de modélisation 3D. Il est possible grâce à ce logiciel de modéliser des environnements ou d'importer des modélisations existantes. Il permet aussi de créer des scènes statiques ou dynamiques en animant les objets de la scène, ou encore le capteur simulé. Afin de faciliter l'analyse qualitative d'algorithmes, il est nécessaire de connaître la position et l'orientation exactes de tous les objets de la scène, et dans notre cas plus précisément la position et l'orientation de la caméra durant la simulation. Un avantage important de Blensor est la facilité avec laquelle la trajectoire vraie du capteur simulé peut être générée. Il est aussi possible, pour chaque instant de la simulation, de connaître les positions et orientations de la caméra.

2.1.1. Premier cas d'étude : scène virtuelle

Cette simulation a pour objectif de valider l'approche proposée pour le pôle caméras 3D dans un contexte favorable. Au cours de cette simulation, une unique caméra 3D est simulée. Le déplacement effectué par la caméra est rectiligne et la scène est favorable dans la mesure où les objets représentés au sein de celle-ci sont de géométrie simple tout en étant réalistes (potelets).

De plus, le nombre d'objets et leur disposition facilitent l'application de nos algorithmes.

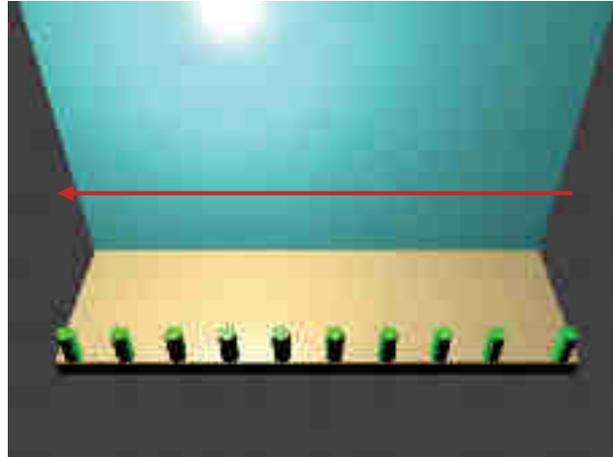


FIGURE 5.2: Premier cas d'étude : déplacement rectiligne (trajectoire représentée en rouge)

Dans cette simulation, la distance parcourue est d'environ 3 mètres. Au cours du déplacement, 50 nuages de points ont été générés et à partir de ceux-ci, 50 orthoimages ont pu être calculées.

2.1.2. Second cas d'étude : St Sulpice

Le second cas d'étude traité est une trajectoire simulée au sein d'un modèle 3D de la place St Sulpice à Paris¹ (Figure 5.3 (a)). Cette simulation nous permet d'évaluer l'approche développée dans un contexte urbain, représentatif de la réalité en termes d'équipements de voirie et de géométrie.

La trajectoire simulée représente un parcours de 120 mètres autour de la place St Sulpice (Figure 5.3 (b)). Celle-ci compte 600 positions et le déplacement entre chacune d'elles est d'environ 20 centimètres. Ainsi à partir de cette simulation, 600 orthoimages ont été générées.

À partir de cette scène, deux jeux de données simulées ont été produits. Le premier a été obtenu à partir de la simulation d'une unique caméra. Le second quant à lui, a été obtenu à partir de la simulation de 3 caméras, positionnées dans une configuration proche de configuration finale.

	Type de trajectoire	Longueur de trajectoire	Nombre d'orthoimages	Remarques
Scène virtuelle	rectiligne	3 mètres	50	scène simple
St Sulpice	rectiligne + curviligne	120 mètres	600	1 caméra + 3 caméras

TABLE 5.1: Tableau récapitulatif des données simulées

1. Données gracieusement mises à disposition par l'équipe de François Goulette (CAOR Mines ParisTech).

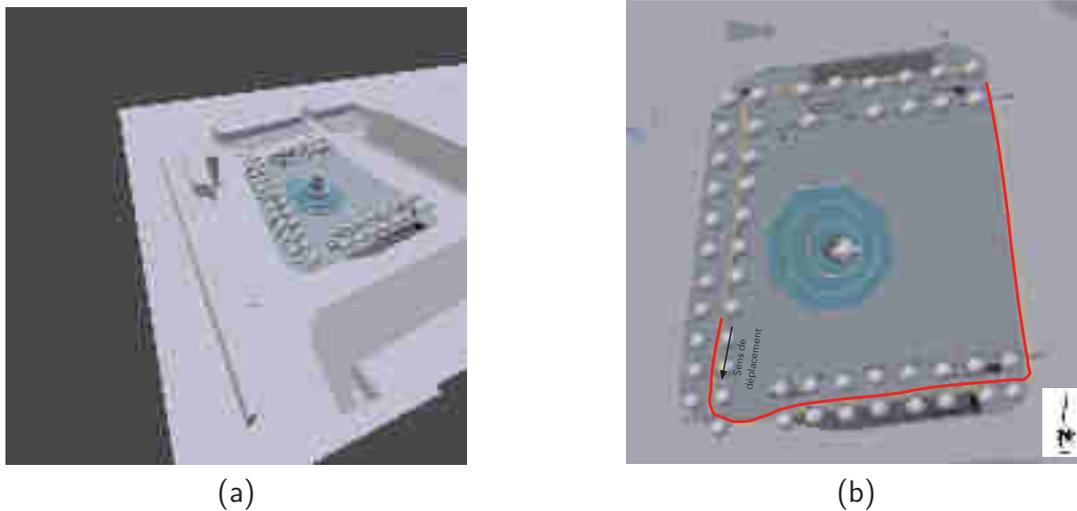


FIGURE 5.3: Second cas d'étude : (a) Modèle 3D de la place St Sulpice - (b) Trajectoire de la caméra (trajectoire représentée en rouge)

2.2. Données réelles

Nous avons réalisé deux campagnes d'acquisitions de données en environnement réel. La première acquisition a été réalisée en intérieur. La seconde expérimentation a été réalisée en extérieur, dans des conditions respectant au mieux le cahier des charges fixé par Terra Mobilita.

2.2.1. Acquisition en intérieur - Conditions de laboratoire

La première acquisition que nous avons effectuée a été réalisée dans les locaux de l'INSA de Strasbourg. Le choix des caméras 3D utilisées pour la réalisation du prototype nous impose une utilisation sans source solaire. Cette acquisition présente donc l'avantage de ne pas être contrainte par l'environnement naturel. Le mobilier urbain a été simulé à l'aide de mobilier de bureaux.

La trajectoire de référence présentée dans la Figure 5.4 a été calculée par topographie conventionnelle à l'aide de deux stations totales (Figure 5.5).

L'utilisation de deux stations totales nous permet d'assurer un suivi continu du mobile durant son déplacement. Il reste toutefois quelques zones qui n'ont pas pu être couvertes par l'une ou l'autre des stations totales, le mobile n'étant visible par aucune d'elles un court instant (Figure 5.6).

Durant cette acquisition, nous avons pu enregistrer respectivement 1560 nuages de point par le biais du pôle caméras 3D, ainsi que 1560 images par l'intermédiaire du pôle caméra *Fish Eye*. Les acquisitions des deux pôles ont été enregistrées de façon synchrone. À partir des nuages de points, 1560 orthoimages ont pu être générées.

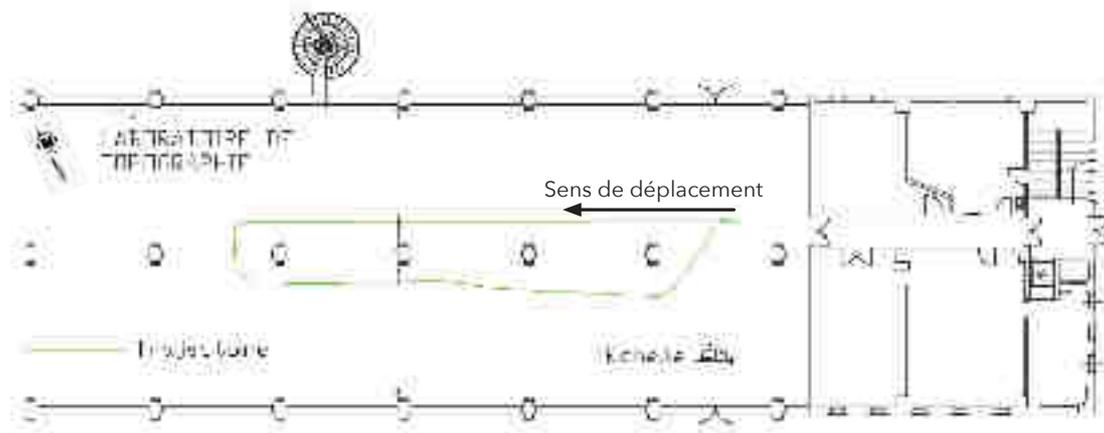


FIGURE 5.4 : Acquisition en intérieur : plan de situation et trajectoire réalisée par le mobile (vert)



(a)



(b)

FIGURE 5.5 : Acquisition en intérieur : environnement de l'acquisition - (a) Station totale 1 - (b) Station totale 2

La distance totale parcourue est de 62 mètres, la distance moyenne entre deux nuages successifs est donc d'environ 4 centimètres. La durée de l'acquisition a été de 8 minutes et 15 s, ce qui nous donne une vitesse moyenne de 0.12 m/s . Cependant, la vitesse de déplacement du mobile étant dépendante de la vitesse de déplacement de l'utilisateur, celle-ci n'est pas constante.

2.2.2. Acquisition en extérieur - Avenue de la Paix à Strasbourg

La seconde acquisition de données réelles a été effectuée en extérieur. Les campagnes d'acquisitions de données réelles ont systématiquement fait intervenir les deux systèmes de vision, à savoir, les caméras 3D, et la caméra de type *Fish Eye*.

L'acquisition réalisée en intérieur avait pour avantage d'offrir des conditions de laboratoire : sol non accidenté, scène instrumentalisée à souhait, faible luminosité ambiante.

En extérieur, il est plus difficile qu'en intérieur de trouver une zone qui soit favorable, aussi bien en termes d'accessibilité, qu'en termes de géométrie. Pour pouvoir s'appuyer sur notre approche de localisation, il faut pouvoir évoluer dans une zone pourvue d'informations, aussi bien pour le

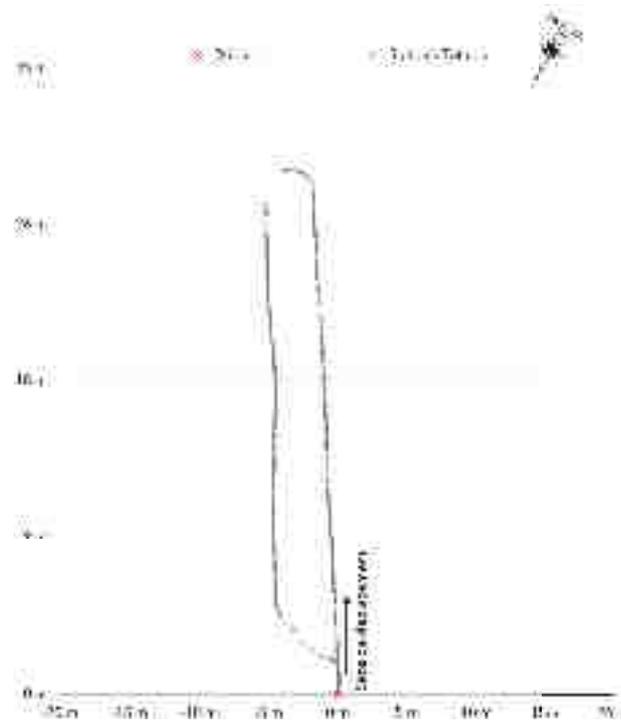


FIGURE 5.6 : Acquisition en intérieur : trajectoire de référence, relevée par les stations totales

pôle caméras 3D que pour le pôle caméra *Fish Eye*. De plus, les conditions idéales d'acquisitions pour chacun des systèmes sont différentes. La grosse difficulté étant la luminosité. Pour la caméra *Fish Eye* il est nécessaire d'avoir une luminosité suffisante, afin de diminuer autant que possible le temps d'exposition. A contrario, pour les caméras 3D que nous utilisons, il est nécessaire de disposer d'une luminosité faible, afin d'assurer un fonctionnement minimal de celles-ci.

Dans ce contexte, l'acquisition a été réalisée au coucher du soleil, ainsi la luminosité était encore suffisante pour la caméra *Fish Eye*, tout en étant suffisamment faible pour que les caméras 3D fonctionnent. Cette acquisition a pu être réalisée le 17 juillet 2014 aux alentours de 21h30, le soleil se couchant à 21h47 ce jour-là.

Cette acquisition a eu lieu Avenue de la Paix à Strasbourg (Figure 5.7). Une trajectoire de référence a été acquise par l'intermédiaire de deux stations totales, mais également un récepteur GNSS (Figure 5.8). L'utilisation de celui-ci avait pour but d'évaluer le fonctionnement d'un système de positionnement par satellite en zone urbaine. Il est à noter que durant l'acquisition, celui-ci a été en mesure de fournir une information de localisation sur environ 70 % de la trajectoire.

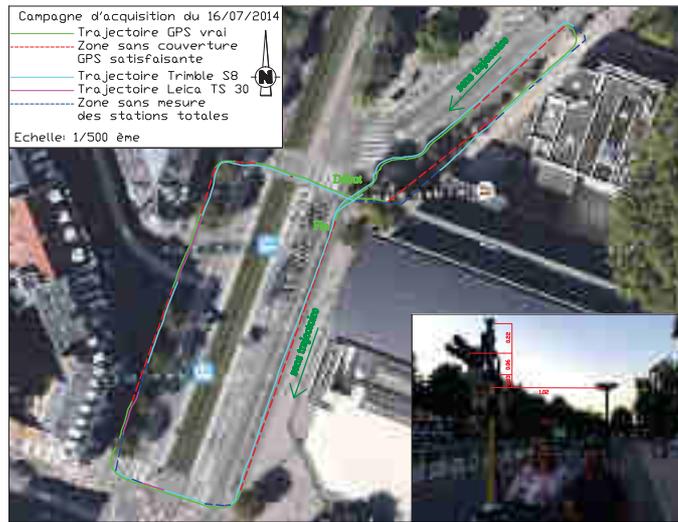
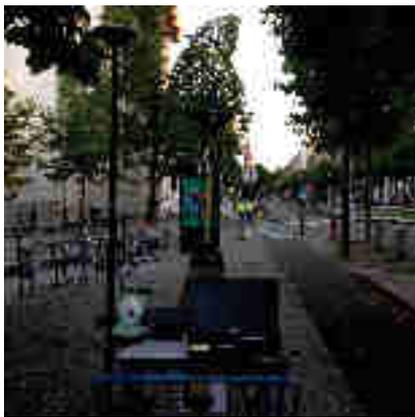


FIGURE 5.7: Acquisition en extérieur : plan de situation et trajectoire réalisée par le mobile



(a)



(b)



(c)

FIGURE 5.8: Acquisition en extérieur : environnement de l'acquisition - (a) et (b) Système d'acquisition - (c) - Afin d'éliminer des vibrations trop importantes, une personne est montée dans le mobile pour maintenir l'ensemble des systèmes imageurs.

La distance parcourue durant l'acquisition est de 291 mètres et 1480 nuages par caméras ont été enregistrés. Nous avons donc pu générer 1480 orthoimages. La distance moyenne entre deux orthoimages successives est donc de 19,5 centimètres. Les différents systèmes de vision sont synchronisés, ainsi 1480 images issues de la caméra *Fish Eye* ont été enregistrées. La distance moyenne entre deux images successives est également de 19,5 centimètres. La durée d'acquisition a été de 12 minutes et 9 secondes, ainsi, la vitesse moyenne de déplacement durant l'acquisition a été de 0.40 m/s .

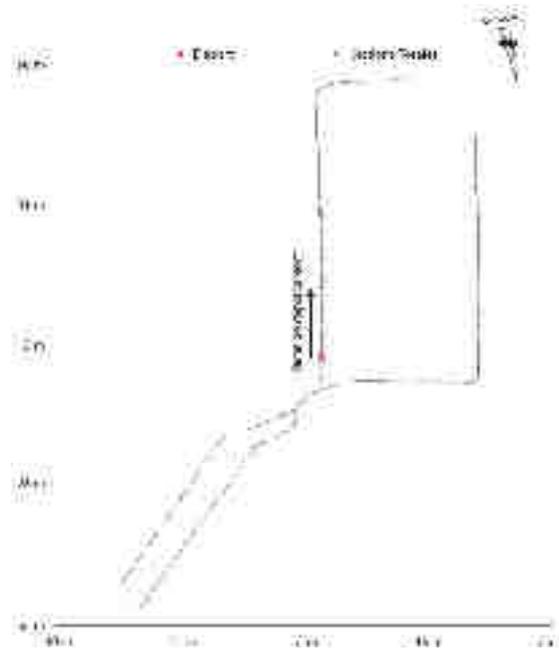


FIGURE 5.9 : Acquisition en extérieur : trajectoire de référence, relevée par les stations totales

L'utilisation de deux stations totales nous a permis de couvrir en grande partie la trajectoire suivie par le mobile. En effet, la configuration de l'environnement a engendré à certains moments des discontinuités dans les mesures de chacune des stations totales (Figure 5.9).

	Type de trajectoire	Longueur de trajectoire	Nombre d'orthoimages et de photos	Durée d'acquisition	Remarques
Intérieur	rectiligne + curviligne	62 mètres	1560	8 min 15 s	caméras 3D + caméra <i>Fish Eye</i>
Extérieur	rectiligne + curviligne	291 mètres	1480	12 min 09 s	caméras 3D + caméra <i>Fish Eye</i>

TABLE 5.2 : Tableau récapitulatif des données réelles, acquises par caméras 3D et caméra *Fish Eye*

3 Résultats

À partir des expérimentations réalisées, nous avons pu tester les différentes méthodologies de localisation présentées dans le Chapitre 4. Les résultats obtenus pour chacun des pôles ont été évalués de façon indépendante. De façon similaire au chapitre précédent, nous présentons dans un premier temps les résultats obtenus pour le pôle caméras 3D, et nous aborderons ensuite les résultats obtenus pour le pôle caméra *Fish Eye*.

3.1. Pôle caméras 3D

Dans cette sous-partie nous présentons les résultats obtenus en terme de capacité de localisation par le pôle caméras 3D, à partir des données simulées et des données réelles présentées précédemment.

3.1.1. Résultats obtenus à partir des données simulées

À partir des simulations présentées au paragraphe 2.1., nous sommes en mesure de calculer plusieurs trajectoires et de présenter de premières conclusions.

a. Premier cas d'étude : scène virtuelle

L'estimation de la transformation entre deux orthoimages repose sur le calcul des paramètres d'une transformation de Helmert 2D. Ce calcul permet également de disposer de plusieurs indicateurs permettant d'évaluer la qualité des transformations successives calculées.

Les simulations nous permettent de connaître parfaitement la position de la caméra pour chacune des acquisitions simulées. Ainsi, nous sommes en mesure de comparer les positions calculées et les positions de référence de la caméra.

Sur la Figure 5.10, nous avons tracé en gris la trajectoire de référence. La trajectoire calculée, en appliquant notre approche, est représentée en rouge.

Afin d'étudier l'erreur en position, nous calculons la distance euclidienne entre les positions de référence et les positions estimées. La distance moyenne obtenue est de 0.02 mètres, soit 2 cm. À la fin du déplacement, l'erreur maximale en position est de 0.04 m, soit 4 cm après 3 m de déplacement.

Nous pouvons observer une répétabilité dans la courbe représentant l'erreur moyenne quadratique de l'unité de poids (Figure 5.11). Ceci s'explique par la répétabilité de la scène, en effet, les orthoimages traitées présentent une redondance. Les zones 1,2 et 3 de la Figure 5.10 correspondent aux orthoimages présentées Figure 5.12.

À ce stade, nous pouvons déjà tirer une première conclusion : notre approche est répétable. Ainsi, lorsque nous fournissons des données identiques en entrée de l'algorithme, la transformation calculée est identique (3 zones identiques, 3 portions de trajectoires identiques). De plus, l'évolution de l'erreur sur la position estimée par rapport à la position vraie confirme que l'approche choisie permet d'évaluer le déplacement d'une caméra 3D dans un environnement pourvu de mobilier.



FIGURE 5.10 : Premier cas d'étude : trajectoire calculée (rouge) et trajectoire vraie (grise)

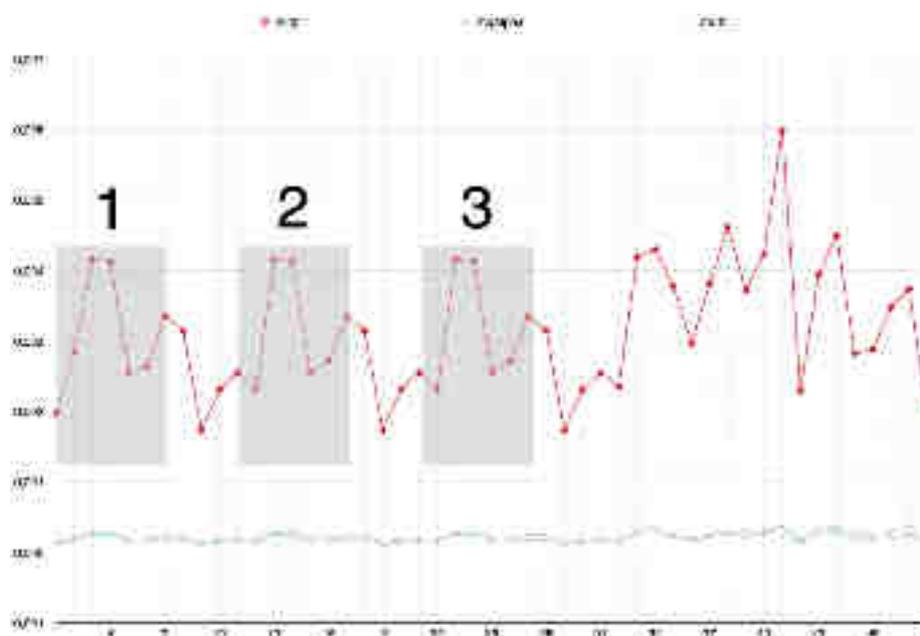


FIGURE 5.11 : Premier cas d'étude : erreurs moyennes quadratique de l'unité de poids pour chaque transformation calculée

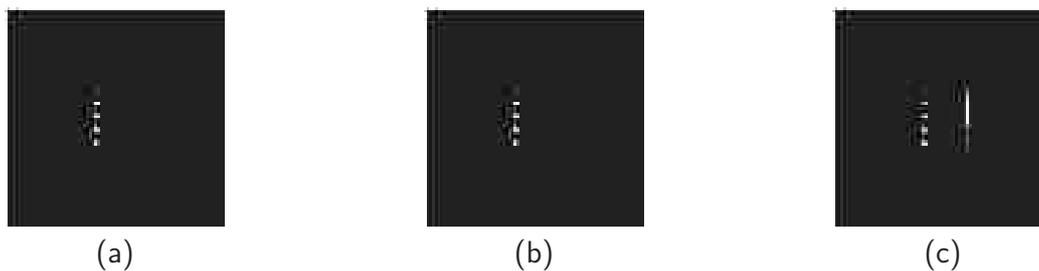


FIGURE 5.12 : Orthoimages correspondant aux différentes zones : (a) Zone 1 - (b) Zone 2 - (c) Zone 3

Cependant, il convient de noter une dérive dans la position calculée. Effectivement, l'approche développée étant une approche incrémentale, nous pouvons observer un phénomène d'accumulation des erreurs tout au long de la trajectoire. La scène modélisée ici est toutefois favorable par sa répétabilité et sa configuration. Il est alors apparu nécessaire d'évaluer notre approche sur une scène plus réaliste, qui ne se limite pas à un parcours rectiligne.

b. Second cas d'étude : St Sulpice

La seconde simulation réalisée lors de notre étude concerne une scène urbaine existante et reconstruite en 3D de la place St Sulpice à Paris. Comme nous l'avons présenté auparavant (paragraphe 2.1.2.), deux simulations ont été réalisées. Une première avec une unique caméra, et une seconde où trois caméras ont pu être simulées.

Une caméra

Dans un premier temps, nous avons simulé les acquisitions réalisées par une unique caméra se déplaçant dans la scène. Sur la Figure 5.13, nous avons reporté en gris la trajectoire de référence, et en rouge la trajectoire estimée. Enfin, en vert, nous avons représenté les positions pour lesquelles il n'a pas été possible d'effectuer l'estimation de position, faute de points détectés dans l'orthoimage courante. Dans ce cas, il a été décidé de considérer la moyenne des deux dernières transformations calculées.

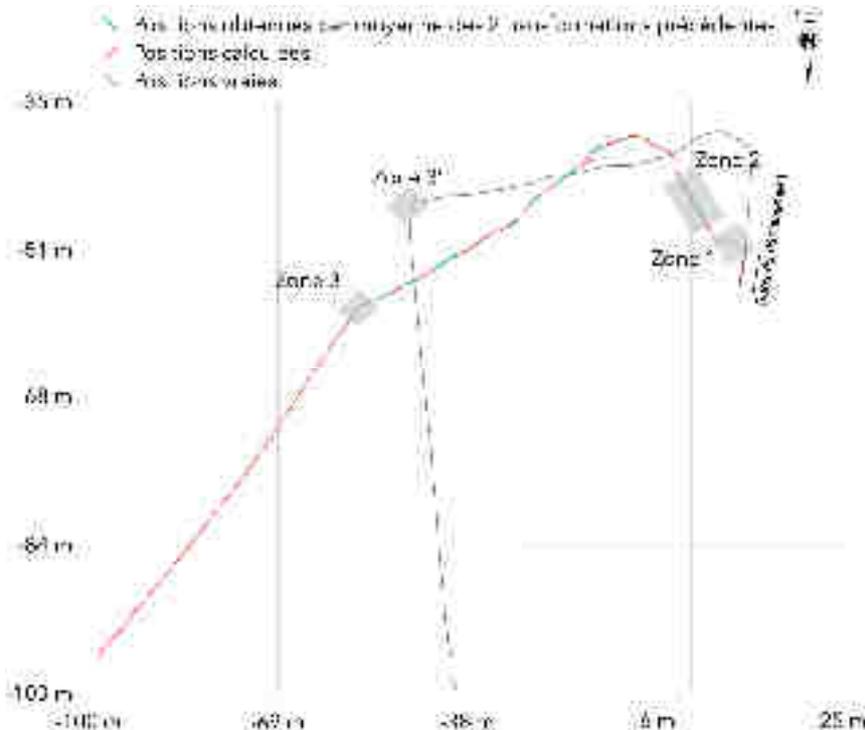


FIGURE 5.13 : Second cas d'étude : trajectoire estimée (rouge) - trajectoire vraie (grise)

Un champ de vision limité

Cette première simulation nous a permis de mettre en évidence l'impossibilité de n'utiliser qu'une unique caméra. En effet, le champ de vue du modèle de caméras 3D, correspondant à notre système et simulé pour cette première étude, est trop réduit. Sur la Figure 5.13, nous pouvons observer en vert les positions obtenues par calcul de la moyenne des deux dernières transformations calculées. Comme nous l'avons mentionné dans le Chapitre 4, cette solution consiste à exploiter les dernières transformations, lorsqu'il est impossible de calculer la transformation en cours (exemple : les zones 1,2 et 3 sur la Figure 5.13). Nous pouvons observer que cette situation est fréquente le long de la trajectoire estimée. En effet, sur la totalité du parcours effectué, plus de 20 % des positions sont obtenues par ce procédé. L'utilisation de cette solution présente un avantage certain, car, elle permet de conserver une continuité dans la trajectoire estimée. Cependant, elle présente également un inconvénient majeur, qui est que tout changement de direction ou d'allure ne pourra être évalué si ce changement s'effectue à ce moment là.

Trois caméras

Afin d'évaluer l'apport de plusieurs caméras sur la qualité de la trajectoire, nous avons à nouveau réalisé une simulation dans la scène de St Sulpice. Cette fois-ci, trois caméras ont été simulées et positionnées l'une par rapport à l'autre, de manière à augmenter la surface couverte au sol (Figure 5.14).

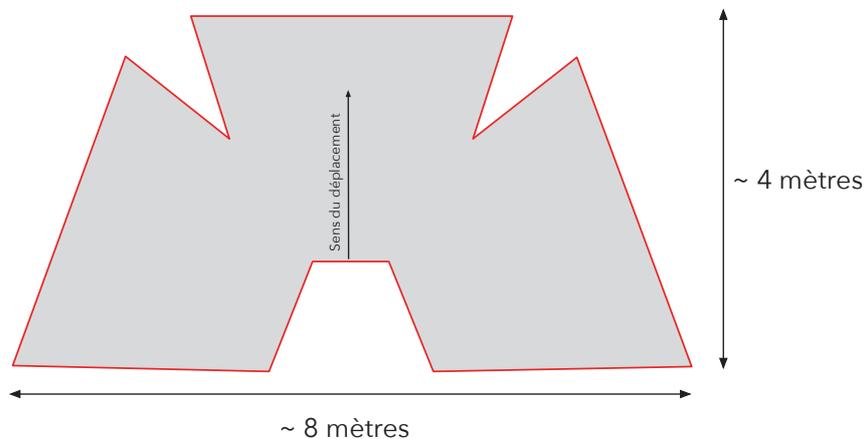


FIGURE 5.14 : Empreinte au sol obtenue grâce aux 3 caméras simulées

Sur la Figure 5.15, nous avons à nouveau reporté en gris la trajectoire de référence, en rouge la trajectoire estimée, et en vert les transformations estimées à partir des transformations précédentes.

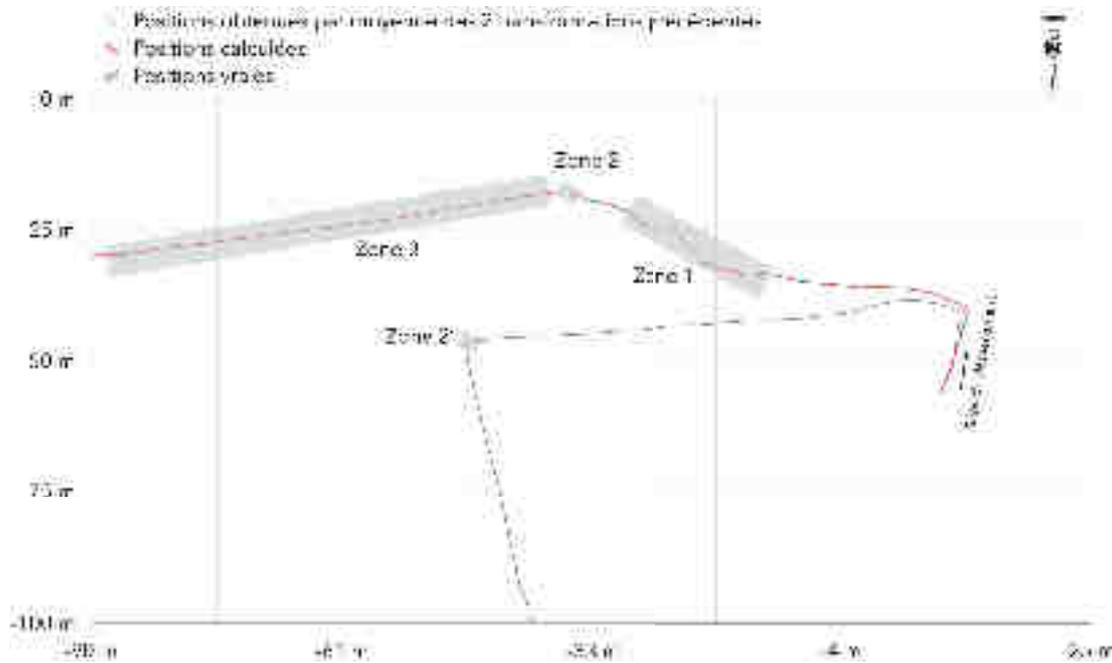


FIGURE 5.15 : Second cas d'étude avec 3 caméras : trajectoire estimée (rouge) - trajectoire vraie (grise)

Une visibilité améliorée

L'utilisation de trois caméras permet d'améliorer considérablement le champ de vision du système. Sur l'ensemble de la trajectoire, le nombre de positions obtenues par le calcul de la moyenne des deux transformations précédentes n'est plus que de 10 %. Cependant, nous sommes tout de même confrontés à ce problème de visibilité dans certaines zones, comme dans la zone numérotée 2 sur la Figure 5.15. Effectivement, nous observons plusieurs transformations obtenues par calcul de la moyenne des 2 dernières transformations, faute d'observations suffisantes. Ainsi, la rotation que nous devrions calculer n'est pas observée comme elle le devrait. Dans ce contexte, il nous semble opportun d'exploiter d'autres informations fournies par les caméras 3D. Comme nous le verrons dans la partie concernant les résultats obtenus à partir de données réelles, nous avons choisi d'exploiter l'information infrarouge fournie par les caméras 3D. Cette information sera exploitée afin de coloriser les orthoimages, ainsi lorsque la scène ne comportera pas d'objet, il sera possible d'exploiter certaines informations colorimétriques.

Malgré le problème constaté au niveau de la zone 2, nous pouvons observer que la trajectoire estimée dans la zone 3 reste conforme à la trajectoire réelle. Sur la Figure 5.16, nous représentons la trajectoire de référence en gris et la trajectoire estimée en rouge. Les positions obtenues par le calcul de la moyenne des 2 transformations précédentes sont à nouveau représentées en vert.

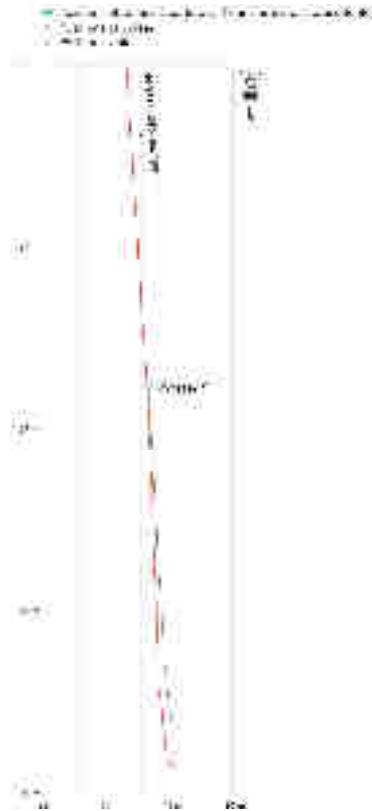


FIGURE 5.16 : Second cas d'étude avec zoom sur la zone 3 : trajectoire estimée (rouge) - trajectoire vraie (grise)

L'erreur de position a été calculée sur la base de la distance euclidienne entre les positions de référence et les positions estimées par notre approche. La distance moyenne obtenue est de l'ordre de 20 cm . Nous avons représenté sur la Figure 5.17 l'évolution de cette erreur en fonction de la position courante.

Nous pouvons observer sur cette courbe que l'erreur de positionnement augmente à partir de la position 110. Celle-ci coïncide avec la zone 4, mise en avant sur la Figure 5.16. Nous pouvons observer pour cette zone que la première position est calculée par l'intermédiaire de la moyenne des transformations précédentes. À partir de cette position, l'erreur augmente de façon constante jusqu'à la fin de la trajectoire. Ceci est cohérent vue l'approche incrémentale mise en place. L'erreur maximale atteinte à la fin du trajet est de 1.08 m , ce qui représente environ 2% de la distance totale parcourue sur cette section (soit 53.2 m).

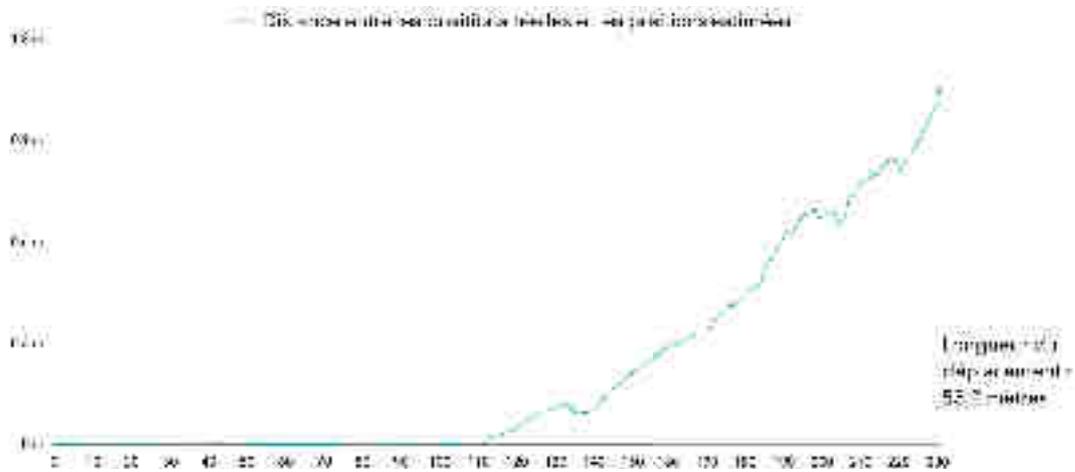


FIGURE 5.17: Second cas d'étude : évolution de l'écart entre les positions réelles et les positions estimées

Influence de la géométrie et des éléments de la scène

Malgré l'utilisation de trois caméras censées augmenter le champ de vision de notre système, nous sommes confrontés à un problème dû à la géométrie de la scène. En effet, nous pouvons observer au niveau de la zone 1 sur la Figure 5.15 une configuration de scène problématique pour notre approche. Cette situation correspond à la zone encadrée en rouge dans la Figure 5.18(a).

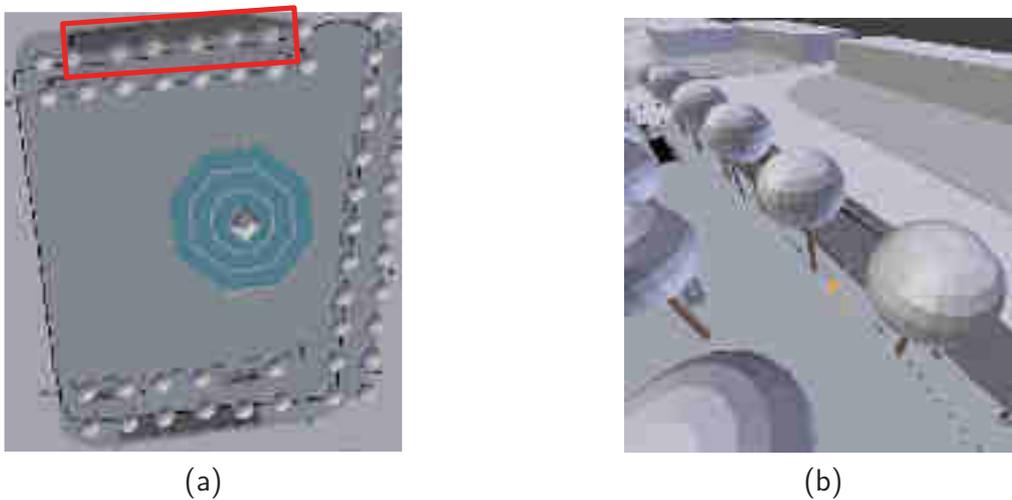


FIGURE 5.18: Scène correspondant à la zone 1 dans la trajectoire estimée : (a) Vue de dessus de la scène - (b) Zoom sur la zone 1 (en rouge dans (a))

La Figure 5.19 présente deux acquisitions successives réalisées dans cette zone 1, où les positions des caméras simulées sont représentées en rouge, vert et gris.

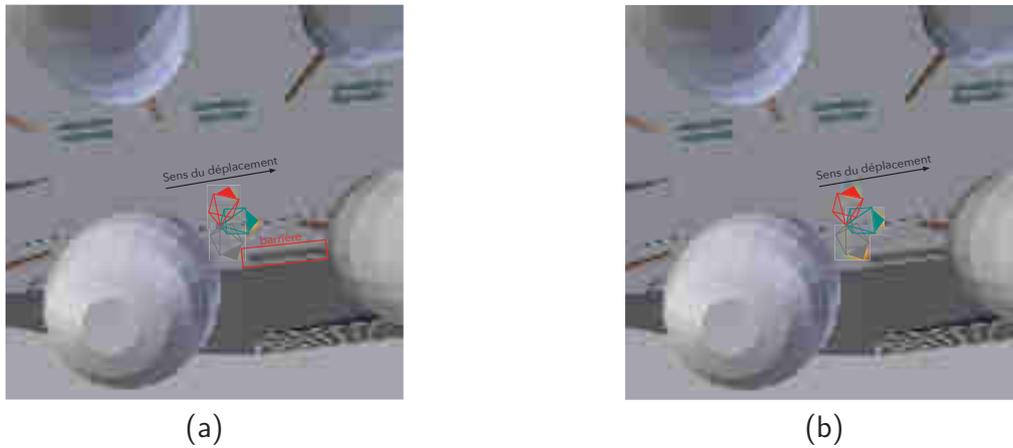


FIGURE 5.19 : Deux acquisitions successives réalisées dans la zone 1 : (a) Position 1 - (b) Position 2

À partir de ces deux acquisitions, nous sommes en mesure de générer les deux orthoimages présentées Figure 5.20.



FIGURE 5.20 : Orthoimages générées à partir des deux acquisitions successives présentées dans la Figure 5.19

La Figure 5.20 montre que les orthoimages sont très semblables. Compte tenu de l'approche que nous avons développée, la forme des orthoimages générées dans cette zone ne nous permet pas d'évaluer les déplacements du mobile à partir de celles-ci. En effet, lors de l'étape de mise en correspondance, les vecteurs de déplacement obtenus représentent l'espace entre les barreaux de la barrière, mais pas le déplacement du mobile. De plus, compte tenu de l'aspect incrémental de notre approche, la qualité de l'ensemble de la trajectoire estimée va s'en ressentir.

3.1.2. Résultats obtenus à partir des données réelles

Après avoir réalisé une première évaluation de l'approche développée pour le pôle caméras 3D grâce à des données simulées, nous présentons à présent l'ensemble des résultats de trajectographie obtenus à partir des données acquises via notre prototype.

Comme nous l'avons introduit au paragraphe précédent (paragraphe b.), l'information infrarouge fournie par les caméras 3D est également exploitée dans le cas du traitement à partir de données réelles.



FIGURE 5.21 : Exemple d'orthoimage colorisée avec l'information infrarouge fournie par les caméras (Passage piéton)

Sur la Figure 5.21, nous présentons une orthoimage colorisée à partir de l'information infrarouge fournie par les caméras 3D. Cette orthoimage est issue d'un nuage de points acquis au niveau d'un passage piéton. Nous pouvons observer l'absence de mobilier urbain dans la scène, cependant les peintures au sol sont porteuses d'informations. L'utilisation de la colorisation par l'infrarouge nous permet d'exploiter cette information. En complétant le jeu de données avec ces orthoimages colorisées avec l'infrarouge, nous minimisons le nombre de situations où les orthoimages ne permettent pas d'estimer le déplacement de la plateforme mobile.

a. Acquisition en intérieur - Conditions de laboratoire

Nous rappelons que ce premier jeu de données a été obtenu lors d'une acquisition en intérieur, réalisée au sein de l'INSA de Strasbourg (paragraphe 2.2.1. et Tableau 5.2).

Sur la Figure 5.22, nous avons représenté en rouge la trajectoire estimée à partir de notre approche. Les transformations calculées à partir des transformations précédentes sont représentées

en vert. Enfin, la trajectoire obtenue à partir des données des stations totales est représentée en gris. Les zones 1 et 2 seront analysées plus en détail, car sur ces positions, des phénomènes qui méritent d'être soulignés, seront explicités.

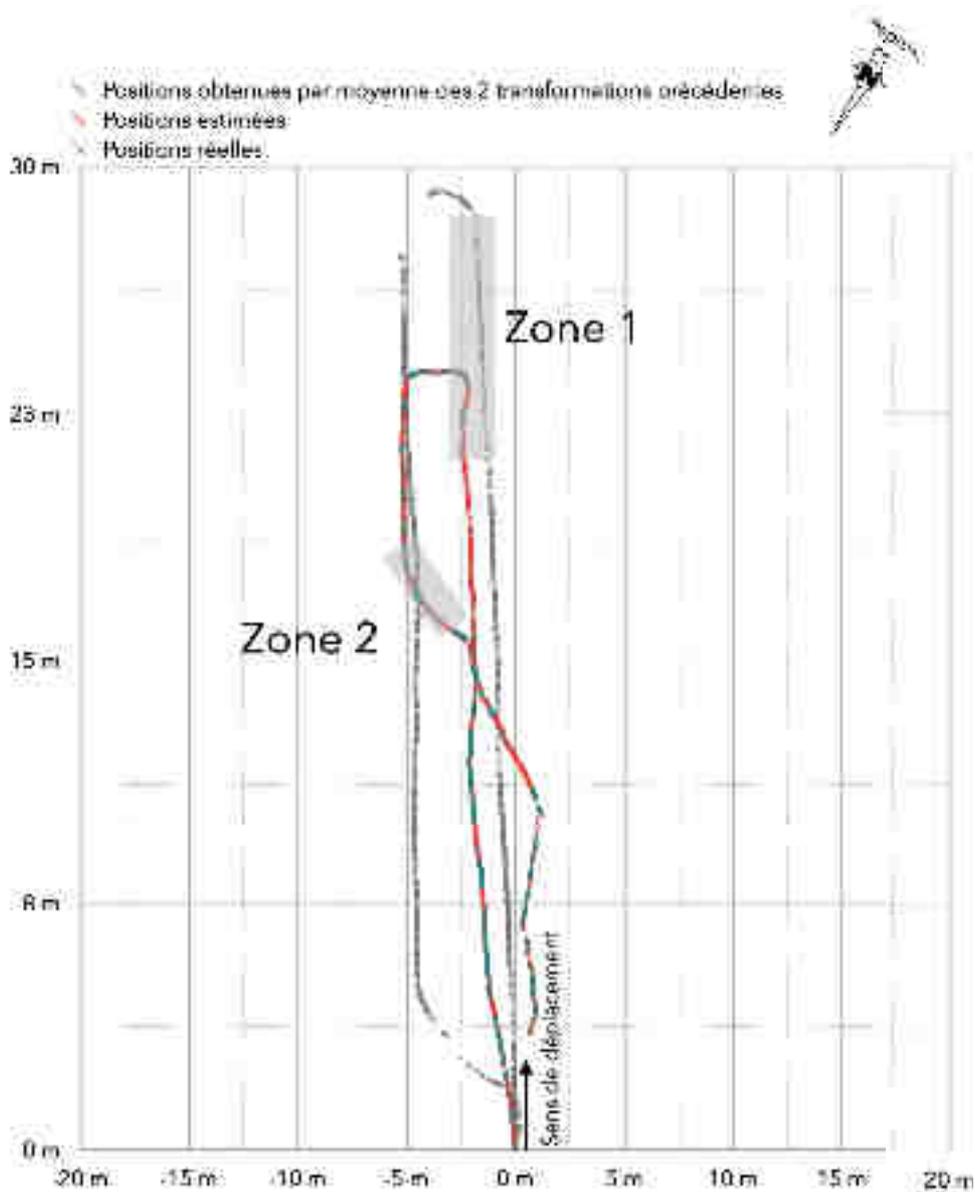


FIGURE 5.22 : Acquisition en intérieur : trajectoire estimée (rouge) - trajectoire vraie (grise)

Des données bruitées

Sur l'ensemble des positions calculées, 25 % sont obtenues à partir de transformations estimées par calcul de la moyenne des deux transformations précédentes. Cette valeur est bien plus

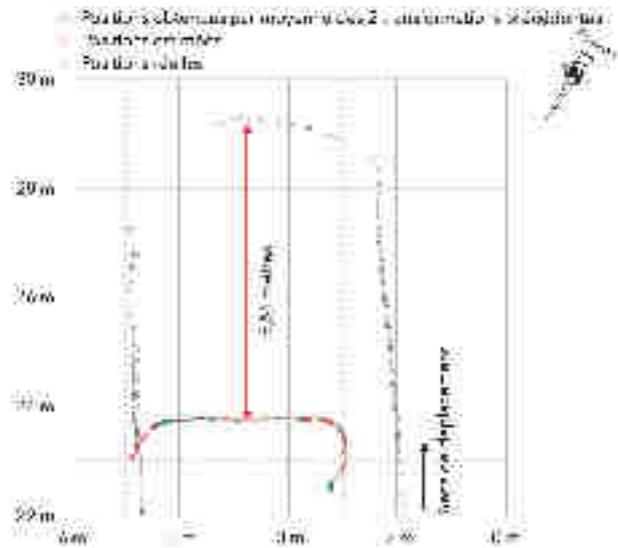


FIGURE 5.23 : Acquisition en intérieur : zoom sur la zone 1

importante que dans le cas des données simulées, où seuls 10 % des positions avaient été obtenues par calcul de moyenne (cas de St Sulpice avec 3 caméras). Plusieurs raisons peuvent expliquer ce phénomène. Premièrement, la scène dans laquelle nous avons évolué a été instrumentalisée par nos soins. Nous avons utilisé du mobilier de bureau, cependant il s'est avéré, après acquisition, que le placement de celui-ci n'a pas toujours été réalisé de façon pertinente (zone "vide" dans la scène). Deuxièmement, et c'est la raison principale, les données fournies par les caméras sont bien plus bruitées que les données obtenues par simulation. Ce bruit de mesure a plusieurs impacts sur notre approche, aussi bien au niveau de la détection de points d'intérêts qu'au niveau de la mise en correspondance de ceux-ci.

Un filtrage trop strict

Nous avons mis en évidence, sur la Figure 5.23, la zone 1, car elle fait l'objet d'une importante erreur en termes de distance parcourue. En effet, les transformations estimées par notre approche ne permettent pas d'évaluer le déplacement exact réalisé par la mobile : nous pouvons observer une différence de plus de 5 mètres entre le déplacement réel du mobile, et le déplacement estimé par notre approche.

Ce phénomène trouve son explication lors de l'étape d'élimination des mises en correspondances aberrantes. Comme nous l'avons vu au cours du chapitre 4, à la fin de l'étape de mise en correspondance est appliqué un filtre consistant à éliminer les mises en correspondance ne reflétant pas un réel déplacement de la plateforme mobile. Nous rappelons ici que le bruit de mesure consécutif à la technologie des caméras utilisées peut entraîner un "faux déplacement" entre deux acquisitions successives.

Ainsi, lorsque le filtre mis en place détecte une distance moyenne trop faible (inférieure à 2 *cm*), nous ne considérons pas de déplacement. Il arrive que nous éliminions des déplacements réels,

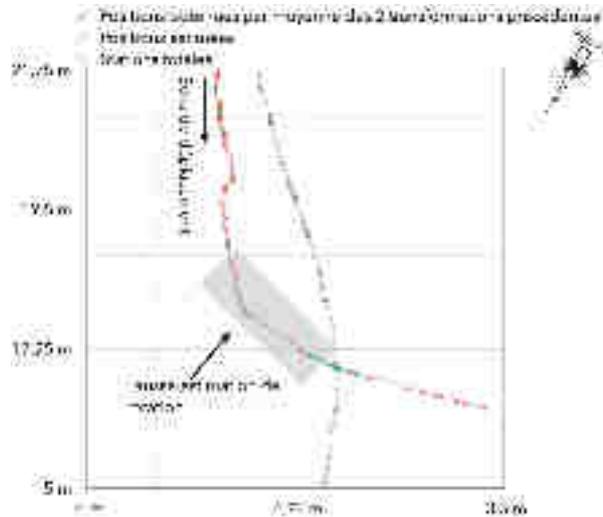


FIGURE 5.24 : Acquisition en intérieur : zoom sur la zone 2

par exemple lorsque la vitesse de déplacement du mobile est faible. Cela nous amène à penser que nous ne sommes pas en mesure de distinguer le bruit de mesure et les faibles déplacements.

Une extrapolation trompeuse

La Figure 5.24 présente un zoom sur la zone 2. Dans cette zone, nous pouvons remarquer que notre approche estime une rotation qui n'a pas eu lieu.

Ce phénomène est dû à de trop nombreuses positions obtenues par extrapolation (représentées en vert). En effet, nous pouvons observer à cet endroit l'utilisation de plus de 20 transformations qui sont obtenues par le calcul de la moyenne des deux transformations précédentes. L'application successive de celles-ci entraîne donc une dérive à rapporter à l'aspect incrémental de notre approche.

b. Acquisition en extérieur - Avenue de la Paix à Strasbourg

À nouveau, nous avons représenté en rouge la trajectoire estimée à partir de notre approche (Figure 5.25), et en vert les transformations calculées à partir des transformations précédentes. La trajectoire obtenue à partir des données des stations totales est représentée en gris. Ici, trois zones représentant des portions de trajectoires méritent d'être d'analysées plus en détail.

Des conditions d'acquisition non optimales

Les contraintes concernant l'acquisition des données étaient fortes (luminosité, accessibilité, équipement de la scène,...). Bien que nous ayons pris toutes les précautions nécessaires pour réaliser cette campagne, les données ont été impactées par les conditions d'acquisition, et plus particulièrement par la luminosité ambiante.

Sur l'ensemble de la trajectoire, 40 % des positions calculées sont obtenues à partir de transformations estimées par extrapolation. Nous pouvons noter que cette valeur est plus élevée que



FIGURE 5.25 : Acquisition en extérieur : trajectoire estimée (rouge) - trajectoire vraie (grise)

pour l'acquisition en intérieur. Comme déjà mentionné, l'extérieur n'est pas un environnement favorable pour les caméras 3D utilisées sur notre prototype. Ainsi les nuages de points fournis par les caméras sont de qualité moindre, ce qui engendre des difficultés lors de l'étape de détection des points d'intérêts, ainsi que lors de l'étape de mise en correspondance.

Ce grand nombre de transformations extrapolées s'explique également par l'agencement du mobilier urbain dans la scène et par la vitesse de déplacement du mobile. En effet, en comparaison avec l'acquisition réalisée en intérieur, notre vitesse de déplacement a été beaucoup plus rapide (0.12 m/s en intérieur contre 0.40 m/s en extérieur). Ainsi, le recouvrement entre deux orthoimages successives est moindre.

Si nous nous concentrons sur la zone 1, nous pouvons observer (Figure 5.26) que l'algorithme développé est mis en difficulté.

D'une part, nous ne sommes pas en mesure de trouver la valeur de distance effectivement parcourue par la plateforme mobile (20.44 m au lieu de 54.12 m). Comme nous l'avons précisé, la vitesse de déplacement a eu un impact négatif sur les résultats obtenus. Ainsi, entre certaines orthoimages, le déplacement réalisé est de plus de 10 centimètres, tandis que pour d'autres, le déplacement n'est que de quelques centimètres. Dans le cas où le déplacement est trop rapide, le recouvrement entre orthoimages est insuffisant pour appliquer notre méthodologie.

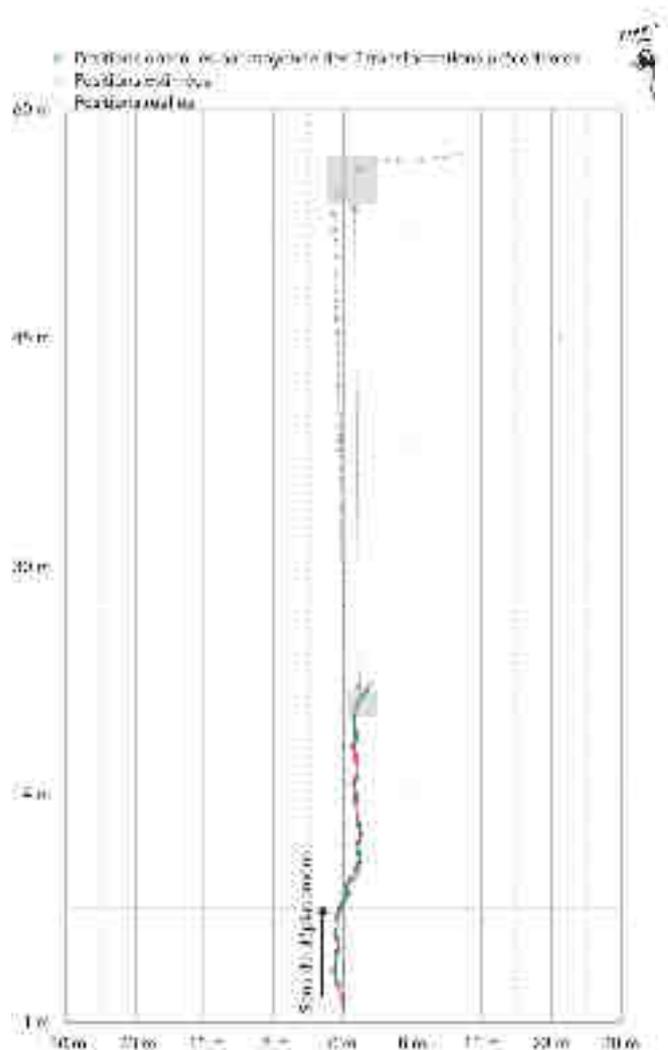


FIGURE 5.27 : Acquisition en extérieur zoom sur la zone 2

D'autre part, le fait que le nombre de transformations extrapolées soit trop important (plus d'un tiers des positions lors de la rotation) empêche l'estimation de la rotation effectuée par la plateforme mobile. Ces extrapolations sont elles mêmes dues à un manque d'information au sein des orthoimages utilisées pour évaluer le déplacement.

Nous pouvons constater à nouveau l'influence de l'extrapolation des transformations sur la qualité de la trajectoire estimée sur la Figure 5.27. L'étude des images infrarouges fournies par les caméras 3D nous permet de re-situer notre acquisition dans la scène. À ce titre, nous tenons d'ailleurs à faire remarquer une difficulté relative à l'évaluation de notre approche : les stations totales et notre solution de capteurs ne sont pas synchronisés. Cela complique le rattachement des mesures des moyens de contrôle aux positions estimées par notre approche.

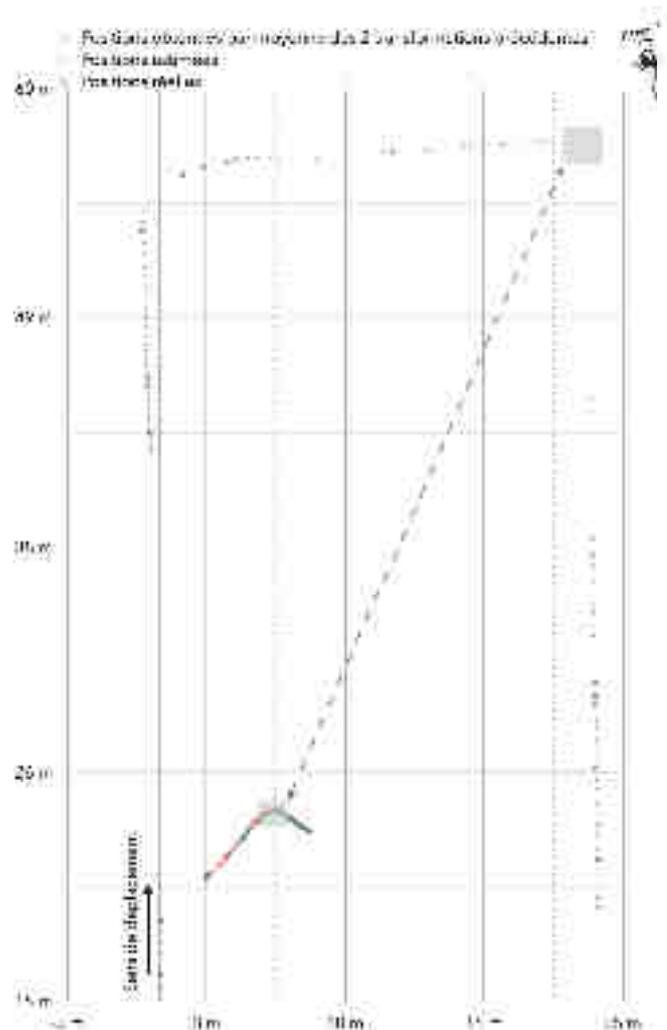


FIGURE 5.28 : Acquisition en extérieur : zoom sur la zone 3

Une approche malgré tout fonctionnelle

Bien que notre approche présente des limitations, il est intéressant de noter que lorsque le nombre de points d'intérêt est suffisant, elle est malgré tout fonctionnelle. C'est pourquoi nous avons souhaité mettre en avant une de ces zones, à l'exemple de la zone 3 (Figure 5.28).

Nous pouvons observer dans cette zone une rotation. L'étude des images infrarouges nous permet de situer cette section de la trajectoire estimée dans la scène et de la rattacher à la trajectoire fournie par les stations totales. Bien que "mal située" par rapport à la trajectoire réelle, du fait des erreurs d'estimation précédentes, nous pouvons observer que la rotation effectuée dans cette section est conforme à celle réalisée par le mobile.

3.2. Bilan des résultats obtenus avec le pôle caméras 3D

Les premiers résultats obtenus avec le pôle caméras 3D nous permettent de tirer un premier bilan. Les résultats obtenus grâce aux données simulées nous ont permis de tirer de premières conclusions sur l'approche mise en place pour le pôle caméras 3D. En effet, la première simulation réalisée nous a permis de valider l'approche choisie. Nous avons pu démontrer qu'il est possible d'évaluer le déplacement d'une caméra 3D par l'exploitation des orthoimages.

La seconde simulation, réalisée au sein du modèle 3D de la place St Sulpice a mis en évidence la nécessité d'utiliser trois caméras 3D afin de couvrir un champ de vision plus important. Cependant, malgré l'utilisation de ces trois caméras, il est apparu nécessaire d'utiliser plus d'informations que celles contenues dans les orthoimages. En effet, nous avons pu observer que dans certaines zones, il est indispensable de calculer la moyenne des transformations précédentes pour obtenir une estimation de la trajectoire. Ceci est dû à l'absence d'objets dans la scène. Dans ce cas de figure, les orthoimages, calculées sur la base de l'information d'altitude, ne contiennent donc pas d'information. Les caméras 3D fournissant également une information infrarouge, il nous a semblé opportun de générer des orthoimages à partir de cette information. Ainsi, notre approche exploitera deux jeux d'orthoimages : celles colorisées avec l'information infrarouge, et celles colorisées avec l'information d'altitude.

Les expérimentations en conditions réelles et le traitement des données obtenues nous ont également permis de tirer un certain nombre de conclusions. Tout d'abord, nous avons observé qu'il est difficile de différencier les faibles déplacements du mobile et le bruit de mesure des caméras. Malgré des opérations de filtrage, notre approche ne permet pas, dans un cas particulier d'immobilité, d'estimer la distance totale réellement parcourue par le mobile. Deuxièmement, malgré l'utilisation de l'information infrarouge, il apparaît que dans un certain nombre de situations, il n'est pas possible d'estimer une transformation entre deux orthoimages, soit parce qu'il n'y a pas suffisamment d'information dans la scène, soit parce que le recouvrement entre deux orthoimages est trop faible (vitesse de déplacement trop grande). Malgré tout, nous avons pu observer que lorsque les données exploitées sont suffisamment porteuses d'informations, et que le recouvrement entre orthoimages est suffisant, l'algorithme développé est fonctionnel.

Dans le cas des données simulées, comme dans le cas des données réelles, l'approche que nous avons choisie de développer est une approche incrémentale, la position courante de notre mobile étant obtenue à partir des positions et transformations précédentes. Ainsi, la moindre erreur dans l'évaluation d'une transformation va engendrer une erreur lors du calcul de la position en cours. Cette erreur va alors se propager sur toutes les positions suivantes. Il convient ainsi de prévoir une solution permettant de contrôler la transformation calculée.

Malgré l'utilisation des plusieurs caméras, notre système a une vision très locale de la scène. Non seulement il est complexe de travailler avec plus de deux orthoimages successives, mais en plus, la mise en correspondance ne peut se faire que "orthoimage-à-orthoimage". Il convient d'envisager une approche plus globale. Dans ce contexte, et comme nous l'avions imaginé, il est

apparu nécessaire d'ajouter un second pôle de vision à notre système. Le paragraphe suivant présente ce second pôle qui est celui de la caméra *Fish Eye*.

3.3. Pôle caméra *Fish Eye*

L'objectif de cette sous partie est de présenter les résultats obtenus à partir de l'approche développée pour le pôle caméra *Fish Eye*. Ces résultats ont été obtenus à partir des jeux de données réelles présentés dans les paragraphes 2.2.1. et 2.2.2..

Comme présenté dans le Chapitre 4, lors du calcul de la position du mobile à partir du pôle caméra *Fish Eye* exclusivement, une étape de calcul de position moyenne est réalisée (paragraphe 4.5.). Deux paramètres interviennent dans ce calcul : un premier paramètre permet de déterminer l'intervalle entre les images à utiliser et un second paramètre permet de déterminer le nombre de transformations antérieures à considérer pour le calcul de la position de la plateforme mobile. Afin d'évaluer l'influence de ces différents paramètres, nous avons réalisé plusieurs tests sur chacune des acquisitions réalisées.

3.3.1. Acquisition en intérieur - Conditions de laboratoire

Sur la Figure 5.29, nous pouvons observer les différents résultats de trajectographie obtenus, à partir du pôle caméra *Fish Eye*, grâce aux données enregistrées lors de l'acquisition en intérieur. Nous avons souhaité évaluer l'incidence de l'intervalle entre les images à utiliser. Afin de représenter l'ensemble de nos résultats, nous avons utilisé une notation reprenant chacun des paramètres pour l'ensemble des trajectoires représentées. Ainsi, la courbe nommée " 1_7" représente une trajectoire où l'intervalle entre les images est de 1, et le nombre de transformations utilisées pour le calcul de la position est 7. De façon similaire, pour la courbe notée " 2_7", nous n'avons considéré qu'une image sur trois, et 7 transformations.

Nous pouvons observer que l'intervalle choisi entre les images utilisées a un impact sur la trajectoire estimée. Comme nous l'avons précisé auparavant, le facteur d'échelle nécessaire au calcul de ces trajectoires est fixé arbitrairement. Cependant, d'un point de vue qualitatif, nous pouvons observer que les trajectoires obtenues sont de bonne qualité car relativement fidèles à la forme de la trajectoire observée par tachéométrie (au facteur d'échelle près). À nouveau, il est difficile d'un point de vue quantitatif d'en évaluer la précision, compte tenu de la non synchronisation des moyens de contrôles et des moyens d'acquisition.

3.3.2. Acquisition en extérieur - Avenue de la Paix à Strasbourg

La Figure 5.30 présente les trajectoires estimées à partir des données acquises en extérieur. Nous avons souhaité ici évaluer l'impact du paramètre représentant le nombre d'images antérieures à considérer dans le calcul de la position courante. Nous pouvons observer que ce paramètre a un impact minime, l'ensemble des trajectoires ayant une forme assez similaire.

FIGURE 5.29 : Acquisition en intérieur : trajectoires obtenues avec la caméra *Fish Eye*

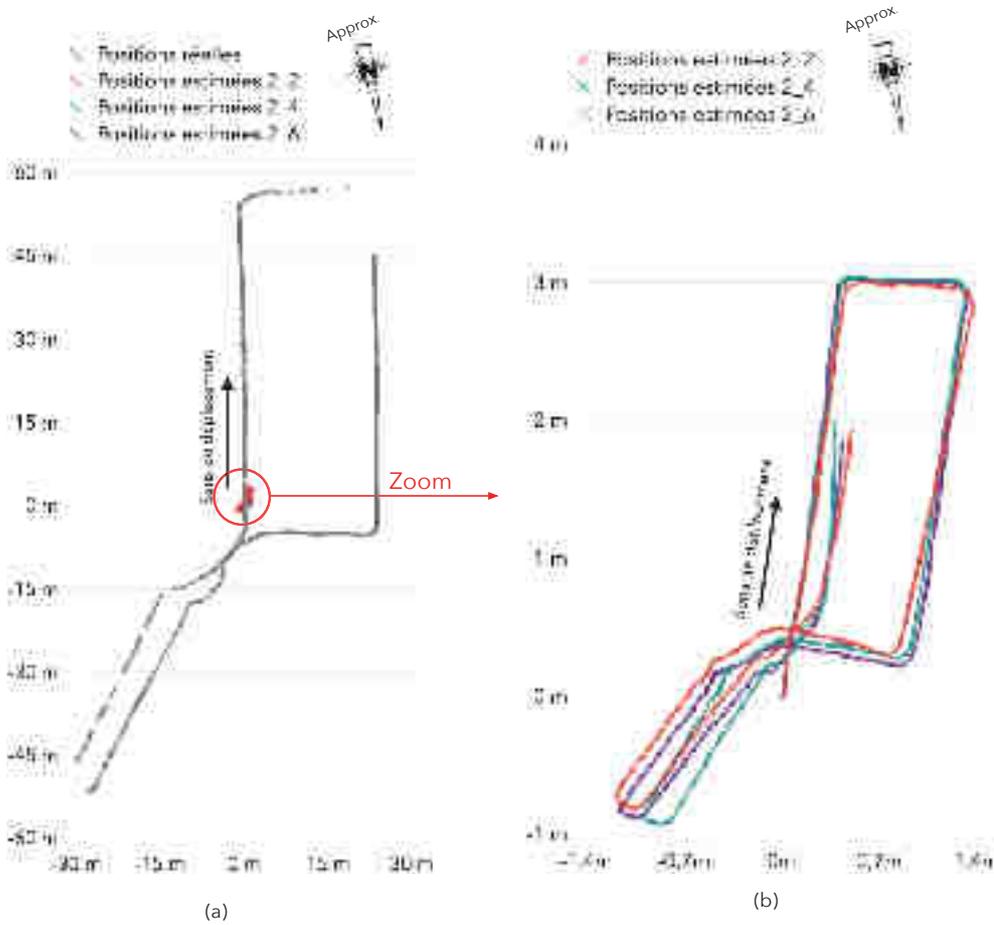


FIGURE 5.30: Acquisition en extérieur : (a) Trajectoires superposées à la trajectoire vraie - (b) Trajectoires obtenues avec la caméra *Fish Eye*

3.4. Bilan des résultats obtenus avec le pôle caméra *Fish Eye*

L'approche développée pour la caméra *Fish Eye* a pu être évaluée à partir d'expérimentations réalisées en conditions réelles. Comme nous l'avons imaginé lors de la conception du système, la caméra *Fish Eye* permet de fournir une trajectoire continue qui souffre d'une dérive moindre, par rapport à celle observée pour le pôle caméras 3D, au cours du déplacement. En effet, sa vision à long terme de la scène permet l'observation et l'exploitation d'informations visibles sur un long intervalle de temps. L'influence de différents paramètres a pu être évaluée. Cependant, le principal inconvénient de la méthodologie développée pour ce pôle de vision est la non détermination du facteur d'échelle nécessaire au calcul de la trajectoire réelle.

3.5. Deux pôles de vision pour une trajectoire

Le système final proposé comprend donc deux pôles de vision. Dans les deux sous parties précédentes, nous avons présenté les résultats obtenus pour chacun de ces pôles. Nous souhaitons dans cette partie les mettre en parallèle. Pour ce faire, nous avons reporté sur une unique figure les résultats obtenus par chacun des pôles de vision, durant une même acquisition, à savoir l'acquisition en intérieur. En effet, les résultats obtenus exclusivement pour le pôle caméras 3D en extérieur rendent la comparaison avec la trajectoire réelle peu pertinente.

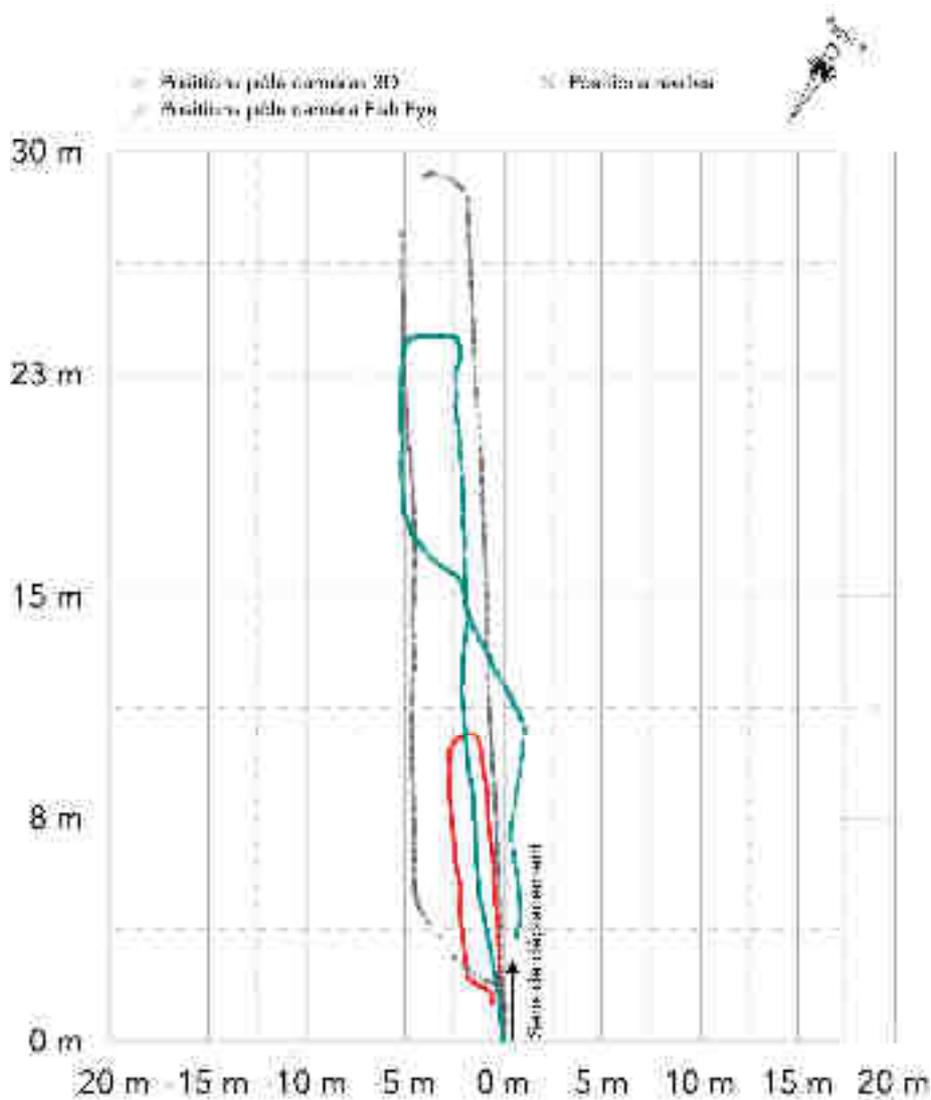


FIGURE 5.31 : Acquisition en intérieur : Trajectoires des deux pôles de vision superposées à la trajectoire vraie

Sur la Figure 5.31, nous pouvons observer les résultats obtenus par les deux systèmes de vision, grâce aux données enregistrées lors de l'acquisition en intérieur. Sur cette figure, nous avons représenté en gris les positions vraies du mobile durant le déplacement, en vert les positions obtenues grâce à l'approche développée autour du pôle de caméras 3D, et enfin en rouge la trajectoire obtenue avec le pôle de caméra *Fish Eye*.

La trajectoire calculée à partir des données fournies par la caméra *Fish Eye* est obtenue à un facteur d'échelle près, ceci rendant plus complexe la comparaison des trajectoires, cependant nous pouvons formuler un ensemble d'observations.

Tout d'abord, nous pouvons remarquer que la trajectoire estimée grâce à l'approche développée pour la caméra *Fish Eye* est plus stable que celle obtenue au travers du pôle caméras 3D. Ceci peut s'expliquer par une vision plus globale de la scène, en effet, le champ de vision très large de cette caméra lui permet d'observer les informations de la scène et notamment les façades de bâtiments, durant des périodes plus longues. A contrario, le pôle caméras 3D a une vision très locale de la scène et ceci est dommageable dans différentes situations que nous avons évoquées auparavant.

L'approche développée pour la caméra *Fish Eye* permet d'obtenir une trajectoire à un facteur d'échelle près, tandis que le pôle caméras 3D présente l'avantage de fournir cette information.

Il apparaît alors évident que ces deux pôles de vision peuvent jouer des rôles complémentaires. Il semble naturel de proposer une fusion des trajectoires fournies par chacun d'eux.

3.6. Bilan général

Si l'on reprend les critères établis en début de thèse visant à aboutir à un système de localisation innovant basé sur un ou plusieurs capteurs à bas coût, nous pouvons évaluer les différentes solutions proposées et développées. Tout d'abord en ce qui concerne le critère d'**accessibilité**, celui-ci est respecté dans la mesure où notre système est installé sur une plateforme mobile dont la largeur n'excède pas les 80 cm.

Le **coût** était un critère important à prendre en considération. L'ensemble du système développé répond parfaitement à celui-ci. Les modèles de caméras choisis, aussi bien pour les caméras 3D, que pour la caméra *Fish Eye*, permettent de limiter le coût global du système.

En ce qui concerne l'**intégration** avec les capteurs d'acquisition présents sur une plateforme mobile, la compacité du système développé permet de dire que cet aspect est également validé. Pour ce qui est **de la productivité et de la précision**, si l'on considère les pôles de vision de façon indépendante, ils ne répondent pas à ce critère. Cependant, la fusion des deux systèmes laisse à penser que nous serons en mesure d'atteindre le niveau exigé.

La **complexité** était un critère important à prendre en compte, celle-ci était de deux ordres. D'une part, le système imaginé ne devait pas être trop complexe en termes de réalisation et de

mise en œuvre, et c'est le cas de la solution proposée. D'autre part, la complexité algorithmique ne devait pas être trop importante afin d'assurer une estimation de la trajectoire de façon aisée. La représentation sous forme d'orthoimages permet de réduire la quantité de données à traiter, et donc de respecter ce critère.

Finalement, pour ce qui est de l'aspect **temps réel**, le pôle de caméras 3D permet de fournir une information de localisation en temps réel. Ce n'est actuellement pas le cas pour le pôle caméra Fish Eye. Bien que les approches développées pour ces deux systèmes soient proches d'un point de vue algorithmique, les choix de développement, surtout en ce qui concerne les langages sont différents. Cependant, il est tout à fait envisageable de rendre le pôle caméra Fish Eye exécutable en temps réel.

Ce bilan permet d'ouvrir de nombreuses perspectives d'amélioration, la plus notable étant la fusion des deux pôles de vision de notre système.

Conclusions du chapitre 5

Le chapitre 5 a permis de présenter les différents résultats obtenus avec les deux pôles de vision composant notre solution finale. Ce chapitre a également permis de présenter la complémentarité des deux systèmes. Enfin, nous avons pu évaluer notre approche par rapport aux critères établis en début de thèse. Nous pouvons établir le bilan suivant :

- La première simulation de données réalisée nous a permis de valider l'approche de localisation que nous avons souhaité mettre en place pour le pôle caméras 3D.
- La seconde simulation réalisée a mis en avant la nécessité d'utiliser un autre type de données, à savoir l'information infrarouge. Celle-ci nous a également permis de valider l'utilisation de plusieurs caméras 3D sur le prototype final.
- La première acquisition réalisée en intérieur avec le prototype développé a permis de valider notre approche de localisation lorsque les conditions sont favorables. Cela signifie que la scène doit être suffisamment porteuse d'information si nous souhaitons disposer d'orthoimages exploitables. Cependant, nous avons pu observer que le nombre de transformations extrapolées, c'est-à-dire issues d'un calcul de moyenne des déplacements antérieurs, est plus important que lors des simulations. Ceci s'explique du fait de la qualité des nuages qui est moindre, comparée à celle atteinte avec les données simulées.
- La seconde acquisition réalisée en extérieur a permis de mettre en lumière plusieurs limitations de notre approche. Premièrement, les caméras 3D utilisées, bien qu'ayant fonctionné durant l'acquisition, ont fourni des nuages de qualité moindre en raison notamment de la luminosité ambiante, ce qui a directement impacté la qualité des orthoimages. La vitesse de déplacement durant cette acquisition a également été problématique, rendant parfois impossible l'évaluation des transformations, faute de recouvrement suffisant entre orthoimages.
- La solution alternative consistant à calculer une moyenne des transformations en cas d'absence de points d'intérêt peut parfois poser problème. En effet, nous avons pu observer que lorsque cette situation se répète un nombre de fois important, nous pouvons induire une erreur sur la trajectoire estimée. Cette solution a cependant pour avantage de permettre de conserver une continuité dans la trajectoire estimée.

- L'aspect incrémental de notre approche est problématique, surtout lorsque ne serait-ce qu'une seule transformation estimée entre deux orthoimages est fautive. Cette erreur aura un impact sur l'ensemble de la trajectoire fournie par notre algorithme. Il est nécessaire de trouver un critère de qualité permettant de valider chacune des transformations estimées par notre approche.
- Malgré l'emploi de 3 caméras, le pôle caméras 3D a une vision très locale de la scène du fait des champs de vision des caméras, mais aussi de leur orientation. Ainsi, il est difficile d'obtenir un recouvrement important entre les orthoimages successives et de travailler sur un ensemble d'orthoimages. À ce stade, le traitement s'opère uniquement d'orthoimage à orthoimage.
- Les résultats obtenus avec la caméra *Fish Eye* sont prometteurs. Nous avons pu voir que les trajectoires obtenues grâce à cette caméra sont conformes au déplacement effectué par le mobile.
- La caméra *Fish Eye* a une vision globale de la scène, ce qui permet d'assurer une continuité dans la trajectoire estimée. Néanmoins, la mise à l'échelle de la trajectoire doit être envisagée (via l'exploitation des caméras 3D).
- La plupart des critères établis dans le cahier des charges de départ ont été pris en compte dans la réalisation de notre solution.
- Les perspectives d'évolution, et notamment la fusion des résultats fournis par les deux pôles, laissent à penser que nous serons en mesure de répondre à la totalité des critères énumérés en début de thèse.

À partir de ces remarques, nous sommes en mesure de conclure, dans le chapitre suivant, sur l'ensemble des contributions du travail de cette thèse, mais également d'évoquer les nombreuses perspectives qu'offre l'ensemble de cette étude.

Conclusion générale

1	Synthèse des résultats obtenus à l'issue de la thèse	157
1.1.	Volet 1 : Conception d'un système de localisation	158
1.2.	Volet 2 : Développement d'un algorithme de localisation	159
2	Perspectives	160
2.1.	Évolution de la solution de capteurs	161
2.2.	Évolution de l'algorithme de localisation	161
3	Contributions et encadrements	162
3.1.	Contributions	162
3.2.	Encadrements	162

1 Synthèse des résultats obtenus à l'issue de la thèse

Le travail de cette thèse est réalisé dans le contexte de la cartographie mobile terrestre et plus particulièrement dans le cadre du projet Terra Mobilita, dont l'objectif est de fournir de nouvelles solutions pour le relevé des voiries et de l'espace public de façon plus automatisée. Ces solutions doivent également être moins coûteuses que les méthodes actuelles, et doivent permettre des numérisations 3D de qualité.

L'objectif de ce travail de thèse était de définir un nouveau système de capteurs ayant la capacité de localisation en environnement urbain. Afin de répondre à cette problématique, deux objectifs principaux se sont démarqués : la conception d'une solution de capteur(s) innovante ; et le développement d'un algorithme de localisation basé sur l'utilisation de cette solution. Ces objectifs ont été atteints.

Le premier objectif concernait la réalisation d'un prototype matériel, innovant, à bas coût et fonctionnel. Ce prototype devait être en mesure de fournir les données nécessaires à la localisation d'une plateforme mobile en environnement urbain. Notre travail a abouti à la réalisation de ce prototype, composé de deux pôles de vision différents. Le premier pôle comporte trois caméras 3D, le second pôle quant à lui est composé d'une caméra *Fish Eye*. Malgré la complexité de mise en œuvre rencontrée pour synchroniser l'ensemble de ces capteurs, il a été possible d'enregistrer les données et de réaliser plusieurs acquisitions.

Le second objectif concernait le développement d'un algorithme de localisation basé sur les données acquises par la solution de capteurs proposée. Pour ce faire, nous avons mis en place deux méthodologies distinctes pour chacun des pôles de notre système. Celles-ci sont toutes deux basées sur l'odométrie visuelle. Nous avons introduit à cette occasion une façon originale de représenter

les données acquises par les caméras 3D sous forme d'orthoimages. Cette représentation présente de nombreux avantages, dont la simplification des données à traiter, tout en conservant une information tridimensionnelle. Les différentes expérimentations menées ont permis de décrire les atouts et les limitations d'une approche d'odométrie visuelle reposant sur ces orthoimages.

La méthodologie développée et évaluée pour le pôle caméra *Fish Eye* a montré que ce capteur pourrait prendre le relais lorsque l'algorithme de localisation développé autour du pôle caméras 3D est mis en difficulté.

1.1. Volet 1 : Conception d'un système de localisation

Un état de l'art a permis de dresser un état des lieux du problème de localisation, en définissant les différents environnements auxquels nous pouvons être confrontés. Ensuite, pour chacun de ces environnements, les limitations des approches couramment employées ont été mises en évidence. En milieu urbain et plus particulièrement dans le cas de canyons urbains, les approches basées sur la localisation par satellites, par exemple, sont contraintes par la visibilité des satellites. L'ajout de capteurs (centrale inertielle, odomètres, etc.) au GNSS permet de corriger partiellement ce problème, grâce à la fusion des données fournies par l'ensemble de ces capteurs. Évidemment, la qualité de cette localisation va dépendre de l'investissement réalisé qui reste toutefois relativement lourd dans le cas du choix d'une telle solution.

Dans ce contexte, les capteurs de vision font de bons candidats pour remplacer tout ou partie des solutions existantes. De nombreuses techniques et méthodes de localisation ont pu être développées autour des capteurs de vision. Depuis une dizaine d'années, une nouvelle génération de capteurs est apparue : les caméras 3D. Celles-ci permettent l'acquisition rapide de données tridimensionnelles et colorimétriques. Leur prix les rend attrayantes auprès du grand public, mais également auprès de la communauté scientifique.

Dans un second temps, une démarche de conception a été suivie pour aboutir à la conception et au développement d'un prototype matériel innovant basé sur l'intégration de caméras 3D. Il a fallu réaliser une analyse du contexte dans lequel la solution devra évoluer et élaborer un ensemble de concepts susceptibles de répondre à la problématique posée. L'étude conceptuelle et les critères d'évaluation définis ont permis d'évaluer chacun de ces concepts et d'aboutir à la proposition d'une solution finale.

Cette solution est composée de deux pôles de vision tenant chacun un rôle bien spécifique. Un premier pôle composé de trois caméras 3D a été conçu. Celles-ci sont positionnées sur le mobile de façon à pouvoir observer le mobilier urbain. Leur orientation, vers le sol, chacune avec une direction différente, leur permet d'avoir une vision locale et précise de la scène. Le second pôle du prototype est composé d'une caméra de type *Fish Eye*. L'objectif spécifique de ce type de caméra a rendu nécessaire la mise en place d'une étape de calibration. Cette caméra est placée au dessus des caméras 3D, et orientée vers le ciel afin d'apporter une vue globale de l'environnement urbain. L'ensemble des pôles caméras 3D et *Fish Eye* sont disposés à environ deux mètres du sol sur le mobile dont nous souhaitons évaluer la trajectoire. Compte tenu des avantages de la fusion

de données acquises par chacun de ces pôles, notre souhait, à long terme, est de fusionner les trajectoires obtenues par chacun de ces deux pôles.

1.2. Volet 2 : Développement d'un algorithme de localisation

Le second volet de la thèse se focalise sur la mise au point d'une méthodologie permettant d'aboutir à la localisation d'un mobile en milieu urbain. Deux démarches différentes ont pu être développées autour de chacun des systèmes de vision de notre solution de capteurs. Bien qu'étant quelque peu différentes, ces deux démarches reposent sur une approche commune : l'odométrie visuelle. Apparue dans les années 80, l'odométrie visuelle consiste à évaluer le déplacement d'un mobile à partir d'images enregistrées par un système de vision placé sur le mobile.

L'originalité de l'approche proposée dans cette thèse repose sur l'utilisation d'orthoimages. Cette représentation consiste à projeter orthogonalement, pour chaque nuage de points enregistré, chaque point du nuage sur le plan représentant le sol. Cette représentation rend possible l'utilisation de techniques classiques de traitement d'images. La trajectoire de la plateforme mobile s'appuie donc sur le calcul du déplacement opéré entre deux orthoimages successives. Pour y parvenir, la chaîne de traitement se décline en quatre étapes : la détection de points d'intérêts, la mise en correspondance de ces points, l'élimination des correspondances aberrantes et enfin l'estimation de la transformation liant les deux orthoimages. Cette représentation des données bénéficie de plusieurs avantages. Le principal avantage est la simplification des données à traiter. En effet, l'utilisation de données en deux dimensions permet d'exploiter des techniques de traitement d'images éprouvées, mais également d'amoindrir le volume de données à traiter. D'autre part, la colorisation grâce à l'altitude permet tout de même de conserver l'information tridimensionnelle de la scène.

Compte tenu de l'originalité des données traitées, il a été nécessaire de réaliser une étude concernant un ensemble de détecteurs de points d'intérêt. Celle-ci a permis de sélectionner le détecteur le plus adapté à notre contexte. Dans une seconde étape, nous avons confronté différentes approches de mise en correspondance. Nous avons également établi une étape de filtrage permettant d'éliminer les mises en correspondance erronées. Cette étape exploite la contrainte non-holonomique du véhicule sur lequel est placé le système. Enfin, nous avons pu aboutir au calcul de la transformation 2D effectuée entre deux orthoimages successives.

La méthodologie développée autour de la caméra *Fish Eye* est très proche de l'approche développée pour les caméras 3D. La principale différence réside dans l'existence d'une étape finale d'extrapolation. En effet, la vision à long terme de la caméra nous permet de travailler sur un ensemble d'images plutôt qu'entre images successives.

Afin d'évaluer le potentiel du pôle caméras 3D pour un calcul de trajectoire et évaluer l'approche de localisation développée, nous avons mené plusieurs expérimentations. D'une part, des données simulées ont été obtenues à partir d'un logiciel simulant l'acquisition et le déplacement de caméras 3D dans un environnement virtuel. Ces premières expérimentations ont permis d'aboutir à un résultat autorisant de premières conclusions. L'approche développée pour le pôle caméras 3D a pu être validée dans un contexte simplifié. Afin de valider notre approche dans un contexte

proche de la réalité, nous avons réalisé une seconde simulation au sein d'une modélisation réaliste de l'environnement urbain. Celle-ci a permis de mettre en évidence la nécessité d'utiliser plusieurs caméras 3D, mais également la nécessité d'exploiter l'information infrarouge fournie par le modèle de caméra.

Une fois le prototype opérationnel, nous avons réalisé deux acquisitions. La première acquisition a été réalisée en intérieur et a permis de valider notre approche lorsque les conditions sont favorables, c'est à dire lorsque la scène est porteuse de suffisamment d'informations. Une seconde acquisition a été réalisée en extérieur en milieu urbain. Celle-ci a permis de mettre en avant certaines limitations de notre approche. Premièrement, sans grande surprise, nous avons constaté que le modèle de caméras utilisé n'était pas adapté à un fonctionnement en extérieur. Bien que les caméras aient été fonctionnelles durant l'acquisition, la qualité des données fournies par celles-ci a été moindre par rapport à une utilisation en intérieur, ce qui a évidemment impacté la qualité des orthoimages générées. De plus, la vitesse de déplacement durant l'acquisition a également été problématique. En effet, le recouvrement entre orthoimages a parfois été insuffisant pour aboutir à une estimation de transformation correcte. Enfin, l'aspect incrémental de notre approche implique que la moindre erreur d'estimation dans une transformation va avoir un impact direct sur l'ensemble de la trajectoire.

Les acquisitions réalisées grâce au prototype mobile développé ont également permis d'évaluer la méthodologie de localisation développée autour du pôle caméra *Fish Eye*. Les résultats de trajectoire obtenus à partir de ce pôle présentent une allure similaire à la trajectoire réelle. Un des inconvénients de cette approche est l'utilisation d'une seule optique, nous ne sommes donc pas en mesure de déterminer le facteur d'échelle nécessaire au calcul final de la trajectoire. De plus, les développements entrepris ne sont actuellement pas compatibles avec une approche temps réel. En effet, les temps de calcul requis pour aboutir à l'estimation de la trajectoire sont trop importants. Enfin, il est à noter que les conditions d'acquisition influencent également la qualité de la trajectoire. D'une part, il est nécessaire d'être en présence d'une luminosité suffisante et d'autre part, l'orientation de la caméra *Fish Eye* compromet toute opération de localisation lors de passages sous les arbres, ou encore dans les tunnels.

En résumé, la recherche réalisée au cours de cette thèse a permis de proposer d'une part une solution de capteurs innovante et à bas coût permettant de produire rapidement des données peu volumineuses, en vue de localiser un mobile en environnement urbain. D'autre part, nous avons proposé deux méthodologies de localisation propres à chacun des pôles de vision composant notre système. Nous avons développé une méthodologie originale de localisation basée sur l'odométrie visuelle appliquée à des orthoimages. Suite à ces constats, de nouvelles perspectives s'ouvrent à nous.

2 Perspectives

Au terme de ce travail, deux axes de poursuite des travaux se profilent. Le premier axe concerne les évolutions possibles de la solution de capteurs créée. Le second axe d'amélioration concerne

les développements et les méthodologies mises en place.

2.1. Évolution de la solution de capteurs

La première évolution consisterait à adjoindre de nouveaux capteurs aux caméras 3D. Par exemple, l'utilisation d'odomètres permettrait de répondre rapidement à la question de déplacement ou d'immobilité de la plateforme. Ainsi, nous pourrions rendre plus robuste le filtre détectant s'il y a eu un déplacement ou non.

La technologie des caméras 3D est en constante évolution. Tout laisse à penser que la qualité des nuages fournis par celles-ci va s'améliorer, notamment en ce qui concerne leur utilisation en extérieur ou leur portée. La commercialisation de la Kinect V2, et les premières évaluations que nous avons pu entreprendre autour de ce capteur laissent entrevoir une amélioration notable de la qualité des données fournies par la caméra [Lachat et al., 2015]. Nous pouvons également mentionner l'apparition de caméras 3D au sein de divers appareils, et notamment des téléphones portables (exemple : Projet Tango, porté par l'entreprise Google). Il est alors envisageable de déployer nos développements vers de telles plateformes.

2.2. Évolution de l'algorithme de localisation

Il existe plusieurs axes d'évolution concernant l'aspect algorithmique de l'approche de localisation développée. À la vue des différents avantages et inconvénients de chacun des deux pôles de vision, il est possible d'envisager différentes interactions entre eux.

La principale évolution à envisager à court terme concerne la fusion des données fournies par chacun des systèmes de vision. En effet, lors de l'élaboration du concept développé, il était prévu de fusionner le pôle caméras 3D et le pôle caméra *Fish Eye*. Nous avons pu observer que la qualité des résultats obtenus par le pôle caméras 3D est directement liée au champ de vision couvert par les caméras, à la géométrie de la scène et à la qualité des nuages de points. De façon conforme à ce que nous avons imaginé, nous pouvons remarquer que le résultat obtenu à partir de ce système offre une certaine précision localement, mais présente une tendance à la dérive importante. Du fait de la vision globale offerte par l'objectif *Fish Eye* utilisé, la trajectoire calculée par ce pôle semble stable et pourrait venir corriger le calcul de trajectoire fourni par les caméras 3D, à un facteur d'échelle près. Nous avons également pu observer que les résultats obtenus pour le pôle caméra *Fish Eye* sont prometteurs, mais incomplets en ce qui concerne le facteur d'échelle.

Les résultats de calculs de trajectoire obtenus avec les caméras 3D ont mis en évidence un certain nombre de limitations. Une de ces limitations concerne la capacité du pôle caméras 3D à fournir une vision "continue" de l'environnement. Ainsi, lorsque le nombre de mises en correspondance entre deux orthoimages est nul ou insuffisant, il n'est pas possible d'évaluer la transformation entre celles-ci. Afin de pallier ce problème, nous avons mis en place une solution consistant à moyenniser les n dernières transformations calculées dans l'optique d'assurer une continuité dans la trajectoire. Cependant, nous avons pu observer que cette solution présentait des limites, puisque les déviations par rapport à la trajectoire vraie se répercutaient par transmission d'erreur sur le

résultat final. Dans ce contexte, l'utilisation combinée des systèmes de vision permettrait de fournir une alternative. Dans le cas où la transformation entre deux orthoimages, donc deux positions du mobile, ne serait pas estimable par le pôle caméras 3D, il serait envisageable de considérer la transformation obtenue à partir de la caméra *Fish Eye*.

Il est réciproquement possible de faire intervenir les caméras 3D dans l'approche développée autour de la caméra *Fish Eye*. En effet, comme nous avons pu le voir précédemment, la trajectoire obtenue à partir de la caméra *Fish Eye* est calculée à un facteur d'échelle près. L'obtention de ce facteur d'échelle peut être réalisée par différentes approches, notamment par l'utilisation de points, dont les coordonnées 3D sont connues grâce aux caméras 3D.

Finalement, pour le pôle caméras 3D comme pour le pôle caméra *Fish Eye*, il est possible de travailler sur le calcul de fermeture de boucle et de répartition des écarts [Ho & Newman, 2007], [Angeli et al., 2008]. Dans notre cas, il serait intéressant d'appliquer un protocole d'acquisition visant à revenir sur un endroit déjà parcouru, de façon à identifier l'écart de positionnement et de répartir cet écart sur les positions définissant la trajectoire. Cette approche permettrait entre autres de corriger une erreur de positionnement, et de la répartir sur l'ensemble de la trajectoire.

3 Contributions et encadrements

3.1. Contributions

Les contributions de cette thèse ont fait l'objet de trois communications avec actes dans des congrès internationaux :

- **Mittet, M.A.**, Grussenmeyer, P., Landes, T., Yang Y., Bernard N. (2013). Mobile outdoor relative localization using calibrated RGB-D cameras. In Proceedings of 8th International Symposium on Mobile Mapping Technology, Tainan (Taiwan), 6 pages.
- **Mittet, M.A.**, Landes, T., Grussenmeyer, P. (2014). Localization using RGB-D cameras orthoimages. In : ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. ISPRS Technical Commission V Symposium, Riva Del Garda. Vol XL-5. p.425-432.
- Lachat, E., Macher, H., **Mittet, M.A.**, Landes, T., Grussenmeyer, P. (2015). First experiences with Kinect V2 sensor for close range 3D modelling. In : ISPRS-International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences. 3D Virtual Reconstruction and Visualization of Complex Architectures. Vol XL-5/W4. p-93-100.

3.2. Encadrements

Plusieurs projets menés et encadrés au cours de cette thèse ont permis de faire évoluer nos travaux. Le premier projet de fin d'études encadré a permis de valider nos recherches concernant l'utilisation de l'odométrie visuelle sur la base d'orthoimages [Yang, 2012]. Les autres encadre-

ments ont concerné la calibration et l'utilisation d'une caméra *Fish Eye* dans le cadre d'une tâche de localisation en environnement urbain ([Vincent, 2013], [Royère, 2014]).

- Projet de Recherche Technologique (1 jour par semaine d'octobre à janvier) :
 - **Geoffrey Vincent** : Calibration des systèmes omnidirectionnels (2012/2013).
 - **Quentin Choquart** : Calibration d'une caméra *Fish eye* (2013/2014).

- Projet de fin d'études (6 mois, de février à juillet) :
 - **Yifan Yang, étudiant en 5^{ème} année de Télécom St Etienne** : Algorithme de localisation par caméra 3D (2012).
 - **Geoffrey Vincent, étudiant en 5^{ème} année de l'INSA de Strasbourg** : Apport des caméras *Fish Eye* (hypergonnes) au positionnement en milieu urbain (2013).
 - **Mikaël Royère, étudiant en 5^{ème} année de l'INSA de Strasbourg** : Localisation en environnement urbain basée sur l'utilisation d'un objectif *Fish Eye* (2014).

Table des figures

1.1	Exemples de systèmes de cartographie mobile : (a) L3D2 prototype développé au sein du CAOR de MINES ParisTech ; (b) Système mobile de l'université de Floride (FLREC Geomatics) ; (c) Système maritime LandScope ; (d) Google Street View Trekker ; (e) Google Street View Trolley ; (f) Système iMMS de Viametris	2
1.2	Composition d'un système de cartographie mobile [Yoo, 2011]	2
1.3	Diagramme de décomposition fonctionnelle du projet Terra Mobilita. (Source : Site Terra Mobilita)	5
1.4	Approches développées pour la tâche de localisation, positionnées selon leur appartenance à l'une ou l'autre des communautés	7
2.1	Principe de la trilatération où pd_1 , pd_2 et pd_3 représentent les distances séparant le récepteur (la voiture) des émetteurs (satellites)	13
2.2	Phénomène du masquage des satellites pour un récepteur (véhicule).	15
2.3	Mesure de l'angle d'élévation des masques [Nahimana, 2009]	16
2.4	Phénomène de canyon urbain	16
2.5	Phénomène de multi-trajet(en rouge)	17
2.6	Système à plateformes stabilisées [King, 1998]	19
2.7	Système à plateformes liées	20
2.8	Modèle cinématique de type unicycle	23
2.9	Modèle cinématique de type tricycle (où CIR signifie Centre Instantanée de Rotation)	25
2.10	Mode non couplé	28
2.11	Mode faiblement couplé	29
2.12	Mode fortement couplé	29
2.13	Principe de la triangulation active	34
2.14	Principe de mesure du temps de vol direct	35
2.15	Principe de mesure du temps de vol indirect	36
2.16	Exemples de caméras 3D : (a) Kinect de Microsoft ; (b) SR-4000 de MESA Imaging ; (c) PMDCamCube3.0 de PMD Technologies GmbH	37
3.1	Différentes étapes de notre démarche de conception	40
3.2	Exemple de scène enregistrée durant l'acquisition terrain (Rue Mauconseil, Fontenay-sous-Bois)	43
3.3	Modélisation du premier concept établi	44
3.4	Fonctionnement du premier concept	45
3.5	Évaluation du premier concept	46
3.6	(a) Modélisation du second concept - (b) Motorisation des caméras	46

3.7	Exemple d'objets pouvant être détectés (surlignés en rouge)	47
3.8	Évaluation du second concept	48
3.9	Image extraite du jeu de données utilisé dans le cadre du concept 3 (Technische Universität Munchen)	48
3.10	Différence entre deux cartes de profondeur successives sur la scène présentée Figure 3.9	49
3.11	Différence bornée entre deux cartes de profondeur successives	49
3.12	Évaluation du troisième concept	50
3.13	Présentation du concept final	51
3.14	Évaluation du concept final	52
3.15	Différents éléments de la caméra Xtion Pro de Asus	53
3.16	Principe de fonctionnement du système Asus Xtion Pro	53
3.17	Calcul des coordonnées 3D des points acquis à partir de la carte de profondeur	54
3.18	Exemple de données obtenues à partir de la Xtion Pro : (a) Carte de profondeur - (b) Image RGB - (c) Nuage de points (Source : https://graphics.stanford.edu/~mdfisher/Kinect.html)	54
3.19	Empreinte au sol obtenue à partir de la configuration choisie (vue du dessus)	55
3.20	Position des caméras (vue de dessus)	56
3.21	Champs de vision des trois caméras à partir de la configuration choisie	56
3.22	(a) Modélisation de la configuration finale du système de caméras 3D - (b) Prototype réalisé	57
3.23	Mire de calibration pour le contrôle du positionnement des caméras 3D	58
3.24	(a/b/c) Nuages de point issus de chacune des caméras - (d) Nuage fusionné	60
3.25	Principe de l'orthoprojection développée dans notre approche	61
3.26	Géométrie d'acquisition du système développé	61
3.27	(a) Image d'orthoprojection colorisée suivant l'altitude des points - (b) Image d'orthoprojection colorisée en fonction des données infrarouge fournies par les caméras 3D (à partir du nuage présenté Figure 3.24)	62
3.28	Champ de vision de la caméra omnidirectionnelle souhaité	63
3.29	(a/b) Exemples d'images obtenues à partir d'un objectif <i>Fish Eye</i> - (c) Lentille <i>Fish Eye</i> (Fujinon)	64
3.30	(a) Capteur panoramique RingCam [Cutler et al., 2002] - (b) Image panoramique produite par la caméra RingCam - (c) Caméra LadyBug (PointGrey)	65
3.31	(a) Caméra catadioptrique - (b) Image omnidirectionnelle à l'intérieur de la cathédrale d'Amiens obtenue avec un objectif catadioptrique.	66
3.32	(a) Caméra <i>MvBlueCougar_125a</i> de Matrix Vision équipée de l'objectif Fujinon <i>FE 185 CO57HA-1</i> - (b) Exemple d'image capturée avec cette configuration	66
3.33	Le modèle de sténopé [Hartley & Zisserman, 2003]	67
3.34	Coordonnées pixel (u, v) et métriques (x, y) d'un pixel p	68
3.35	Dispositif de prise de cliché en vue de la calibration [Vincent, 2013]	70
3.36	Mode opératoire pour la détermination de l'équation de projection [Bakstein & Pajdla, 2002]	71
3.37	Tableau de valeurs fourni par Fujinon	72
3.38	Logo d'OpenCV	75

3.39	Logo de PCL (<i>Point Cloud Library</i>)	75
3.40	Logo de ROS (<i>Robot Operating System</i>)	76
3.41	Principe de communication dans ROS (www.ros.org)	76
3.42	Principe de communication dans ROS (www.ros.org)	77
3.43	Communication mise en place au sein de notre prototype	77
3.44	Assemblage de l'ensemble des pôles de vision	79
3.45	Prototype en condition d'acquisition	79
3.46	Schéma de l'architecture matérielle du prototype	80
4.1	Principe de l'odométrie visuelle	86
4.2	Étapes de l'odométrie visuelle ([Scaramuzza & Fraundorfer, 2011])	87
4.3	Transformation entre les repères \mathcal{R}^* et \mathcal{R}	89
4.4	Schéma représentant le principe de notre approche de localisation	90
4.5	Méthodologie complète de localisation développée (pour chaque étape principale, le paragraphe associé est indiqué entre parenthèses)	92
4.6	Exemple de détection de points d'intérêt pour chacune des catégories de détecteurs mentionnées : (a) Détecteurs de coins (Harris [Harris & Pike, 1987]) - (b) Détecteurs basés sur les différence d'intensité (Hessian Laplace [Mikolajczyk & Schmid, 2004])	94
4.7	Exemple de détection de points d'intérêt (en vert) avec le détecteur de Shi & Tomasi	95
4.8	Exemple de détection de points d'intérêt (en vert) avec le détecteur SIFT	97
4.9	Exemple de détection de points d'intérêt (en vert) avec le détecteur SURF	97
4.10	Exemple de détection de points d'intérêt (en vert) avec le détecteur ORB	98
4.11	Données utilisées pour l'évaluation des détecteurs : (a) Scène 3D modélisée pour la simulation de nuage de points - (b) Orthoimage obtenue à partir de cette scène	99
4.12	Répétabilité des détecteurs après rotations	100
4.13	Répétabilité des détecteurs après translations	100
4.14	Évaluation du temps de détection pour chaque détecteur de notre étude : (a) Ratio $temps_de_détection/nb_points_détectés$ (b) Nombre de points détectés sur l'orthoimage présentée Figure 4.11(b) et temps de détection	101
4.15	Application des détecteurs de points d'intérêt évalués sur l'orthoimage présentée sur la Figure 4.11(b) (points détectés en vert)	102
4.16	Scène simulée utilisée pour évaluer les différentes approches de mise en correspondance (flèche rouge : sens du déplacement)	105
4.17	Résultat de la mise en correspondance grâce au suivi de points	106
4.18	Résultat de la mise en correspondance grâce à l'appariement	106
4.19	Principe de l'algorithme RANSAC. Illustration des étapes du paradigme de RANSAC pour la détection de ligne 2D.	107
4.20	Mouvement du système rapporté au CIR pendant un déplacement	109
4.21	Histogramme des valeurs de θ des mises en correspondance présentées dans la Figure 4.17	110
4.22	Application des détecteurs de point d'intérêts évalués sur une image <i>Fish Eye</i> acquise avec la caméra de notre système	116
4.23	Principe de la géométrie épipolaire	117

4.24	Procédure de calcul de position moyenne, utilisant les "n" précédentes images pour le calcul de la matrice position Ck_i . On notera Ck_{N_b} la dernière position à calculer. . .	119
5.1	Capteurs disponibles dans Blensor et paramétrages	123
5.2	Premier cas d'étude : déplacement rectiligne (trajectoire représentée en rouge) . . .	124
5.3	Second cas d'étude : (a) Modèle 3D de la place St Sulpice - (b) Trajectoire de la caméra (trajectoire représentée en rouge)	125
5.4	Acquisition en intérieur : plan de situation et trajectoire réalisée par le mobile (vert)	126
5.5	Acquisition en intérieur : environnement de l'acquisition - (a) Station totale 1 - (b) Station totale 2	126
5.6	Acquisition en intérieur : trajectoire de référence, relevée par les stations totales . .	127
5.7	Acquisition en extérieur : plan de situation et trajectoire réalisée par le mobile	128
5.8	Acquisition en extérieur : environnement de l'acquisition - (a) et (b) Système d'acquisition - (c) - Afin d'éliminer des vibrations trop importantes, une personne est montée dans le mobile pour maintenir l'ensemble des systèmes imageurs.	128
5.9	Acquisition en extérieur : trajectoire de référence, relevée par les stations totales . .	129
5.10	Premier cas d'étude : trajectoire calculée (rouge) et trajectoire vraie (grise)	131
5.11	Premier cas d'étude : erreurs moyennes quadratique de l'unité de poids pour chaque transformation calculée	132
5.12	Orthoimages correspondant aux différentes zones : (a) Zone 1 - (b) Zone 2 - (c) Zone 3	132
5.13	Second cas d'étude : trajectoire estimée (rouge) - trajectoire vraie (grise)	133
5.14	Empreinte au sol obtenue grâce aux 3 caméras simulées	134
5.15	Second cas d'étude avec 3 caméras : trajectoire estimée (rouge) - trajectoire vraie (grise)	135
5.16	Second cas d'étude avec zoom sur la zone 3 : trajectoire estimée (rouge) - trajectoire vraie (grise)	136
5.17	Second cas d'étude : évolution de l'écart entre les positions réelles et les positions estimées	137
5.18	Scène correspondant à la zone 1 dans la trajectoire estimée : (a) Vue de dessus de la scène - (b) Zoom sur la zone 1 (en rouge dans (a))	137
5.19	Deux acquisitions successives réalisées dans la zone 1 : (a) Position 1 - (b) Position 2	138
5.20	Orthoimages générées à partir des deux acquisitions successives présentées dans la Figure 5.19	138
5.21	Exemple d'orthoimage colorisée avec l'information infrarouge fournie par les caméras (Passage piéton)	139
5.22	Acquisition en intérieur : trajectoire estimée (rouge) - trajectoire vraie (grise)	140
5.23	Acquisition en intérieur : zoom sur la zone 1	141
5.24	Acquisition en intérieur : zoom sur la zone 2	142
5.25	Acquisition en extérieur : trajectoire estimée (rouge) - trajectoire vraie (grise)	143
5.26	Acquisition en extérieur : zoom sur la zone 1	144
5.27	Acquisition en extérieur zoom sur la zone 2	145
5.28	Acquisition en extérieur : zoom sur la zone 3	146
5.29	Acquisition en intérieur : trajectoires obtenues avec la caméra <i>Fish Eye</i>	149

5.30	Acquisition en extérieur : (a) Trajectoires superposées à la trajectoire vraie - (b) Trajectoires obtenues avec la caméra <i>Fish Eye</i>	150
5.31	Acquisition en intérieur : Trajectoires des deux pôles de vision superposées à la trajectoire vraie	151
B.1	Le filtre de Kalman	178
C.1	Modélisation du quatrième concept établi	181
C.2	Modélisation du cinquième concept établi	182
C.3	Modélisation du sixième concept établi	182
D.1	Empreinte au sol obtenue à partir de la configuration	183
D.2	Champs de vision des caméras à partir de la configuration	184
D.3	Empreinte au sol obtenue à partir de la configuration	184
D.4	Champs de vision des caméras à partir de la configuration	185
D.5	Empreinte au sol obtenue à partir de la configuration	185
D.6	Champs de vision des caméras à partir de la configuration	186
D.7	Empreinte au sol obtenue à partir de la configuration	186
D.8	Champs de vision des caméras à partir de la configuration	187
E.1	Mouvement du système rapporté au CIR pendant un déplacement	189



Liste des tableaux

2.1	Catégorie des milieux urbains [Malicorne et al., 2001]	15
2.2	Visibilité en milieux urbains	16
2.3	Classification de la précision des centrales inertielles [Cramer, 1997]	22
3.1	Coordonnées de chacune des caméras dans le repère de la mire. Ω , Φ , κ représentent les angles de rotation en degrés et T_x , T_y , T_z les translations.	59
3.2	Tableau comparatif des écarts et résidus de 3 clichés réalisés autour des bâtiments de l'INSA de Strasbourg	73
4.1	Tableau de comparaison des méthodes de détection de points d'intérêt sur une image <i>Fish Eye</i> acquise en extérieur	115
5.1	Tableau récapitulatif des données simulées	124
5.2	Tableau récapitulatif des données réelles, acquises par caméras 3D et caméra <i>Fish Eye</i>	129

Annexes

Les systèmes de localisation par satellites

1 Le système GPS

Le système de localisation par satellite le plus ancien et le plus largement diffusé aujourd'hui est le système GPS développé par les Etats-Unis. En 1958, apparaît aux USA le système de navigation TRANSIT/NNSS (Navy Navigation Satellite System) uniquement pour des besoins militaires. Reposant sur la mesure de l'effet Doppler, le système TRANSIT présentait une couverture satellitaire insuffisante et une précision trop faible. C'est pour cette raison que l'US Navy entreprend des recherches plus avancées sur un nouveau projet qui porte le nom de Timation. Parallèlement à ces travaux, l'US Air Force développe un autre concept spatial, le System 621, qui permettra un positionnement tridimensionnel.

Le 17 Février 1973, sur décision du gouvernement américain, l'US Air Force est désignée pour regrouper ces deux axes de recherches en un projet unique de navigation globale. C'est ainsi que le système NAVSTAR/GPS (NAVigation Satellite Timing And Ranging) voit le jour. Les premiers tests concluants sont réalisés en 1977. Il est alors décidé de lancer une constellation de satellites pré-opérationnelle (Block I), composée de sept satellites. Celle-ci est mise en service afin de valider l'ensemble du système. Le système NAVSTAR/GPS est déclaré opérationnel par le US DOD (Department of Defense) en février 1994. Depuis, il a toujours été en service et constitue le système GNSS actuellement utilisé par la majorité des récepteurs GNSS.

La constellation des satellites GPS comprend au moins 24 satellites pour être pleinement opérationnelle. Aujourd'hui, on compte 32 satellites en orbite, dont 30 opérationnels. Les satellites GPS transmettent des signaux sur deux porteuses de la bande L : L1 (1575,42 MHz) et L2 (1227,6 MHz), modulées en phases par différents codes binaires appelés code pseudo-aléatoire (PRN : *pseudo range noise*). La porteuse L1 est dédiée au signal civil, non codé, et donc utilisable par n'importe quel utilisateur, tandis que la L2 transmet un signal militaire qui est codé.

Le système GPS est composé de trois segments :

- **Le segment spatial** : ce segment est composé d'une constellation de satellites en orbite. Ces satellites sont répartis sur six orbites planes quasi-circulaires, ayant toutes une inclinaison de 55° par rapport au plan équateur. L'altitude est d'environ 20200 km et la période d'environ 12 h .
- **Le segment de contrôle** : ce segment était tout d'abord constitué de cinq stations au sol (*Monitor Stations*) se trouvant sur le territoire américain : Colorado Springs, Ascension Island, Diego Garcia, Kwajalein, Hawaii. Six autres stations ont vu le jour en 2005. La station de Colorado Springs constitue la station maîtresse de contrôle (*MCS : Master Control Station*) du système GPS. Les stations de contrôle suivent tous les satellites en vue, en enregistrant leurs données de navigation. Ces informations sont traitées au centre de calcul *MCS* pour déterminer les orbites des satellites et mettre à jour le message de navigation de chaque satellite.
- **Le segment utilisateur** : Ce segment est composé des récepteurs pouvant recevoir les signaux des satellites du système GPS. Ces récepteurs fournissent la position, la vitesse et le temps précis à l'utilisateur à tout moment et n'importe où sur le globe terrestre (air, mer, terre). Ils sont variés dans leur conception, et leur précision varie selon leur fonction.

2 Le système GLONASS

Le système russe GLONASS a été développé par l'ex-URSS en 1976, a peu près à la même époque que le système GPS. A partir de 1982, les premiers lancements de satellites ont eu lieu, et ceci jusqu'en 1995. A partir de cette date, le système a connu des problèmes de maintenance qui l'ont limité à six satellites, et ceci jusqu'en 2000. Cependant le gouvernement russe reprend les choses en main, et aux débuts des années 2008, nous trouvons 14 satellites opérationnels en orbite. Les satellites de cette constellation sont placés sur trois orbites inclinées de $64,8^\circ$ par rapport au plan équateur, l'altitude est d'environ 19100 km , et la période d'environ $11\text{ h}15$.

Comme le GPS, GLONASS produit deux types de signaux sur deux porteuses de la bande L. Le système est basé sur le même principe de fonctionnement que le GPS avec trois segments : un segment spatial, un segment de contrôle, et un segment utilisateur. Le segment de contrôle est composé d'un centre de commande et de surveillance, et de stations réparties sur le territoire russe. Le segment utilisateur est constitué des équipements destinés à recevoir et faire le traitement des signaux du système. Les systèmes GPS et GLONASS peuvent être utilisés de façon combinée par un récepteur GNSS. Cependant l'utilisation du système GLONASS n'est pas très répandue pour les applications civiles, on le retrouve plus souvent comme "complément" de la disponibilité du système GPS.

3 Le système COMPASS

En 2000, la Chine a démarré à son tour son propre système de localisation par satellites. Ce système appelé CNSS (Compass Navigation Satellite System) ou Beidou Navigation System est disponible en Chine et dans les pays limitrophes. La première phase du projet a permis le lancement de cinq satellites géostationnaires. A terme, le système COMPASS prévoit cinq satellites

géostationnaires plus une constellation de 30 satellites. Comme pour le GPS et GLONASS, le système COMPASS prévoit un service civil, et un service militaire.

4 Le système GALILEO

Afin de faire face aux "problèmes de mobilité et de transport", la commission européenne lance le projet GALILEO, en 1994. Lors du lancement du projet, l'un des objectifs de la commission européenne était de créer GALILEO afin de ne plus dépendre des systèmes militaires que sont le GPS (américain), et GLONASS (russe). Encore en cours de mise en place, *GALILEO* prévoit de nombreuses améliorations par rapport à ses prédécesseurs. L'avantage significatif de GALILEO est son interopérabilité avec le système GPS, ceci apportant "un réel avantage en terme de précision et en termes de sécurité en cas de défaillance de l'un des systèmes" [Galileo, 2003].

Le système offrira cinq services dont les performances dépendront des usages de chacun :

- **Le service ouvert** (*OS : Open Service*) : destiné aux applications grand public, il sera gratuit et couvrira toute la planète.
- **Le service commercial** (*CS : Commercial Service*) : de performances équivalentes au service ouvert, il lui sera ajouté des données de correction et une information sur l'intégrité des signaux. Son accès sera restreint, par cryptage, aux applications commerciales.
- **Le service de sûreté de vie** (*SOL : Safety Of Life*) : de performances identiques au service de base avec une information additionnelle de son intégrité élevée. Il sera restreint aux groupes d'utilisateurs tels que les hopitaux, les pompiers, les transports...
- **Le service public régulé** (*PRS : Public Regulated Service*) : à destination des utilisateurs relevant d'une mission du service public. Il sera contrôlé par l'Union Européenne et les gouvernements des états membres. Il sera crypté par des codes spécifiques, et reposera sur deux signaux dédiés, afin d'en garantir la disponibilité.
- **Le service d'aide à la recherche et au sauvetage** (*SAR : Search And Rescue*) : ce service permettra la diffusion et la localisation de messages de détresse.

Les 4 premiers satellites ont été mis en orbite en 2011 et 2012. Les tests effectués sur ces derniers ont montré qu'ils étaient opérationnels et ont permis de valider l'efficacité de l'ensemble du dispositif en orbite comme au sol.

Le lancement des satellites 5 et 6 a eu lieu le 22 août 2014, cependant selon Arianespace les satellites n'ont pas atteint la bonne orbite. Ceci ne remet pas en cause GALILEO, et les satellites 7 et 8 devraient être lancés dans le courant de l'année. Au total, la constellation GALILEO devrait contenir 30 satellites, lancés au rythme de 6 à 8 par an. L'objectif est que 16 satellites au moins soient en orbite d'ici 2016.

Le segment sol de GALILEO comporte deux centres de contrôle (*GCC*) implantés à Oberpfaffenhofen (Allemagne) et Fucino (Italie), ils sont en charge du suivi, du contrôle et de la gestion de la mission du système. On compte également 40 stations de surveillance (*GSS : Galileo Sensor Stations*) à travers le monde qui enregistrent les données de navigation de tous les satellites avant transmission au GCC.

Annexe

B

Filtre de Kalman

Le filtre de Kalman est un prédicteur-estimateur, il définit une estimation récursive, optimale au sens d'un critère quadratique de l'état du système à partir d'une séquence d'observations. Il met à jour un estimateur qui tient compte de toutes les informations disponibles jusqu'à l'instant k considéré. Nous considérons que le système est discret, c'est-à-dire que l'évolution du système n'est pas prise en compte de façon continue par le calculateur, mais que les mesures ne sont disponibles qu'à certains instants. Le modèle dynamique du système peut alors s'écrire [Kalman, 1960] :

$$X_{k+1} = A_k X_k + B_k U_k + v_k \quad (\text{B.1})$$

avec A_k la matrice de transition décrivant l'évolution du système d'un état à l'autre et B_k la matrice de commande; v_k représente le bruit du processus, il est supposé blanc, gaussien (distribution normale de covariance $Q_k : v_k \sim \mathcal{N}(0, Q_k)$). Le modèle de mesure est alors défini tel que :

$$Y_k = C_k X_k + w_k \quad (\text{B.2})$$

où C_k est la matrice de mesures et w_k le bruit de mesure de type blanc, gaussien (distribution normale de covariance $R_k : w_k \sim \mathcal{N}(0, R_k)$). L'hypothèse, selon laquelle les bruits d'état v_k et de mesure w_k sont des variables aléatoires dont les distributions gaussiennes sont connues, permet de simplifier le formalisme des équations d'évolutions et d'observations.

Le filtre de Kalman repose sur trois étapes fondamentales : l'initialisation, la prédiction (estimation *a priori*) et la correction (estimation *a posteriori*) (Figure B.1). Les paramètres du système sont estimés dans un premier temps à partir des valeurs de l'étape précédente. Ils sont ensuite corrigés par des mesures dans l'étape de correction. Ces deux étapes tournent de façon récursive après l'initialisation. L'ensemble de ces étapes seront décrites plus en détails par la suite.

a. L'initialisation

L'état initial du filtre est une variable gaussienne, indépendante des bruits du système, de moyenne et de matrice de covariance connues :

$$\begin{aligned} \hat{X}_{0|0} &= E(X_0) \\ P_{0|0} &= E[(X_0 - \hat{X}_{0|0})(X_0 - \hat{X}_{0|0})^T] = \text{var}[X_0] \end{aligned} \quad (\text{B.3})$$

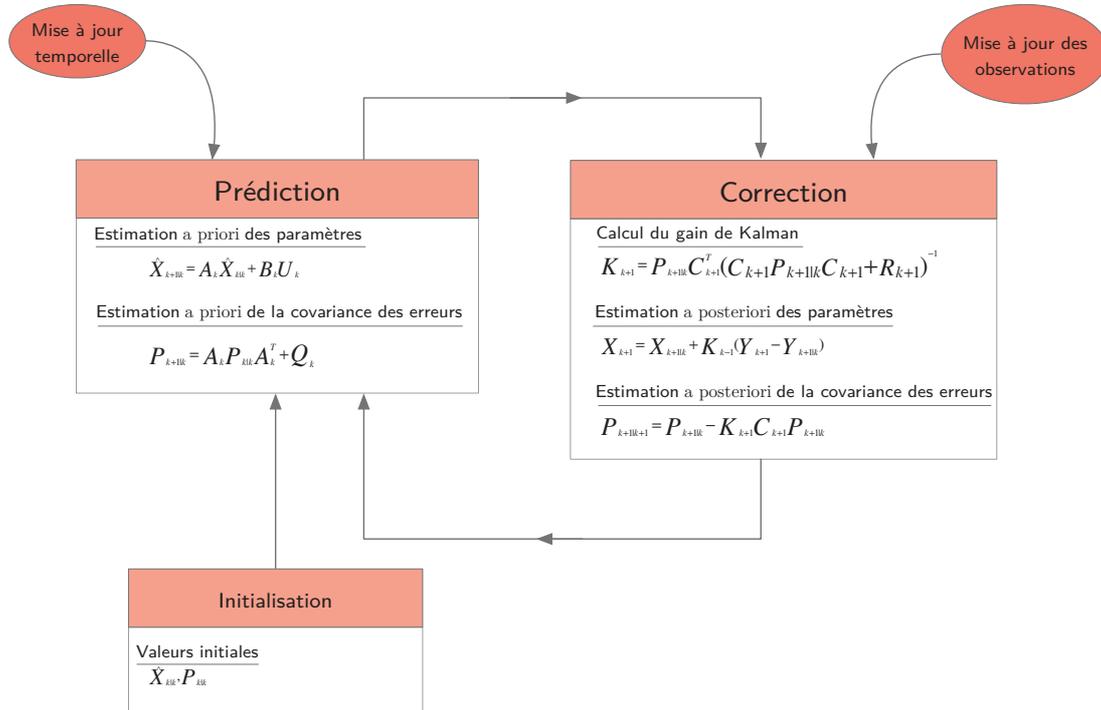


FIGURE B.1 : Le filtre de Kalman

- Prédiction temporelle

Cette étape permet la prédiction de l'état du système et sa précision à $k + 1$ en se basant sur l'état à l'instant k et sur le modèle d'évolution du système. Nous notons $\hat{X}_{k+1|k}$ cette prédiction et $P_{k+1|k}$ la matrice de covariance associée :

$$\begin{aligned} \hat{X}_{k+1|k} &= A_k \hat{X}_{k|k} + B_k U_k \\ P_{k+1|k} &= A_k P_{k|k} A_k^T + Q_k \\ \hat{Y}_{k+1|k} &= C_{k+1} \hat{X}_{k+1|k} \end{aligned} \quad (\text{B.4})$$

où :

- $\hat{X}_{k|k}$: le vecteur d'état,
- A_k : la matrice de transition entre les instants $k - 1$ et k ,
- Q_k : la matrice de covariance du processus (bruit du système).

- Correction et mise à jour des mesures

Une fois la mesure Y_{k+1} disponible, l'état prédit peut alors être corrigé par ce que l'on nomme l'innovation $\hat{S}_{k+1|k}$ pondérée par le gain K_{k+1} . Nous pouvons alors en déduire l'estimation de

l'état $\hat{X}_{k+1|k+1}$ ainsi que la matrice de covariance $P_{k+1|k+1}$ associée :

$$\begin{aligned}
 \hat{S}_{k+1|k} &= Y_{k+1} - Y_{k+1|k} \\
 S_{k+1} &= C_{k+1}P_{k+1|k}C_{k+1}^T + R_{k+1} \\
 K_{k+1} &= P_{k+1|k}C_{k+1}^T(S_{k+1|k})^{-1} \\
 \hat{X}_{k+1|k+1} &= \hat{X}_{k+1|k} + K_{k+1}\hat{S}_{k+1|k} \\
 P_{k+1|k+1} &= (I - K_{k+1}C_{k+1})P_{k+1|k}
 \end{aligned} \tag{B.5}$$

où :

- K_{k+1} : le gain de Kalman compris entre 0 et 1,
- $P_{k|k}$: la matrice de covariance,
- Y_{k+1} : le vecteur de mesures,
- C_{k+1} : la matrice d'observations,
- R_k : la matrice de covariance des mesures.

Le gain de Kalman, toujours compris entre 0 et 1, joue un rôle très important dans l'étape de correction. Il permet de connaître la confiance que l'on peut apporter à la mesure par rapport à la prédiction. Il devient d'autant plus important que la covariance R_k du bruit de mesure tend vers une matrice nulle, ce qui privilégie la mesure devant la prédiction. A l'inverse, si la covariance $P_{k|k+1}$ prédite tend vers 0, et donc que la matrice Q_k tend elle-même vers 0, le gain de Kalman tend vers 0 aussi, et dans ce cas le filtre privilégie la prédiction devant la mesure.

Annexe

C

Autres concepts élaborés

Concept 4

Ce concept repose sur l'utilisation d'une unique caméra 3D (Figure C.1). Nous souhaitons étudier ici la possibilité de mettre en relation les nuages de points fournis par la caméra 3D.



FIGURE C.1: Modélisation du quatrième concept établi

Aspect fonctionnel

Dans le cadre de ce concept, la solution imaginée consiste à détecter dans les nuages de points les objets de l'environnement urbain. Une fois que cette première étape de détection est réalisée, il est prévu de mettre en correspondance les différents objets détectés. Finalement, nous appliquerons une approche de type ICP [Besl & McKay, 1992] afin d'évaluer la transformation entre les nuages successifs fournis par les caméras. L'Iterative Closest Point est le premier des algorithmes de recalage d'un nuage de points sur un autre. Cette approche est composée de 3 étapes : 1. Appariement des paires de points les plus proches, 2. Estimation des paramètres de transformation, 3. Test et ajout des résidus.

Concept 5

Ce concept est proche du premier concept présenté dans ce manuscrit. La principale différence repose sur le nombre de caméras utilisées. En effet, il est prévu de placer cinq caméras à l'avant

du mobile, afin d'avoir un champ de vue le plus large possible.

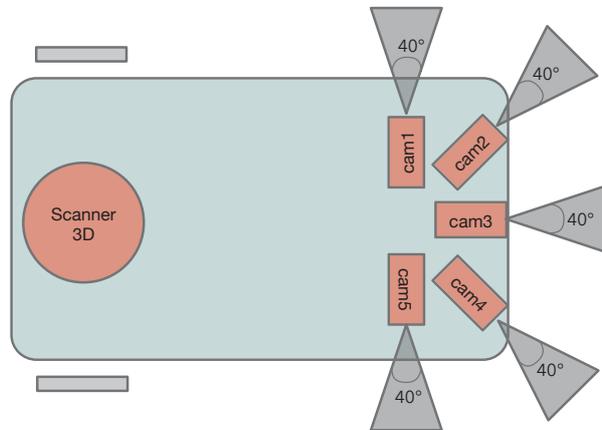


FIGURE C.2: Modélisation du cinquième concept établi

Aspect fonctionnel

A nouveau le système imaginé fonctionnera en mode "stop & go". Il est prévu d'étudier la possibilité de mettre en relation les nuages de points fournis par le scanner, et par les différentes caméras 3D. Connaissant les positions relatives de chacune des caméras, il est possible d'obtenir un unique nuage de points. Il est alors envisagé de réaliser une corrélation entre le nuage de points fourni par le scanner, et celui obtenu par l'intermédiaire des caméras 3D.

Concept 6

Ce concept repose sur l'utilisation d'un ensemble de capteurs utilisés classiquement dans la tâche de localisation. Il est envisagé de développer une approche de SLAM classique autour de l'ensemble des ces capteurs.

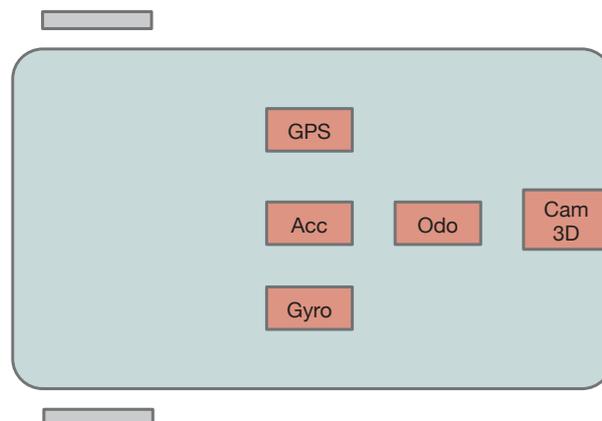


FIGURE C.3: Modélisation du sixième concept établi

Différentes configurations étudiées pour le placement des caméras 3D

1 2x2 caméras

Cette configuration est composée de 4 caméras réparties en deux groupes. Un premier groupe est orienté vers l'avant du mobile, tandis que le second est orienté vers l'arrière du mobile.

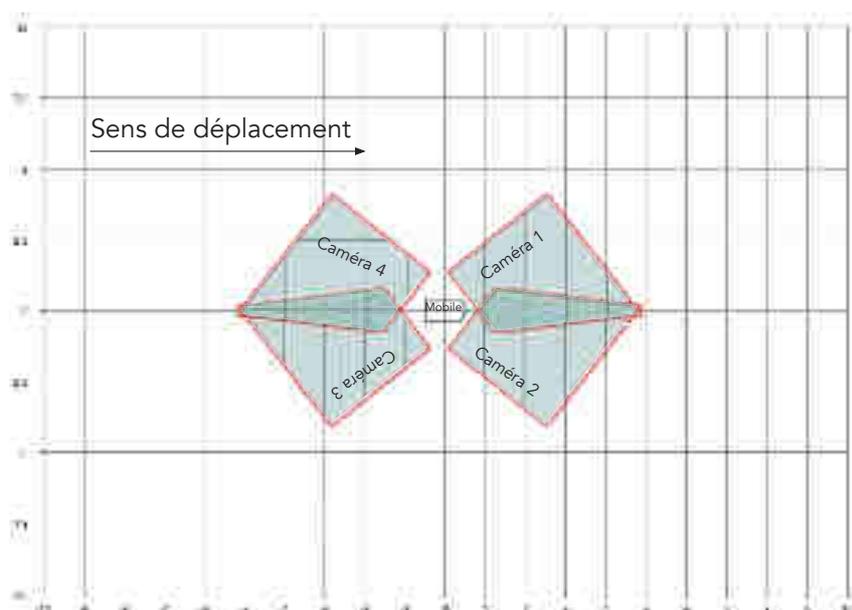


FIGURE D.1 : Empreinte au sol obtenue à partir de la configuration

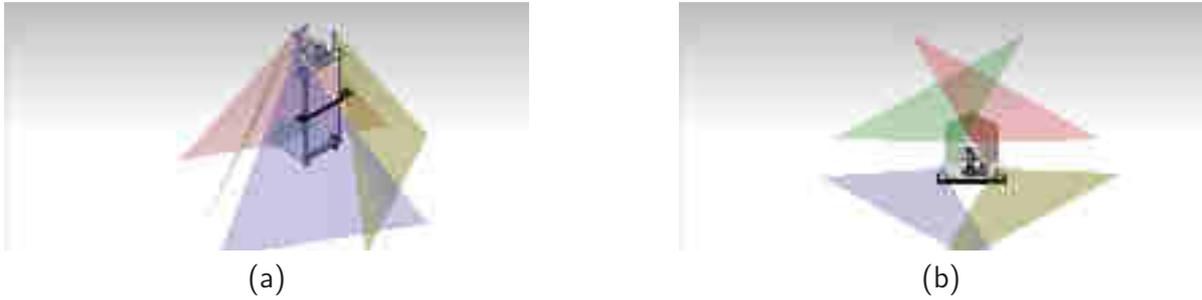


FIGURE D.2: Champs de vision des caméras à partir de la configuration

2 3 caméras

Pour cette configuration, l'utilisation de 3 caméras est requise. Nous trouvons deux caméras orientées vers l'avant du mobile et une troisième caméra orientée vers l'arrière du mobile.

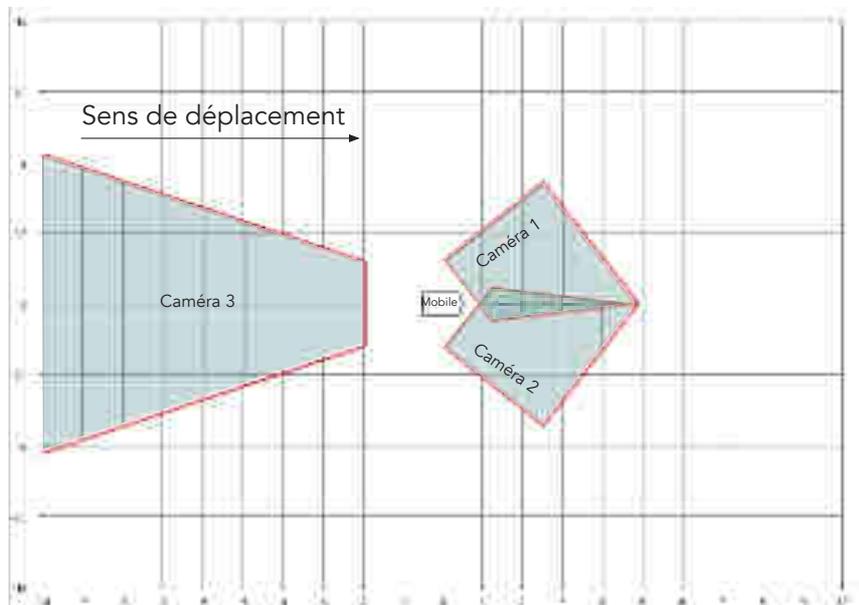


FIGURE D.3: Empreinte au sol obtenue à partir de la configuration

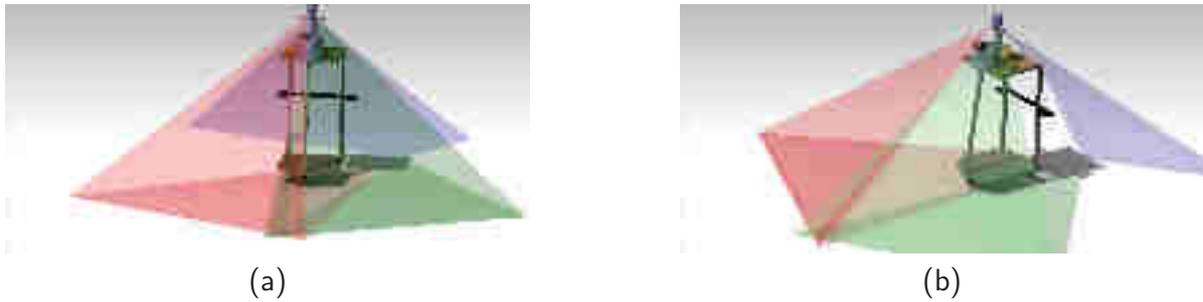


FIGURE D.4: Champs de vision des caméras à partir de la configuration

3 4 caméras

Pour cette configuration, l'utilisation de 4 caméras est requise. Nous trouvons deux caméras orientées vers l'avant du mobile et deux caméras orientées vers l'arrière du mobile.

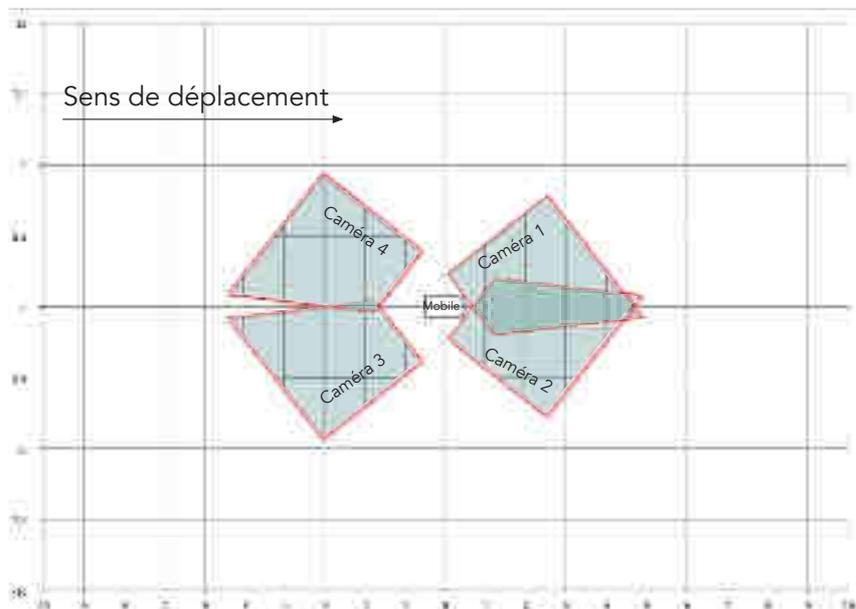


FIGURE D.5: Empreinte au sol obtenue à partir de la configuration

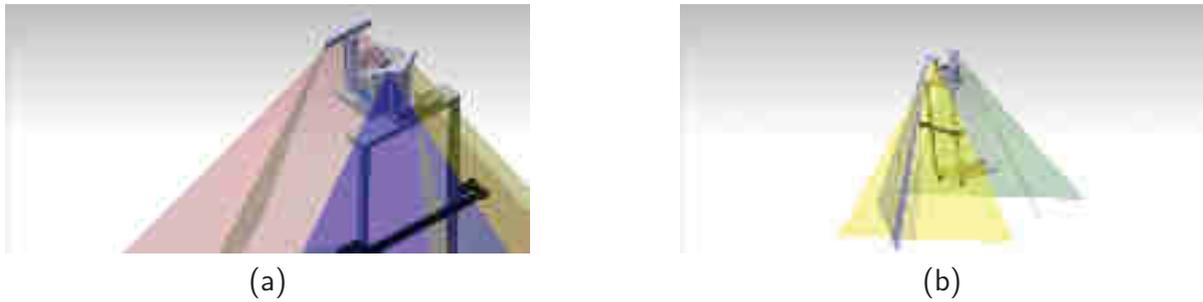


FIGURE D.6: Champs de vision des caméras à partir de la configuration

4 4 caméras verticales

Pour cette configuration, l'utilisation de 4 caméras est requise. Nous trouvons deux caméras orientées vers l'avant du mobile et deux caméras orientées vers l'arrière du mobile. Pour cette configuration les caméras ont été orientées à la verticale, ceci permet d'améliorer la compacité de la solution de capteurs.

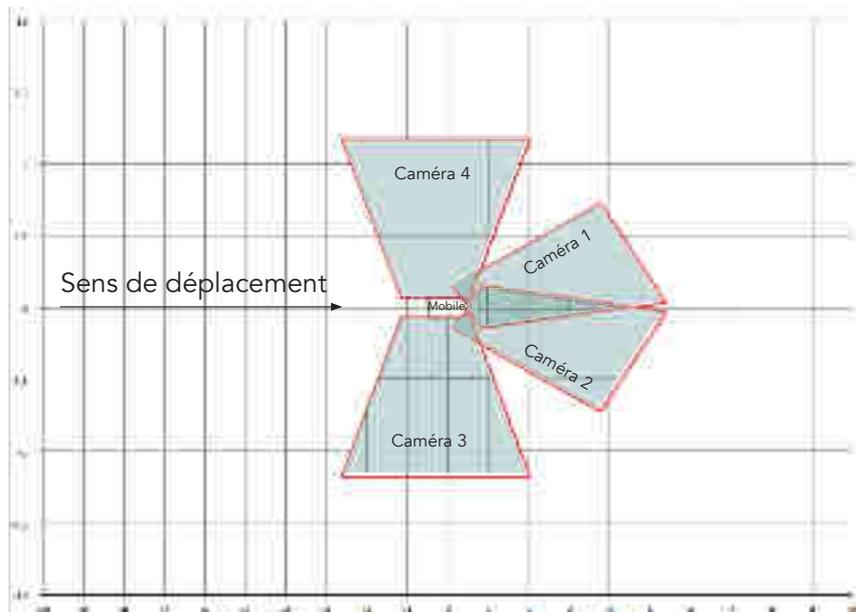


FIGURE D.7: Empreinte au sol obtenue à partir de la configuration

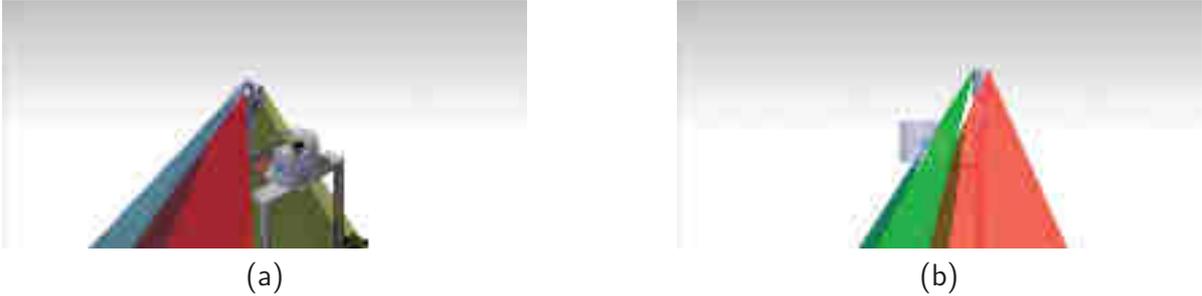


FIGURE D.8 : Champs de vision des caméras à partir de la configuration

Annexe

E

Démonstration du calcul de θ

On reprend la Figure 4.20 :

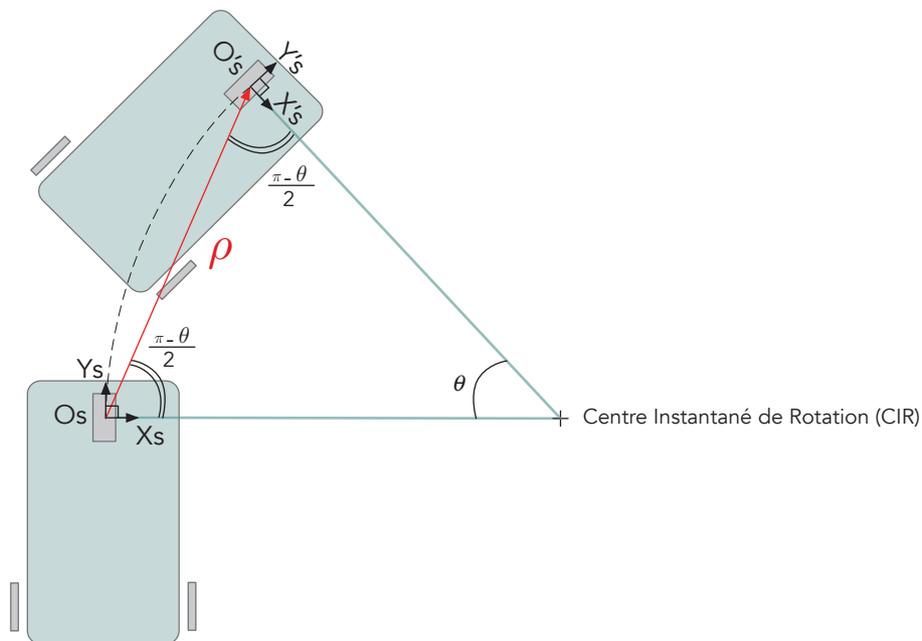


FIGURE E.1 : Mouvement du système rapporté au CIR pendant un déplacement

Nous notons p et p' ($p = [x, y, 1]^T$ et $p' = [x', y', 1]^T$), en coordonnées homogènes, des points présents dans deux orthoimages successives et mis en correspondance. On peut alors écrire :

$$\begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta & \rho \sin\frac{\theta}{2} \\ -\sin\theta & \cos\theta & \rho \cos\frac{\theta}{2} \\ 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (\text{E.1})$$

Si on pose $\frac{\theta}{2} = \Phi$, on aura alors :

$$\rho = \frac{x' - x \cos 2\Phi - y \sin 2\Phi}{\sin \Phi} \quad (\text{E.2})$$

$$\rho = \frac{y' + x \sin 2\Phi - y \cos 2\Phi}{\cos \Phi} \quad (\text{E.3})$$

À partir de (E.1) et (E.2), on peut écrire :

$$\frac{x' - x \cos 2\phi - y \sin 2\phi}{\sin \phi} = \frac{y' + x \sin 2\phi - y \cos 2\phi}{\cos \phi} \quad (\text{E.4})$$

On a alors :

$$\cos \phi (x' - x \cos 2\phi - y \sin 2\phi) = \sin \phi (y' + x \sin 2\phi - y \cos 2\phi) \quad (\text{E.5})$$

On sait que $\sin 2a = 2 \sin a \cos a$ et $\cos 2a = \cos^2 a - \sin^2 a$, on a :

$$\cos \phi (x' - x(\cos^2 \phi - \sin^2 \phi) - y(2 \sin \phi \cos \phi)) = \sin \phi (y' + x(2 \sin \phi \cos \phi) - y(\cos^2 \phi - \sin^2 \phi))$$

En développant, on obtient :

$$\begin{aligned} x' \cos \phi - x \cos^3 \phi + x \cos \phi \sin^2 \phi - 2y \sin \phi \cos^2 \phi - y' \sin \phi - 2x \sin^2 \phi \cos \phi + y \cos^2 \phi \sin \phi - y \sin^3 \phi &= 0 \\ \Rightarrow x' \cos \phi + x(-\cos^3 \phi + \cos \phi \sin^2 \phi - 2 \sin^2 \phi \cos \phi) + y(-2 \sin \phi \cos^2 \phi + \cos^2 \phi \sin \phi - \sin^3 \phi) - y' \sin \phi &= 0 \\ \Rightarrow x' \cos \phi + x(-\cos^3 \phi - \sin^2 \phi \cos \phi) + y(-\sin \phi \cos^2 \phi - \sin^3 \phi) - y' \sin \phi &= 0 \\ \Rightarrow x' \cos \phi + x(-\cos \phi(\cos^2 \phi + \sin^2 \phi)) + y(-\sin \phi(\sin^2 \phi + \cos^2 \phi)) - y' \sin \phi &= 0 \\ \Rightarrow x' \cos \phi - x \cos \phi - y \sin \phi - y' \sin \phi &= 0 \\ \Rightarrow \cos \phi (x' - x) + \sin \phi (-y - y') &= 0 \\ \Rightarrow \frac{\sin \phi}{\cos \phi} &= \frac{x' - x}{y' + y} \\ \Rightarrow \tan \phi &= \frac{x' - x}{y' + y} \\ \Rightarrow \phi &= \arctan \frac{x' - x}{y' + y} \\ \Rightarrow \theta &= 2 * \arctan \frac{x' - x}{y' + y} \end{aligned}$$

Annexe

F

Études sur les techniques de localisation par la vision

Étudiant : Yifan YANG, 5^{ème} année de Télécom St Etienne

Intitulé du mémoire : Algorithme de localisation par caméra 3D

Lieu : Trimble

Date de réalisation : mars à août 2012



Internship summary report

Algorithms for visual odometry

Trainee: YANG Yifan
School supervisor: DENIS Loic
Company supervisor: MITTET Marie-Anne
Dates: 05/03/2012-31/08/2012

School: Télécom St-Etienne
Study level: 5th year

Recommendation

Ready for release Ready to go in development phase Requires additional work

Presentation

The subject of the internship is related to the thesis of Marie-Anne. The subject of the thesis is about the localization of a mobile in urban environment with 3D cameras. The tractography based on 2D cameras is a mastered technique. The main goal of the internship is to explore and evaluate some techniques used with 2D cameras, with 3D cameras.

Execution

The internship was divided in different parts:

- First, for discovering 3D cameras, Yifan worked on calibration. The previous year, a first work was done by another intern on the kinect calibration. So, Yifan developed an application usable directly by the user. The parameters of calibration can be written in a configuration file. They are applied directly on the raw point cloud.
- Secondly, to discover point cloud, and these treatments, Yifan worked on the segmentation plan in a 3D scene. To do this, he used PCL (Point Cloud Library)
- Then Yifan studied the library ROS. This allowed us to quickly test the developed algorithms. In addition, ROS contains OpenCV and PCL.
- To discover ROS, and confronted our algorithms, Yifan studied RGBDSLAM. This is a package in ROS which is specialized in SLAM and, 3D scene reconstruction.

- Using various studies developed since the beginning of the internship, Yifan had implemented a module of planar trajectory. He placed a camera at 2 meters from the ground, whose field of view is oriented towards the ground. From the point cloud obtained, he has established orthoprojections images. The study of transformations between images allowed us to study the trajectory of the mobile. To validate the computed trajectory, Yifan used a TX5. He placed a target on the camera, after that he took a first scan at the beginning of the displacement. Then he took a second scan at the end. With this two scans, Yifan has computed the distance between the beginning and the end of the displacement. This was a first evaluation of the algorithms. To exploit this result, we tried to do a straight displacement.



Figure 1 - Camera

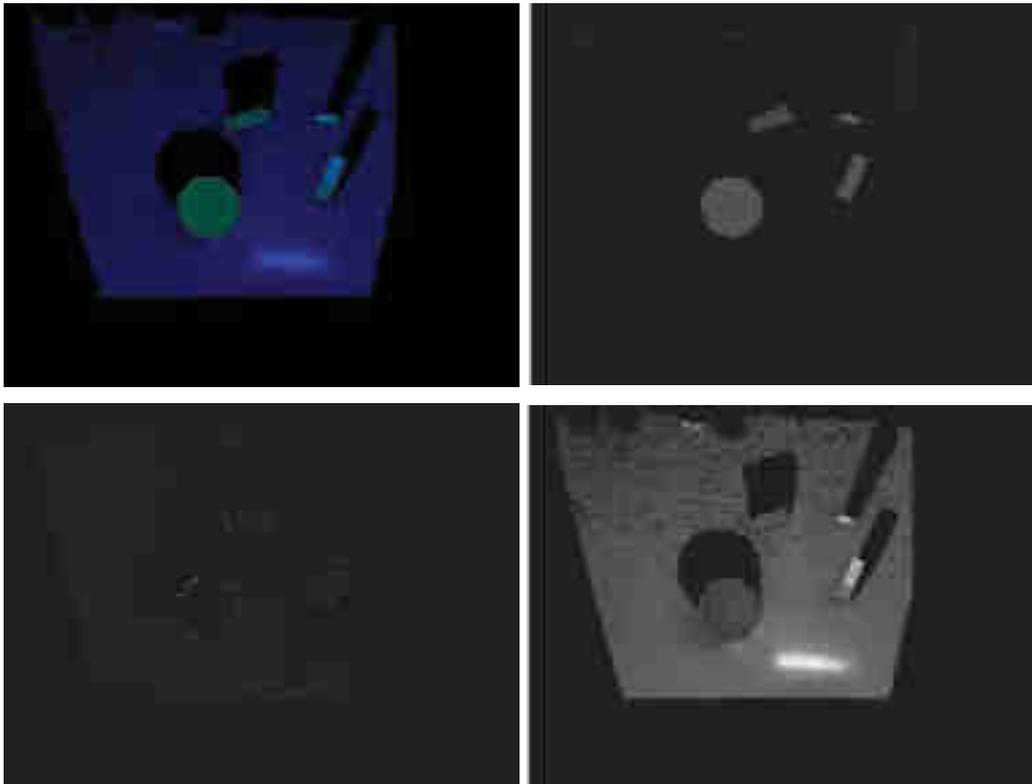


Figure 2 - Orthoprojections

Conclusions

First, Yifan finalized the work of Nicolas, today; it's possible to calibrate a Kinect, with a configuration file, to personalize the calibration.

Secondly, Yifan has tested and validated an important part of my thesis. However, the result today is not as accurate as we would like. Several tracks are to be tested, including the generation of images orthoprojection, or refine the transformation computation.

Perspectives

The status of this work is still that it requires development.

Points to develop:

- Test to determinate why there is an error on the computed trajectory
- Work on the orthoprojection, find the better way to project point cloud
- Optimize the code to use it in real time
- Test strictly the code, and evaluate it

G

Étude de calibration d'une optique *Fish Eye*

Étudiant : Geoffrey Vincent, 5^{ème} année de l'INSA de Strasbourg

Intitulé du mémoire : Apport des caméras *Fish Eye* (hypergonnes) au positionnement en milieu urbain

Lieu : INSA de Strasbourg

Date de réalisation : février à juillet 2013



Résumé de projet de fin d'études

L'UTILISATION DE LA VISION en robotique mobile a toujours séduit les chercheurs du domaine à cause de l'analogie que l'on peut faire avec la façon dont l'humain se localise. Dans ce cas, les données acquises sont des images issues d'une caméra embarquée.

Ce projet de fin d'études est en partenariat avec la société Trimble. Cette société fait partie des leaders mondiaux dans le domaine de la topographie, en fabriquant GPS tachéomètres et laser scanner... Malgré la grande maîtrise actuelle de la méthode de positionnement par GPS, celle-ci relève toujours des problèmes concernant son bon fonctionnement ou sa précision en milieu urbain comme les obstructions du signal ou les multi-trajets. C'est pourquoi Trimble oriente une partie de ses recherches sur la localisation de mobiles en milieux urbains sans l'aide d'aucun GPS.

Ceci a conduit Trimble à financer, avec l'INSA de Strasbourg, la thèse de Marie-Anne Mittet intitulée : *Caméra 3D pour la localisation d'un système mobile en environnement urbain*. Les travaux de ce projet de fin d'études sont reliés avec cette thèse et doivent approfondir ce qu'il est possible de faire avec une lentille fish-eye pour aider le mobile à se déplacer.

Le mobile envisagé pour la thèse est un couplage entre une caméra 3D et une lentille fish-eye. Depuis que Microsoft a démocratisé les caméras 3D¹ avec sa célèbre Kinect, les applications dans le milieu scientifique ne cessent de fleurir. Pourtant l'idée de coupler une caméra 3D avec une lentille fish-eye pour du positionnement urbain semble être inédite dans la littérature. La lentille fish-eye fut adoptée dans ce projet pour son grand champ de vision².

Le matériel mis à disposition pour cette étude est une lentille fish-eye EF 8-15mm f/4L USM de Canon à projection à angle équisolide comme celle-ci 1 :

1. en anglais : *Time-of-flight (ToF) camera/Range Imaging (RIM)*
2. en anglais : *Field of View (FOV)*



FIGURE 1 – La lentille fish-eye EF 8-15mm f/4L USM de Canon.

Cette lentille est associée à une chambre canon EOS 5D mark II de résolution 4368×2912 pixels dont le capteur CCD fait $24 \times 36mm$. Le laboratoire de physique de l'INSA a aussi mis à disposition un banc optique afin d'optimiser les conditions de prises de vues. Les prises de vues sont effectuées avec une focale de $8mm$, un zoom à l'infini, un ISO de 100, une grande ouverture du diaphragme et un faible temps d'exposition pour éviter les effets de flou.

La caméra 3D est tournée vers le sol car la portée de ces caméras est faible (de l'ordre de $10m$). En étant tournée vers le sol la caméra obtient un nuage de points de tous les éléments au ras du sol comme les poteaux ou les marques de peinture. En théorie, il devrait être possible au mobile de se positionner seulement avec le nuage de points issu de la caméra 3D. Pourtant, il est possible pour un problème lié à l'environnement, si par exemple le mobile traverse une zone sans aucun élément à détecter, que le mobile perde la connaissance de son environnement. La lentille fish-eye devrait alors intervenir pour fournir les informations angulaires nécessaires au mobile pour qu'il puisse recalculer la bonne trajectoire. La lentille fish-eye est donc pensée tournée vers le haut pour recueillir le maximum d'informations possibles sur les façades ou bordures des toits comme illustré dans la figure 2.

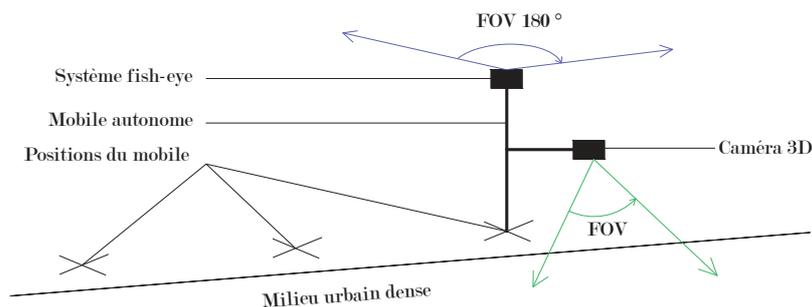


FIGURE 2 – Mobile robotique comprenant une caméra 3D et une caméra fish-eye.

La problématique générale est donc d'étudier ce qu'il est envisageable de faire avec une lentille fish-eye pour compléter en informations les données obtenues par la caméra

3D. Bien entendu, le mieux serait d'arriver à faire de la *structures de mouvement*³ en temps réel. Cet idéal vers lequel nous tendons sera étudié ainsi que ses limites techniques.

Faire de la photogrammétrie avec une fish-eye est un défi à cause principalement du fait qu'une fish-eye ne respecte pas le modèle sténopé basique. Selon la précision désirée, plusieurs modèles ont été proposés prenant en compte, plus ou moins fidèlement, la façon dont se déplace le faisceau de lumière pour constituer l'image. L'estimation de la valeur des paramètres de ce modèle pour une caméra est réalisée par calibrage.

Ce rapport propose donc dans une première partie une méthode de calibration novatrice pour les lentilles fish-eyes. Cette méthode permet d'obtenir les éléments suivants :

- La projection interne propre à la lentille (entre equisolide, orthographique, orthoscopique et stéréographique).
- La position du point principal sur le capteur CCD.
- Les coefficients de distorsions radiales et tangentiels (dues aux défauts de fabrication de la lentille) en accord avec le modèle de Duane C. Brown.

Pour la réalisation, une interface utilisateur a été développée sous le langage Matlab et une mire tridimensionnelle a été construite comme on voit sur la figure 3.

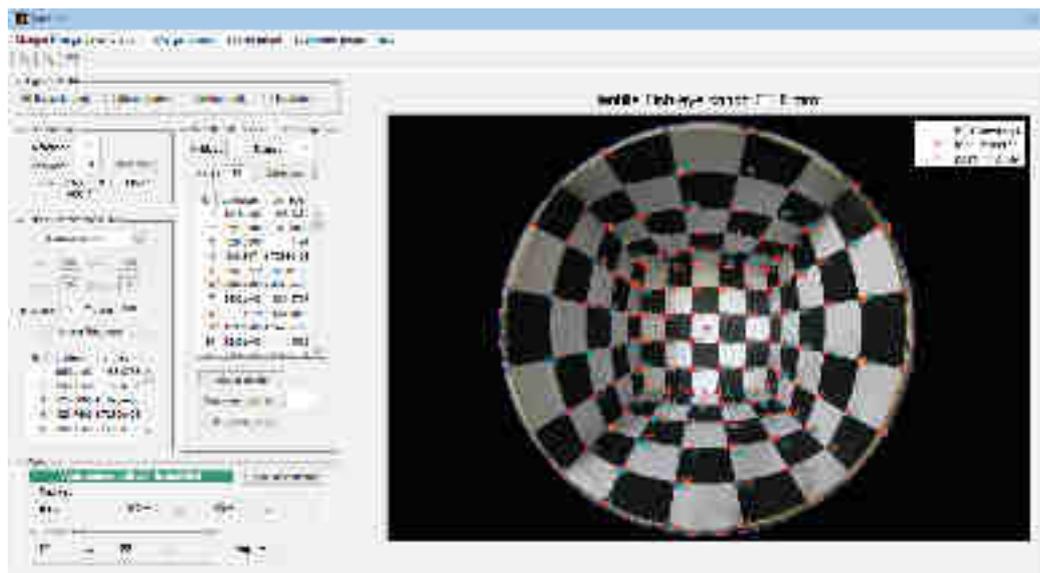


FIGURE 3 – Une photo de la mire 3D prise par la fish-eye et traitée par le GUI.

Le point principal est obtenu en appliquant un RANSAC sur une image à forte valeur d'exposition. Les coordonnées objets des points de la mire théorique sont calculées en amont, ce qui permet de calculer des coordonnées images théoriques en sélectionnant le modèle qui convient le mieux, puis les coordonnées images des points de la mire sont détectées dans la photo. Ceci permet de résoudre le modèle de Brown qui renvoie K_1 , K_2 et K_3 les coefficients de distorsion radiale et P_1 et P_2 les coefficients de distorsion tangentielle.

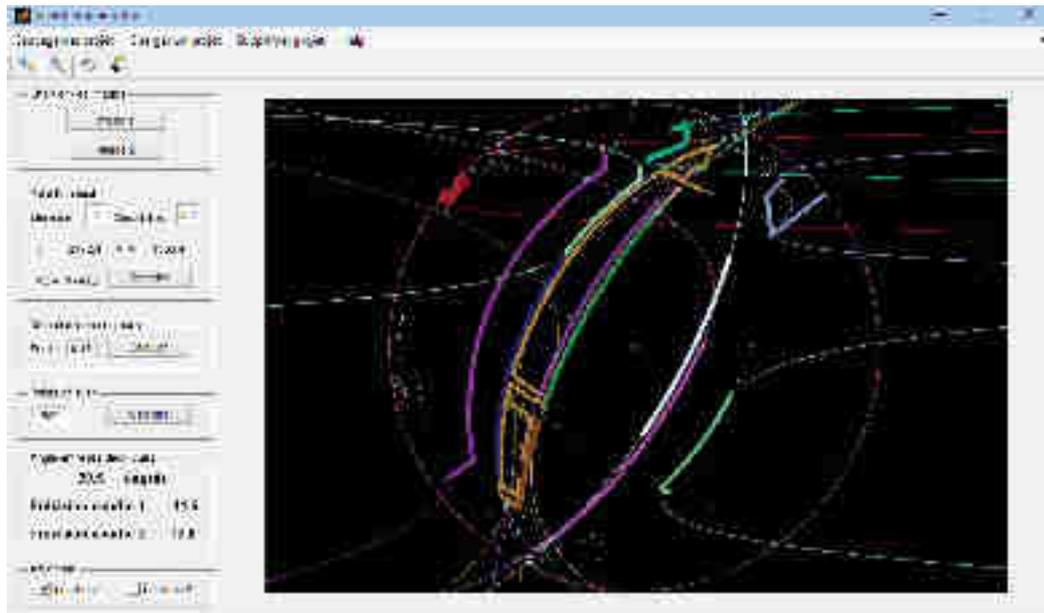
3. De l'anglais : *Structure from motion*

On peut ensuite passer à l'étude des structures de mouvement dans la deuxième partie du rapport. La structure de mouvement désigne le processus d'estimation de structures tridimensionnelles à partir de séquences d'images qui peuvent être couplées avec des données de mouvements locaux. Pour effectuer du SFM avec une lentille fish-eye, il faudrait effectuer une linéarisation sur son équation de transformation interne. La principale limitation provient de la complexité et de la non linéarité de la transformation. La recherche sur la linéarisation du système semblait être un problème intéressant à traiter et à explorer, mais malheureusement le temps alloué pour ce projet de fin d'études n'a pas permis d'aller au bout de ces recherches. Des méthodes alternatives ont donc été envisagées comme le modèle d'imagerie unifié afin que les images fish-eye soient reprojétées et donc conformes à la géométrie épipolaire. Une deuxième méthode envisagée est le positionnement grâce aux points de fuite détectés dans le champ de vision de la lentille.

Le modèle d'imagerie unifié⁴ fournit un cadre puissant pour envisager le passage entre différents types de caméras telles que la perspective standard et de nombreux types de fish-eye. Il est donc possible de recalculer les paramètres internes de cette caméra nouvellement calculée et qui est donc maintenant à projection centrale. Mais on constate aussi que la qualité de l'image résultante est aussi beaucoup dégradée compte tenu de ce que nous avons en entrée. On peut alors s'interroger sur l'intérêt d'utiliser cette caméra virtuelle car elle n'apporte pas plus d'information qu'une image qui aurait été prise par une lentille classique.

La dernière partie du rapport concerne la détection de points de fuite. La connaissance des points de fuites entre poses permettent d'obtenir les valeurs de certaines rotations. La méthode proposée ici consiste donc à extraire les contours dans une image segmentée automatiquement, on détermine ensuite les courbes passant au mieux par ces contours et on calcule les points d'intersection de ces courbes. En faisant cela dans deux images, on obtient alors deux couples de deux points de fuite dont les droites issues forment un angle. Cette algorithmique a aussi été implémentée dans un GUI comme illustré dans la figure 4 :

4. De l'anglais *Unified imaging*



(a)



(b)

(c)

(d)

FIGURE 4 – Sur 4(a) on voit le résultat de la détection et segmentation automatique des lignes de fuite par le GUI. Puis le résultat de la position des points de fuite sur différentes prises de vue 4(b) 4(c) et 4(d).

La méthode proposée ici n'est pas complètement automatique, l'utilisateur intervient surtout pour régler des paramètres optimaux des différents algorithmes : seuil dans la détection de Canny et nombre minimum de pixels pour qu'un nuage soit considéré comme une courbe.

La méthode générale fonctionne bien en milieu de laboratoire. Les clichés sont pris dans des positions favorables et il n'y a pas d'obstruction qui créerait des zones d'ombre. En milieu urbain, les configurations peuvent vite se compliquer avec les arbres ou le manque de luminosité. De plus la configuration des bâtiments importe beaucoup car ici nous avons

besoin de beaucoup de lignes de fuites parallèles entre elles. Ce programme sera tout de suite moins robuste en ville si les arrêtes des toits ne sont pas rectilignes.

Étant donné de ce que recherchent les industriels et du matériel devenant de moins en moins couteux (lentilles fish-eye comme caméras 3D), les études sur le calcul des structures de mouvement d'un mobile en temps réel devraient se multiplier. Ce rapport aura étudié les principales composante des lentilles hypergones tout en apportant des méthodes innovantes sur le plan de la calibration et de la détermination de valeurs de rotation entre poses.

Annexe

H

Études annexes sur l'apport d'une optique *Fish Eye*

Étudiant : Mikaël Royère, 5^{ème} année de l'INSA de Strasbourg

Intitulé du mémoire : Localisation en environnement urbain basée sur l'utilisation d'un objectif *Fish Eye*

Lieu : INSA de Strasbourg

Date de réalisation : février à juillet 2014

Localisation en environnement urbain basée sur l'utilisation d'un objectif *fish-eye*

PFE présenté par : **Mikael Royere**
Société d'accueil : **ICUBE**
Directeur de PFE : **Pierre Grussenmeyer**
Correcteur : **Eddie Smigiel**



1. Contexte et objectifs de l'étude

Le projet consiste à étudier les possibilités de localisation en milieu urbain à l'aide d'une seule caméra dont l'axe optique est dirigé vers le ciel. L'intérêt est de proposer une solution alternative aux dispositifs standards de localisation souvent basés sur l'utilisation d'un récepteur GPS couplé à une centrale inertielle. Le système proposé permet de s'affranchir des désagréments dont sont sujets les systèmes GPS tels que les masques et les multi-trajets altérant la précision de positionnement. Afin de couvrir un champ de vision maximum, un objectif grand angle (figure 1) dénommé *fish-eye* ou caméra hypergone est fixé sur la caméra. Le champ de vision de l'objectif utilisé dans l'étude est de 185° dans toutes les directions, soit supérieur à une demi sphère. (figure 2).



Figure 1 : Objectif *fish-eye* Fujinon fixé sur une caméra industrielle Matrix Vision



Figure 2 : Photographie prise devant l'INSA l'objectif dirigé vers le haut

Le sujet est à mi chemin entre deux disciplines: la photogrammétrie, discipline plus ancienne basée sur les équations de colinéarités et la vision par ordinateur, discipline apparue lors de l'explosion de l'image numérique qui s'intéresse particulièrement à la modélisation et la localisation en robotique. De nombreux algorithmes sont tirés de la deuxième discipline car l'utilisation des coordonnées homogènes couplées au calcul matriciel permet des développements aisés.

La détermination de la position d'un appareil photographique peut se décomposer en 3 étapes:

- La calibration du système réalisée en amont permettant de fixer les paramètres d'orientation interne (chapitre 2)
- La mise en correspondance d'entités communes entre deux images successives par détection de points d'intérêts. (chapitre 3)
- Le calcul de l'orientation externe de la caméra par rapport à un repère défini (chapitre 4)

2. Calibration du système

Un champ de vision aussi conséquent engendre des fortes distorsions qui courbent les lignes droites. Par conséquent, l'orientation relative de la caméra ne répond pas aux mêmes principes qu'un objectif standard. Il est donc nécessaire de réaliser une calibration particulière du dispositif. Deux types de calibration peuvent être réalisés:

- Une calibration que l'on peut qualifier de géométrique car elle utilise les caractéristiques de la lentille pour déterminer une équation de projection (figure 4). L'équation de projection permet le passage des coordonnées images aux coordonnées objets grâce à la liaison entre la distance radiale mesurée sur l'objectif et l'angle d'incidence entre le rayon de lumière mesuré et le rayon passant par le point principal. La détermination du point principal et des paramètres de distorsions est ensuite déterminée sur la base des travaux applicables aux objectifs standards. Cette méthode est précise mais la non linéarité de l'équation de projection pour le calcul des transformations limite son usage.

- Une calibration dite analytique développée pour les caméras omnidirectionnelles qui ont un champ de vision supérieur à une demi sphère. Ce type de calibration ne nécessite pas la connaissance de l'équation de projection car elle modélise la calibration par projection des points images sur une sphère virtuelle de rayon 1 à l'aide d'une fonction polynomiale. L'intérêt de cette calibration est la facilité de mise en œuvre et la compatibilité avec la plupart des algorithmes de transformations utilisés pour des objectifs standards. Cependant, la projection sur une sphère est moins précise que la première méthode utilisant l'équation de projection.



Figure 3 : Représentation sur une sphère virtuelle des coordonnées images

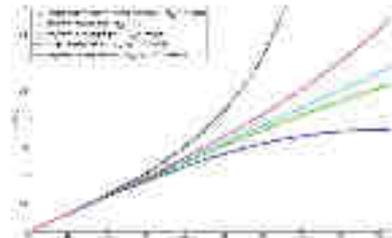


Figure 4 : Distance radiale en fonction de l'angle d'incidence pour chaque équation de projections

Pour les besoins de l'étude, les deux types de calibrations sont complémentaires, tandis que la deuxième calibration sert de base pour le filtrage et le calcul de la transformation, la première calibration sert à compenser la trajectoire et permet de modéliser en trois dimensions les points d'intérêts.

La calibration est une étape à ne pas négliger avant tout type de calcul car elle conditionne la précision du modèle. Les calibrations sont effectuées à l'aide d'une mire à quelques centimètres de l'objectif afin de couvrir au maximum la surface du capteur. Afin de contrôler la calibration dans des conditions de prises de vues identiques à l'utilisation ultérieure, une interface graphique a été mise au point. Elle permet de comparer les résidus sur les points d'appuis par détermination de la matrice caméra à l'aide de points d'appuis connus en coordonnées. L'interface graphique a permis de conforter les résultats obtenus avec les calibrations précédentes.

3. Étude des détecteurs de points d'intérêts

Le but est de limiter l'intervention de l'utilisateur dans le calcul de la trajectoire, une des étapes automatiques la plus importante est la détection d'entité commune entre deux images successives. Les entités peuvent être communes en terme de forme (détecteur de coins basé sur la détection de contours) et d'intensité, (détecteur basé sur la différence de gradient de couleur), il est même possible de détecter des similitudes sur des régions d'intérêts (détecteurs de régions d'intérêts). Un détecteur viable doit répondre à plusieurs caractéristiques: les plus importantes sont la répétabilité (un même point d'intérêt doit être détecté dans plusieurs images), la distinction (les caractéristiques doivent permettre de le différencier), la précision et la rapidité de détection. La description du point d'intérêt définit ses caractéristiques (facteur d'échelle, direction...) par rapport à ses plus proches voisins. Cette étape est essentielle car elle permet de différencier les points détectés lors de l'étape d'appariement.

Les méthodes de détection de points d'intérêts sont nombreuses et complémentaires. Les plus connues sont les détecteurs basés sur les contours *Harris*, *Susan*, *Fast*, *Shi Tomasi*... les méthodes basées sur les différences d'intensité *SIFT*, *SURF*, *PCA-SIFT*... et les méthodes basées sur les régions d'intérêts *MSER*, *IBR*...

Une comparaison des détecteurs appliqués à une image prise avec un objectif fish-eye a été réalisée. Le but étant de vérifier si les caractéristiques valables pour un objectif standard sont aussi valables pour un objectif fish-eye. Les résultats de l'étude et l'évaluation des détecteurs disponibles dans le logiciel matlab ont conduit au choix du détecteur SURF dans les travaux suivants.

4. Mise en place d'une solution de détermination de la trajectoire d'un mobile

Dans la littérature, plusieurs principes de localisation utilisant un seul objectif sont présentés, les plus connus sont *l'odométrie visuelle* et *la localisation et cartographie simultanée (VSLAM)*. L'odométrie visuelle est une méthode qui détermine la position d'un objectif image par image par utilisation de la contrainte épipolaire entre un couple d'images. L'intérêt de cette méthode est qu'elle peut être utilisée en temps réel et qu'elle n'est pas gourmande en ressource machine. Elle présente cependant quelques inconvénients puisque la précision de la détermination de la position de l'objectif dépend de la précision des calculs précédents, ceci induit une dérive du système au cours du

déplacement ou une erreur franche de la trajectoire si une erreur de convergence survient entre un couple d'images. Le principe VSLAM permet de réduire cette dérive par la détection de fermetures de boucles et l'utilisation d'un filtrage de Kalman. Il est cependant plus difficile à développer et le temps de calcul est accru. Par conséquent, l'odométrie visuelle est la technique utilisée dans l'étude.

L'algorithme de localisation est développé sur le logiciel matlab, l'étude a permis la mise à disposition d'un code source et d'une interface graphique (toolbox)(figure 5). L'algorithme peut être décomposé en 4 étapes : détection, filtrage, transformation et compensation.

Détection :

Afin de réduire le temps de traitement et de détecter une somme acceptable de points d'intérêts, le détecteur SURF implémenté dans Matlab a été utilisé, ce détecteur présente de nombreux avantages tels que sa vitesse d'exécution, la répétabilité, le nombre et la précision des points détectés. L'appariement des points détectés est réalisé à l'aide de la méthode des plus proches voisins implémentés également dans le logiciel de programmation.

Filtrage:

Une des étapes les plus délicates est la suppression des mises en correspondances erronées entre les images. L'algorithme itératif RANSAC est une solution très utilisée, il est cependant coûteux en temps lorsque le nombre d'inconnues est élevé. Dans notre étude, la caméra est toujours dirigée vers le ciel, le problème peut donc être ramené dans le plan afin de limiter le nombre d'inconnues à 3. De plus les caractéristiques du chariot permettent l'utilisation du centre instantané de rotation (ICR) qui supprime 2 inconnues supplémentaires. L'inconnue restante est l'angle ϑ correspondant à la rotation autour de l'ICR (figure 5). Il est ainsi possible de calculer l'angle ϑ pour chaque couple de points puis de tracer un histogramme des valeurs. La valeur maximale de l'histogramme correspond à la valeur la plus probable de ϑ , les mises en correspondance erronées correspondent aux valeurs dont la différence avec la valeur la plus probable est supérieure à un seuil fixé.



Figure 5 : Déplacement du chariot autour du centre instantané de rotation ICR (Scaramuzza)

Transformation:

Le calcul de transformation entre les différents couples d'images permet l'obtention de l'orientation externe, soit les trois paramètres de rotation et les trois paramètres de position du point principal. Le calcul de transformation est basé sur la contrainte épipolaire dont la représentation analytique est la matrice essentielle ou la matrice fondamentale suivant si les données sont normalisées ou non (prise en compte des paramètres internes de la caméra). La matrice essentielle est calculée à l'aide de l'algorithme à 8 points de Longuet-Higgins, ensuite les paramètres de translation et de rotation formant la matrice de transformation sont extraits de cette matrice. Enfin, la matrice de position est calculée à l'aide de la matrice de transformation et la matrice position précédente. Le calcul est réitéré dès qu'une image est ajoutée.

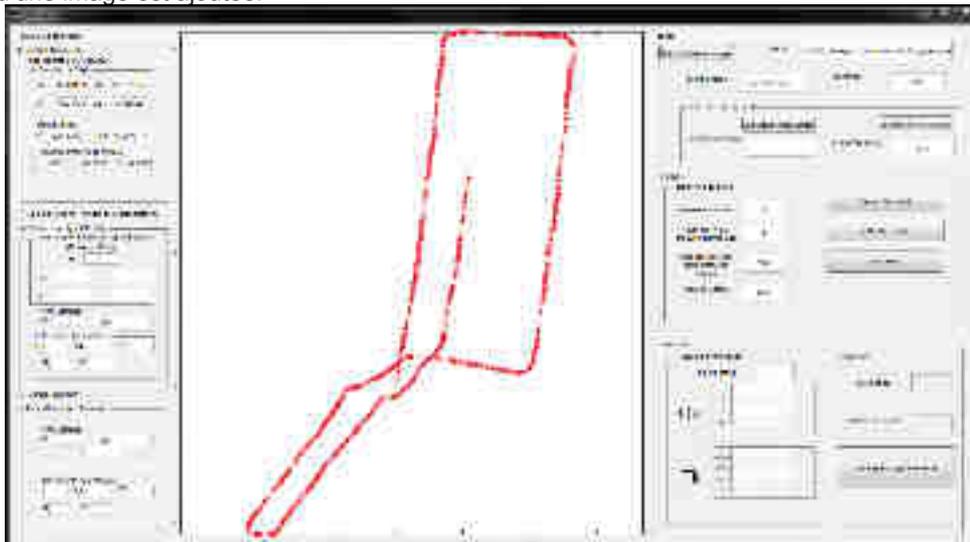


Figure 6 : Toolbox de trajectographie « TrajectoVo »

Compensation:

La compensation est réalisée à l'aide de la surabondance des données de transformation. Pour une position à calculer, l'intérêt est de définir une fenêtre avec plusieurs positions calculées précédemment. Ainsi, la position d'une image sera calculée plusieurs fois, le nombre d'images précédentes à prendre en compte est choisi suivant les possibilités de mise en correspondances entre les images les plus espacées. Les nombreuses transformations permettent d'effectuer une moyenne pondérée de la position en fonction de la convergence du calcul et de l'éloignement entre deux caméras. Ce mode opératoire permet de détecter si un calcul de transformation ne converge pas et de supprimer les aberrations.

5 Acquisition des données tests et évaluation du résultat

Afin d'évaluer qualitativement la localisation, des campagnes de mesures ont été réalisées dans différentes conditions de prise de vue, les campagnes ont permis l'obtention de 2 trajectoires en intérieur dans le laboratoire de topographie et 2 trajectoires en extérieur. Pour ces 4 campagnes de mesure, une trajectoire de comparaison a été mesurée à l'aide d'une station totale en mode tracking. Afin d'évaluer l'orientation du chariot et de montrer les faiblesses de la mesure GPS en milieu urbain, un récepteur GPS a été utilisé en complément des mesures tachéométriques lors d'une campagne de mesure en extérieur.

Localisation	Laboratoire de topographie (25/03)	Rue Gaspard Monge (09/05)	Laboratoire de topographie (15/07)	Avenue de la Paix (17/07)
Nombre de clichés	1164	792	1560	1480
Distance parcourue	40m	90m	62m	290m
Moyens d'acquisition supplémentaires	Station totale	Station totale	2 stations totales ('position+ orientation)	2 stations totales 1 récepteur GPS

Différents tests ont été effectués en faisant varier les paramètres de calcul tels que le nombre d'images précédentes utilisées pour calculer la transformation ou les paramètres de détection. L'étude a montré que l'augmentation de la distance entre deux prises de vues augmente la précision sur la trajectoire lorsque la rotation est faible (il est donc préférable de laisser un intervalle entre deux images à calculer). Cependant, cette méthode n'est pas applicable lorsque le chariot effectue une rotation entre deux prises de vues consécutives. Il serait intéressant de faire varier cet intervalle entre deux calculs de transformation en rapport avec la rotation.

6 Conclusion et perspectives

Les deux premières études consistant à la calibration et la comparaison des détecteurs ont permis l'obtention d'une base solide pour l'étude principale visant à déterminer la trajectoire d'un objectif fish-eye. L'étude sur la calibration a fait ressortir deux moyens de calibrer un objectif fish-eye présentant chacun leurs avantages et leurs inconvénients. Une interface graphique de contrôle de la calibration a été développée, les travaux sur la comparaison des détecteurs ont permis d'évaluer l'incidence de l'utilisation d'un objectif fish-eye. La comparaison a été faite sur la base de l'homographie entre deux images. Afin de déterminer l'homographie, une interface graphique a été créée. La calibration et la comparaison des détecteurs de points d'intérêts ont permis l'obtention d'une base de calcul solide pour l'étude principale.

Un outil permettant la localisation d'un objectif fish-eye a été réalisé. L'allure de la trajectoire calculée est en adéquation avec les données mesurées par tachéométrie. Il est cependant difficile de comparer la trajectoire précisément car l'utilisation d'un seul objectif implique un calcul de trajectoire à un facteur d'échelle près.

En perspective, il serait nécessaire de rendre le calcul plus robuste aux aléas de mesures et continuer le travail entrepris sur la compensation du système. Le problème du facteur d'échelle en distance est un frein qui pourrait être éliminé en couplant avec un autre système de mesure. Enfin, il pourrait être intéressant d'appliquer les principes utilisés en VSLAM car le champ de vision élevé de la caméra permet des recouvrements importants, donc des possibilités de détection de fermetures de boucles importantes. Afin de comparer précisément les données, il serait nécessaire de coupler les mesures tachéométriques et photographiques dans le même référentiel temps. Enfin, le système est destiné à réaliser une localisation relative, un autre moyen doit permettre la localisation absolue de la trajectoire. Ce moyen devrait permettre de compenser une nouvelle fois la trajectoire calculée.



Bibliographie

- Angeli, A., Filliat, D., Doncieux, S., & Meyer, J. A. (2008). Fast and incremental method for loop-closure detection using bags of visual words. *IEEE Transactions on Robotics*, 24(5), 1027–1037. *Cité page 162*
- Arieli, Y., Freedman, B., Machline, M., & Shpunt, A. (2010). Depth mapping using projected patterns. U.S. Patent No. 8,150,142. Washington, DC : U.S. Patent and Trademark Office. *3 citations pages 4, 37, et 52*
- Arpin, C. (2003). Global Positioning System (GPS) errors and limitations for vehicle tracking. Technical report, Boomerang Tracking Inc, Canada. *Cité page 13*
- Arya, S., Mount, D. M., Netanyahu, N. S., Silverman, R., & Wu, A. Y. (1998). An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM. JACM*, 45(6), 891–923. *Cité page 103*
- Baker, S. & Matthews, I. (2003). Lucas-Kanade 20 years on : a unifying framework . *International journal of Computer Vision*, 56(3), 221–255. *Cité page 104*
- Bakstein, H. & Pajdla, T. (2002). Panoramic mosaicing with a 180 field of view lens. In *Proceedings of the IEEE Workshop on Omnidirectional Vision 2002. Held in conjunction with ECCV'02*, (pp. 60–67). *4 citations pages 70, 71, 72, et 165*
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006). Surf : Speeded up robust features. In *Proceedings of the European Conference on Computer Vision. ECCV'06*, (pp. 404–417). Heidelberg Springer Berlin. *5 citations pages 93, 97, 103, 115, et 116*
- Bernard, N. (2011). Caméra 3D - Améliorations métrologiques par calibration et post-traitement. Projet de Fin d'étude, INSA de Strasbourg. *Cité page 111*
- Besl, P. J. & McKay, N. D. (1992). Method for registration of 3-D shapes. *IEEE transactions on pattern analysis and machine intelligence*, 14(2), 239–256. *Cité page 181*
- Blender (1995). Blender - Free and open source 3D animation suite. <http://www.blender.org>. *Cité page 123*
- Borenstein, J., Everett, H. R., & Liqiang, F. (1996). Where am I? Sensors and methods for mobile robot positioning. *University of Michigan*, 119(120), 15. *Cité page 26*
- Borenstein, J. & Feng, L. (1996). Gyrodometry : A New Method for Combining Data from Gyros and Odometry in Mobile Robots. In *Proceedings of IEEE International Conference on Robotics and Automation (vol. 1)*, (pp. 423–428). *Cité page 30*

- Bouguet, J. Y. (2001). Pyramidal Implementation of the Lucas Kanade Feature Tracker Description of the algorithm. Technical report. *Cité page 104*
- Brown, D. C. (1971). Close-range camera calibration. *Photogrammetric Engineering*, 37, 855–866. *2 citations pages 69 et 70*
- Calonder, M., Lepetit, V., Strecha, C., & Fua, P. (2010). BRIEF : Binary Robust Independent Elementary Features. In *Proceedings of the European Conference on Computer Vision. ECCV'10*, (pp. 778–792). Springer Berlin Heidelberg. *2 citations pages 93 et 98*
- Chapoulie, A., Rives, P., & Filliat, D. (2004). A spherical representation for efficient visual loop closing. In *Computer Vision Workshops. ICCV'04*, (pp. 335–342). *Cité page 32*
- Civera, J., Grasa, O. G., Davison, A. J., & Montiel, J. M. M. (2009). 1-point RANSAC for EKF-based Structure from Motion. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS'09*, (pp. 3498–3504). *Cité page 31*
- Cox, D. B. (1978). Integration of GPS with inertial navigation systems. *Navigation*, 25(2), 236–245. *Cité page 27*
- Cramer, M. (1997). GPS/INS integration. In *Proceedings of Photogrammetric Week'97*, (pp. 3–12). *2 citations pages 22 et 169*
- Cutler, R., Rui, Y., Gupta, A., Cadiz, J. J., Tashev, I., He, L. W., & Silverberg, S. (2002). Distributed meetings : A meeting capture and broadcasting system. In *In Proceedings of the 10th ACM international conference on Multimedia*, (pp. 503–512). *3 citations pages 64, 65, et 165*
- Dario Piatti (2010). *Time-of-Flight cameras : tests, calibration and multi-frame registration for automatic 3D object reconstruction*. PhD thesis, Politecnico di Torino. *Cité page 52*
- David (2015). DAVID - 3D Scanner. <http://www.david-3d.com/en>. *Cité page 53*
- Dewitt, B. A. (1996). Initial approximations for the three-dimensional conformal coordinate transformation. *Photogrammetric engineering and remote sensing*, 62(1), 79–84. *Cité page 58*
- Duquenne, F., Botton, S., Peyret, F., Bétaille, D., & Willis, P. (2005). *Les GPS : localisation et navigation par satellites*. Lavoisier. *Cité page 12*
- Durrant-Whyte, H. & Bailey, T. (2006). Simultaneous localization and mapping : part I. *Robotics & Automation Magazine*, 13(2), 99–110. *Cité page 31*
- Dusha, D. & Mejias, L. (2010). Attitude observability of a loosely-coupled gps/visual odometry integrated navigation filter. In *Proceedings of Australasian Conference on Robotics and Automation. ACRA'10*. Gordon Wyeth and Ben Upcroft. *Cité page 6*
- Eade, E. & Drummond, T. (2006). Scalable Monocular SLAM. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, (pp. 469–476). *Cité page 31*

- Eberhardt, H., Klumpp, V., & Hanebeck, U. D. (2010). Density Trees for Efficient Nonlinear State Estimation. In *Proceedings of 13th Conference on Information Fusion. FUSION. IEEE*, (pp. 1–8). IEEE. *Cité page 103*
- Ellekilde, L. P., Huang, S., Valls Miro, J., & Dissanayake, G. (2007). Dense 3D map construction for indoor search and rescue. *Journal of Field Robotics*, 24(1-2), 71–89. *Cité page 37*
- Faugeras, O. (1993). *Three-dimensional computer vision : A Geometric Viewpoint* (MIT Press ed.). MIT Press. *Cité page 67*
- Fischler, M. A. & Bolles, R. C. (1981). Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395. *Cité page 107*
- Friedman, J. H., Bentley, J. L., & Finkel, R. A. (1977). An Algorithm for Finding Best Matches in Logarithmic Expected Time. *ACM Transactions on Mathematical Software (TOMS)*, 3(3), 209–226. *Cité page 103*
- Furukawa, Y. & Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8), 1362–1376. *Cité page 32*
- Galileo (2003). GALILEO Design Consolidation. Technical report. *Cité page 175*
- Geyer, C. & Daniilidis, K. (2000). A unifying theory for central panoramic systems and practical implications. In *Proceedings of the European Conference on Computer Vision. ECCV'00*, (pp. 445–461). *Cité page 71*
- Ghilani, C. D. (2010). *Adjustment Computations – Spatial Data Analysis*. Wiley Series in Surveying and Boundary Control. *Cité page 112*
- Grewal, M. S., Weill, L. R., & Andrews, A. P. (2007). *Global Positioning Systems, Inertial Navigation, and Integration*. New York, USA : John Wiley & Sons, Inc. *Cité page 27*
- Grisetti, G., Grzonka, S., Stachniss, C., Pfaff, P., & Burgard, W. (2007). Efficient Estimation of Accurate Maximum Likelihood Maps in 3D. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS'07*, (pp. 3472–3478). *Cité page 32*
- Gschwandtner, M., Kwitt, R., Uhl, A., & Pree, W. (2011). Blensor : Blender Sensor Simulation Toolbox. In *Proceedings of the 7th International conference on Advances in visual computing.*, (pp. 199–208). Springer Berlin Heidelberg. *3 citations pages 99, 105, et 122*
- Guizzo, E. (2011). How Google's Self-Driving Car Works. <http://spectrum.ieee.org/automaton/robotics/artifi>. *Cité page 3*
- Hahnel, D., Triebel, R., Burgard, W., & Thrun, S. (2003). Map building with mobile robots in dynamic environments. In *Proceedings of IEEE International Conference on Robotics and Automation. ICRA'03*, (pp. 1557–1563). *Cité page 11*

- Harris, C. & Stephens, M. (1988). A combined corner and edge detector. In *Proceedings of Alvey Vision Conference*, volume 15, (pp.50). *3 citations pages 94, 95, et 115*
- Harris, C. G. & Pike, J. M. (1987). 3D Positional Integration from Image Sequences. *Image and Vision Computing*, 6(2), 87–90. *5 citations pages 84, 93, 94, 98, et 166*
- Hartley, R. & Zisserman, A. (2003). *Multiple view geometry in computer vision*. (Cambridge ed.). Cambridge university press. *4 citations pages 67, 117, 118, et 165*
- Ho, K. L. & Newman, P. (2007). Detecting loop closure with scene sequences. *International Journal of Computer Vision*, 74(3), 261–286. *2 citations pages 30 et 162*
- Horn, B. K., & Schunck, B. G. (1981). Determining optical flow. In *Proceedings of 1981 Technical Symposium East. International Society for Optics and Photonics.*, (pp. 319–331). *Cité page 103*
- Hullo, J.-F. (2009). Étude de la technique de Corrélation Epipolaire Dense (CED), application au chantier archéologique de Kilwa, Arabie Saoudite. Projet de Fin d'étude, INSA de Strasbourg. *Cité page 117*
- Iborra, A., Pastor, J. A., Alvarez, B., Fernandez, C., & Merono, J. M. F. (2003). Robots in radioactive environments. *IEEE Robotics and Automation Magazine*, 10(4), 12–22. *Cité page 3*
- Ioannides, R. T. & Strangeways, H. J. (2002). Improved ionospheric correction for dual frequency GPS. In *Proceeding of the 27th General Assembly of IUGG*, number 1, (pp. 4–7). *Cité page 14*
- Janoch, A., Karayev, S., Jia, Y., Barron, J. T., Fritz, M., Saenko, K., & Darrell, T. (2013). A Category-Level 3-D Object Dataset : Putting the Kinect to Work. In *Proceedings of the IEEE International Conference on Computer Vision. ICCV'11*, (pp. 141–165). *Cité page 53*
- Jazwinski, A. H. (1970). *Stochastic processes and filtering theory*. Academic Press, New York. *Cité page 31*
- Kalman, R. (1960). A new approach to linear filtering and prediction problems. *Journal of Fluids Engineering*, 82(1), 35–45. *2 citations pages 27 et 177*
- Kelly, A. (1994). Modern inertial and satellite navigation systems. Technical report, Robotics Institute Carnegie Mellon University. *Cité page 18*
- Kelly, A. (2013). *Mobile Robotics : Mathematics, Models, and Methods*. Cambridge University Press. *Cité page 18*
- Kim, J. & Sukkarieh, S. (2005). 6DoF SLAM aided GNSS/INS navigation in GNSS denied and unknown environments. *Journal of Global Positioning Systems*, 4(1-2), 120–128. *Cité page 6*
- King, A. D. (1998). Inertial Navigation – Forty Years of Evolution. *GEC review*, 13(3), 140–149. *2 citations pages 19 et 164*

- Kneip, L., Chli, M., & Siegwart, R. (2011). Robust Real-Time Visual Odometry with a Single Camera and an IMU. In *Proceedings of British Machine Vision Conference. BMVC'11*, volume 23, (pp. 1–11). *Cité page 108*
- Kolb, A., Barth, E., & Koch, R. (2008). ToF-Sensors : New Dimensions for Realism and Interactivity. In *Proceedings of Computer Vision and Pattern Recognition Workshops. IEEE Computer Society Conference.*, (pp. 1–6). *Cité page 36*
- Lachat, E., Macher, H., Mittet, M.-a., Landes, T., & Grussenmeyer, P. (2015). First Experiences With Kinect V2 Sensor for Close Range 3D Modelling. *ISPRS - International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, XL-5/W4*(February), 93–100. *Cité page 161*
- Landes, T. (2011). Calculs d'erreurs et de compensation I. INSA de Strasbourg. *Cité page 113*
- Landes, T. & Grussenmeyer, P. (2011). Les principes fondamentaux de la lasergrammétrie terrestre : systèmes et caractéristiques (partie 1/2). *Revue XYZ de l'Association Française de Topographie (AFT)*, 128, 37–49. *Cité page 36*
- Leonard, J. & Newman, P. (2003). Consistent, Convergent, and Constant-Time SLAM. In *Proceedings of International Joint Conference on Artificial Intelligence. IJCAI'03*, (pp. 1143–1150). *Cité page 31*
- Leonard, J. J. & Durrant-Whyte, H. F. (1991). Mobile robot localization by tracking geometric beacons. *IEEE Transactions on Robotics and Automation*, 7(3), 376–382. *Cité page 10*
- Liu, B., Adams, M., & Ibanez-Guzmán, J. (2005). Multi-aided Inertial Navigation for Ground Vehicles in Outdoor Uneven Environments. In *Proceedings of IEEE International Conference on Robotics and Automation*, (pp. 4703–4708). *Cité page 30*
- Longuet-Higgins, H. (1987). A computer algorithm for reconstructing a scene from two projections. In *Readings in Computer Vision : Issues, Problems, Principles, and Paradigms* (pp. 61–62). *3 citations pages 84, 108, et 118*
- Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. *3 citations pages 93, 96, et 103*
- Lucas, B. & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of International Joint Conference on Artificial Intelligence. IJCAI'81*, (pp. 674–679). *2 citations pages 103 et 104*
- Maimone, M., Cheng, Y., & Matthies, L. (2007). Two Years of Visual Odometry on the Mars Exploration Rovers. *Journal of Field Robotics*, 24(3), 169–186. *2 citations pages 3 et 84*
- Malicorne, M., Bousquet, M., Calmettes, V., & Macabiau, C. (2001). Effects of masking angle and multipath on Galileo performances in different environments. In *Proceedings of the 8th Saint Petersburg International Conference on Integrated Navigation Systems*, (pp. 271–280). *2 citations pages 15 et 169*

- Matas, J., Chum, O., Urban, M., & Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and vision computing*, 22(10), 761–767. Cité page 115
- Matthies, L. & Shafer, S. A. (1987). Error modelling in stereo navigation. *IEEE Journal of Robotics and Automation*, 3(3), 239–248. Cité page 84
- McGlone, J. C. (2013). *Manual of photogrammetry, 6th ed.* American society for photogrammetry and remote sensing (ASPRS). Cité page 117
- Mikolajczyk, K. & Schmid, C. (2004). Scale & affine invariant interest point detectors. *International journal of computer vision*, 60(1), 63–86. 3 citations pages 93, 94, et 166
- Mikolajczyk, K. & Schmid, C. (2005). Performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10), 1615–1630. Cité page 98
- Miksik, O. & Mikolajczyk, K. (2012). Evaluation of local detectors and descriptors for fast feature matching. In *Proceedings of the 21st International Conference on Pattern Recognition. ICPR'12*, (pp. 2681–2684). Cité page 98
- Moravec, H. P. (1980). *Obstacle avoidance and navigation in the real world by a seeing robot rover.* PhD thesis, Stanford. 4 citations pages 32, 84, 93, et 94
- Moré, J. J. (1978). The Levenberg-Marquardt algorithm : implementation and theory. In *Numerical analysis* (pp. 105–116). Cité page 70
- Muja, M. & Lowe, D. (2009). Fast Approximate Nearest Neighbors with Automatic Algorithm Configuration. In *Proceedings of Computer Vision and Applications. VISAPP'09*, volume 2, (pp. 331–340). 3 citations pages 103, 105, et 115
- Nahimana, D. F. (2009). *Impact des multitrajets sur les performances des systèmes de navigation par satellite : contribution à l'amélioration de la précision de localisation par modélisation bayésienne.* PhD thesis, Ecole Centrale de Lille. 3 citations pages 15, 16, et 164
- Nassar, S. (2003). *Improving the inertial navigation system (INS) error model for INS and INS/DGPS applications.* PhD thesis, University of Calgary. Cité page 22
- Newcombe, R. A., Davison, A. J., Izadi, S., Kohli, P., Hilliges, O., Shotton, J., & Fitzgibbon, A. (2011). KinectFusion : Real-time dense surface mapping and tracking. *Proceedings of the 10th IEEE International symposium Mixed and augmented reality. ISMAR'11*, 127–136. Cité page 53
- Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6), 756–77. 2 citations pages 85 et 118
- Nistér, D., Naroditsky, O., & Bergen, J. (2004). Visual odometry. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, volume 1, (pp. 652–659). 3 citations pages 84, 85, et 108

- Ohno, K., Tsubouchi, T., Shigematsu, B., Maeyama, S., & Yuta, S. I. (2003). Outdoor navigation of a mobile robot between buildings based on DGPS and odometry data fusion. In *Proceedings of IEEE International Conference on Robotics and Automation*, (pp. 1978–1984). Cité page 30
- Olson, C. F., Matthies, L. H., Schoppers, M., & Maimone, M. W. (2000). Robust Stereo Ego-motion for Long Distance Navigation. In *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition*, (pp. 453–458). Cité page 85
- Olson, C. F., Matthies, L. H., Schoppers, M., & Maimone, M. W. (2003). Rover navigation using stereo ego-motion. *Robotics and Autonomous Systems*, 43(4), 215–229. Cité page 85
- Olynik, M. C. (2002). Temporal characteristics of GPS error sources and their impact on relative positioning. Technical report, University of Calgary. Cité page 13
- Ortin, D. & Montiel, J. M. M. (2001). Indoor robot motion based on monocular images. *Robotica*, 19(3), 331–342. 2 citations pages 30 et 108
- Oskiper, T., Chiu, H.-P., Zhu, Z., Samaresekera, S., & Kumar, R. (2011). Stable vision-aided navigation for large-area augmented reality. In *IEEE Virtual Reality Conference (VR)*, (pp. 63–70). Cité page 84
- Puig, L., Bermúdez, J., Sturm, P., & Guerrero, J. J. (2012). Calibration of omnidirectional cameras in practice : A comparison of methods. *Computer Vision and Image Understanding*, 116(1), 120–137. Cité page 70
- Remondino, F. & Stoppa, D. (2013). *ToF range-imaging cameras*. 68121 Heidelberg, Germany : Springer. Cité page 33
- Rosten, E. & Drummond, T. (2006). Machine learning for high-speed corner detection. In *Proceedings of the European Conference on Computer Vision. ECCV'06*, number Springer Berlin Heidelberg, (pp. 430–443). Springer Berlin Heidelberg. 2 citations pages 93 et 98
- Royère, M. (2014). Localisation en environnement urbain basée sur l'utilisation d'un objectif fish-eye. Projet de Fin d'étude, INSA de Strasbourg. 4 citations pages 72, 73, 114, et 163
- Rublee, E., Rabaud, V., Konolige, K., & Bradski, G. (2011). ORB : An efficient alternative to SIFT or SURF. In *Proceedings of IEEE International Conference on Computer Vision. ICCV'11*, (pp. 2564–2571). 4 citations pages 93, 97, 98, et 105
- Samet, H. & Webber, R. E. (1985). Storing a collection of polygons using quadtrees. *ACM Transactions on Graphics*, 4(3), 182–222. Cité page 103
- Scaramuzza, D. (2007). *OCamCalib : Omnidirectional Camera Calibration Toolbox for Matlab*. 2 citations pages 71 et 72
- Scaramuzza, D. (2008). *Omnidirectional vision : from calibration to robot motion estimation*. PhD thesis, INRIA Sophia Antipolis. 2 citations pages 71 et 74

- Scaramuzza, D. (2011). 1-point-ransac structure from motion for vehicle-mounted cameras by exploiting non-holonomic constraints. *International journal of computer vision*, 95(1), 74–85. 2 citations pages 85 et 108
- Scaramuzza, D. & Fraundorfer, F. (2011). Visual Odometry : Part I - The First 30 Years and Fundamentals. *IEEE Robotics and Automation Magazine*, 18(4), 80–92. 2 citations pages 87 et 166
- Scaramuzza, D., Fraundorfer, F., & Siegwart, R. (2009). Real-time monocular visual odometry for on-road vehicles with 1-point RANSAC. In *Proceedings of IEEE International Conference on Robotics and Automation*, (pp. 4293–4299). Ieee. 6 citations pages 10, 32, 84, 108, 111, et 116
- Schwarz, K. P., Chapman, M. A., Cannon, M. W., & Gong, P. (1993). An integrated INS/GPS approach to the georeferencing of remotely sensed data. *Photogrammetric engineering and remote sensing*, 59(11), 1667–1674. Cité page 1
- Se, S., Lowe, D., & Little, J. (2001). Vision-based mobile robot localization and mapping using scale-invariant features. In *Proceedings of IEEE International Conference on Robotics and Automation*, volume 2, (pp. 2051–2058). Ieee. Cité page 96
- Shi, J. & Tomasi, C. (1994). Good features to track. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR'94*, volume 94, (pp. 593–600). 4 citations pages 93, 94, 95, et 115
- Siegwart, R., Nourbakhsh, I. R., & Scaramuzza, D. (2011). *Introduction to autonomous mobile robots (second edition)*. MIT press. 2 citations pages 24 et 25
- Skrypnik, I. & Lowe, D. G. (2004). Scene Modelling , Recognition and Tracking with Invariant Image Features. In *Proceedings of IEEE and ACM International Symposium on Mixed and Augmented Reality. ISMAR'04*, (pp. 110–119). Cité page 96
- Soutschek, S., Penne, J., Hornegger, J., & Kornhuber, J. (2008). 3-D gesture-based scene navigation in medical imaging applications using time-of-flight cameras. In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops. CV-PRW'08*, (pp. 1–6). Cité page 37
- Stovall, S. H. (1997). Basic inertial navigation. Technical report, NAWCWPNS TM 8128. Naval Air Warfare Center Weapons Division. Cité page 18
- Strang, G. & Borre, K. (1997). *Linear Algebra, Geodesy, and Gps*. Siam. Cité page 112
- Tardif, J. P., Pavlidis, Y., & Daniilidis, K. (2008). Monocular visual odometry in urban environments using an omnidirectional camera. In *Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems. IROS'08*, (pp. 2531–2538). Cité page 85
- Tassinari, R. (2003). *Pratique de l'analyse fonctionnelle (3ème édité ed.)*. Cité page 40
- Thrun, S. (2002). Robotic Mapping : A Survey. *Exploring artificial intelligence in the new millennium*, 1–35. 2 citations pages 11 et 31

- Titterton, D. & Weston, J. (2004). *Strapdown inertial navigation technology*. Cité page 18
- Tomasi, C. & Kanade, T. (1991). *Detection and Tracking of Point Features*. Number April. Pittsburgh. 2 citations pages 104 et 105
- Triggs, B., McLauchlan, P. F., Hartley, R. I., & Fitzgibbon, A. W. (2000). Bundle adjustment—a modern synthesis. In *Vision algorithms : theory and practice*, volume 34099 (pp. 298–372). Springer Berlin Heidelberg. Cité page 32
- Vincent, G. (2013). Apport des caméras Fish-eye (hypergonies) au positionnement en milieu urbain. Projet de Fin d'étude, INSA de Strasbourg. 3 citations pages 70, 163, et 165
- Weingarten, J. W., Gruener, G., & Siegwart, R. (2004). A State-of-the-Art 3D Sensor for Robot Navigation. In *Proceedings of IEEE International Conference on Intelligent Robots and Systems. IROS'04*, number 3, (pp. 2155–2160). Cité page 37
- Wolf, D. F. & Sukhatme, G. S. (2005). Mobile robot simultaneous localization and mapping in dynamic environments. *Autonomous Robots*, 19(1), 53–65. Cité page 11
- Woodman, O. J. (2007). An introduction to inertial navigation. *University of Cambridge, Computer Laboratory, Tech. Rep. UCAMCL-TR-696*, 14, 15. Cité page 21
- Yang, Y. (2012). Algorithme de localisation par caméra 3D. Projet de Fin d'étude, Télécom St Etienne. 3 citations pages 51, 88, et 162
- Ying, X. & Hu, Z. (2004). Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model. In *Proceedings of the European Conference on Computer Vision. ECCV'04*, (pp. 442–455). Cité page 71
- Yoo, H. J. (2011). *Analyse et conception de scanners laser mobiles dédiés à la cartographie 3D d'environnements urbains*. PhD thesis, École Nationale Supérieure des Mines de Paris. 2 citations pages 2 et 164

Marie-Anne MITTET

Caméras 3D pour la localisation d'un système mobile en environnement urbain

Résumé :

L'objectif de la thèse est de développer un nouveau système de localisation mobile composé de trois caméras 3D de type Kinect et d'une caméra additionnelle de type *Fish Eye*. La solution algorithmique est basée sur l'odométrie visuelle et permet de calculer la trajectoire du mobile en temps réel à partir des données fournies par les caméras 3D. L'originalité de la méthodologie réside dans l'exploitation d'orthoimages créées à partir des nuages de points acquis en temps réel par les trois caméras. L'étude des différences entre les orthoimages successives acquises par le système mobile permet d'en déduire ses positions successives et d'en calculer sa trajectoire.

Mots-clés : Localisation, caméras 3D, orthoimages, caméra Fish-Eye, odométrie visuelle, robotique, géomatique.

Résumé en anglais :

The aim of the thesis is to develop a new kind of localization system, composed of three 3D cameras such as Kinect and an additional Fisheye camera. The localization algorithm is based on Visual Odometry principles in order to calculate the trajectory of the mobile platform in real time from the data provided by the 3D cameras. The originality of the processing method lies within the exploitation of orthoimages processed from the point clouds that are acquired in real time by the three cameras. The positions and trajectory of the mobile platform can be derived from the study of the differences between successive orthoimages.

Keywords : Localization, 3D cameras, orthoimages, Fish Eye camera, visual odometry, robotics, geomatics.