



HAL
open science

Estimation continue de la pose d'un équipement tenu en main par fusion des données visio-inertielles pour les applications de navigation piétonne en milieux urbains

Nicolas Antigny

► To cite this version:

Nicolas Antigny. Estimation continue de la pose d'un équipement tenu en main par fusion des données visio-inertielles pour les applications de navigation piétonne en milieux urbains. Sciences de l'ingénieur [physics]. Université Loire Bretagne, 2018. Français. NNT : . tel-01935037v1

HAL Id: tel-01935037

<https://hal.science/tel-01935037v1>

Submitted on 26 Nov 2018 (v1), last revised 6 Dec 2018 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

THESE DE DOCTORAT DE

L'ÉCOLE CENTRALE DE NANTES
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
*Mathématiques et Sciences et Technologies
de l'Information et de la Communication*
Spécialité : *Signal, Image, Vision*

Par

Nicolas ANTIGNY

**Estimation continue de la pose d'un équipement tenu en main par
fusion des données visio-inertielles pour les applications de navigation
piétonne en milieux urbains**

Thèse présentée et soutenue à Nantes, le 18 octobre 2018

Unités de recherche :

Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux
(IFSTTAR)

Institut de Recherche en Sciences et Technique de la Ville (IRSTV – Centrale Nantes)

Rapporteurs avant soutenance :

Alain Lambert	Maître de conférences, HDR, Université Paris-Sud
Pascal Vasseur	Professeur des universités, IUT de Rouen

Composition du Jury :

Président et rapporteur Pascal Vasseur	Professeur des universités, IUT de Rouen
Rapporteur Alain Lambert	Maître de conférences, HDR, Université Paris-Sud
Directeur de thèse Valérie Renaudin	Directrice de recherche, IFSTTAR
Co-encadrant de thèse Myriam Servières	Maître de conférences, Ecole Centrale de Nantes

Remerciements

En premier lieu, je tiens à exprimer ma gratitude à Valérie Renaudin et Myriam Servières, respectivement directrice et co-encadrante de thèse, pour leur confiance et leur encadrement au cours de ces trois années.

Je tiens également à remercier Muriel Pressigout et Guillaume Moreau, membres de mon comité de suivi de thèse, pour leur disponibilité et leur suivi de mes travaux.

Je remercie Alain Lambert et Pascal Vasseur d'avoir accepté de relire et de juger ces travaux de thèse en tant que rapporteur.

J'adresse une pensée particulière aux membres de l'équipe GEOLOC, avec qui j'ai partagé ces trois années de doctorat. Je remercie donc les doctorants Christophe Combettes, Taia Alaoui, Mahdi Abid et Johan Perul pour les nombreuses discussions que nous avons pu avoir ; Miguel Ortiz pour son aide sur les questions techniques ; Aigul Khamitova, Céline Ragoïn, Mathieu De Sousa, Gilles Le Roux et Julien Le Scornec pour la bonne humeur qu'ils ont apporté au laboratoire.

Je tiens à saluer Jean-Yves Martin, Jean-Marie Normand, Vincent Tourre et Hugues Digonnet de l'École Centrale de Nantes pour leurs conseils avisés et les échanges que nous avons pu avoir lors de mes heures de monitorat. Je souligne également le travail de Hakim Cheikh et Alexis Dupuis, qui ont contribué à l'avancée de cette thèse à travers leur stage à l'IRSTV.

D'une certaine manière, une thèse est une aventure, la mienne m'a mené pendant 3 mois à l'université de Kyushu au Japon. Je remercie les membres du LIMU pour leur accueil et plus particulièrement Hideaki Uchiyama pour son accompagnement durant cette période. (Domo arigato gozaimashita)

Sur une note plus personnelle, je tiens à remercier mes parents de m'avoir soutenu depuis toutes ces années dans mes choix et dans mon parcours en enseignement supérieur.

C'est avec une tendresse particulière que je conclus ces remerciements avec Sofia, qui m'a accompagné, soutenu et motivé tout au long de cette thèse.

Publications et distinction

Articles de journal

1. **Fusion of 3D GIS, Vision, Inertial and Magnetic Data for Improved Urban Pedestrian Navigation and Augmented Reality Applications**, N. Antigny, M. Servieres, V. Renaudin, Journal of The Institute of Navigation, Septembre 2018.
2. **Pedestrian Step Length Estimation-aided Hand-held Monocular Visual Odometry for Loosely Coupling with PDR in Urban Environments**, N. Antigny, H. Uchiyama, M. Servieres, V. Renaudin, D. Thomas, R. Taniguchi. (en cours de soumission à MDPI Sensors)
3. **Visual SLAM : State-of-the-Art and Benchmarking in the Context of Pose Estimation for Pedestrian Applications**, A. Dupuis, N. Antigny, M. Servieres, V. Renaudin. (en cours de soumission à IEEE Transactions on Visualization and Computer Graphics)

Articles de conférence avec actes

1. **Hybrid Visual and Inertial Position and Orientation Estimation based on Known Urban 3D Models**, N. Antigny, M. Servieres, V. Renaudin, *7th IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN 2016)*, Alcalá de Henares, Espagne, Octobre 4-7, 2016.
2. **Pedestrian Track Estimation with Handheld Monocular Camera and Inertial-Magnetic Sensor for Urban Augmented Reality**, N. Antigny, M. Servieres, V. Renaudin, *8th IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN 2017)*, Sapporo, Japon, Septembre 18-21, 2017.
3. **[Poster] An Inertial, Magnetic and Vision based Trusted Pose Estimation for AR and 3D Data Qualification on Long Urban Pedestrian Displacements**, N. Antigny, M. Servieres, V. Renaudin, *16th IEEE International Symposium on Mixed and Augmented Reality (ISMAR 2017)*, Nantes, France, Octobre 9-13, 2017.
4. **Continuous Pose Estimation for Urban Pedestrian Mobility Applications on Smart-handheld Devices**, N. Antigny, M. Servieres, V. Renaudin, *9th IEEE International Conference on Indoor Positioning and Indoor Navigation (IPIN 2018)*, Nantes, France, Septembre 24-27, 2018.

Distinction

1. **Runner-up Best Paper** à la conférence IPIN 2017.

Résumé

Pour assister la navigation piétonne dans les espaces urbains et intérieurs, une estimation précise de la pose (i.e. la position 3D et l'orientation 3D) d'un équipement tenu en main constitue un point essentiel dans le développement d'outils d'aide à la mobilité (e.g. applications de réalité augmentée). Dans l'hypothèse où le piéton n'est équipé que d'appareils grand public, l'estimation de la pose est limitée à l'utilisation de capteurs à faible coût intégrés dans ces derniers (i.e. un récepteur GNSS, une unité de mesure inertielle et magnétique et une caméra monoculaire). De plus, les espaces urbains et intérieurs, comprenant des bâtiments proches et des éléments ferromagnétiques, constituent des zones difficiles pour la localisation et l'estimation de la pose lors de grands déplacements piétons. Cependant, le développement récent et la mise à disposition d'informations contenues dans des Systèmes d'Information Géographiques 3D constituent une nouvelle source de données exploitable pour la localisation et l'estimation de la pose.

Pour relever ces défis, cette thèse propose différentes solutions pour améliorer la localisation et l'estimation de la pose des équipements tenus en main par le piéton lors de ses déplacements en espaces urbains et intérieurs. Les solutions proposées intègrent l'estimation de l'attitude basée inertielle et magnétique, l'odométrie visuelle monoculaire mise à l'échelle grâce à l'estimation des déplacements du piéton, l'estimation absolue de la pose basée sur la reconnaissance d'objets SIG 3D parfaitement connus et la mise à jour en position de la navigation à l'estime du piéton. Toutes ces solutions s'intègrent dans un processus de fusion permettant d'améliorer la précision de la localisation et d'estimer en continu une pose qualifiée de l'appareil tenu en main. Cette qualification est nécessaire à la mise en place d'un affichage en réalité augmentée sur site.

Pour évaluer les solutions proposées, des données expérimentales ont été recueillies au cours de déplacements piétons dans un espace urbain avec des objets de référence et des passages intérieurs.



Abstract

To support pedestrian navigation in urban and indoor spaces, an accurate pose estimate (i.e. 3D position and 3D orientation) of an equipment held in hand constitutes an essential point in the development of mobility assistance tools (e.g. Augmented Reality applications). On the assumption that the pedestrian is only equipped with general public devices, the pose estimation is restricted to the use of low-cost sensors embedded in the latter (i.e. an Inertial and Magnetic Measurement Unit and a monocular camera). In addition, urban and indoor spaces, comprising closely-spaced buildings and ferromagnetic elements, constitute challenging areas for localization and sensor pose estimation during large pedestrian displacements. However, the recent development and provision of data contained in 3D Geographical Information System constitutes a new wealth of data usable for localization and pose estimation.

To address these challenges, this thesis proposes solutions to improve pedestrian localization and hand-held device pose estimation in urban and indoor spaces. The proposed solutions integrate inertial and magnetic-based attitude estimation, monocular Visual Odometry with pedestrian motion estimation for scale estimation, 3D GIS known object recognition-based absolute pose estimation and Pedestrian Dead-Reckoning updates. All these solutions are fused to improve accuracy and to continuously estimate a qualified pose of the handheld device. This qualification is required to validate an on-site augmented reality display.

To assess the proposed solutions, experimental data has been collected during pedestrian walks in an urban space with sparse known objects and indoors passages.

Table des matières

Table des matières	xi
Liste des figures	xiii
Liste des tableaux	xvii
1 Introduction	1
1.1 Contexte général	2
1.2 Problématiques et objectifs	2
1.3 Contributions	3
1.4 Organisation du mémoire	3
2 Définition des référentiels, des capteurs à disposition du piéton et des notations utilisées	5
2.1 Introduction	6
2.2 Définition des référentiels de la navigation piétonne	6
2.3 Représentations de l'orientation	9
2.4 Système de perception à disposition du piéton	11
2.5 Conclusion	17
3 Méthodes usuelles de localisation du piéton et d'estimation de la pose	19
3.1 Introduction	20
3.2 Méthodes usuelles de localisation du piéton	20
3.3 Méthodes usuelles d'estimation de la pose	23
3.4 Conclusion	27
4 Pose absolue estimée par reconnaissance d'objets issus d'un SIG 3D	29
4.1 Introduction	30
4.2 Système d'Information Géographique	30
4.3 Choix des objets de référence	31
4.4 Détection et identification des objets de référence	33
4.5 Estimation de la pose de la caméra	36
4.6 Estimation des erreurs	40
4.7 Expérimentation : estimation de la pose basée sur la reconnaissance d'un objet de référence et affichage en réalité augmentée sur site	43
4.8 Conclusion	47
5 Odométrie visuelle monoculaire mise à l'échelle par PDR	49
5.1 Introduction	50
5.2 Odométrie visuelle monoculaire	51
5.3 Estimation des longueurs de pas du piéton	52
5.4 Estimation du facteur d'échelle	53
5.5 Détermination de la position sur l'axe vertical	54
5.6 Expérimentation : estimation de la localisation par odométrie visuelle monoculaire mise à l'échelle	56

5.7 Conclusion	67
6 Fusion de données et estimation continue de la pose d'un équipement tenu en main	69
6.1 Introduction	70
6.2 Estimation des angles d'attitudes	70
6.3 Navigation à l'estime et corrections absolues	70
6.4 Estimation de la pose en statique avec couplage sur l'orientation	72
6.5 Estimation de la pose en dynamique	73
6.6 Expérimentation : corrections absolues d'un processus PDR	74
6.7 Expérimentation : estimation continue de la pose	78
6.8 Conclusion	80
7 Conclusion générale et perspectives	83
7.1 Conclusion et résumé des contributions	84
7.2 Perspectives	85

Liste des figures

2.1	Visualisation du référentiel <i>ECEF</i> (en bleu) et du référentiel <i>NED</i> (en vert) défini comme le référentiel <i>Navigation</i>	6
2.2	Visualisation du référentiel <i>Piéton</i> . Image : www.sci-sport.com	7
2.3	Visualisation du référentiel <i>Body</i> associé à la plateforme ULISS. Source : Datasheet ULISS.	8
2.4	Visualisation du référentiel <i>Objet</i> lié au modèle 3D d'un objet de référence.	8
2.5	Représentation de la réception des signaux satellitaires en milieu urbain.	12
2.6	Estimation des distorsions de la lentille de la caméra : (a) Distorsion tangentielle; (b) Distorsion radiale; (c) Distorsion complète.	13
2.7	Correction des distorsions de la lentille de la caméra : (a) Image avec distorsions; (b) Image avec distorsions corrigées.	13
2.8	Visualisation de l'association entre le référentiel <i>Caméra</i> et le plan <i>Image</i>	14
2.9	Plateforme d'acquisition : la plateforme ULISS et la caméra monoculaire Garmin VIRB.	16
2.10	Visualisation des repères associés au body et à la caméra.	17
4.1	Les cinq niveaux de détails définis dans CityGML. Source : [BILJECKI et al., 2016].	31
4.2	Exemple de station vélo de la ville de Nantes.	32
4.3	(a) Implantation des stations vélo de la ville de Nantes; (b) Implantation des stations vélo dans le centre ville de Nantes.	32
4.4	(a) Côtes (en mm) du terminal de la station vélo; (b) Modèle 3D de la station vélo réalisé avec un logiciel de CAO.	33
4.5	(a) Points d'intérêt HOG extraits; (b) Densité des points d'intérêt HOG extraits.	34
4.6	(a) Reconnaissance de caractères par OCR; (b) Mise en correspondance avec l'objet d'intérêt contenu dans le SIG.	35
4.7	Points SURF (en rouge) et points géométriques (en vert) du terminal de la station vélo sur une image extraite de la vidéo acquise par la caméra tenue en main par le piéton.	36
4.8	Illustration du passage entre les référentiels <i>Objet</i> , <i>Caméra</i> et <i>Image</i>	37
4.9	(a) Lien entre le référentiel <i>Objet</i> et le référentiel <i>Navigation</i> (i.e. NED); (b) Vue sur le plan horizontal du terminal et des orientations des référentiels <i>Objet</i> et <i>Navigation</i>	38
4.10	Acquisition de la position des bords du terminal de la station vélo avec récepteur GNSS géodétic double fréquence et antenne.	39
4.11	Visualisation dans <i>Google Earth</i> des localisations du terminal de la station vélo renseignées dans "DataNantes" (en rouge), déterminées via orthophotos (en orange) et déterminées par mesures GNSS en différentiel (en vert).	40
4.12	Visualisation de la position 3D estimée sans erreur introduite (en bleu) et de 100 positions 3D estimées avec l'introduction d'une erreur pseudo-aléatoires uniformément répartis de [-5; +5] pixels sur chacun des points \mathbf{x}_i	41
4.13	Distribution sur 100.000 itérations de l'erreur de positionnement du centre optique de la caméra pour l'introduction d'une erreur pseudo-aléatoire uniformément répartie de [-5; +5] pixels sur les coordonnées des points \mathbf{x}_i	42
4.14	Erreur de projection.	43

4.15	Acquisition avec récepteur GNSS géodésique, antenne pour déterminer la position des points de référence et canne à bulle.	44
4.16	Positions extraites de l'estimation de la pose basé sur la reconnaissance d'un objet de référence pour l'acquisition en statique sur les points de références.	45
4.17	(a) Vue 2D (i.e. sur le plan horizontal) des positions estimées; (b) Vue 3D des positions estimées.	45
4.18	Reprojection en réalité augmentée sur site : (a) Modèle 3D de l'objet d'intérêt; (b) Dalle LOD2.	46
4.19	Reprojection en réalité augmentée sur site : Dalle LOD2 corrigée.	47
4.20	Reconnaissance de l'objet de référence robuste à l'occlusion et reprojection de son modèle 3D.	47
5.1	Diagramme bloc de la solution d'estimation du facteur d'échelle dans l'odométrie visuelle monoculaire.	50
5.2	Triangulation.	52
5.3	Interpolation des longueurs de pas du piéton estimées.	53
5.4	Trajectoire de référence pour le parcours de 0.7 km lors de l'acquisition du piéton M2 et points de passages de référence à proximité des objets de référence sur le parcours (stations 44 et 77).	56
5.5	Matériel expérimental du piéton, antenne hélicoïdale attachée au poignet et récepteur GNSS géodésique dans le sac à dos durant les acquisitions.	57
5.6	Acquisition avec la plateforme expérimentale (ULISS + caméra) en main, l'antenne hélicoïdale attachée au poignet et le récepteur GNSS géodésique dans le sac à dos, la plateforme PERSY attachée au pied pour déterminer la position lors des acquisitions, et les points de référence au sol.	59
5.7	(a) Classification de l'activité du piéton : Statique (en rose), Marche avec portage en mode <i>Texting</i> (en vert); (b) Estimation des longueurs de pas du piéton M1.	60
5.8	(a) Trajectoires de référence au DGNSS (en vert) et à PERSY (en bleu), trajectoire estimée avec l'odométrie visuelle mise à l'échelle (en rouge); (b) Erreur de positionnement de l'odométrie visuelle mise à l'échelle à la solution DGNSS (en vert) et à la solution PERSY (en bleu). Les résultats sont présentés pour les piétons <i>W1</i> , <i>M1</i> et <i>M2</i> (du haut vers le bas).	61
5.9	Trajectoires estimées lors de petits déplacements du piéton avec le positionnement par GPS (en bleu), par navigation à l'estime du piéton (en rouge), par odométrie visuelle monoculaire à l'échelle (en vert).	62
5.10	Pourcentage d'erreur de positionnement en fonction de la distance parcourue pour les différents piétons.	63
5.11	(a) Détection des pics sur le signal de l'accéléromètre sur l'axe vertical; (b) Détection des phases statiques sur le signal de l'accéléromètre.	64
5.12	Estimation de la vitesse avec application des ZUPT et estimation des variations de la position sur l'axe vertical.	64
5.13	(a) Mesures brutes de la pression durant l'acquisition; (b) Mesure de la température durant l'acquisition en milieu extérieur.	65
5.14	Hauteur estimée durant l'acquisition.	66
5.15	(a) Visualisation dans le modèle 3D de la trajectoire estimée par odométrie visuelle mise à l'échelle; (b) Visualisation dans <i>Google Earth</i> de la trajectoire estimée par odométrie visuelle mise à l'échelle.	66
6.1	Visualisation du principe de la navigation à l'estime du piéton.	71
6.2	Diagramme bloc de la solution de fusion entre les mesures inertielles et magnétiques, et les informations extraites de la vision monoculaire.	73
6.3	Synchronisation entre les données vision et les données IMMU.	74

6.4	Trajectoire de référence GPS standalone avec ULISS (en rouge) et trajectoire de référence DGNSS avec Septentrio (en vert) sur le parcours de 2.9 km en milieu urbain.	76
6.5	Points de référence, trajectoire de référence DGNSS (en vert) et trajectoires obtenues pour la navigation à l'estime du piéton sans correction (en rouge) et avec corrections (en orange) sur le parcours de 2.9 km en milieux urbains.	77
6.6	Erreurs de positionnement du processus de navigation à l'estime du piéton et corrections appliquées.	77
6.7	Trajectoire de référence pour le parcours de 0.5 km lors de l'acquisition du piéton <i>M2</i> et points de passages de référence à proximité d'un objet de référence sur le parcours (stations vélo 44).	78
6.8	Points de référence (en blanc), trajectoire de référence DGNSS (en rouge), trajectoire de référence PERSY (en vert) et trajectoire estimée avec le processus de fusion des capteurs (en bleu) pour l'acquisition du piéton <i>M2</i>	80
7.1	(a) Implantation des stations vélo (en rouge) de la ville de Nantes; (b) Implantation des stations vélo (en rouge), des stations de bus et de tramway (en bleu) de la ville de Nantes.	85

Liste des tableaux

4.1	Précision de la position extrait de l'estimation de la pose basée sur la reconnaissance d'un objet de références.	44
5.1	Mesures de la hauteur du piéton et de la hauteur de l'équipement tenu en main. . .	55
5.2	Informations sur les différents piétons et leur acquisition.	58
5.3	Erreurs de positionnement.	60
6.1	Informations sur le piéton et sur l'acquisition.	74
6.2	Résultats d'acquisition.	75
6.3	Erreurs de positionnement.	77
6.4	Informations sur les différents piétons et leur acquisition.	79
6.5	Erreurs de positionnement.	79

Chapitre 1

Introduction

*« Indocti discant et ament
meminisse periti »*

Charles-Jean-François Hénault

Sommaire

1.1 Contexte général	2
1.2 Problématiques et objectifs	2
1.3 Contributions	3
1.4 Organisation du mémoire	3

1.1 Contexte général

Motivées par des objectifs de réduction de la pollution et de la congestion routière dans les villes, d'amélioration de la santé publique et de protection de l'environnement, les autorités publiques s'efforcent de promouvoir la mobilité douce, via les modes de transports actifs, et l'usage des transports publics. Cela se caractérise par l'augmentation de l'espace réservé aux piétons au sein des villes, par le développement des services de transports publics et par l'augmentation des stations en libre service pour les vélos et les voitures électriques. C'est dans ce contexte de mobilité urbaine et de navigation piétonne que se positionnent les travaux présentés dans cette thèse.

Avec l'augmentation du nombre de piétons possédant un dispositif mobile (i.e. smartphones et tablettes) intégrant des capteurs bas-coût, de plus en plus de travaux s'intéressent à l'utilisation des données complémentaires issues de ces capteurs pour estimer la localisation du piéton durant ses déplacements. Les travaux présentés dans cette thèse se placent dans cette continuité, et cherchent à améliorer l'estimation de la pose en fusionnant les mesures des capteurs intégrés dans un équipement tenu en main, avec des estimations de la dynamique du piéton durant ses déplacements et des données spatiales. En effet, avec le développement récent des Systèmes d'Information Géographiques (SIG) 3D et la mise à disposition de leurs contenus, de nouvelles données peuvent être intégrées dans les différentes solutions de navigation piétonne pour améliorer la localisation et l'estimation de la pose.

1.2 Problématiques et objectifs

Dans le contexte de mobilité urbaine et de navigation piétonne, le développement d'outils d'aide à la mobilité gagnerait à obtenir une meilleure intégration des différentes mesures issues des systèmes de navigation par satellites (GNSS), des unités de mesures inertielles et magnétiques (IMMU), et des caméras monoculaires intégrés dans les équipements grand public du piéton. Ainsi, dans le but de développer ces outils, un des objectifs de cette thèse est de proposer des solutions autonomes de localisation et d'estimation de la pose. Ces solutions devront se limiter à l'utilisation des capteurs et des données à disposition du piéton, et ne dépendre d'aucune connexion ou déploiement de réseaux de capteurs. L'objectif étant que ces solutions puissent être proposées en continu, quel que soient la complexité des lieux à travers lesquels le piéton se déplace. Par conséquent, les solutions reposant sur l'utilisation de capteurs spécifiques (e.g. LIDAR, caméra stéréoscopique, etc.) ou sur le déploiement d'infrastructure (e.g. WiFi, Ultra Large Bande (UWB), Bluetooth, IDentification par Radio Fréquence (RFID), etc.), ne pourront être appliquées dans le contexte de cette thèse.

L'estimation précise de la pose d'un équipement tenu en main constitue un enjeu majeur pour le développement d'outils d'aide à la mobilité. Cependant, les espaces urbains représentent encore des zones difficiles pour la localisation du piéton et l'estimation de la pose d'équipements tenus en main. C'est particulièrement le cas dans les canyons urbains et les espaces intérieurs où les systèmes de navigation par satellites souffrent d'effets d'atténuation, de réflexions et de blocages qui détériorent fortement la précision du positionnement [LATEGAHN et al., 2013; YU et al., 2016; ZANDBERGEN et BARBEAU, 2011]. A cela s'ajoute la présence d'éléments ferromagnétiques, générant des perturbations magnétiques, mais aussi d'éléments et de facteurs environnementaux entraînant des occlusions, des réflexions spéculaires et des changements d'illumination, dégradant les estimations basées sur la vision. Ces estimations sont encore plus difficiles en utilisant des ressources informatiques limitées et des capteurs bas-coûts de type *Micro Electro Mechanical System*-(MEMS) intégrés dans les équipements grand public [KHAN et al., 2013]. La fusion de ces données complémentaires, issues de différents processus, constitue l'un des objectifs de cette thèse.

L'estimation absolue et qualifiée de la pose d'un dispositif tenu dans la main permettrait d'améliorer les applications d'aide à la mobilité, notamment en proposant une visualisation d'informations en réalité augmentée. Dans le contexte de navigation piétonne en milieu urbain, les capteurs

sont portés en main, cette dernière pouvant effectuer des mouvements libres indépendamment du corps. Ainsi, les solutions de l'état de l'art nécessitant que les capteurs soient fixés sur une partie du corps (i.e. pied, torse, tête) ne peuvent être facilement adaptées ou appliquées dans le contexte de cette thèse. Avec la mise à disposition de données, contenues dans les Systèmes d'Information Géographique, de nouvelles solutions peuvent être mises en place pour la localisation du piéton et l'estimation de la pose d'un équipement tenu en main. La capacité à intégrer ces nouvelles données dans les traitements constitue un autre objectif de cette thèse. L'estimation absolue de la pose, permettant de proposer un affichage en réalité augmentée qualifié directement sur site, constitue un point essentiel au développement de nouvelles solutions d'aide à la navigation piétonne.

1.3 Contributions

Pour relever les défis évoqués, les travaux présentés dans cette thèse ont été menés pour tirer profit de la connaissance de données complémentaires, i.e. les vitesses angulaires, les accélérations, le champ magnétique, les images/vidéos acquises au moyen des capteurs intégrés dans les équipements grand public à disposition du piéton et la connaissance détaillée d'objets de référence représentatifs des données contenues dans un Système d'Information Géographique 3D. Ainsi, trois innovations sont proposées dans cette thèse :

- Une estimation de la pose basée sur la reconnaissance d'objets de référence stockés dans un SIG 3D, permettant une correction de la navigation à l'estime du piéton ;
- Une mise à l'échelle de l'odométrie visuelle monoculaire par l'estimation des longueurs de pas du piéton ;
- Une fusion des données visio-inertielles pour estimer en continu une pose qualifiée de l'équipement tenu en main par le piéton.

L'ensemble de ces innovations permet de mettre en place une solution autonome pouvant être valorisée par un outil d'aide à la navigation du piéton en milieux urbains et intérieurs. De plus, grâce à une estimation qualifiée de la pose, la visualisation d'informations via un affichage en réalité augmentée sur site est alors possible quand cette dernière est définie comme étant suffisamment précise. Outre l'usage de la réalité augmentée pour des applications de navigation piéton, ces solutions ouvrent également la voie à de nouvelles applications comme la caractérisation des données contenues dans les SIG 3D directement sur site.

1.4 Organisation du mémoire

Après avoir introduit dans ce chapitre le contexte, les problématiques et les objectifs de cette thèse, ce manuscrit se compose de six chapitres.

Le second chapitre est dédié à la définition du système de perception tenu en main par le piéton lors de ses déplacements en milieu urbain. Les référentiels utilisés et les notions mathématiques associées sont détaillés en début de chapitre. Les différents capteurs embarqués dans les équipements grand public du piéton sont ensuite présentés. La configuration matérielle utilisée pour les acquisitions, la calibration et les mesures des capteurs embarqués sont détaillés en fin de chapitre.

Le troisième chapitre présente un état de l'art des méthodes usuelles de localisation et d'estimation de la pose basée sur la vision et sur les mesures inertielles et magnétiques. Une attention est portée sur leur applicabilité dans le contexte de la navigation piétonne en milieu urbain avec les capteurs grand public tenus en main. Les solutions, quelles soient hybrides ou non, reposant sur les GNSS, la vision, les capteurs inertiels et magnétiques sont investiguées.

Les chapitres suivants sont relatifs aux innovations proposées dans cette thèse. Les solutions proposées sont introduites, détaillées et évaluées dans leur chapitre respectif.

Le quatrième chapitre est consacré à la reconnaissance d'objets de référence contenus dans un Système d'Informations Géographiques 3D. Cette reconnaissance permet d'estimer de manière absolue la pose d'une caméra monoculaire dans un référentiel de navigation global. Le processus de reconnaissance d'objets de référence et d'estimation de la pose est détaillé et évalué expérimentalement lors d'acquisitions en milieu urbain. L'estimation absolue de la pose est ainsi utilisée, d'une part pour améliorer la localisation du piéton en milieu urbain, d'autre part pour permettre la reprojexion d'informations en réalité augmentée sur la scène réelle observée.

Le cinquième chapitre est consacré à la mise en place d'une solution d'odométrie visuelle avec un équipement intégrant une caméra monoculaire tenue en main. L'estimation des longueurs de pas du piéton, proposée dans les solutions de navigation à l'estime du piéton, est utilisée pour résoudre l'ambiguïté d'échelle de l'odométrie visuelle monoculaire. Cette mise à l'échelle permet d'estimer en continu la pose de l'équipement tenu en main, de manière relative à une pose initiale ou à une pose absolue, lors des déplacements du piéton en milieu urbain.

Le sixième chapitre est consacré à l'estimation continue de la pose et à la fusion des données issues des différents traitements présentés dans cette thèse. La pose absolue estimée grâce à la reconnaissance d'un objet de référence permet de corriger un processus de navigation à l'estime. Le couplage permet ensuite d'estimer en continu une pose qualifiée de l'équipement tenu en main par le piéton lors de ses déplacements en milieu urbain. Des acquisitions sont réalisées sur différents parcours en milieux urbains par plusieurs piétons pour mettre en évidence l'apport de ces solutions.

Ce manuscrit est conclu par le récapitulatif des principales contributions des travaux menés au cours de cette thèse. Différentes pistes de recherches ouvertes à l'issue de ceux-ci sont également proposées.

Chapitre 2

Définition des référentiels, des capteurs à disposition du piéton et des notations utilisées

« Une fois qu'on a dépassé la mesure, il n'y a plus de limite. »

Euripide

Sommaire

2.1 Introduction	6
2.2 Définition des référentiels de la navigation piétonne	6
2.2.1 Référentiel <i>Navigation</i>	6
2.2.2 Référentiel <i>Piéton</i>	7
2.2.3 Référentiel <i>Inertiel</i> et référentiel <i>Body</i>	7
2.2.4 Référentiel <i>Caméra</i> et plan <i>Image</i>	8
2.2.5 Référentiel <i>Objet</i>	8
2.3 Représentations de l'orientation	9
2.3.1 Angles d'Euler	9
2.3.2 Matrices de rotation	9
2.3.3 Quaternions	10
2.4 Système de perception à disposition du piéton	11
2.4.1 Récepteur GNSS	11
2.4.2 Caméra monoculaire	12
2.4.3 Unité de mesures inertielles et magnétiques	14
2.5 Conclusion	17

2.1 Introduction

Ce chapitre définit dans un premier temps les différents référentiels utilisés dans le contexte de la navigation piétonne. Dans un second temps, les notations et les conversions mathématiques, utilisées dans ces travaux de thèse pour représenter la pose, sont introduites et détaillées. Les différents capteurs intégrés dans les équipements grand public à disposition du piéton sont ensuite présentés. Ces derniers sont utilisés pour acquérir les données et ainsi mettre en place les différentes solutions présentées dans cette thèse. Les étapes préliminaires nécessaires à la mise en place des solutions, telle que la calibration des capteurs, sont détaillées.

2.2 Définition des référentiels de la navigation piétonne

2.2.1 Référentiel *Navigation*

Le référentiel *Navigation*, noté avec le symbole n , est un système de coordonnées géographiques utilisé pour représenter l'espace 3D dans lequel évolue le piéton. Comme l'orientation de l'équipement tenu en main par le piéton est estimée dans le référentiel *Nord-Est-Down* (*NED*), ce dernier est défini comme référentiel *Navigation*. Comme présenté sur la Figure 2.1, le référentiel *NED* est un système de coordonnées géographiques défini à partir d'un plan local tangent en un point de la Terre. Ce plan tangent est une approximation locale de la Terre permettant d'appliquer les principes de la géométrie euclidienne. Les trois composantes $\{X_n, Y_n, Z_n\}$ de ce référentiel sont définies comme suit :

- Le vecteur X_n est orienté vers le *Nord* géographique ;
- Le vecteur Y_n est orienté vers l'*Est* ;
- Le vecteur Z_n est orienté à la perpendiculaire du plan tangent, ce qui est différent du vecteur orienté vers le centre de la Terre.

Le système de coordonnées géographiques *Nord-Est-Down* (*NED*) est similaire au système de coordonnées géographiques *Earth-Centered, Earth-Fixed* (*ECEF*). Ce dernier est utilisé pour les positions déterminées par les systèmes de localisation par satellites. Tout comme le référentiel *NED*, le référentiel *ECEF*, noté avec le symbole e , est un système de coordonnées cartésiennes.

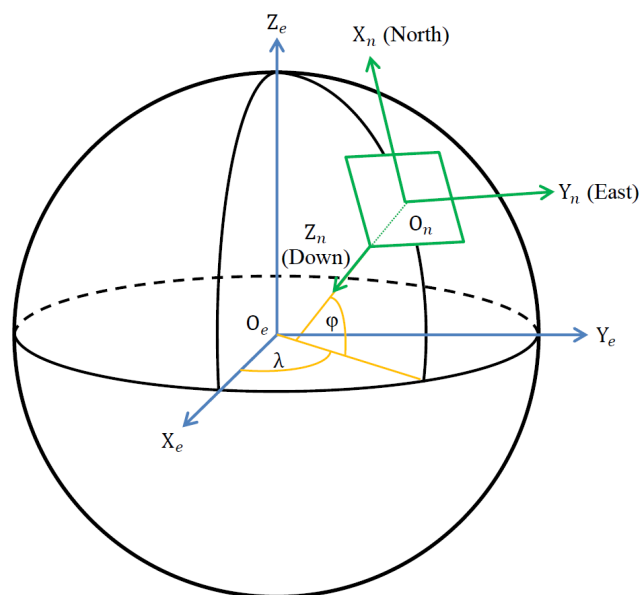


FIGURE 2.1 – Visualisation du référentiel *ECEF* (en bleu) et du référentiel *NED* (en vert) défini comme le référentiel *Navigation*.

2.2.2 Référentiel *Piéton*

Le référentiel *Piéton*, étiqueté avec le symbole p , est un référentiel 3D lié au centre de masse du piéton. Les trois composantes $\{X_p, Y_p, Z_p\}$ de ce référentiel sont définies comme suit :

- Le vecteur X_p est défini sur l'axe Dorso-Ventral;
- Le vecteur Y_p est défini sur l'axe Droite-Gauche;
- Le vecteur Z_p est défini sur l'axe Rostro-Caudal.

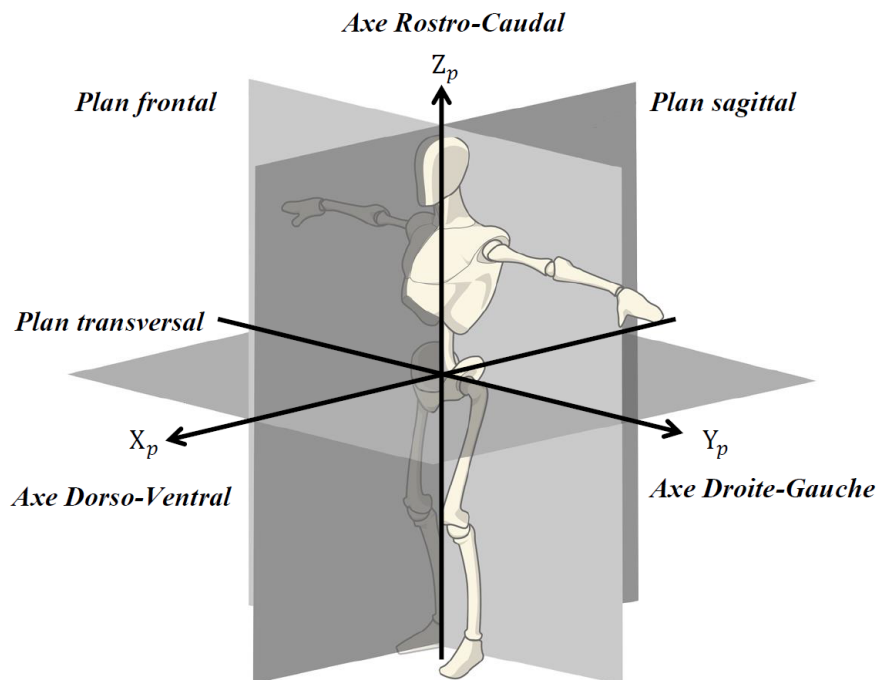


FIGURE 2.2 – Visualisation du référentiel *Piéton*. Image : www.sci-sport.com

Il est à noter que ce référentiel ne sera pas directement utilisé dans les travaux présentés dans cette thèse.

2.2.3 Référentiel *Inertiel* et référentiel *Body*

Le référentiel *Inertiel*, noté avec le symbole in , est un référentiel 3D défini pour l'unité de mesures inertielles intégrée dans les équipements du piéton. Ce dernier est un référentiel supposé galiléen, i.e. les lois de la mécanique newtonienne s'y appliquent. Ce référentiel est attaché au centre de la Terre. Les trois composantes $\{X_{in}, Y_{in}, Z_{in}\}$ de ce référentiel sont définies comme suit :

- Le vecteur X_{in} est orienté vers le Soleil à l'équinoxe;
- Le vecteur Y_{in} est défini de manière à ce que le repère orthonormé $\{X_{in}, Y_{in}, Z_{in}\}$ soit direct;
- Le vecteur Z_{in} est l'axe de rotation de la Terre.

Le référentiel *Body*, noté avec le symbole b , est un référentiel 3D orthonormé direct lié à un point de la plateforme utilisée pour les acquisitions et intégrant les capteurs inertiels et magnétiques [ORTIZ et al., 2017]. Cette dernière est introduite dans la sous-section 2.4.3 et une visualisation du référentiel est présentée sur la Figure 2.3.

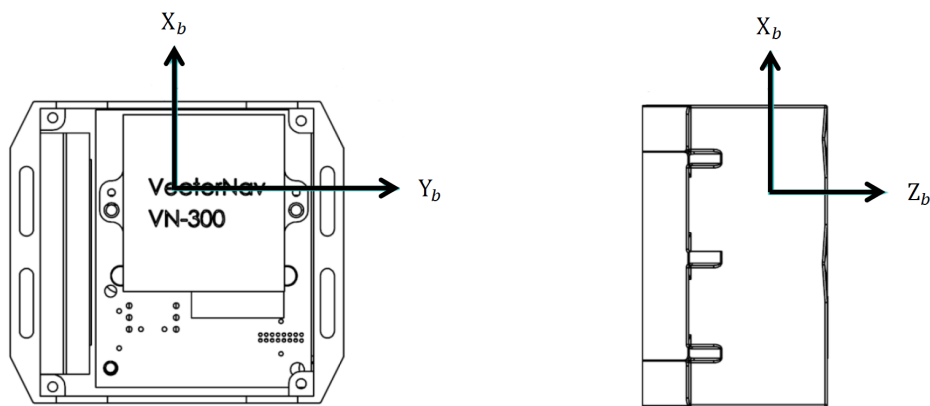


FIGURE 2.3 – Visualisation du référentiel *Body* associé à la plateforme ULISS. Source : Datasheet ULISS.

2.2.4 Référentiel *Caméra* et plan *Image*

Le référentiel *Caméra*, noté avec le symbole c , est un référentiel 3D lié au centre optique de la caméra. Le plan *Image*, noté avec le symbole i , est un référentiel 2D (i.e. en pixels) lié aux images acquises par la caméra. Le détail des composantes, ainsi que le lien entre ces deux référentiels est donné dans la sous-section 2.4.2. Une visualisation du référentiel *Caméra* et du plan *Image* est présentée sur la Figure 2.8.

2.2.5 Référentiel *Objet*

Le référentiel *Objet*, étiqueté avec le symbole o , est un référentiel 3D lié au modèle 3D des objets de référence. Les trois composantes $\{X_o, Y_o, Z_o\}$ de ce référentiel sont définies comme suit :

- Le vecteur X_o est défini sur l'axe droite-gauche;
- Le vecteur Y_o est défini sur l'axe face avant-face arrière;
- Le vecteur Z_o est défini sur l'axe bas-haut.

Une visualisation de ce repère est présenté sur la Figure 2.4.

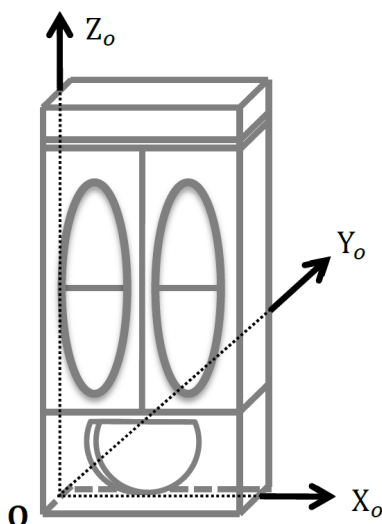


FIGURE 2.4 – Visualisation du référentiel *Objet* lié au modèle 3D d'un objet de référence.

2.3 Représentations de l'orientation

Dans cette section sont présentés les trois paramétrages utilisés dans les différentes solutions présentées dans cette thèse pour représenter l'orientation :

- Les angles d'Euler;
- Les matrices de rotation;
- Les quaternions.

2.3.1 Angles d'Euler

Les angles d'Euler sont utilisés pour représenter l'orientation d'un élément par rapport à un référentiel. Trois types rotations sont définies :

- La *roulis*, noté *roll* avec le symbole ϕ ;
- La *tangage*, noté *pitch* avec le symbole θ ;
- La *lacet*, noté *yaw* avec le symbole ψ .

2.3.2 Matrices de rotation

Dans un espace euclidien à 3 dimensions, la matrice de rotation permet de représenter les changements d'orientation à opérer pour passer d'un référentiel à un autre. Ainsi, l'orientation d'un objet muni d'un repère orthonormé noté a dans un espace muni d'un repère orthonormé noté z , peut être définie par la matrice de rotation \mathbf{R}_a^z exprimant les vecteurs de la base a dans la base z . En se basant sur les angles d'Euler précédemment introduits, les matrices de rotations représentant chacune de ces rotations sont définies de la manière suivante :

$$\mathbf{R}_\psi = \begin{pmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix}; \mathbf{R}_\theta = \begin{pmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{pmatrix}; \mathbf{R}_\phi = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{pmatrix} \quad (2.1)$$

La matrice de rotation \mathbf{R}_a^z se définit alors comme la séquence de trois rotations consécutives :

$$\mathbf{R}_a^z = \mathbf{R}_\psi \mathbf{R}_\theta \mathbf{R}_\phi \quad (2.2)$$

$$\mathbf{R}_a^z = \begin{bmatrix} \cos\theta \times \cos\psi & \sin\phi \times \sin\theta \times \cos\psi - \sin\psi \times \cos\phi & \cos\phi \times \sin\theta \times \cos\psi + \sin\phi \times \sin\psi \\ \cos\theta \times \sin\psi & \sin\psi \times \sin\theta \times \sin\phi + \cos\psi \times \cos\phi & \cos\phi \times \sin\theta \times \sin\psi - \sin\phi \times \cos\psi \\ -\sin\theta & \cos\theta \times \sin\phi & \cos\theta \times \cos\phi \end{bmatrix}$$

Le passage matrice de rotation-angles d'Euler dépend bien évidemment du choix de la séquence des angles. Pour la séquence *yaw-pitch-roll*, le passage d'une matrice de rotation vers les angles d'Euler est donné par :

$$\begin{cases} \phi = \arctan_2(\cos\theta \times \sin\phi, \cos\theta \times \cos\phi) \\ \theta = -\arctan(-\sin\theta) \\ \psi = \arctan_2(\cos\theta \times \sin\psi, \cos\theta \times \cos\psi) \end{cases} \quad (2.3)$$

Il est à noter que la fonction $\arctan_2(y, x)$ correspond à la fonction inverse de tangente qui prend en compte les signes de x et y .

Les matrices de rotation présentent également un certain nombre de propriétés qui seront utilisées dans la mise en place des solutions présentées dans cette thèse, à savoir :

- Le déterminant d'une matrice de rotation est unitaire :

$$\det(\mathbf{R}) = 1 \quad (2.4)$$

- La matrice de rotation est inversible.
- La multiplication d'une matrice de rotation par sa transposée équivaut à l'identité :

$$\mathbf{R}\mathbf{R}^T = \mathbf{I}_3 \quad (2.5)$$

2.3.3 Quaternions

L'ensemble des quaternions, noté \mathbb{H} en référence au mathématicien William Hamilton, est un ensemble basé sur le corps des nombres réels. Un quaternion $\mathbf{q} \in \mathbb{H}$ se compose d'un quadruplet de réels :

$$\mathbf{q} = (q_1, q_2, q_3, q_4)^T \quad (2.6)$$

Une autre notation est également adoptée pour décrire un quaternion :

$$\mathbf{q} = (q_1, \mathbf{u}_q^T)^T \quad (2.7)$$

Avec :

- La partie scalaire ou réelle du quaternion : q_1 ;
- La partie vectorielle ou imaginaire du quaternion : $\mathbf{u}_q = (q_2, q_3, q_4)^T$.

L'espace des quaternions permet la multiplication, notée avec le symbole \otimes . Ainsi, pour deux quaternions $\mathbf{q}, \mathbf{p} \in \mathbb{H}$, la multiplication est définie comme suit :

$$\begin{pmatrix} q_1 \\ \mathbf{u}_q \end{pmatrix} \otimes \begin{pmatrix} p_1 \\ \mathbf{u}_p \end{pmatrix} = \begin{pmatrix} q_1 p_1 - \langle \mathbf{u}_q, \mathbf{u}_p \rangle \\ q_1 \mathbf{u}_p + p_1 \mathbf{u}_q + \mathbf{u}_q \wedge \mathbf{u}_p \end{pmatrix} \quad (2.8)$$

Plusieurs éléments sont également définis pour un quaternion :

- $\bar{\mathbf{q}}$, le conjugué d'un quaternion \mathbf{q} ;

$$\bar{\mathbf{q}} = (q_1, -\mathbf{u}_q)^T \quad (2.9)$$

- $\|\mathbf{q}\|$, la norme euclidienne d'un quaternion \mathbf{q} .

$$\|\mathbf{q}\| = \sqrt{q_1^2 + q_2^2 + q_3^2 + q_4^2} = \sqrt{q_1^2 + \|\mathbf{u}_q\|^2} \quad (2.10)$$

Les quaternions représentant les angles d'Euler précédemment introduits sont définis de la manière suivante :

$$\mathbf{q}_\psi = \begin{pmatrix} \cos \frac{\psi}{2} \\ 0 \\ 0 \\ \sin \frac{\psi}{2} \end{pmatrix}; \mathbf{q}_\theta = \begin{pmatrix} \cos \frac{\theta}{2} \\ 0 \\ \sin \frac{\theta}{2} \\ 0 \end{pmatrix}; \mathbf{q}_\phi = \begin{pmatrix} \cos \frac{\phi}{2} \\ \sin \frac{\phi}{2} \\ 0 \\ 0 \end{pmatrix} \quad (2.11)$$

Pour la séquence *yaw-pitch-roll*, le passage d'un quaternion vers les angles d'Euler est donné par :

$$\begin{cases} \phi = \arctan_2(2 \times (q_3 q_4 + q_1 q_2), q_1^2 - q_2^2 - q_3^2 + q_4^2) \\ \theta = -\arctan(2 \times (q_2 q_4 + q_1 q_3)) \\ \psi = \arctan_2(2 \times (q_2 q_3 + q_1 q_4), q_1^2 + q_2^2 - q_3^2 - q_4^2) \end{cases} \quad (2.12)$$

2.4 Système de perception à disposition du piéton

2.4.1 Récepteur GNSS

Les systèmes de géolocalisation par satellites se sont largement démocratisés dans les équipements grand public du piéton. Ils reposent sur l'utilisation de satellites dédiés à la navigation. Ces derniers gravitent autour de la Terre, et diffusent continuellement des signaux pouvant être captés par un récepteur GNSS. Connaissant la vitesse de propagation de ces signaux dans l'espace et dans les couches atmosphériques de la Terre, la différence entre les instants d'émission et de réception permet d'estimer la distance entre le satellite et le récepteur. Avec la réception des signaux d'au moins quatre satellites différents, la position absolue du récepteur peut être estimée par trilatération. Les signaux de trois satellites permettent de déterminer la position du récepteur et le signal d'un quatrième satellite permet de résoudre l'inconnue de synchronisation entre les horloges du récepteur et du satellite. Cette solution est définie sous l'appellation *Global Navigation Satellite System* (GNSS). Pour mettre en place cette solution, il existe plusieurs constellations de satellites, certaines ayant une couverture globale, d'autres ayant une couverture locale. En 2017, les principales constellations mises en place ayant une couverture globale sont :

- La constellation de satellites *GPS* (Global Positioning System), comptant 30 satellites mis en place par les Etats-Unis¹;
- La constellation de satellites *GLONASS* (GLObalnaïa NAVigatsionnaïa Spoutnikovaïa Sistéma), comptant 24 satellites mis en place par la Russie²;
- La constellation de satellites *Galileo*, comptant 15 satellites mis en place par l'Europe³;
- La constellation de satellites *Beidou/Compass*, comptant 20 satellites mis en place par la Chine⁴.

En environnement dégagé, les systèmes de navigation par satellites constituent une solution suffisamment précise pour la localisation et l'aide à la navigation du piéton, i.e. précision décimétrique. Cependant, cette localisation au moyen des récepteurs embarqués dans les équipements grand public peut souffrir d'erreurs [BOAZ et al., 2011; ZANDBERGEN et BARBEAU, 2011]. Les principales sources d'erreurs sont :

- La dérive des horloges des récepteurs, due à la qualité matérielle;
- La différence d'écoulement du temps entre les satellites et le récepteur, due à la relativité;
- Les variations de temps de propagation des signaux satellitaires, dues à la traversée des différents milieux rencontrés (e.g. couches de l'atmosphère terrestre).
- La dégradation des signaux satellitaires, due à la propagation à travers les différents milieux rencontrés, aux réflexions (e.g. surfaces des bâtiments en canyon urbain) et aux blocages (e.g. milieu intérieur). La Figure 2.5 illustre le problème de réflexion en canyon urbain.
- La réception d'un nombre insuffisant de signaux satellitaires, due à la couverture satellitaire;
- La résolution des équations d'estimation de la position, due à la géométrie de réception des signaux satellitaires;

Malgré cela, différentes solutions peuvent être mises en place pour éliminer une partie de ces erreurs et ainsi obtenir une meilleure précision. Une solution classique consiste à utiliser un réseau de stations sur Terre, pour lesquelles les coordonnées sont parfaitement connues et invariantes, e.g. le réseau RGP-IGN⁵ pour la France. Les signaux GNSS sont collectés par une station de base et par un récepteur intégré dans la plateforme mobile du piéton. Les données GNSS sont ensuite post-traitées en mode différentiel afin d'améliorer la précision des coordonnées. Cette solution est définie sous l'appellation de Differential GNSS (DGNSS).

1. <https://www.gps.gov/systems/>
2. <https://www.glonass-iac.ru/en/>
3. <https://www.gsc-europa.eu/>
4. <http://en.chinabeidou.gov.cn/>
5. <http://rgp.ign.fr/DONNEES/diffusion/>

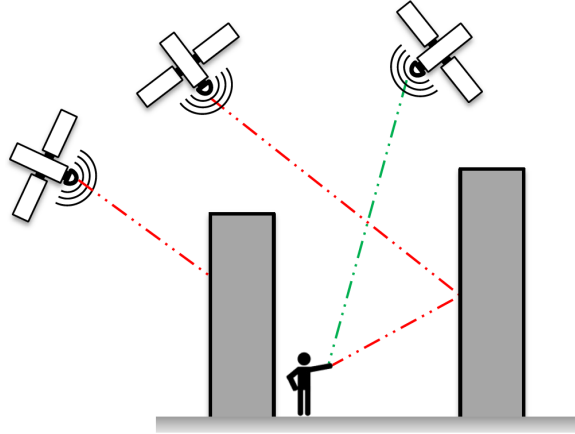


FIGURE 2.5 – Représentation de la réception des signaux satellitaires en milieu urbain.

2.4.2 Caméra monoculaire

La caméra utilisée pour les enregistrements vidéo lors des acquisitions est une caméra “Garmin VIRB 30 Ultra”. A la différence de la plupart des caméras intégrées dans les équipements grand public qui ont un autofocus, la caméra est paramétrée avec une distance focale fixe et une résolution d’image de 1920×1080 pixels. Cette résolution a été choisie de manière à correspondre à la résolution standard des équipements grand public. Les vidéos sont acquises avec une fréquence d’images de 60 Hz et sont sous-échantillonnées à une fréquence d’images de 10 Hz. Cette réduction de la fréquence d’échantillonnage des images est faite principalement pour réduire le temps de calcul des traitements basés sur la vision par ordinateur. La caméra intègre également un récepteur GPS. Ce dernier permet de synchroniser les images acquises par la caméra avec les données acquises par l’unité de mesures inertielles et magnétiques grâce aux marqueurs de temps GPS.

Calibration : correction des distorsions de la lentille et estimation des paramètres intrinsèques

Pour corriger les distorsions d’images et obtenir une projection réaliste des éléments de la scène observée sur le plan de l’image, une correction des distorsions de la lentille est nécessaire [ZHANG, 2000]. Comme présenté dans [BRADSKI et KAEHLER, 2008], la lentille d’une caméra présente deux distorsions principales :

- Une distorsion radiale;
- Une distorsion tangentielle.

La calibration de la caméra est effectuée, pour une résolution d’image fixe, en modélisant la caméra comme une caméra à sténopé [HARTLEY et ZISSERMAN, 2003] et en utilisant une mire de calibration [BOUGUET, 2008]. Cela permet d’exprimer les coordonnées de la caméra dans un espace normalisé [ZHANG, 2000]. Comme présenté sur la Figure 2.6, la calibration permet d’estimer les distorsions de la lentille de la caméra, et ainsi corriger les distorsions des images acquises par la caméra (cf. Figure 2.7).

La calibration permet également de déterminer la matrice des paramètres intrinsèques \mathbf{K} de la caméra, nécessaire pour passer du référentiel *Caméra* au plan *Image*. Une visualisation des deux référentiels est présentée sur la Figure 2.8.

$$\mathbf{K} = \begin{bmatrix} f_x & s & u \\ 0 & f_y & v \\ 0 & 0 & 1 \end{bmatrix} \quad (2.13)$$

Avec :

- (f_x, f_y) , les distances focales de la caméra.
- (u, v) , les coordonnées du point principal de la caméra sur le plan *Image*;

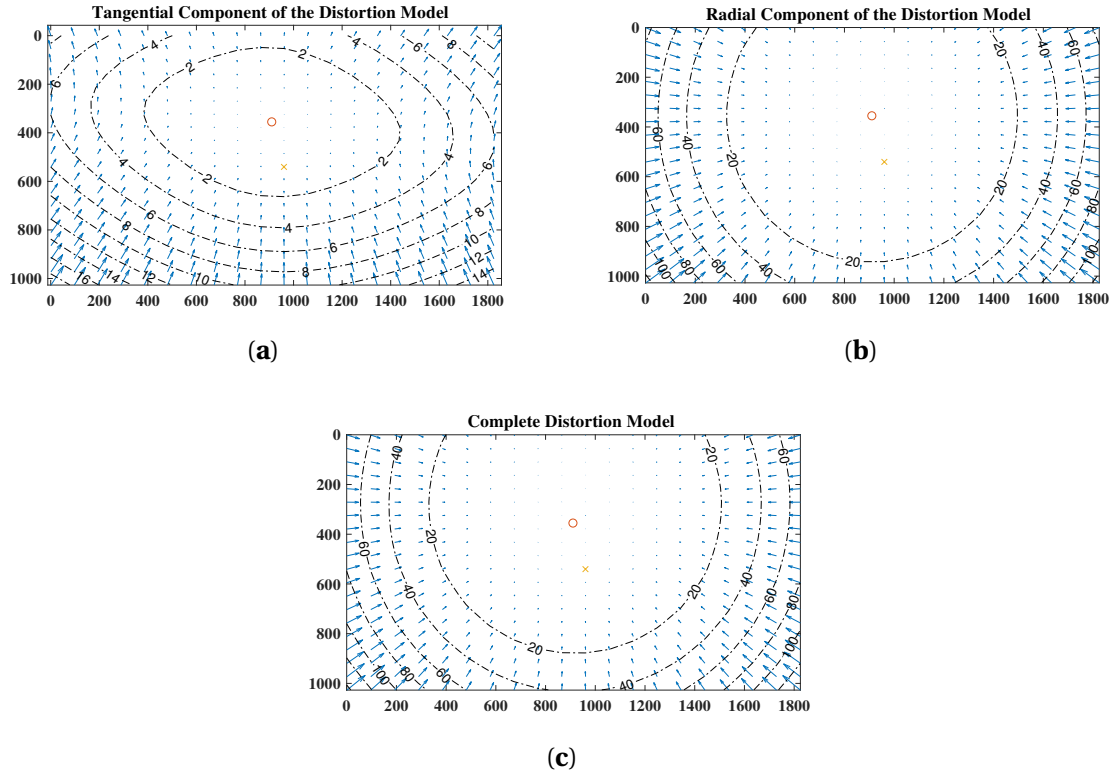


FIGURE 2.6 – Estimation des distorsions de la lentille de la caméra : (a) Distorsion tangentielle; (b) Distorsion radiale; (c) Distorsion complète.



FIGURE 2.7 – Correction des distorsions de la lentille de la caméra : (a) Image avec distorsions; (b) Image avec distorsions corrigées.

— s , le facteur d'orthogonalité du repère *Image*;

Les pixels n'étant pas parfaitement carrés, deux distances focales sont définies, i.e. sur l'axe horizontal et sur l'axe vertical. Ces distances correspondent au ratio entre la distance focale de la lentille de la caméra et les dimensions des pixels sur chacun des axes.

Comme défini dans [HARTLEY et ZISSERMAN, 2003], la transformation linéaire, permettant de passer des points \mathbf{X}_c dans le référentiel *Caméra* aux points \mathbf{x}_i dans le plan *Image*, est définie par les équations suivantes. Il est à noter que les formules sont exprimées en coordonnées homogènes, représentées avec le symbole ($\tilde{\cdot}$).

$$\tilde{\mathbf{x}}_i = \mathbf{K} \left[\mathbf{I} \mid 0 \right] \tilde{\mathbf{X}}_c \quad (2.14)$$

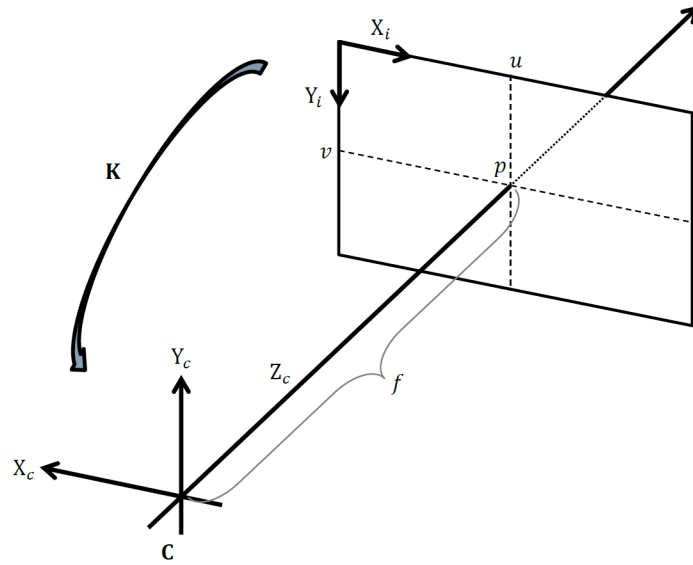


FIGURE 2.8 – Visualisation de l’association entre le référentiel *Caméra* et le plan *Image*.

2.4.3 Unité de mesures inertielles et magnétiques

La plateforme utilisée pour les mesures inertielles et magnétiques est la plateforme ULISS (Ubiquitous Localization with Inertial Sensors and Satellites) développée par le laboratoire IFSTTAR-GEOLoc [ORTIZ et al., 2017]. Cette plateforme intègre des capteurs similaires à ceux des smartphones (i.e. des capteurs bas-coût). A la différence de la plupart des équipements mobiles grands publics, cette plateforme permet de disposer de données brutes, sans les filtres généralement appliqués aux signaux des capteurs. Cette plateforme intègre une unité de mesure inertielle et magnétique (IMMU) à trois axes, un baromètre, un récepteur GPS à haute sensibilité (HS-GPS) et une antenne. Tous les repères des capteurs sont orthogonaux et co-alignés avec la même origine. C’est pourquoi un référentiel unique est défini pour tous les capteurs de la plateforme ULISS, i.e. le référentiel *Body* défini dans sous-section 2.2.3.

Gyromètre

Un gyromètre de type MEMS, i.e. *Vibratory Gyroscopes*, est un capteur permettant de mesurer la vitesse angulaire, i.e. la vitesse de rotation autour d’un axe. Le gyromètre repose sur l’observation des forces de Coriolis, i.e. les variations d’accélération de Coriolis d’un élément vibrant. Un élément du gyromètre possède un mode de vibration propre. Lors d’un mouvement de rotation, l’accélération de Coriolis $\mathbf{a}_{Coriolis}$ va être modifiée suivant la vitesse angulaire ω_{gyro} issue de ce mouvement.

$$\mathbf{a}_{Coriolis} = 2 \times \omega_{gyro} \wedge \mathbf{v}_r \quad (2.15)$$

Avec :

- $\mathbf{a}_{Coriolis}$, l’accélération de Coriolis;
- \mathbf{v}_r , la vitesse relative de l’élément en vibration par rapport au gyromètre;
- ω_{gyro} , la vitesse angulaire du gyromètre dans le référentiel *Inertiel*.

Pour identifier et modéliser les erreurs stochastiques des gyromètres intégrés dans l’IMMU, i.e. les principales composantes du bruit, des analyses de la variance d’Allan ont été effectuées sur les signaux acquis au cours d’une longue phase statique. Ces sorties permettent de déterminer l’évolution temporelle des erreurs du capteur. Pour la mesure du moment angulaire ω_{gyro}^b exprimé dans le référentiel *Body*, un biais de mesure \mathbf{b}_ω et un bruit blanc \mathbf{n}_ω ont été identifiés. La modélisation du signal du gyromètre est :

$$\mathbf{y}_g = \omega_{gyro}^b + \mathbf{b}_\omega + \mathbf{n}_\omega \quad (2.16)$$

Avec :

- \mathbf{y}_g , la mesure du gyromètre;
- ω_{gyro}^b , la vitesse angulaire dans le référentiel *Inertiel* et exprimée dans le référentiel *Body*;
- \mathbf{b}_ω , le biais du gyromètre .
- \mathbf{n}_ω , un bruit blanc gaussien centré de variance σ_ω^2 ;

La vitesse angulaire est exprimée en $rad.s^{-1}$ ou $d^\circ.s^{-1}$.

Accéléromètre

Un accéléromètre de types MEMS, i.e. *Pendulous and Vibrating-Beam Accelerometer*, est un capteur permettant de mesurer la force spécifique. L'accéléromètre repose sur le principe d'estimation du déplacement d'une masse attachée à un ressort par rapport à un point d'équilibre. Lors d'un mouvement, l'accélération \mathbf{a}_{acc} est proportionnelle au déplacement de la masse m .

$$\mathbf{a}_{acc} = \mathbf{g} - \frac{\mathbf{K}}{m}\mathbf{x} \quad (2.17)$$

Avec :

- \mathbf{x} , le déplacement de la masse m ;
- \mathbf{a}_{acc} , l'accélération de la masse m dans le référentiel *Inertiel*;
- \mathbf{g} , la gravité terrestre;
- \mathbf{K} , la raideur du ressort.

La force spécifique f_{acc} dans le référentiel *Inertiel* est alors déduite :

$$\mathbf{f}_{acc} = \mathbf{a}_{acc} - \mathbf{g} \quad (2.18)$$

De manière similaire au gyromètre, pour identifier et modéliser les erreurs stochastiques des accéléromètres intégrés dans l'IMMU, des analyses de la variance d'Allan ont été effectuées sur les signaux acquis au cours d'une longue phase statique. Ces sorties permettent de déterminer l'évolution temporelle des erreurs du capteur. Pour la mesure de la force spécifique \mathbf{f}_{acc}^b exprimée dans le référentiel *Body*, un biais de mesure \mathbf{b}_a et un bruit blanc \mathbf{n}_a ont été identifiés. La modélisation du signal de l'accéléromètre est :

$$\mathbf{y}_a = \mathbf{f}_{acc}^b + \mathbf{b}_a + \mathbf{n}_a \quad (2.19)$$

Avec :

- \mathbf{y}_a , la mesure de l'accéléromètre;
- \mathbf{f}_{acc}^b , la force spécifique dans le référentiel *Inertiel* et exprimée dans le référentiel *Body*;
- \mathbf{b}_a , le biais de l'accéléromètre .
- \mathbf{n}_a , un bruit blanc gaussien centré de variance σ_a^2 ;

La force spécifique est exprimée en $m.s^{-1}$ ou en g . Il est à noter que le g utilisé pour la mesure de la force spécifique ne fait pas partie du système international (SI), qui utilise ce symbole pour le gramme.

Magnétomètre

Un magnétomètre est un capteur permettant de mesurer un champ magnétique. Le magnétomètre repose sur le changement de propriétés électriques d'un élément lorsque celui-ci est placé dans un environnement où le champ magnétique fluctue.

Afin de mesurer le champ magnétique externe \mathbf{B} , une calibration du magnétomètre est requise. Pour déterminer les paramètres de calibration qui doivent être appliqués au magnétomètre, un environnement sans sources artificielles de champ magnétique est nécessaire [RENAUDIN et al., 2010]. Cela permet de mesurer le champ géomagnétique de la Terre \mathbf{B}_{Earth} , et ainsi d'atténuer les effets de fer doux et durs, i.e. *soft-iron* et *hard-iron*, induits par les éléments ferromagnétiques de la plateforme d'acquisition dans laquelle est intégré le magnétomètre. Ainsi, une phase de calibration est appliquée sur site avant le début de chaque expérience pour déterminer les paramètres qui doivent être appliqués au magnétomètre pour mesurer le champ magnétique \mathbf{B} . Le facteur d'échelle, la non-orthogonalité des axes et le biais sont supprimés après cette phase de calibration. Par conséquent, seul un bruit blanc \mathbf{n}_m est considéré dans le modèle de mesure du champ magnétique \mathbf{B}_{cal} . La modélisation du signal du magnétomètre est :

$$\mathbf{y}_m^b = \mathbf{m}^b + \mathbf{n}_m \quad (2.20)$$

Avec :

- \mathbf{y}_m^b , la mesure du magnétomètre;
- \mathbf{m}^b , le champ magnétique local exprimé dans le référentiel *Body*;
- \mathbf{n}_m , un bruit blanc gaussien centré de variance σ_m^2 ;

Le champ magnétique est exprimée en T (Tesla) ou en G (Gauss).

Plateforme d'acquisition

Comme présenté dans les sous-sections précédentes, la plateforme d'acquisition représentative des équipements grand public du piéton se compose d'une IMMU et d'une caméra monoculaire. Les deux dispositifs sont fixés rigidement ensemble, comme illustré sur la Figure 2.9. Les récepteurs GPS intégrés dans ces derniers permettent de synchroniser les images acquises par la caméra avec les données acquises par l'IMMU grâce aux marqueurs de temps GPS.



FIGURE 2.9 – Plateforme d'acquisition : la plateforme ULISS et la caméra monoculaire Garmin VIRB.

Du fait que l'estimation de l'orientation, basée sur les données inertielles et magnétiques, est exprimée sous forme de quaternion \mathbf{q}_b^n , i.e. l'orientation du référentiel *Body* exprimée dans le référentiel *Navigation*, une conversion de la matrice de rotation estimée \mathbf{R}_c^n est nécessaire pour se replacer dans le même référentiel. En se basant sur le montage de la plateforme d'acquisition présenté sur la Figure 2.10, i.e. la caméra montée avec l'unité de mesures inertielles et magnétiques, la matrice de rotation \mathbf{R}_b^c du référentiel *Body* exprimée dans le référentiel *Caméra* est alors définie. Il est défini que le référentiel *Caméra* est co-aligné avec le référentiel *Body*, et qu'ils sont reliés par la transformation suivante :

$$\mathbf{R}_b^c = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (2.21)$$

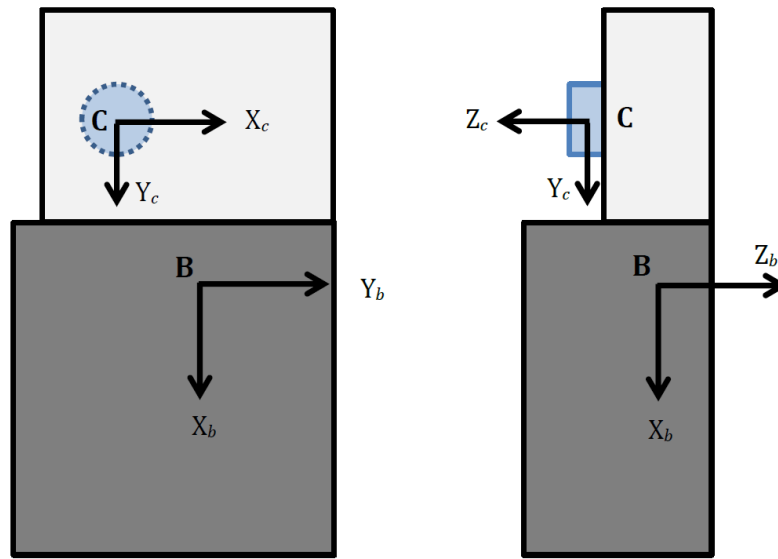


FIGURE 2.10 – Visualisation des repères associés au body et à la caméra.

2.5 Conclusion

Dans le contexte de la navigation piétonne en milieu urbain, les référentiels, les représentations de l'orientation et les conversions utilisés dans les différentes solutions proposés ont été introduit. Les différents capteurs intégrés dans les équipements grand public à disposition du piéton ont également été listés. La plateforme expérimentale intégrant ces différents capteurs, leur calibration et modèle de mesure ont ensuite été présentés et détaillés. Il est donc à noter que les solutions présentées dans la suite de ce mémoire seront limitées à l'usage exclusif des capteurs de la plateforme expérimentale définie dans ce chapitre. De plus, afin de correspondre à l'utilisation des équipements mobiles du piéton et aux usages de ces derniers, la plateforme d'acquisition est tenue en main lors des déplacements en milieu urbain. En aucun cas la plateforme expérimentale n'est fixée sur le corps du piéton. Ces limitations sont d'autant plus importantes que certaines solutions de l'état de l'art présentées dans le chapitre suivant ne respectent pas certaines hypothèses nécessaires pour mettre en place ces solutions, et ne peuvent donc pas être appliquées dans la cadre des travaux présentés dans cette thèse.

Chapitre 3

Méthodes usuelles de localisation du piéton et d'estimation de la pose

« Le piéton ne va nulle part - mais, en revanche, il est partout, partout où il ne devrait pas être. C'est une sorte d'ennemi qui a été créé pour rendre la circulation difficile. »

Sacha Guitry

Sommaire

3.1 Introduction	20
3.2 Méthodes usuelles de localisation du piéton	20
3.2.1 Localisation basée sur la vision	20
3.2.2 Localisation basée sur les données inertielles et magnétiques	22
3.2.3 Localisation basée sur des solutions hybrides	22
3.3 Méthodes usuelles d'estimation de la pose	23
3.3.1 Estimation de la pose sans la vision	23
3.3.2 Estimation de la pose par la vision	23
3.4 Conclusion	27

3.1 Introduction

Ce chapitre présente un état de l'art des méthodes usuelles de localisation du piéton et d'estimation de la pose d'un équipement, quel que soit la nature des capteurs mis en jeux. Une attention particulière est portée quant aux hypothèses sur lesquelles reposent ces méthodes, ainsi que les contraintes nécessaires à leurs applications. Dans le contexte de navigation piétonne en milieu urbain avec un équipement tenu en main, leurs applicabilités constitue un point essentiel de l'analyse de ces méthodes. Il est donc important de noter que de plus en plus de piétons disposent d'équipements mobiles (e.g. smartphones, tablettes et objets connectés) intégrant des capteurs, souvent bas coût, de type *Micro Electro Mechanical System* (MEMS). Ces équipements se composent principalement d'une caméra monoculaire à l'arrière et d'une unité de mesures inertielles et magnétiques tri-axial (IMMU) [KHAN et al., 2013; LATEGAHN et al., 2013]. Certains équipements peuvent également embarquer un baromètre, plusieurs caméra à l'arrière et une caméra monoculaire à l'avant de l'équipement. Il est à noter que des solutions reposant sur la communication avec une infrastructure (e.g. antennes relais téléphoniques, bornes Wi-Fi, Ultra Large Bande (UWB), etc.) ne sont pas détaillées dans ce chapitre car elles ne répondent pas au critère d'autonomie défini pour ces travaux de thèse. La suite du chapitre se concentre sur les différentes solutions autonomes permettant d'estimer la localisation du piéton, de faire un suivi de ses déplacements et d'estimer la pose d'un équipement tenu en main dans le contexte de la navigation piétonne.

Afin de présenter les méthodes usuelles de localisation du piéton en milieu urbain et d'estimation de la pose d'un équipement tenu en main, l'état de l'art est décomposé en deux parties. La première partie est consacrée aux méthodes de localisation du piéton avec un équipement mobile tenu en main. La seconde partie est consacrée aux méthodes d'estimation de la pose. Dans cette seconde section, un certain nombre d'approches présentées sont proposées dans le contexte de la réalité augmentée en milieu extérieur, qui constitue l'un des objectifs des travaux de cette thèse.

3.2 Méthodes usuelles de localisation du piéton

Dans le contexte de la navigation piétonne, l'estimation de la localisation est classiquement réalisée grâce à la réception des signaux GNSS [ZANDBERGEN et BARBEAU, 2011]. Cependant, comme cela est introduit dans la sous-section 2.4.1, dans les espaces urbains et intérieurs, le positionnement par GNSS est très imprécis ou même impossible. En tirant profit des capteurs intégrés dans les équipements grand public, des méthodes alternatives ont été proposées dans le domaine de la navigation piétonne pour résoudre le problème de localisation, qu'elles soient basées sur la vision, l'inertiel et le magnétisme, la pression ou sur une hybridation de ces dernières. Ces solutions reposant sur l'extraction d'informations permettant de remonter à la position ou d'en faire une estimation, et ce, en mettant à profit les capteurs à disposition.

3.2.1 Localisation basée sur la vision

Pour la localisation du piéton, la vision est classiquement utilisée avec des techniques basées sur le contenu des images, telle que *Content-Based Image Retrieval* (CBIR) [CUMMINS et NEWMAN, 2011; HALLQUIST et ZAKHOR, 2013; LYNEN et al., 2015; MORITA et al., 2005; VENTURA et al., 2014] ou sur la reconnaissance d'objets [ANATI et al., 2012]. La localisation peut également être réalisée en utilisant des techniques basées sur l'apprentissage automatique, telle que *Bag of Visual Words* (BoVW) [SIVIC et ZISSERMAN, 2003]. Ainsi, dans le contexte de la navigation piétonne dans les zones où la réception des signaux GNSS est difficile, [RUOTSALAINEN et al., 2010] propose une vue d'ensemble des méthodes basées sur la vision pour l'estimation de la position et du mouvement.

Localisation basée sur la reconnaissance de texte

Un premier type d'approche pour estimer une localisation approximative du piéton consiste à reconnaître des chaînes de caractères dans les images acquises par la caméra. L'extraction des chaînes de caractères est faite grâce à des méthodes de reconnaissance optique de caractères (*Optical Character Recognition-OCR*) [SMITH, 2007]. Ces chaînes de caractères sont ensuite comparées avec une base de données contenant des mots-clés ou des chaînes de caractères plus spécifiques associées à leur localisation. Une mise en correspondance permet alors de déduire une localisation approximative du piéton [IM et al., 2018; MISHRA et al., 2013].

Il est à noter que cela nécessite au préalable d'avoir renseigné une base de données avec les chaînes de caractères et leur localisation associée, ce qui n'est pas évident à mettre en place à grande échelle, et encore moins à maintenir dans le temps. Les contraintes d'utilisation d'équipements grand public, ne nécessitant pas de connexions, limitent ces solutions à utiliser de petites bases de données.

Localisation basée sur la reconnaissance d'informations contenues dans des bases de données

Un second type d'approche pour estimer la localisation du piéton consiste à comparer des informations extraites des images acquises par la caméra avec celles contenues dans une base de données. Cette dernière contenant des images géolocalisées ou associées à leurs poses globales [HALLQUIST et ZAKHOR, 2013; TAKACS et al., 2008; VENTURA et HOLLERER, 2012]. Elle peut également contenir des informations sur des éléments dont la localisation est connue, des informations sur le contenu d'une scène, etc.

Généralement la comparaison se fait en utilisant des points d'intérêt (POI), reposant sur des détecteurs et des descripteurs locaux. Ces points d'intérêt sont ensuite comparés et mis en correspondance avec ceux contenus dans une base de données. Ces mises en correspondances sont définies, après exclusion des valeurs aberrantes, par système de vote en fonction de critères spécifiques [ZHANG et KOSECKA, 2006]. [KNOPP et al., 2010] améliore la mise en correspondance avec une base de données et la reconnaissance de lieux, en mettant en place un critère de confusion permettant de ne pas prendre en compte les parties des images présentant un niveau élevé de mauvaises mises en correspondance.

Pour la reconnaissance et la mise en correspondance, il existe plusieurs représentants des caractéristiques locales de l'image qui sont reconnus pour être robuste à des changements d'échelle et/ou des transformations affines. Parmi les plus utilisés, on distingue les points d'intérêt reposant sur la géométrie, et ceux reposant sur la texture. Parmi les détecteurs et les descripteurs de points d'intérêt les plus utilisés, on peut citer Harris [HARRIS et STEPHENS, 1988], SIFT [LOWE, 1999, 2004], FAST [ROSTEN et DRUMMOND, 2006], SURF [BAY et al., 2008], BRISK [LEUTENEGGER et al., 2011], ORB [RUBLEE et al., 2011] et FREAK [ALAHY et al., 2012]. Le choix est généralement motivé par un compromis entre robustesse, efficacité et coût de calcul. Une comparaison de plusieurs de ces détecteurs et descripteurs, ainsi que différents tests, sont présentés dans [KUMAR, 2014; MIKOLAJCZYK et SCHMID, 2004, 2005; SCHAEFFER, 2013]. Selon le nombre de mises en correspondance, ces points d'intérêt permettent la détection et la reconnaissance d'éléments connus dans la scène observée [ANATI et al., 2012; FRITZ et al., 2006], permettant ainsi d'en déduire la localisation. Cependant, la précision de ces mises en correspondance se dégrade fortement sous l'effet des changements d'apparence causés par les différences d'illumination. Pour améliorer la robustesse face aux différences saisonnières et aux changements d'éclairage jour-nuit, la base de données doit contenir des images capturées en toutes saisons [NASEER et al., 2015; VALGREN et LILIENTHAL, 2008, 2010], et sous différents éclairages de la journée [JAIN et al., 2017; MILFORD et WYETH, 2012; ZHOU et al., 2016].

Pour estimer la localisation en comparant les images acquises à une banque d'images, certains développent leur propre base de données [CAPPELLE et al., 2007], tandis que d'autres utilisent une banque d'images telle que *Google Street View* ou *Microsoft Streetside*, dont les images sont géoréférencées [HALLQUIST et ZAKHOR, 2013; MAJDIK et al., 2015; YU et al., 2016]. Ainsi, afin d'être ef-

ficaces, ces approches nécessitent des bases avec un volume important de données. Il est évident que de telles bases de données ne sont pas toujours disponibles, pas toujours facile à préparer, et peut aussi ne pas être applicable aux appareils mobiles du piétons en raison de leurs ressources informatiques limitées.

[SALARIAN et ANSARI, 2015] propose d'exploiter la connaissance approximative de la localisation, obtenue par GNSS, pour envoyer une requête à *Google Street View* et ainsi générer les images à proximité du point de vue de la caméra. Les images générées permettent ensuite de raffiner la localisation grâce aux techniques de vision précédemment évoquées. [ANATI et al., 2012; MISHRA et al., 2013] proposent de réduire l'espace de recherche dans une base de données en utilisant la reconnaissance d'objet ou de contenu textuel. Si une unité de mesures inertielles est disponible, les données acquises (i.e. les mesures d'orientation) permettent de réduire encore plus l'espace de recherche dans une base de données d'images géolocalisées, et ainsi de réduire le temps de calcul pour trouver une correspondance [CAPPELLE et al., 2007; SALARIAN et ANSARI, 2015].

3.2.2 Localisation basée sur les données inertielles et magnétiques

Dans le contexte de la navigation piétonne, la navigation à l'estime du piéton (*Pedestrian Dead-Reckoning* -PDR), reposant sur une unité de mesures inertielles et magnétiques, est une technique largement utilisée pour déterminer en continu la localisation en milieu urbain et intérieur [FISCHER et al., 2013; JIMENEZ et al., 2009]. La localisation est déterminée de manière incrémentale en calculant le déplacement par rapport à une initialisation. Pour cela, les signaux inertiels sont généralement intégrés selon une mécanisation dite *strap-down* (i.e. les capteurs sont fixés sur le corps). Cette méthode n'est valable que si l'erreur accumulée est fréquemment étalonnée ou réinitialisée, e.g. en utilisant des moments de vitesse nulle. Cependant, lorsque l'équipement intégrant les capteurs est tenu en main durant la marche, les moments de vitesse nulle se produisent rarement. Par conséquent, au lieu de faire une double intégration des accélérations mesurées, un modèle de longueur de pas et une estimation du cap de marche sont alors utilisés pour estimer les déplacements du piéton et sa localisation [ANTIGNY et al., 2017; RENAUDIN et al., 2012]. Une solution pour estimer les longueurs de pas du piéton avec les capteurs tenus en main est présenté dans la section 5.3

3.2.3 Localisation basée sur des solutions hybrides

Pour la navigation à l'estime, [HIDE et al., 2010; MOURIKIS et ROUMELIOTIS, 2007; RUOTSALAINEN et al., 2010] proposent une navigation inertielle assistée par la vision en utilisant un filtre de Kalman. Cela permet d'estimer la localisation, le cap de marche, les informations de vitesse, mais aussi les erreurs et la dérive du positionnement. [HESCH et al., 2014] améliore la localisation en proposant un estimateur de l'imprécision de mesure. [JIANG et al., 2004] présente une solution de suivi hybride où les mesures du gyromètre sont utilisées pour prédire l'orientation et la position des lignes dans les images, et la dérive du gyromètre est corrigée par le suivi de la vision. Leur système combine un détecteur de contours pour une localisation précise, des mesures gyroscopiques pour traiter les mouvements rapides, des mesures d'accéléromètre et de magnétomètre pour éviter la dérive. [KLEIN et DRUMMOND, 2003] utilise les données de l'unité de mesures inertielles pour fournir une première estimation d'orientation et ainsi corriger les images.

Ainsi, la plupart de ces approches hybrides permettent d'améliorer la localisation grâce à une compensation mutuelle des erreurs et à une estimation de la dérive des capteurs. En environnements urbains, ces solutions ne permettent d'obtenir qu'une précision de positionnement de plusieurs mètres. Dans un objectif seul de localisation pour la navigation piétonne, une précision métrique est suffisante, mais est généralement insuffisante pour proposer un affichage qualifié en réalité augmentée sur site.

3.3 Méthodes usuelles d'estimation de la pose

La seconde partie de l'état de l'art est consacrée aux méthodes d'estimation de la pose, principalement dans le contexte de la réalité augmentée. Comme introduit dans la section précédente, la pose peut être déterminée grâce à une estimation de la localisation, et ensuite complétée avec une mesure de l'orientation. Il est important de noter que la pose complète comprend les six degrés de liberté que sont la localisation 3D et l'orientation 3D. Ainsi, les méthodes de localisation précédemment présentées ne peuvent être mises à profit que si la localisation 3D est estimée.

Plusieurs approches ont été développées pour calculer la pose d'un équipement mobile à l'aide de capteurs intégrés. Il est à noter que pour estimer la pose, nombre de ces approches utilisent des dispositifs spécifiques. Il est cependant possible de les classer en tenant compte de leurs entrées.

- Les méthodes ne reposant pas sur la vision;
- Les méthodes basées sur la vision uniquement;
- Les méthodes hybrides.

3.3.1 Estimation de la pose sans la vision

Les premiers systèmes pour estimer la pose en environnement extérieur reposent sur l'utilisation d'un récepteur GPS pour la localisation et d'un système estimant l'orientation de la tête [FEINER et al., 1997; HOLLERER et al., 1999]. [AZUMA et al., 1999] propose une approche de fusion permettant de combiner la vitesse angulaire d'un gyromètre avec l'orientation d'un magnétomètre intégré dans un appareil portable. [LEE et al., 2012] a développé l'application *CityView AR* fonctionnant sur smartphones et permettant de superposer des informations virtuelles sur une scène réelle. Cette dernière repose sur l'utilisation d'une unité de mesures inertielles intégrée dans les appareils mobiles et sur une estimation de la localisation. Ces approches permettent d'obtenir une estimation de la pose. Cependant un grand nombre de ces méthodes ne permettent d'atteindre qu'une précision plutôt faible et reposent sur des équipements pouvant être assez éloigné de ceux à disposition du piéton.

3.3.2 Estimation de la pose par la vision

Odométrie visuelle (VO)

De manière similaire à la navigation à l'estime du piéton, basée sur des mesures inertielles et magnétiques, l'odométrie visuelle (VO) est une première approche basée sur la vision, permettant d'estimer la pose relative d'un dispositif intégrant une ou plusieurs caméras, ainsi que des capteurs complémentaires. L'odométrie visuelle est largement utilisée pour la localisation et l'aide à la navigation de véhicules, robots et drones possédant une configuration de capteurs restreinte [AQEL et al., 2016; NISTÉR et al., 2004; SCARAMUZZA et FRAUNDORFER, 2011].

Dans des environnements inconnus, la pose de la caméra est déterminée de façon incrémentale en traitant un flux vidéo image par image. Une mise en correspondance de caractéristiques entre les images consécutives, suivi d'une étape de filtrage, permet d'estimer une pose relative à partir des correspondances valides. Cependant, un premier problème est que les poses relatives sont déterminées avec une ambiguïté d'échelle [AQEL et al., 2016], i.e. l'amplitude du déplacement est inconnue. Pour résoudre cette ambiguïté d'échelle, il est important de différencier les configurations matérielles utilisées. En effet, en fonction du type de caméra et des capteurs disponibles, la solution mise en place pour résoudre cette ambiguïté d'échelle n'est pas la même. On distingue ainsi les configurations suivantes :

- L'odométrie visuelle reposant sur l'usage d'une caméra monoculaire;
- L'odométrie visuelle reposant sur l'usage d'une caméra stéréoscopique;
- L'odométrie visuelle reposant sur l'usage d'une caméra et de capteurs complémentaires.

En fonction de la configuration utilisée, les différentes solutions proposées dans la littérature sont présentées dans les paragraphes suivants.

Malgré la résolution de l'ambiguïté d'échelle, l'estimation de l'échelle souffre d'erreurs lors de longs déplacements, dégradant la précision du suivi de la pose relative, constituant un second problème de l'odométrie visuelle [STRASDAT et al., 2010].

Odométrie visuelle monoculaire Dans le cas où la configuration matérielle est limitée à l'utilisation d'une caméra monoculaire, les poses relatives peuvent être localement estimées à l'échelle si des éléments, dont les dimensions sont connues, sont identifiés dans les images acquises par la caméra. Ces éléments peuvent être des objets 3D ou des motifs 2D connus [KNORR et KURZ, 2016; NÜTZI et al., 2011; WEISS et SIEGWART, 2011]. [DAVISON et al., 2007] initialise l'échelle en utilisant une mire de taille connue. [KLEIN et MURRAY, 2007] initialise également l'échelle en effectuant un déplacement de la caméra dont l'amplitude est connue.

Les différentes approches d'odométrie visuelle ne peuvent cependant pas éviter la dérive de l'échelle lorsqu'une caméra monoculaire est uniquement disponible [WANG et al., 2017]. La dérive peut être corrigée en utilisant la localisation basée sur l'apparence, en utilisant des repères connus [VENTURA et al., 2014] ou en utilisant la fermeture de boucle (*loop closure*) lorsque le dispositif mobile revisite un lieu [STRASDAT et al., 2010]. Bien que la fermeture de boucle soit une solution intéressante, elle reste ponctuelle et ne constitue pas une solution envisageable pour développer une solution d'aide à la navigation où le piéton irait d'un point A à un point B, sans repasser par un même lieu.

En tant qu'approche récente dédiée aux piétons, [KNORR et KURZ, 2016] utilise le visage comme objet connu lors de phases statiques. Sur un smartphone, deux caméras sont normalement installées : une caméra orientée vers l'utilisateur et une caméra orientée vers le monde. La caméra orientée vers l'utilisateur permet de capturer le visage, tout en exécutant l'odométrie visuelle avec la caméra orientée vers le monde. Lorsque la pose relative des deux caméras est calibrée, l'échelle de l'odométrie visuelle peut être calculée à partir de la taille du visage. Cette configuration est raisonnable pour les équipements tenus en main, cependant, du fait que le piéton est majoritairement en phase dynamique, l'approche n'est pas applicable dans le contexte de navigation piétonne.

Il convient de noter que la méthode proposée par [TOMAZIC et SKRJANC, 2015] pour l'estimation à l'échelle de l'odométrie visuelle monoculaire peut être similaire à l'approche proposée dans section 5.4. La solution utilise un podomètre et une longueur moyenne de pas [LEPPÄKOSKI et al., 2012]. La caméra doit cependant être fixée sur le corps du piéton, or cette contrainte n'est pas valide pour le contexte de ces travaux où le piéton tient son équipement en main.

Odométrie visuelle stéréoscopique Dans le cas où la configuration matérielle comprend une caméra stéréoscopique, la disparité entre les deux centres optiques permet d'estimer les distances entre la caméra et les différents éléments constitutifs de l'environnement. La connaissance de ces distances permet ensuite de résoudre l'ambiguïté d'échelle du processus d'odométrie visuelle en calculant la distance parcourue entre les images consécutives [WANG et al., 2017]. Il est à noter que lorsque des éléments éloignés sont utilisés, la précision de l'estimation de la pose est dégradée.

Odométrie visuelle et capteurs complémentaires Dans le cas où la configuration matérielle comprend un ou plusieurs autres capteurs en plus de la caméra, l'estimation de l'échelle peut être améliorée, tant sur son exactitude que sur sa stabilité. Ainsi, pour calculer et corriger l'échelle dans le processus d'odométrie visuelle, une unité de mesures inertielle et magnétiques ou un capteur de distance (e.g. LIDAR, télémètre IR, caméra RGB-D, projecteur de motif IR et caméra IR, etc.) sont généralement utilisés [CORKE et al., 2007; KNORR et KURZ, 2016; NÜTZI et al., 2011; WEISS et SIEGWART, 2011].

Dans l'estimation de l'échelle, une approche classique [NÜTZI et al., 2011] consiste à ajuster et à mettre à l'échelle la trajectoire de la caméra en estimant la trajectoire de l'unité de mesures

inertielles. La trajectoire de cette dernière est basée sur la double intégration de l'accélération après élimination des éléments de gravité. La double intégration étant sensible au bruit, l'erreur s'accumule rapidement, même si cette dernière est faible pendant un court intervalle de temps. Puisque l'unité de l'accéléromètre est métrique, cela permet d'estimer l'échelle. Cette méthode reste cependant sensible au biais dynamique qui est difficile à estimer.

Localisation et cartographie simultanées (SLAM)

La problématique de la localisation et de la cartographie simultanées (SLAM) est l'un des sujets de recherche les plus actifs depuis sa formulation dans les années 1980. De manière similaire à l'odométrie visuelle, le SLAM a pour objectif d'estimer de manière cohérente la trajectoire d'un appareil, mais également de reconstruire une carte de l'environnement parcouru. Ces deux points sont indispensables en particulier dans le domaine de la robotique mobile. Le SLAM est également une tâche préliminaire à de nombreux autres domaines. Il est ainsi devenu une solution de référence pour résoudre des problèmes tels que la navigation autonome, la planification de trajectoire, la réalité augmentée sur appareils mobiles [SCHOPS et al., 2014], etc.

Le SLAM visuel (vSLAM) [DAVISON et al., 2007; TAKETOMI et al., 2017], utilisant uniquement une ou plusieurs caméras, fait partie des méthodes de localisation et de cartographie les plus représentatives. Les caméras permettent d'extraire de nombreuses données sur l'environnement et constituent des capteurs relativement peu cher et facile à embarquer. Le SLAM visuel-inertiel (viSLAM) [NÜTZI et al., 2011], utilisant une unité de mesures inertielles, fait également partie des méthodes commençant à émerger ces dernières années.

Les nombreuses méthodes et configurations matérielles requises, ainsi que les nombreuses applications font du SLAM un domaine difficile à appréhender. Choisir la méthode la mieux adaptée à une application nécessite de connaître les limitations du système utilisé. Bien que le SLAM soit considéré comme une solution de référence, sa mise en œuvre peut soulever de nombreux problèmes en termes de configuration matériel, de limites de calcul et de mémoire. De fait, le SLAM devient moins attractif pour résoudre certains problèmes, notamment ceux étudiés dans ces travaux de thèse.

Estimation de la pose basée sur l'utilisation de données SIG et de modèles 3D

Pour estimer la pose, une solution est de se baser sur la reconnaissance d'objet dont le modèle 3D est connu. Les correspondances entre les points 3D du modèle et leurs projections 2D sur les images [MARCHAND et al., 2016] servent ainsi à estimer la pose en calculant la transformée géométrique pour passer de l'un à l'autre. L'une des approches les plus anciennes pour estimer la position et l'orientation est la Transformée Linéaire Directe (*Direct Linear Transform-DLT*) [SUTHERLAND, 1974] qui calcule la transformation pour un ensemble donné de points 2D ou 3D dans le même plan [HARTLEY et ZISSERMAN, 2003]. Travaillant sous des contraintes linéaires, certaines inexactitudes sont introduites. Les algorithmes *Perspective n-Points* (PnP) sont une alternative pour résoudre le problème d'estimation de la pose à partir de N (2D-3D) correspondances de points, et qui prend en compte les contraintes non linéaires. Ils se basent sur un calcul itératif pour estimer la transformée géométrique. Bien que les algorithmes P3P soient une solution bien connue, une approche PnP qui utilise plus de points est généralement préférée. En effet, la précision de position et d'orientation augmente avec le nombre de points. *POSIT* [DAVID et al., 2003; DEMENTHON et DAVIS, 1995] est l'un de ces algorithmes utilisés pour résoudre le problème PnP.

Dans la sous-section 4.5.1, l'approche présentée repose en partie sur l'utilisation de l'algorithme *Efficient PnP* (EPnP) [MORENO-NOGUER et al., 2007] pour estimer la pose de la caméra par rapport au modèle 3D de l'objet de référence, ainsi que les erreurs angulaires qui seront utilisées pour la définition du modèle d'estimation. L'algorithme EPnP est un processus d'estimation non itératif avec une meilleure précision et une complexité beaucoup plus faible que les méthodes non itératives, lui permettant d'être plus rapide. Cette solution pourrait également être remplacée par des algorithmes plus récents permettant d'atteindre une meilleure précision. Parmi les algo-

rithmes PnP les plus récents, on peut citer OPnP [ZHENG et al., 2013], UPnP [KNEIP et al., 2014a] et MLPnP [URBAN et al., 2016].

D'autres méthodes, comme le système de vision SCERPO [LOWE, 1987] ou le *Virtual Visual Servoing* (VVS) [COMPORT et al., 2006; MARCHAND et CHAUMETTE, 2002], tirent également profit de la connaissance d'un objet en utilisant directement son modèle 3D pour estimer une pose relative à ce dernier.

Pour estimer la position et l'orientation, des travaux plus récents proposent une intégration des mesures inertielles et GNSS avec le contenu d'un système d'information géographique. [BAATZ et al., 2011] exploite les informations 3D de bâtiments et les points de fuite pour supprimer la rotation 3D et afficher la position et l'orientation de la caméra dans un système de coordonnées globales. [ARTH et al., 2015] et [LIANG et al., 2015] exploitent une carte cadastrale 2D avec un modèle d'élévation pour générer une base de données d'images géoréférencées 2.5D. [CAPPELLE et al., 2007] et [GHINAMO et al., 2015] utilisent une connaissance préalable de l'environnement 3D et initialisent une caméra virtuelle pour construire des images synthétiques de l'environnement. Les mises en correspondance des images de la scène réelle avec les images de la scène virtuelle permettent de retrouver la pose de la caméra grâce à la pose associée à la caméra virtuelle. Pour faciliter le suivi de la vision, [MIN et al., 2012] utilise un SIG 3D avec un récepteur GNSS et une unité de mesures inertielles.

Estimation de la pose basée sur des solutions hybrides

Les solutions hybrides, intégrant généralement un processus de fusion, ont été introduites pour compenser l'inconvénient des approches basées sur l'apparence qui échouent lors des mouvements de rotation rapide en raison du flou généré et du grand déplacement dans les images [JIANG et al., 2004; SATOH et al., 2001; YOU et al., 1999]. Ces dernières tirent profit des capteurs à disposition dans les systèmes établis. Les erreurs de suivi s'accumulent rapidement au fil du temps, les approches fondées sur la fusion ont par conséquent été développées pour compenser chaque inconvénient. Un gyromètre présente l'avantage de pouvoir suivre avec précision les mouvements du capteur à haute fréquence pendant de courts intervalles de temps. En statique, un accéléromètre permet de compenser l'estimation de l'orientation globale en utilisant la mesure de la gravité. Ainsi, quelques unes de ces solutions sont présentées.

[YOU et al., 1999] propose un système hybride de suivi du déplacement combinant un suivi optique avec des données inertielles. [KURZ et al., 2014] combine les informations visuelles du patch autour d'un point d'intérêt avec des informations spatiales, basées sur un modèle de l'environnement et les mesures de l'accéléromètre et du magnétomètre. [SATOH et al., 2001] développe l'un des premiers systèmes portables de réalité augmentée, appelé *Town-wear*, utilisant un gyromètre de haute précision pour mesurer l'orientation. La dérive de mesure du gyromètre est ensuite compensée par des données de vision. [RIBO et al., 2002] décrit un système de suivi hybride qui peut être utilisé pour les applications de réalité augmentée en milieu extérieur. Une base de données de points caractéristiques, établie hors ligne, est utilisée avec un ensemble d'images de référence et un modèle 3D. [REITMAYR et DRUMMOND, 2006] présente un système portable pour la réalité augmentée en milieux urbains qui utilise un système de suivi hybride basé sur un détecteur de contour, avec des mesures gyroscopiques pour les mouvements rapides, des mesures de gravité et de champ magnétique pour atténuer la dérive pendant les mouvements à basse fréquence. La solution utilise des images de référence et la sélection d'images en ligne pour réinitialiser automatiquement après des occlusions ou des défaillances. [SCHALL et al., 2009] utilise une plateforme mobile dédiée et un filtre de Kalman est proposé pour fusionner un système GPS différentiel ou cinématique en temps réel (RTK) avec des hauteurs barométriques. Une unité de mesures inertielles, comprenant gyromètres et accéléromètres, et magnétomètres complète le système. Un traqueur d'orientation visuelle est appliquée pour corriger la dérive du magnétomètre.

3.4 Conclusion

Dans le contexte de la navigation piétonne en milieux urbains, différentes solutions de l'état de l'art pour estimer la localisation du piéton et estimer la pose d'un équipement tenu en main ont été présentées. Des choix ont été fait quant à l'autonomie et aux capteurs mis en jeu dans les solutions investiguées. Ces dernières devant respecter la contrainte d'usage des capteurs intégrés dans les équipements grand public à disposition du piéton. Il en ressort un certain nombre de solutions pouvant s'adapter à la navigation piétonne en milieux urbains. Cependant, ces solutions restent sensibles à des contraintes de fonctionnement ou d'installation. De plus, avec l'objectif d'estimer avec précision la pose pour proposer un affichage en réalité augmentée sur site, et ainsi donner la possibilité de faire de la caractérisation de contenu SIG 3D sur site, il apparaît que les solutions se basant sur des modèles 3D existant souffrent d'erreurs de positionnement. Par conséquent, cela implique des erreurs d'estimation du positionnement de plusieurs mètres. Or dans le cadre d'une caractérisation de contenu SIG 3D, aucune correction par alignement entre la scène réelle observée et les modèles 3D ne peut être faite pour palier à ces erreurs de positionnement et de reprojection. Afin de répondre à ces exigences, les travaux développés dans cette thèse reprendront certains concepts présentés dans ce chapitre, tout en proposant des innovations palliant aux limitations des solutions sur lesquelles elles reposent.

Chapitre 4

Pose absolue estimée par reconnaissance d'objets issus d'un SIG 3D

« Connaître, ce n'est point
démontrer, ni expliquer. C'est
accéder à la vision. »

Antoine de Saint-Exupéry

Sommaire

4.1 Introduction	30
4.2 Système d'Information Géographique	30
4.2.1 Définition	30
4.2.2 Constitution	30
4.2.3 Niveaux de détail	31
4.3 Choix des objets de référence	31
4.4 Détection et identification des objets de référence	33
4.4.1 Recherche d'éléments connus à proximité	33
4.4.2 Identification par reconnaissance de caractères	33
4.4.3 Reconnaissance par la texture et association au modèle 3D	35
4.5 Estimation de la pose de la caméra	36
4.5.1 Pose relative dans le référentiel <i>Objet</i>	36
4.5.2 Pose absolue dans le référentiel <i>Navigation</i>	37
4.6 Estimation des erreurs	40
4.6.1 Erreur de position des objets d'intérêt dans le référentiel <i>Navigation</i>	40
4.6.2 Erreur d'estimation des points géométriques dans le référentiel <i>Image</i>	41
4.6.3 Erreurs de reprojection dans le référentiel <i>Image</i>	42
4.7 Expérimentation : estimation de la pose basée sur la reconnaissance d'un objet de référence et affichage en réalité augmentée sur site	43
4.7.1 Description de l'expérimentation	43
4.7.2 Définition de la référence terrain	43
4.7.3 Résultats	44
4.7.4 Réalité augmentée qualifiée sur site	45
4.7.5 Occlusion	46
4.8 Conclusion	47

4.1 Introduction

Ce chapitre présente une solution permettant d'estimer, de manière absolue dans un référentiel global, la pose d'une caméra tenue en main par le piéton, et ce au moyen de la connaissance fine d'objets du mobilier des transports urbains contenus dans un Système d'Information Géographique (SIG) 3D. En effet, les SIG 3D constituent aujourd'hui un ensemble de données extrêmement riche et représentatif de l'environnement du piéton lors de ses déplacements en milieux urbains. La capacité à utiliser la connaissance de certains éléments finement détaillés contenus dans les SIG 3D pour estimer la pose absolue de la caméra constitue la principale innovation de ce chapitre. Cette estimation peut ensuite être utilisée et fusionnée avec des données complémentaires issues des capteurs inertiels et magnétiques intégrés dans l'équipement tenu en main par le piéton. L'estimation précise de la pose de caméra, permettant une visualisation en réalité augmentée sur site, constitue l'une des applications de ce chapitre. Cette visualisation s'intègre facilement dans le contexte de navigation piétonne en milieu urbain, notamment en enrichissant les applications d'aide à la mobilité du piéton. Cela permet de caractériser le niveau de "qualité" et de "précision" des données, et potentiellement de faire remonter des erreurs ou des manques dans le contenu des SIG 3D. Le cas échéant, cela permettrait également d'intégrer de nouveaux éléments SIG 3D dans le processus d'estimation de la pose de la caméra et ainsi d'étendre la couverture de la solution d'aide à la navigation du piéton en milieux urbains.

4.2 Système d'Information Géographique

4.2.1 Définition

Un Système d'Information Géographique est défini comme un système d'information permettant d'intégrer, de stocker et d'afficher des informations géographiques. L'usage des Systèmes d'Information Géographiques constitue la base de la géo-informatique, également appelée géomatique. Les SIG permettent de mettre en relation des données pouvant sembler très éloignées, mais toutes ayant une composante spatiale, qu'elles soient liées à l'aménagement du territoire, à la gestion des infrastructures et des réseaux, aux transports et à la logistique, aux télécommunications, à la recherche ou encore aux assurances. Les SIG permettent ainsi de réunir toutes ces dimensions autour d'un même référentiel, et sont à l'origine de nombreux services et applications basés sur l'analyse et l'édition des données spatiales, que ce soit pour la production de cartes ou la visualisation des données. En effet, les SIG comportent aujourd'hui des informations beaucoup plus riches que celles pouvant être représentées sous forme de cartes, notamment des modèles numériques 3D (e.g. objets, maquettes d'urbanisme, etc.) conçus grâce à la Conception Assistée par Ordinateur (CAO).

4.2.2 Constitution

Les modèles 3D urbains reposent sur une surface de référence : le Modèle Numérique de Terrain (MNT). Il est défini par mesures topographiques (e.g. LIDAR, photogrammétrie), à l'exclusion des bâtiments. Un pas plus ou moins fin est choisi pour le MNT en fonction de l'utilisation du modèle 3D. Les modèles 3D des bâtiments sont ensuite constitués de plusieurs éléments :

- les toitures;
- les façades et les ouvertures;
- les intersections avec le MNT.

Les modèles 3D urbains sont généralement produits, soit par extrusion des empreintes de bâti de la BD Topo de l'IGN, soit par projection des contours de toits sur le MNT. Une distinction est également faite entre les objets surfaciques (i.e. les murs) et les objets volumiques (i.e. les bâtiments). Ces données sont produites par les agences de l'État ou par la communauté avec les in-

formations géographiques volontaires (*Volunteered Geographic Information-VGI*) comme OpenStreetMap [GOODCHILD, 2007],

4.2.3 Niveaux de détail

Dans le cadre de la modélisation des bâtiments et des zones urbaines, l'OpenGeospatial Consortium (OGS¹) a défini la norme CityGML pour représenter la géométrie des bâtiments. Le modèle des données est un modèle évolutif intégrant le concept de niveau de détail (*Level of Detail-LOD*) permettant d'exprimer la résolution ou la complexité géométrique des modèles 3D. Cinq niveaux de détails, numérotés de 0 à 4, sont définis comme suit :

- LOD0 : correspond à une représentation du terrain, également défini comme le Modèle Numérique de Terrain. Ce niveau de détail est utilisé pour des modélisations à l'échelle régionale.
- LOD1 : correspond à une représentation des bâtiments sous forme de bloc à toits plats. Ce niveau de détail est utilisé pour des modélisations à l'échelle d'une ville ou d'une agglomération.
- LOD2 : correspond à une représentation des bâtiments avec leur structure de toit et des surfaces délimitées. Ce niveau de détail est utilisé pour des modélisations à l'échelle d'un quartier. Des éléments de végétation, de mobilier urbain et de réseau de transport peuvent également être modélisés.
- LOD3 : correspond à une représentation extérieure des bâtiments avec une structure détaillée (i.e. façades, toits, balcons, fenêtres, etc.). Ce niveau de détail est utilisé pour des modélisations architecturales de bâtiments ou d'infrastructures.
- LOD4 : correspond à une représentation intérieure et extérieure des bâtiments avec une structure détaillée (i.e. portes, escaliers, etc.). Ce niveau de détail est utilisé pour des modélisations architecturales de bâtiments en cohérence avec le modèle IFC (standard permettant de modéliser les intérieurs et extérieurs de bâtiments)



FIGURE 4.1 – Les cinq niveaux de détails définis dans CityGML.

Source : [BILJECKI et al., 2016].

4.3 Choix des objets de référence

Dans un esprit de promotion de la mobilité active et d'utilisation des transports en commun, l'attention est portée sur le mobilier des transports urbains, déployés sur une partie importante des villes, présentant des caractéristiques simples et facilement identifiables, et susceptible d'être rencontré lors de la mobilité du piéton. De nombreux éléments issus du mobilier des transports urbains peuvent être utilisés pour estimer la pose de la caméra. Ici, le terminal de la station de vélo de la ville de Nantes (cf. Figure 4.2) a été choisi comme premier objet de référence. Il présente des caractéristiques uniques : un modèle géométrique relativement simple et une implantation dans une centaine de lieux à travers la ville (cf. Figure 4.3-(a)). Il est intéressant de noter que l'implantation des stations vélo est plus dense dans le centre-ville et dans les quartiers périphériques (cf.

1. <http://www.opengeospatial.org/standards/citygml>

Figure 4.3-(b)). Cette configuration correspond aux lieux où les signaux satellitaires sont fortement dégradés (e.g. canyons urbains, lieux avec une forte densité de bâtiments ou lieux couverts) et où le piéton est le plus susceptible d'avoir besoin d'outils précis d'aide à la mobilité. Il est également à noter que ce modèle est construit par la société JCDecaux sous le nom générique "Cyclocity"², et est implanté en 2018 dans 28 villes à travers le monde. La solution proposée dans ce chapitre est donc potentiellement utilisable dans toutes ces villes.



FIGURE 4.2 – Exemple de station vélo de la ville de Nantes.

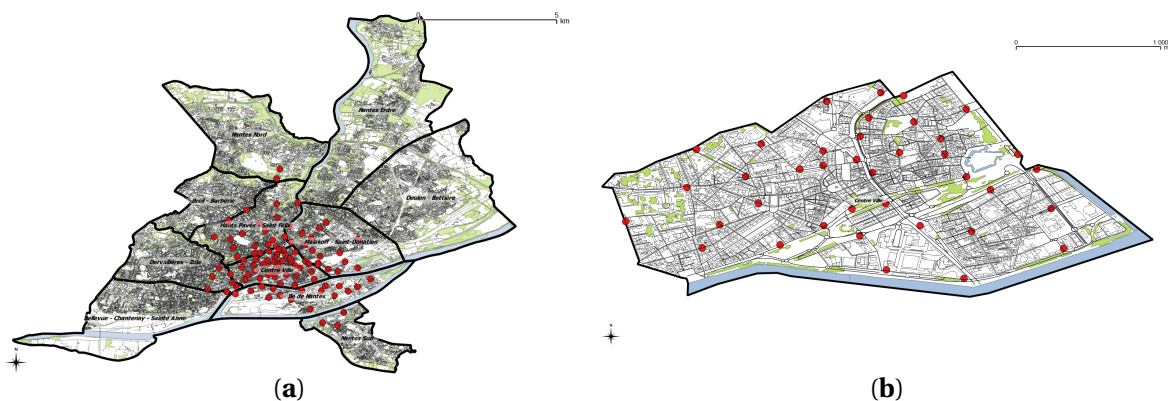


FIGURE 4.3 – (a) Implantation des stations vélo de la ville de Nantes; (b) Implantation des stations vélo dans le centre ville de Nantes.

Afin d'estimer la pose de la caméra tenue en main par le piéton lors de ses déplacements en milieu urbain, les dimensions géométriques exactes de l'objet de référence doivent être déterminées. Des mesures sur le terrain ont été collectées et traitées afin de compléter les informations fournies par la société JCDecaux (cf. Figure 4.4-(a)). Cela permet ainsi de pouvoir réaliser le modèle 3D de l'objet de référence avec un logiciel de Conception Assistée par Ordinateur (cf. Figure 4.4-(b)), e.g. *Solidworks* ou *SketchUp*. Dans ces travaux de thèse, le modèle 3D est ensuite exporté au format .obj pour être utilisé. Ce format comprend des surfaces polygonales décrites par un ensemble de sommets et de faces.

2. <http://www.cyclocity.com>

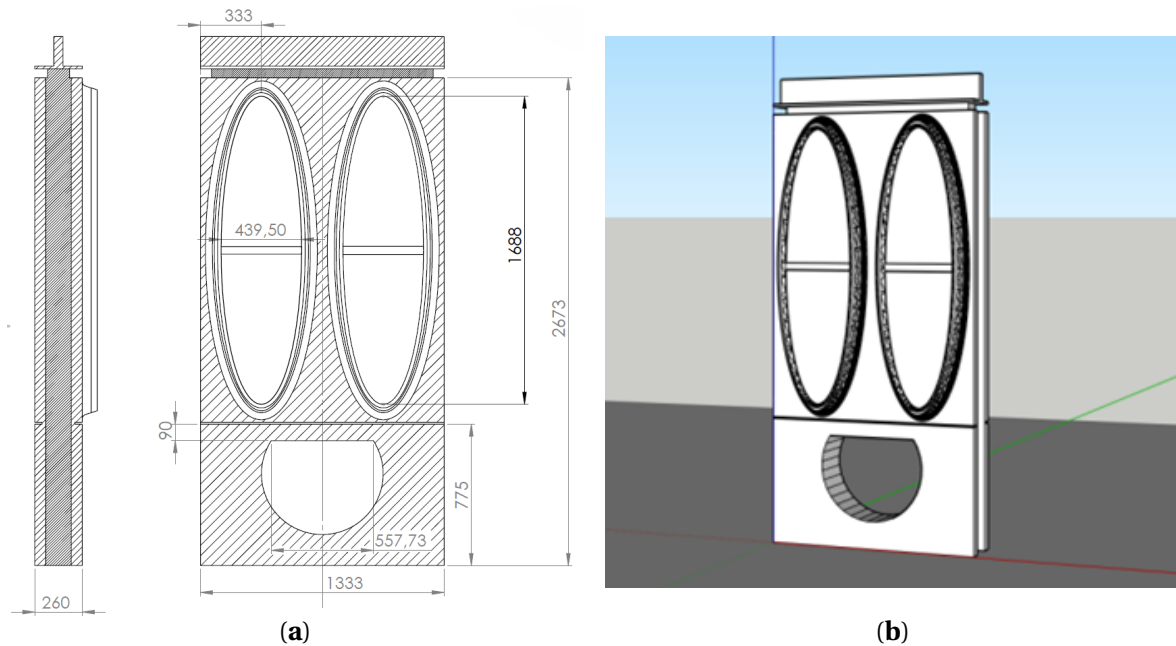


FIGURE 4.4 – (a) Côtes (en mm) du terminal de la station vélo; (b) Modèle 3D de la station vélo réalisé avec un logiciel de CAO.

4.4 Détection et identification des objets de référence

La détection et l'identification d'un objet de référence constitue une partie critique du processus présenté dans ce chapitre [HEISELE et al., 2006; HETTIGE et al., 2015]. Ainsi, afin d'estimer la pose basée sur la reconnaissance d'un objet de référence, plusieurs méthodes ont été étudiées pour retrouver les caractéristiques géométriques de l'objet de référence et optimiser le processus. Les détails des méthodes implémentées sont présentés dans les sous-sections suivantes, ainsi que leurs avantages et inconvénients. Les méthodes ne remplissant pas les critères nécessaires à la mise en place d'une solution autonome ont été écartées de la solution finale.

4.4.1 Recherche d'éléments connus à proximité

L'utilisation de bases de données mises à disposition par les métropoles³ contenant des informations relatives aux stations, i.e. leur nom, numéro et notamment la localisation approximative, permettent une recherche des éléments connus à proximité. Le calcul de la distance entre la localisation approximative des éléments connus et la localisation du piéton permet de limiter le nombre d'éléments à analyser, et ainsi d'accélérer le processus en filtrant les mises en correspondance possibles. Cette étape est donc préliminaire au processus de détection et d'identification d'objets de référence, mais peut également être utilisée pour valider l'objet de référence reconnu.

Il est à noter que la localisation du piéton utilisée pour cette recherche d'éléments connus à proximité peut être définie au moyen d'un système de positionnement par satellites. Les objets de référence étant espacés de plusieurs centaines de mètres, les imprécisions des GNSS, pouvant aller jusqu'à plusieurs dizaines de mètres, n'empêchent pas de déterminer les éléments connus à proximité.

4.4.2 Identification par reconnaissance de caractères

De manière similaire à l'utilisation des GNSS, la reconnaissance de texte et la comparaison avec les noms et numéros contenus dans les bases de données, mises à disposition par les métropoles, permet de valider la bonne reconnaissance d'un objet de référence. L'approche proposée

3. <http://data.nantes.fr/>

ne permet pas d'extraire les points géométriques du modèle 3D, de ce fait, elle n'est pas nécessaire au fonctionnement de la solution finale, mais constitue un outil de validation de l'identification des objets de référence. A la différence de l'approche basée sur le positionnement par satellites, l'approche proposée ici présente l'avantage de pouvoir être utilisée en continu. Cependant, l'environnement urbain contient un nombre important de textes (e.g. enseignes, publicités, panneaux de signalisation, etc.), c'est pourquoi une analyse complète des images serait inutile et coûteuse en temps de calcul. Ainsi, seule la zone de l'image contenant l'objet de référence est analysée, ce qui permet de conserver une certaine rapidité d'exécution de la solution. L'Histogramme de Gradient Orienté (HOG), proposé par [DALAL et TRIGGS, 2005], est utilisé afin d'identifier les zones de textes potentiellement intéressantes pour la reconnaissance de caractères. En effet, HOG permet d'extraire des points ayant des caractéristiques géométriques particulières, ce qui est notamment le cas pour du texte. Ces points caractéristiques sont calculés à partir des histogrammes locaux de l'orientation du gradient sur des zones de l'image. Ils possèdent des points communs avec les points d'intérêt SIFT et les histogrammes d'orientation de contours, mais diffèrent notamment par l'utilisation d'une grille dense à la manière des superpixels. L'idée importante derrière le descripteur HOG est que l'apparence et la forme locale d'un élément dans une image peuvent être décrites par la distribution de l'intensité du gradient ou la direction des contours. La combinaison des histogrammes forme alors le descripteur HOG. Pour de meilleurs résultats, les histogrammes locaux sont normalisés en contraste, en calculant une mesure de l'intensité sur des zones plus larges que les cellules, et en utilisant cette valeur pour normaliser toutes les cellules du bloc. Cette normalisation permet une meilleure résistance aux changements d'illuminations et aux ombres. Une visualisation des points d'intérêt extraits par HOG est présentée sur la Figure 4.5-(a). Une analyse de la densité des points d'intérêt HOG extraits permet ensuite de filtrer les zones de l'image ne présentant pas ou peu de points d'intérêt HOG, et donc pas ou peu de texte. La densité des points d'intérêt HOG est présentée sur la Figure 4.5-(b).

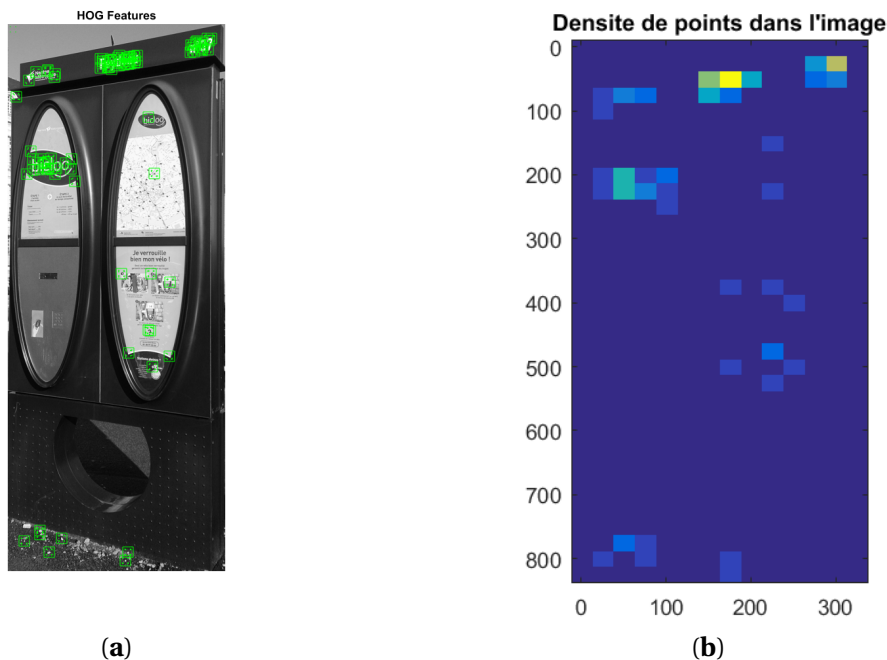


FIGURE 4.5 – (a) Points d'intérêt HOG extraits; (b) Densité des points d'intérêt HOG extraits.

Une fois les zones potentielles de textes identifiées, une reconnaissance optique de caractères (*Optical Character Recognition-OCR*) [SMITH, 2007] est appliquée sur ces zones. Cette reconnaissance, basée sur l'apprentissage des caractères, renvoie un certain nombre de chaîne de caractères. La Figure 4.6-(a) présente une visualisation des chaînes de caractères identifiées par le processus OCR. Les chaînes de caractères identifiées par le processus d'OCR sont ensuite comparées aux données contenues dans le SIG contenant le noms et le numéro de chaque station. Le pro-

cessus OCR n'étant pas parfait et les typographies sur les objets d'intérêt n'étant pas pérennes dans le temps (i.e. effacement, décolllement), les chaînes de caractères estimées peuvent présenter des erreurs. Afin de pallier ce problème et retrouver le nom et/ou le numéro de la station correspondante, une mise en correspondance, basée sur le calcul de la distance de Levenshtein est faite entre les chaînes de caractères estimées par le processus OCR et les chaînes de caractères contenues dans le SIG. La distance de Levenshtein se définit comme la distance entre deux mots $Word_x$ et $Word_y$ correspondant au nombre minimal d'édition à faire pour transformer $Word_x$ en $Word_y$. Ainsi, la chaîne de caractères de la base de données ayant le plus petit score est considérée comme correspondance de la chaîne de caractère identifiée par le processus OCR. La mise en correspondance permet ainsi de retrouver les informations associées à la station vélo identifiée (cf. Figure 4.6-(b)).



(a)

IDOBJ	NOM COMPLET	LIBTYPE	ADRESSE	location
3715	30 Station Abonnement Commerce	Bicloo	Allée Brancas	[47.2131450294256 , -1.55738072263822]
3716	31 Station Accueil Bourse	Bicloo	Allée de la Bourse	[47.2120943808486 , -1.55910527758924]
3717	32 Station Accueil Place Foch	Bicloo	Place Maréchal Foch	[47.219342309251 , -1.55027785651826]
3718	33 Station Accueil Racine	Bicloo	5 Rue Racine	[47.2134832603266 , -1.5631877058962]
3719	34 Station Abonnement Médiathèque	Bicloo	24 Quai de la Fosse	[47.2109126549257 , -1.5618418449525]
3720	35 Station Abonnement Jean V	Bicloo	Place Jean V	[47.212180370735 , -1.56556289207478]
3721	36 Station Accueil Alger	Bicloo	5 Rue d'Alger	[47.210705238574 , -1.5666743136651]
3722	37 Station Accueil Feydeau	Bicloo	Rue Gaston Veil	[47.2114728989176 , -1.55688248342545]
3723	38 Station Abonnement Place Ricard	Bicloo	Chaussée de la Madeleine	[47.2120904234566 , -1.553136666507461]

(b)

FIGURE 4.6 – (a) Reconnaissance de caractères par OCR; (b) Mise en correspondance avec l'objet d'intérêt contenu dans le SIG.

4.4.3 Reconnaissance par la texture et association au modèle 3D

L'estimation de la pose du centre optique de la caméra est basée sur la reconnaissance d'un objet de référence parfaitement connu, et plus spécifiquement sur l'extraction de N points \mathbf{X}_o de l'objet de référence. Pour calculer une pose relative à un objet de référence, la contrainte de $N > 4$ points doit être respectée. Pour estimer la pose de manière robuste, $N = 10$ points correspondant à des caractéristiques géométriques facilement reconnaissables \mathbf{X}_o de l'objet ont été définis sur le modèle 3D de l'objet de référence (cf. cercles verts sur la Figure 4.7 et croix bleues sur le terminal de la station de vélo sur la Figure 4.8).

Un ensemble de points caractéristiques de référence \mathbf{x}_{SURF_ref} , extrait par SURF [BAY et al., 2008] a ensuite été établi hors ligne sur des images correspondant à deux vues différentes de chaque objet identifié (i.e. pour chaque terminal susceptible d'être rencontré par le piéton lors de ses déplacements en milieux urbains). Pour chaque image, les points spécifiques de l'objet \mathbf{x}_{i_ref} correspondant à \mathbf{X}_o ont été sélectionnés manuellement et reliés aux points de caractéristiques SURF extraits. Une estimation de la transformation géométrique entre \mathbf{x}_{SURF_ref} et \mathbf{x}_{i_ref} est ensuite calculée. Pour chaque station, les mises en correspondance entre \mathbf{x}_{i_ref} et \mathbf{x}_{SURF_ref} , ainsi que la transformation géométrique associée, sont alors renseignées dans une base de données spatiales. Pour chaque image extraite de la vidéo acquise par la caméra tenue en main par le piéton lors de ses déplacements, les points de caractéristiques \mathbf{x}_{SURF} sont alors extraits. Une mise en correspondance est appliquée avec la base de données contenant les différents points caractéristiques \mathbf{x}_{SURF_ref} . Cette mise en correspondance est faite en se basant sur le calcul de la *Sum of Squared Differences* (SSD), et en n'autorisant que des correspondances uniques entre

les deux ensembles de points. Une étape de filtrage est ensuite appliquée pour exclure les mises en correspondance aberrantes (*outliers*) à l'aide de l'algorithme M-estimateur SAmple Consensus (MSAC) [TORR et ZISSERMAN, 2000]. L'objet d'intérêt est considéré comme détecté si le seuil, défini expérimentalement, de quinze mises en correspondances (*inliers*) est dépassé. L'objet d'intérêt est ensuite identifié grâce aux informations associées aux points caractéristiques \mathbf{x}_{SURF_ref} . En appliquant la transformée géométrique inverse correspondante, les coordonnées des points \mathbf{x}_i correspondant aux points géométriques \mathbf{X}_o de l'objet d'intérêt sont déterminés.

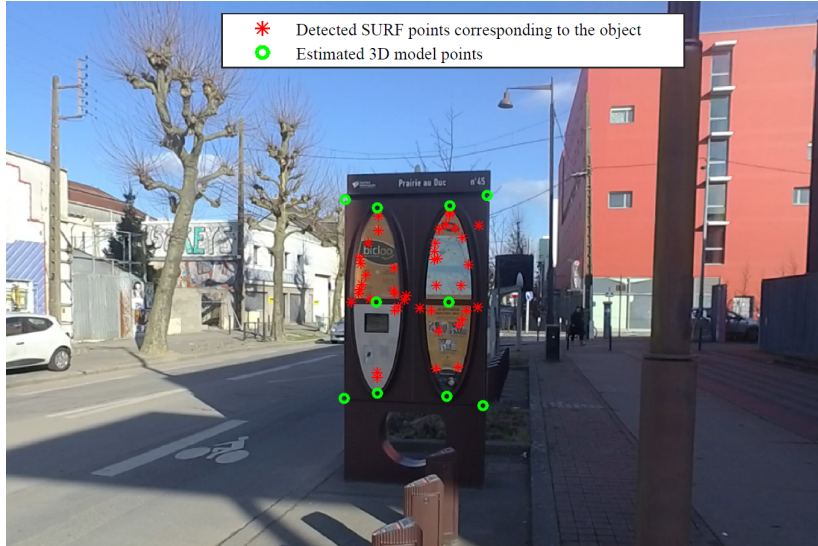


FIGURE 4.7 – Points SURF (en rouge) et points géométriques (en vert) du terminal de la station vélo sur une image extraite de la vidéo acquise par la caméra tenue en main par le piéton.

4.5 Estimation de la pose de la caméra

4.5.1 Pose relative dans le référentiel *Objet*

Connaissant la matrice des paramètres intrinsèques \mathbf{K} de la caméra et les mises en correspondance 2D-3D entre \mathbf{x}_i et \mathbf{X}_o , respectivement les points 2D sur l'image correspondant aux points 3D du modèle de l'objet de référence, la pose du centre optique de la caméra peut être calculée. Les formules sont exprimées en coordonnées homogènes, notées avec un ($\tilde{\cdot}$).

Le référentiel *Image* et le référentiel *Caméra* sont liés par :

- La matrice des paramètres intrinsèques \mathbf{K} de la caméra.

$$\tilde{\mathbf{x}}_i = \mathbf{K} \begin{bmatrix} \mathbf{I} & | & \mathbf{0} \end{bmatrix} \tilde{\mathbf{X}}_c \quad (4.1)$$

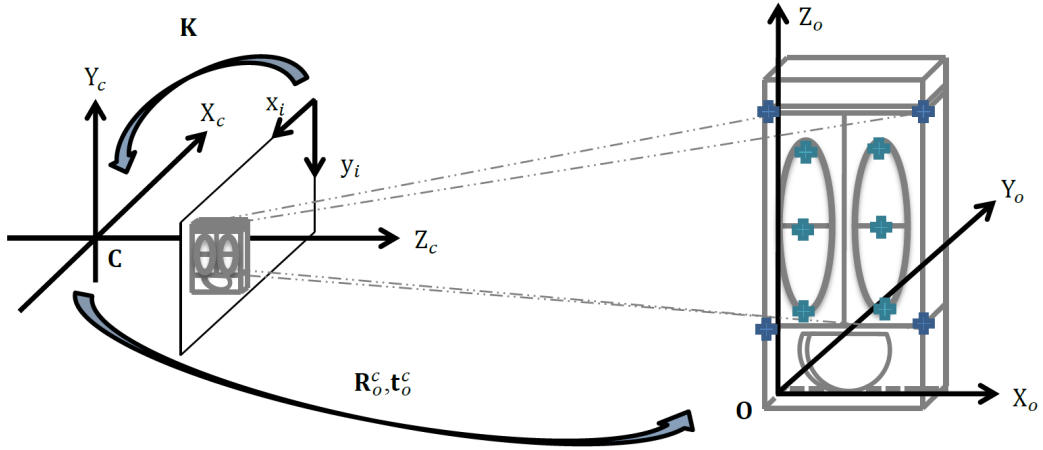
Le référentiel *Caméra* et le référentiel *Objet* sont liés par :

- Une matrice de rotation \mathbf{R}_o^c , donnant l'orientation du référentiel *Objet* par rapport au référentiel *Caméra*;
- Un vecteur de translation \mathbf{t}_o^c , donnant la position de l'origine \mathbf{O} du référentiel *Objet* par rapport à l'origine \mathbf{C} du référentiel *Caméra*.

Il est à noter que ces deux éléments constituent les paramètres extrinsèques de la caméra.

$$\tilde{\mathbf{X}}_c = \begin{bmatrix} \mathbf{R}_o^c & | & \mathbf{t}_o^c \end{bmatrix} \tilde{\mathbf{X}}_o \quad (4.2)$$

La Figure 4.8 propose une visualisation des passages entre les référentiels *Image*, *Caméra* et *Objet*.


 FIGURE 4.8 – Illustration du passage entre les référentiels *Objet*, *Caméra* et *Image*.

Ainsi, le calcul de la pose du centre optique de la caméra dans le référentiel *Objet* peut alors être écrit en reliant directement les points \mathbf{x}_i et \mathbf{X}_o , comme défini dans [HARTLEY et ZISSERMAN, 2003].

Avec :

$$\mathbf{t}_o^c = -\mathbf{R}_o^c \mathbf{C}^o \quad (4.3)$$

La transformation géométrique s'exprime alors sous la forme suivante :

$$\tilde{\mathbf{x}}_i = \mathbf{K} \begin{bmatrix} \mathbf{I} & | & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{R}_o^c & | & -\mathbf{R}_o^c \mathbf{C}^o \end{bmatrix} \tilde{\mathbf{X}}_o \quad (4.4)$$

$$\tilde{\mathbf{x}}_i = \mathbf{K} \begin{bmatrix} \mathbf{R}_o^c & | & -\mathbf{R}_o^c \mathbf{C}^o \end{bmatrix} \tilde{\mathbf{X}}_o \quad (4.5)$$

$$\tilde{\mathbf{x}}_i = \mathbf{K} \mathbf{R}_o^c \begin{bmatrix} \mathbf{I} & | & -\mathbf{C}^o \end{bmatrix} \tilde{\mathbf{X}}_o \quad (4.6)$$

$$\tilde{\mathbf{x}}_i = \mathbf{K} \begin{bmatrix} \mathbf{R}_o^c & | & \mathbf{t}_o^c \end{bmatrix} \tilde{\mathbf{X}}_o \quad (4.7)$$

Pour réaliser l'estimation de pose à l'aide d'une caméra monoculaire calibrée, de la connaissance d'une partie de l'environnement et de mises en correspondances 2D-3D, les algorithmes de type Perspective n-Points (PnP) constituent la solution la plus appropriée [WU, 2016]. L'algorithme Efficient PnP (EPnP) [MORENO-NOGUER et al., 2007] a été choisi pour effectuer cette tâche, car il s'agit d'un processus non itératif, respectant ainsi la contrainte de coût et de temps de calcul des plateformes sur lesquelles est destinée l'implémentation de la solution proposée dans ce chapitre. Cependant d'autres algorithmes de type PnP, comme ceux présentés dans [KNEIP et al., 2014b; MARCHAND et al., 2016; URBAN et al., 2016] peuvent être utilisés en lieu et place de l'algorithme EPnP. L'estimation de la pose du centre optique de la caméra peut également être affinée au moyen d'une optimisation de Gauss-Newton. Le détail de ce calcul itératif est également présenté dans [MORENO-NOGUER et al., 2007].

4.5.2 Pose absolue dans le référentiel *Navigation*

Afin de pouvoir replacer l'estimation de la pose relative au référentiel *Objet* à une estimation absolue dans le référentiel *Navigation*, la position et l'orientation précises de chaque objet de référence, susceptible d'être rencontré lors des déplacements du piéton, doivent être déterminées dans le référentiel *Navigation* (cf. Figure 4.9).

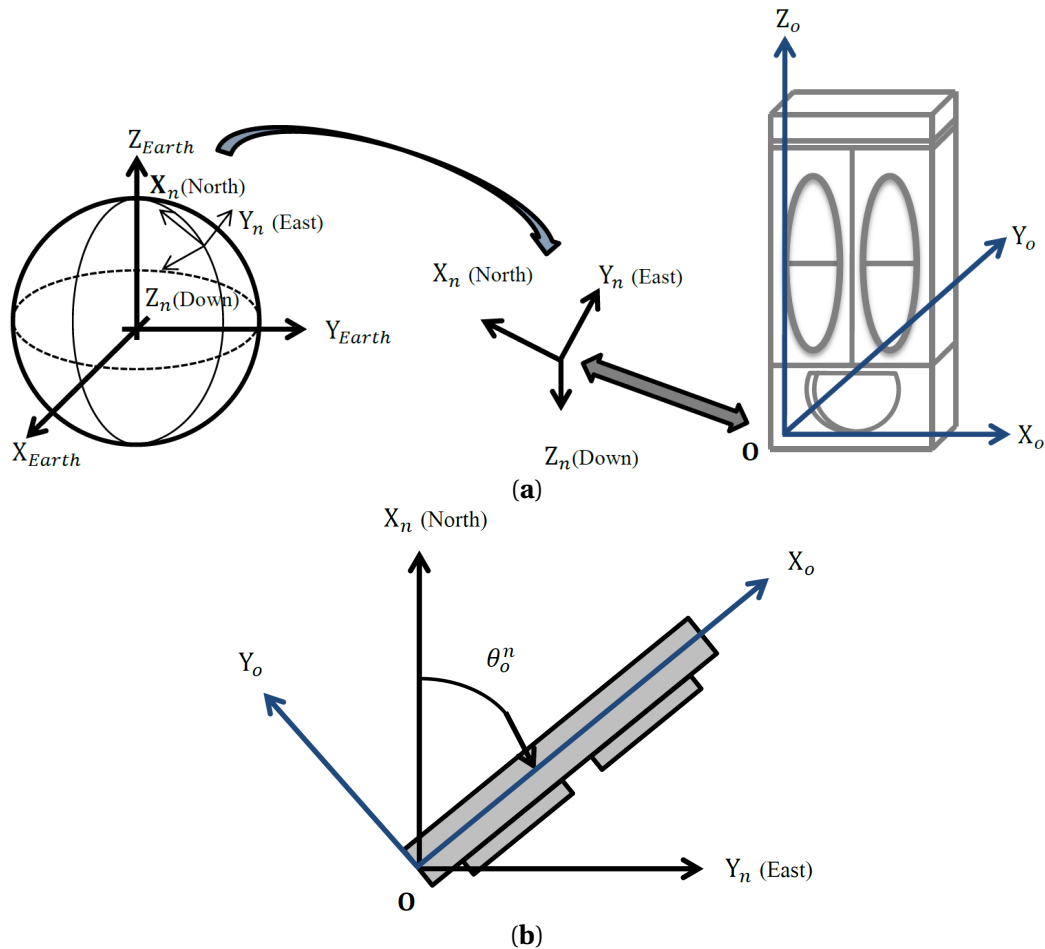


FIGURE 4.9 – (a) Lien entre le référentiel *Objet* et le référentiel *Navigation* (i.e. NED); (b) Vue sur le plan horizontal du terminal et des orientations des référentiels *Objet* et *Navigation*.

Données SIG

Grâce aux politiques "Open Data" adoptées par les métropoles, des SIG sur les transports urbains sont mis à disposition. Comme présenté dans la sous-section 4.4.2, de nombreuses données sont accessibles, notamment les noms, numéros et localisations de chaque station vélo de la métropole de Nantes^{4,5}. Cependant les localisations renseignées dans les bases de données correspondent à l'emplacement de la station dans son ensemble, et non à la localisation du terminal de la station vélo (cf. Figure 4.11). Les localisations sont donc trop imprécises pour être utilisées. De plus, ces SIG ne contiennent pas l'information d'orientation des objets de référence dans quelque référentiel géographique que ce soit.

Orthophotos géoréférencées

Une première approche pour déterminer la position et l'orientation de chaque terminal de station vélo a été de récupérer les positions des bords gauche et droit de chaque terminal au moyen des orthophotos géoréférencées disponibles en ligne (e.g. *Google Earth*, *OpenStreetMap*). Cependant, cette approche est dépendante de la précision des orthophotos et de l'appréciation de l'utilisateur pour déterminer les positions des bords de chaque terminal sur les orthophotos (cf. Figure 4.11). Cette stratégie ne permet pas de quantifier la précision de l'orientation et donc d'estimer les erreurs potentielles dans l'estimation des paramètres extrinsèques du centre optique de la caméra. De plus, les orthophotos géoréférencées souffrent d'imprécisions dans leur posi-

4. <http://data.nantes.fr/>

5. <http://vlsstats.ifsttar.fr/>

tionnement, entraînant une erreur de localisation des objets de référence dans le référentiel *Navigation*. Cette approche permet bien de mettre en place la solution globale, mais peut souffrir d'importantes imprécisions, et donc mettre à mal une estimation qualifiée de la pose permettant un affichage précis en réalité augmentée sur site.

Mesures GNSS en différentiel

Afin de pallier à ces imprécisions, une seconde approche pour déterminer la position et l'orientation de chaque terminal a été de mesurer les emplacements des bords gauche et droit de chaque objet de référence avec un récepteur GNSS géodésique Septentrio⁶ à double fréquence. Ce dernier a été utilisé pour collecter les signaux des satellites GPS et GLONASS pendant une phase statique de plusieurs minutes (cf. Figure 4.10). Pendant ce temps, les signaux des satellites étaient collectés par une "station de base" référencée dans le réseau RGP-IGN⁷. Cela a permis d'estimer et de supprimer les erreurs de propagation des signaux satellitaires dans les couches atmosphériques de la Terre, et ainsi de calculer de manière plus précise les positions du récepteur. Avec une acquisition statique longue en mode différentiel, i.e. de l'ordre de la dizaine de minutes, un post-traitement avec le logiciel libre RTKlib⁸ revendique une position avec une incertitude sub-centimétrique. Cela permet donc de déterminer, avec une précision qualifiée, la position des bords de chaque objet de référence, et ainsi en déduire leur orientation dans le référentiel *Navigation*. Si l'hypothèse d'une erreur de positionnement inférieure au centimètre n'est pas vérifiée, cela introduit une erreur dans l'estimation de la pose. Cette propagation d'erreur est présentée dans la sous-section 4.6.1.



FIGURE 4.10 – Acquisition de la position des bords du terminal de la station vélo avec récepteur GNSS géodésic double fréquence et antenne.

La Figure 4.11 présente les différences de positionnement entre la localisation renseignée dans le SIG mis à disposition par la métropole de Nantes, les localisations des bords d'un terminal estimées avec les orthophotos géoréférencées *Google Earth* et celles déterminées par mesures GNSS sur le terrain et post-traitées en différentiel avec RTKlib. Cela met en avant le fait que les orthophotos géoréférencées ne sont pas assez précises pour estimer les positions et orientations des

6. <http://www.septentrio.com/>

7. <http://rgp.ign.fr/DONNEES/diffusion/>

8. <http://www.rtklib.com/>

objets de référence sur lesquelles reposeraient l'estimation de la pose de la caméra. Comme introduit dans la sous-section 4.2.2, certaines données stockées dans les SIG 3D, comme les modèles 3D de bâtiments qui sont créés par extrusions du bâti de la BD Topo, lui-même étant créé à partir d'interprétations d'orthophotos géoréférencées. C'est également le cas pour les cartographies proposées par OpenStreetMap⁹, qui sont en partie créés et éditées par le grand public.

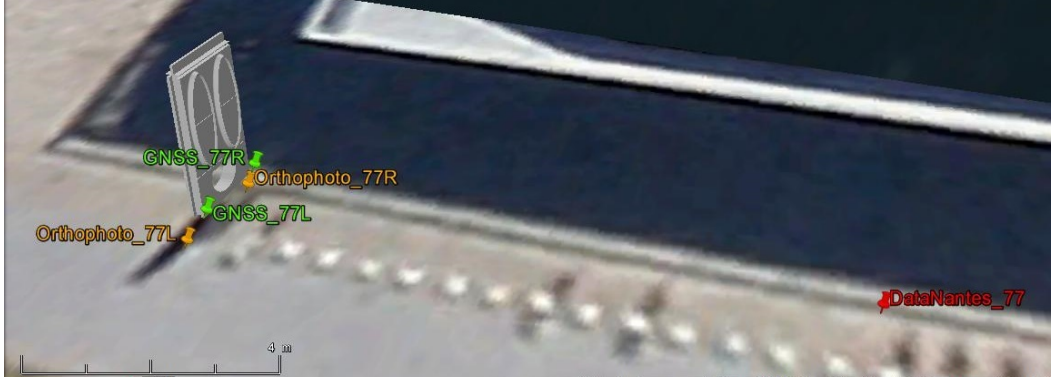


FIGURE 4.11 – Visualisation dans *Google Earth* des localisations du terminal de la station vélo renseignées dans "DataNantes" (en rouge), déterminées via orthophotos (en orange) et déterminées par mesures GNSS en différentiel (en vert).

Changement de repère

La détermination de l'orientation θ_o^n des objets de référence dans le référentiel *Navigation* permet ainsi de définir la matrice de rotation \mathbf{R}_o^n . Cette dernière permet de passer d'une orientation dans le référentiel *Objet* à une orientation dans le référentiel *Navigation*.

$$\mathbf{R}_o^n = \begin{bmatrix} \cos(\theta_o^n) & \sin(\theta_o^n) & 0 \\ \sin(\theta_o^n) & -\cos(\theta_o^n) & 0 \\ 0 & 0 & -1 \end{bmatrix} \quad (4.8)$$

Ainsi, la matrice de rotation \mathbf{R}_c^n , définissant l'orientation du centre optique de la caméra dans le référentiel *Navigation* peut être déterminée avec :

- La matrice de rotation \mathbf{R}_c^o définissant l'orientation du référentiel *Objet* par rapport au référentiel *Caméra*;
- La matrice de rotation \mathbf{R}_o^n définissant l'orientation du référentiel *Objet* par rapport au référentiel *Navigation*.

$$\mathbf{R}_c^n = \mathbf{R}_o^n \mathbf{R}_c^o = \mathbf{R}_o^n (\mathbf{R}_c^o)^T \quad (4.9)$$

Connaissant la position de chaque objet de référence dans le référentiel *Navigation*, et ayant estimé la position du centre optique de la caméra dans le référentiel *Objet*, la position du centre optique de la caméra est calculée dans le référentiel *Navigation* comme suit :

$$\mathbf{C}^n = [\mathbf{R}_c^n \mid \mathbf{t}_c^n] \mathbf{C}^o \quad (4.10)$$

4.6 Estimation des erreurs

4.6.1 Erreur de position des objets d'intérêt dans le référentiel *Navigation*

Afin d'évaluer l'impact des erreurs de positionnement des objets d'intérêt dans le référentiel *Navigation* sur l'estimation de la pose, une erreur de position est considérée dans la détermination des coordonnées des bords d'un terminal. Cela a pour conséquence directe d'introduire une

9. <http://openstreetmap.fr/>

erreur $\epsilon\theta_o^n$ dans l'estimation de l'orientation θ_o^n . Les deux bords du terminal de la station de vélo sont distants de L_{ref} . Si l'hypothèse d'une erreur sub-centimétrique sur la définition des positions par GNSS différentiel n'est pas vérifiée, une erreur ϵL est alors considérée sur la position de chacun des bords du terminal de la station vélo. Ainsi, dans le pire cas, l'erreur d'orientation de l'objet d'intérêt peut être définie comme suit :

$$\epsilon\theta_o^n = \arctan\left(\frac{2 * \epsilon L_{DGNS}}{L_{ref}}\right) \quad (4.11)$$

Cette erreur d'orientation de l'objet de référence implique une erreur de positionnement du centre optique de la caméra définie comme suit :

$$\epsilon C_{DGNS} = t_o^c * \tan(\epsilon\theta_o^n) \quad (4.12)$$

Ainsi, les distances suivantes sont considérées :

- Une distance L_{ref} de 1,35 m, correspondant à la distance entre les deux bords du terminal de la station de vélo;
- Une erreur ϵL_{DGNS} de 2 cm sur la position de chacun des bords du terminal de la station vélo, correspondant à une erreur classique de positionnement par DGNS;
- Une distance t_o^c de 10 m, correspondant à une distance "classique" entre la position du centre optique de la caméra et la position d'un objet de référence.

Cela introduit une erreur ϵC_{DGNS} de 29 cm dans l'estimation de la position du centre optique de la caméra sur le plan horizontal. Cela montre à quel point il est important de connaître précisément la position des objets de référence dans le référentiel *Navigation*.

4.6.2 Erreur d'estimation des points géométriques dans le référentiel *Image*

Afin d'évaluer l'impact qu'aurait une extraction imprécise des points x_i sur la précision de l'estimation de la position du centre optique de la caméra, une erreur pseudo-aléatoire uniformément répartie est introduite sur les coordonnées de chacun des points x_i . Pour rappel, les points x_i sont les points 2D sur les images acquises par la caméra, correspondant aux points géométriques 3D X_o d'un objet de référence. La Figure 4.12 présente une visualisation des positions 3D du centre optique de la caméra estimées avec une erreur pseudo-aléatoire uniformément répartie de $[-5; +5]$ pixels introduite sur les coordonnées des points x_i .

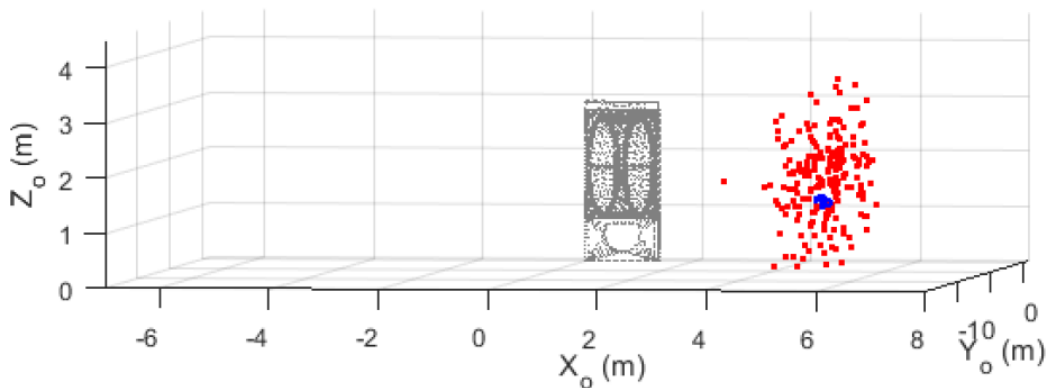


FIGURE 4.12 – Visualisation de la position 3D estimée sans erreur introduite (en bleu) et de 100 positions 3D estimées avec l'introduction d'une erreur pseudo-aléatoire uniformément répartis de $[-5; +5]$ pixels sur chacun des points x_i .

La Figure 4.13 présente la distribution, calculée sur 100000 itérations, de l'erreur de positionnement 3D du centre optique de la caméra pour l'introduction d'une erreur pseudo-aléatoire uniformément répartie de $[-5; +5]$ pixels sur les coordonnées des points x_i . On observe une erreur

médiane de positionnement 3D du centre optique de la caméra de 85 cm avec un écart-type de 48 cm. La précision de l'estimation de la pose dépend donc fortement de la précision des coordonnées des points \mathbf{x}_i extrait des images acquises par la caméra. Cela confirme bien que la reconnaissance d'objet de référence et l'extraction des points caractéristiques géométriques associés constituent est une partie critique de la solution proposée dans ce chapitre.

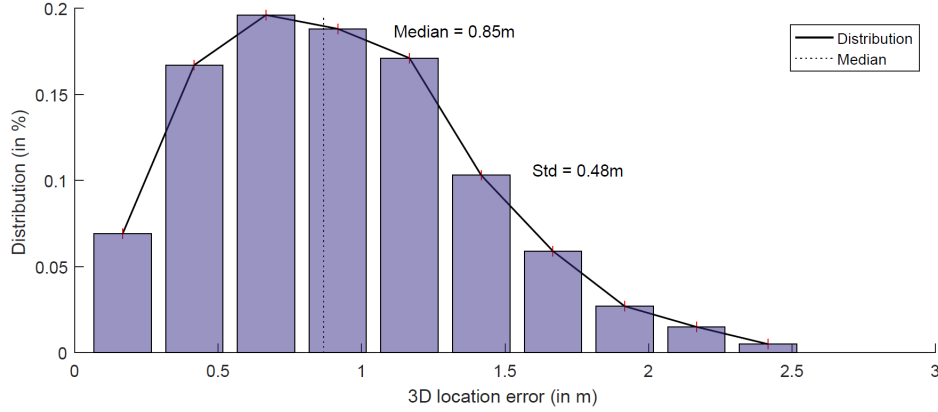


FIGURE 4.13 – Distribution sur 100.000 itérations de l'erreur de positionnement du centre optique de la caméra pour l'introduction d'une erreur pseudo-aléatoire uniformément répartie de $[-5; +5]$ pixels sur les coordonnées des points \mathbf{x}_i .

4.6.3 Erreurs de reprojection dans le référentiel *Image*

Comme défini dans la sous-section 4.5.1, la pose du centre optique de la caméra est calculée grâce à la connaissance de la matrice des paramètres intrinsèques \mathbf{K} de la caméra et des mises en correspondance 2D-3D entre \mathbf{x}_i et \mathbf{X}_o . Une fois cette pose $[\hat{\mathbf{R}}_o^c \mid \hat{\mathbf{t}}_o^c]$ estimée, il est possible de recalculer les coordonnées des points \mathbf{x}_{i_repro} et de les reprojeter sur l'image acquise par la caméra.

$$\tilde{\mathbf{x}}_{i_repro} = \mathbf{K} [\hat{\mathbf{R}}_o^c \mid \hat{\mathbf{t}}_o^c] \tilde{\mathbf{X}}_o \quad (4.13)$$

Comme présenté sur la Figure 4.14, cela permet d'observer les différences de coordonnées entre les points \mathbf{x}_i ayant servi à estimer la pose et les points \mathbf{x}_{i_repro} calculés à partir de la pose estimée. Les deux ensembles de points ont leurs coordonnées très proches, mais des différences résultant d'imprécisions dans le processus d'extraction des points \mathbf{x}_i et dans le processus d'estimation de la pose $[\hat{\mathbf{R}}_o^c \mid \hat{\mathbf{t}}_o^c]$ sont observables.

Afin d'évaluer l'impact de cette différence de coordonnées, l'erreur angulaire induite par ces différences de coordonnées est calculée comme définie dans [HARTLEY et ZISSERMAN, 2003] :

$$\cos(\theta) = \frac{(\mathbf{K}^{-1}\mathbf{x}_i)^T (\mathbf{K}^{-1}\mathbf{x}_{i_repro})}{\sqrt{(\mathbf{K}^{-1}\mathbf{x}_i)^T (\mathbf{K}^{-1}\mathbf{x}_i)} \sqrt{(\mathbf{K}^{-1}\mathbf{x}_{i_repro})^T (\mathbf{K}^{-1}\mathbf{x}_{i_repro})}} \quad (4.14)$$

Il est important de noter que les erreurs de reprojection dépendent de plusieurs facteurs :

- La résolution des images acquises par la caméra, après correction des distorsions ;
- La distance entre l'objet de référence et le centre optique de la caméra

Ainsi, le calcul de l'erreur angulaire permet de détecter les estimations "imprécises" de la pose de la caméra et permet également de valider ou non l'intégration du processus d'estimation de la pose basée sur reconnaissance d'un objet de référence dans la solution globale présentée dans cette thèse. Pour cela, il est nécessaire de définir un seuil à partir duquel l'estimation de la pose est considérée comme "imprécise". Dans notre contexte applicatif, le piéton acquiert des images d'une résolution de 1920×1080 , et la détection d'un objet d'intérêt se fait généralement à une distance inférieure à une quinzaine de mètres. L'estimation de la pose, basée sur la reconnaissance

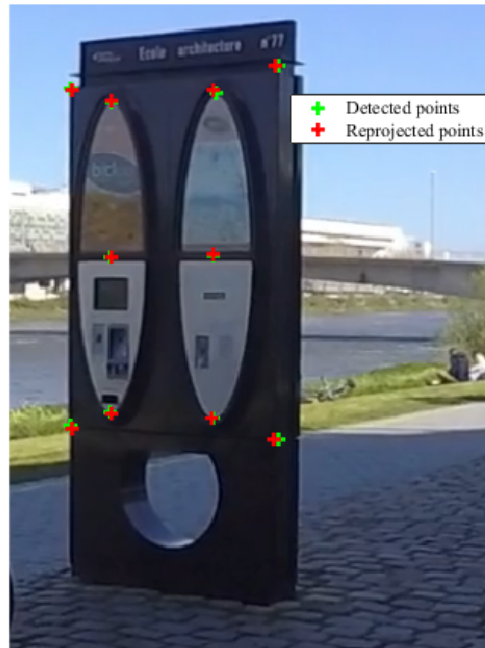


FIGURE 4.14 – Erreur de reprojexion.

d'un objet de référence, n'est prise en compte dans la solution globale que si l'erreur de reprojexion est considérée "suffisamment précise", i.e. qu'elle ne dépasse pas le seuil fixé expérimentalement de :

$$\epsilon_{repro} = \sqrt{(\mathbf{x}_i - \mathbf{x}_{i_repro})^2} < 50 \text{ pixels} \quad (4.15)$$

4.7 Expérimentation : estimation de la pose basée sur la reconnaissance d'un objet de référence et affichage en réalité augmentée sur site

4.7.1 Description de l'expérimentation

Comme il est difficile d'avoir une référence pour la pose d'un équipement tenu en main, une évaluation de l'estimation de la pose est proposée en termes de distance à des positions de référence sur le plan horizontal. Ainsi, afin d'évaluer la précision de l'estimation absolue dans le référentiel *Navigation* de la pose du centre optique de la caméra basée sur la reconnaissance d'objet de référence, ces points de référence ont été défini à proximité des objets de référence. Ces points de référence permettent d'évaluer la précision du processus d'estimation de la pose, en comparant la position réelle de la caméra et la position estimée par le processus d'estimation de la pose basée sur la reconnaissance d'un objet de référence. Le matériel expérimental et plus spécifiquement la caméra a été placée précisément sur la canne où se trouvait l'antenne GNSS lors de la détermination des points de référence présentée dans la sous-section 4.7.2.

4.7.2 Définition de la référence terrain

L'emplacement exact des points de référence a été déterminé à l'aide de la solution GNSS en différentiel présentée dans la section 4.5.2. Cela a permis de déterminer avec une précision subcentimétrique les positions des points de référence dans le référentiel *Navigation*. Comme présenté sur la Figure 4.15, les positions de ces points de référence ont été marquées sur le sol d'un cercle blanc barré.



FIGURE 4.15 – Acquisition avec récepteur GNSS géodésique, antenne pour déterminer la position des points de référence et canne à bulle.

Il est important de noter que les positions des points de référence déterminées par la solution DGNSS sont sur le plan horizontal. Bien que la position de l'équipement sur l'axe vertical, et plus spécifiquement le centre optique de la caméra, soit connue lors de l'acquisition, le sol n'étant pas un plan horizontal parfait entre le point de référence et l'objet de référence, la position sur l'axe verticale n'est pas prise en compte dans l'évaluation de la solution présentée dans ce chapitre. De plus, une erreur de quelques centimètres dans le plan horizontal est à prendre en compte dans l'interprétation des résultats. Cette dernière correspondant à la séparation entre l'antenne GNSS et le centre optique de la caméra. En effet, bien que l'emplacement de l'antenne GNSS était connu avec une précision sub-centimétrique grâce au post-traitement RTKlib, une erreur de 2 cm est classiquement utilisée pour les positions déterminées par DGNSS. De plus, l'emplacement exact du centre optique à l'intérieur de la caméra est connu avec une précision centimétrique.

4.7.3 Résultats

Pour l'acquisition statique sur un point de référence, l'estimation de la pose est évaluée en comparant le positionnement estimé sur le plan horizontal à la position réelle du point de référence défini par DGNSS. Dans l'ensemble de cette thèse, les erreurs de positions sont définies sur le plan horizontal de la manière suivante.

$$\epsilon_D = \sqrt{((X_{Ref} - X_{Solution})^2 + (Y_{Ref} - Y_{Solution})^2)} \quad (4.16)$$

Le Tableau 4.1 présente l'erreur moyenne de positionnement et l'écart type obtenus pour le processus d'estimation de la pose lors de phases statiques sur les points d'intérêt. Les résultats sont présentés pour le processus de vision seul et une visualisation est donnée sur la Figure 4.16.

Point de référence	Waypoint 1	Waypoint 2
Distance à l'objet de référence	8.15 m	7.43 m
Erreur de position moyenne	22 cm	25 cm
Écart type	13 cm	18 cm

TABLEAU 4.1 – Précision de la position extrait de l'estimation de la pose basée sur la reconnaissance d'un objet de références.

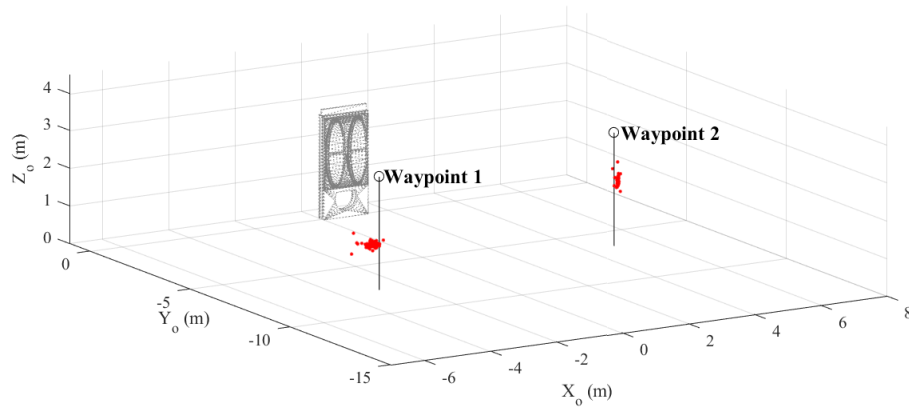


FIGURE 4.16 – Positions extraites de l'estimation de la pose basé sur la reconnaissance d'un objet de référence pour l'acquisition en statique sur les points de références.

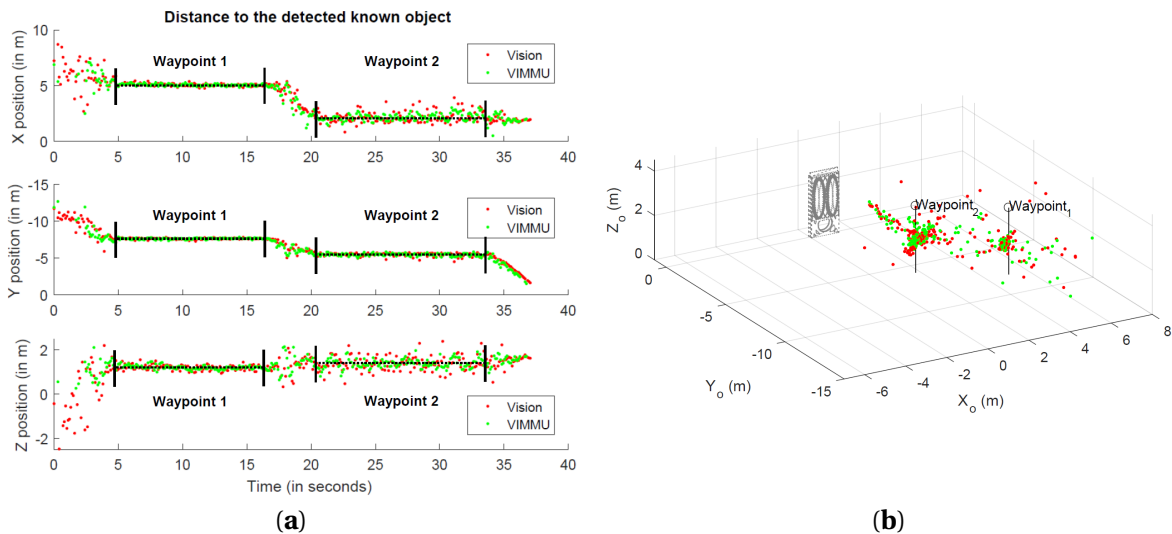


FIGURE 4.17 – (a) Vue 2D (i.e. sur le plan horizontal) des positions estimées; (b) Vue 3D des positions estimées.

4.7.4 Réalité augmentée qualifiée sur site

Dans le cadre de la navigation piétonne et de la réalité augmentée sur site, lorsque l'estimation de la pose est certifiée "suffisamment précise", une reprojektion du modèle 3D de l'objet de référence et du modèle 3D des bâtiments autour est alors possible. La Figure 4.18-(a) présente une visualisation du modèle 3D de l'objet directement sur site. Le modèle 3D étant correctement reprojété sur la scène observée, cela valide le fait que la pose du centre optique de la caméra est estimée correctement. Cette estimation absolue de la pose de la caméra dans le référentiel *Navigation* ouvre la possibilité de reprojeter des informations contenues dans les SIG 3D, et ainsi de pouvoir caractériser leurs niveaux de précision et de détails. La Figure 4.18-(b) présente une visualisation du modèle 3D des bâtiments contenu dans un SIG 3D. Il est à noter que la reprojektion

des bâtiments de la ville peut ne pas être clairement lisible car la reprojection correspond à un découpage d'une dalle 3D (i.e. pas seulement les surfaces visibles). Il est alors facilement observable que les modèles 3D des bâtiments ne sont pas alignés avec la scène réelle observée. Cela est intrinsèquement dû à leur méthode de création, présenté dans la sous-section 4.2.2 et aux problèmes de précision des orthophotos relevés dans la section 4.5.2 sur lesquelles est basé la constitution de la BD Topo.

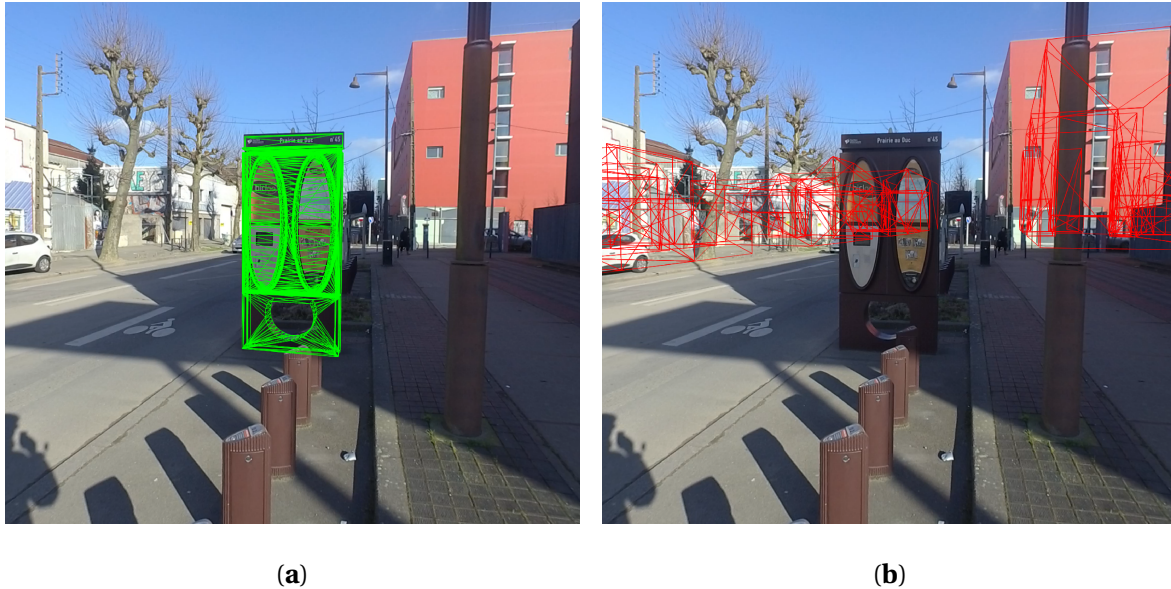


FIGURE 4.18 – Reprojection en réalité augmentée sur site : (a) Modèle 3D de l'objet d'intérêt; (b) Dalle LOD2.

Une correction est alors possible, et cela permet ainsi de quantifier l'erreur des modèles 3D contenus dans les SIG. La correction peut alors être effectuée en translation, auquel cas l'erreur de positionnement des modèles 3D serait l'inverse de celui du centre optique de la caméra pour corriger l'erreur. Une correction en orientation est également possible, mais présente des problèmes quant à la précision de la caractérisation :

- Soit le changement d'orientation est inverse à celui du centre optique de la caméra;
- Soit une rotation est appliquée directement sur la dalle 3D, avec la position du centre optique de la caméra comme axe de rotation.

Pour une estimation de l'orientation des bâtiments, cela pose évidemment le problème de la distance entre les bâtiments et le centre optique de la caméra. Ainsi, une limite de reprojection est fixée pour la caractérisation. De plus, bien que la BD Topo, sur laquelle sont constitués les modèles 3D des bâtiments, souffre d'erreurs, cette dernière présente principalement des erreurs de translation et d'échelle, et non de rotation. La Figure 4.19 présente ainsi une visualisation du modèle 3D des bâtiments corrigé. Une correction de [+3 m; +5 m] a été appliquée sur le modèle 3D des bâtiments pour que la reprojection soit alignée avec la scène réelle observée.

4.7.5 Occlusion

La reconnaissance d'un objet de référence repose sur la détection et la mise en correspondance de nombreux points d'intérêt SURF, répartis sur toute la surface de l'objet de référence. Par conséquent, la détection d'un objet de référence est robuste à une occlusion partielle (e.g. piéton, cycliste, lampadaire, etc.). De cette manière, l'estimation de la pose du centre optique de la caméra et l'affichage en réalité augmentée sur site sont toujours disponibles lorsqu'une partie importante de l'objet d'intérêt est masquée. La Figure 4.20 présente un exemple de visualisation de reprojection du modèle 3D de l'objet de référence alors que ce dernier est partiellement occulté par un piéton qui se place entre l'objet de référence et le centre optique de la caméra. Bien évidemment



FIGURE 4.19 – Reprojection en réalité augmentée sur site : Dalle LOD2 corrigée.

la taille des objets de référence, qui est plus grande que les éléments occultant, permet de fiabiliser les chances de détection. Le choix d'un objet de référence pouvant être facilement occulté, en ferait perdre l'intérêt de son utilisation.



FIGURE 4.20 – Reconnaissance de l'objet de référence robuste à l'occlusion et reprojection de son modèle 3D.

4.8 Conclusion

Dans le contexte de navigation piétonne en milieu urbain, l'approche proposée dans ce chapitre permet d'estimer, de manière absolue dans le référentiel *Navigation*, la pose d'une caméra intégrée dans un équipement tenu en main par le piéton. Pour cela, la connaissance détaillée d'objets contenus dans les Systèmes d'Information Géographique 3D a été mise à profit. La reconnaissance d'un objet de référence dans les enregistrements de la caméra permet d'extraire des points caractéristiques géométriques correspondant au modèle 3D de l'objet. La mise en correspondance entre les points d'intérêt extraits d'une image et le modèle 3D, permet d'estimer la pose

de la caméra dans un repère relatif au modèle 3D de l'objet. Ensuite, la connaissance de la position et de l'orientation de l'objet de référence dans le référentiel *Navigation* permet d'en déduire celle de la caméra dans ce dernier. Cette pose absolue sera également utiliser pour mettre à jour un processus de navigation à l'estime du piéton dans la suite de cette thèse.

Chapitre 5

Odométrie visuelle monoculaire mise à l'échelle par PDR

« L'homme est la mesure de toute chose »

Protagoras

Sommaire

5.1 Introduction	50
5.2 Odométrie visuelle monoculaire	51
5.3 Estimation des longueurs de pas du piéton	52
5.3.1 Détection des instants de pas et estimation des longueurs de pas du piéton	52
5.4 Estimation du facteur d'échelle	53
5.5 Détermination de la position sur l'axe vertical	54
5.6 Expérimentation : estimation de la localisation par odométrie visuelle monoculaire mise à l'échelle	56
5.6.1 Description des acquisitions	56
5.6.2 Définition de la référence terrain	56
5.6.3 Initialisation de la pose	58
5.6.4 Résultats et analyses	59
5.7 Conclusion	67

5.1 Introduction

Ce chapitre présente une approche permettant de résoudre l'ambiguïté d'échelle qui constitue la principale inconnue de l'odométrie visuelle avec une caméra monoculaire tenue en main. Ce travail a été réalisé en collaboration avec Hideaki Uchiyama au sein du Laboratory for Image and Media Understanding (LIMU) de l'Université de Kyushu (Japon). La principale contribution de cette approche est l'utilisation des mouvements du piéton pour résoudre l'ambiguïté d'échelle de l'odométrie visuelle monoculaire. L'estimation du facteur d'échelle est réalisée grâce à l'analyse des données inertielles proposée dans le domaine de la navigation à l'estime du piéton (*Pedestrian Dead Reckoning-PDR*), et plus précisément grâce à l'estimation des longueurs de pas du piéton. A la différence des approches présentées dans la sous-section 3.3.2, la contrainte est mise sur le fait que le piéton utilise un équipement intégrant des capteurs bas coût et que ce dernier est tenu en main dans une position où la caméra de l'équipement du piéton fait face au monde. Ce mode de portage correspond à l'utilisation d'une application d'aide à la mobilité en milieu urbain avec un affichage en réalité augmentée. Pour permettre cela, il est important d'estimer et de suivre les changements de pose d'un équipement tenu en main lors des déplacements du piéton en milieux urbains. La Figure 5.1 présente une vue d'ensemble de l'approche proposée. Cette dernière est divisée en trois parties :

- Partie *Inertielle*;
- Partie *Vision*;
- Partie *Couplage*.

Les parties *Inertielle* et *Vision* sont indépendantes, de fait l'approche proposée peut être catégorisée comme une approche de couplage lâche. Les variables nécessaires pour mettre en place l'approche proposée sont toutes détaillées dans les sections suivantes.

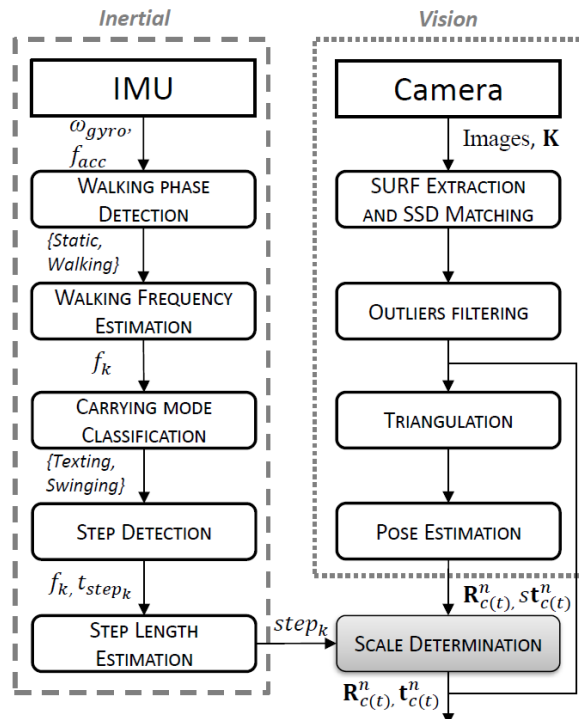


FIGURE 5.1 – Diagramme bloc de la solution d'estimation du facteur d'échelle dans l'odométrie visuelle monoculaire.

La partie *Vision* est dédiée au processus d'odométrie visuelle permettant de suivre les changements de pose de la caméra monoculaire intégrée dans l'équipement tenu en main par le piéton lors de ses déplacements en milieu urbain. Du fait que l'estimation du facteur d'échelle est indépendante de la méthode d'estimation de la pose, n'importe quelle méthode d'odométrie visuelle

ou de SLAM visuel peut être appliquée [AQEL et al., 2016; TAKETOMI et al., 2017]. L'implémentation de l'odométrie visuelle monoculaire mise en place dans la solution proposée est détaillé dans la section 5.2.

La partie *Inertielle* est dédiée à l'estimation des longueurs de pas du piéton durant ses déplacements en milieux urbains. Le processus traite les données en sortie de l'unité de mesures inertielles intégrée dans l'équipement tenu en main par le piéton et estime une longueur de pas en mètre à chacun de ses pas. Les détails du processus d'estimation des longueurs de pas du piéton sont décrits dans la section 5.3.

La partie *Couplage*, correspondant au bloc grisé, est dédiée à la détermination du facteur d'échelle. Cela permet de résoudre l'ambiguïté d'échelle de l'odométrie visuelle monoculaire. Ainsi, l'estimation de la pose, relative à l'initialisation ou à une pose connue de manière absolue, est le résultat final de la solution proposée. Le détail du processus d'estimation du facteur d'échelle est présenté dans la section 5.4.

Afin de compléter l'estimation de la pose, notamment en définissant la position de l'équipement tenu en main par le piéton sur l'axe vertical, plusieurs approches ont été étudiées. Leurs détails sont présentés dans la section 5.5.

Une validation expérimentale de la solution proposée dans ce chapitre est présentée dans la sous-section 5.6.4. Les résultats obtenus, correspondant aux différentes étapes présentées dans ce chapitre, sont également illustrés dans la partie expérimentale.

5.2 Odométrie visuelle monoculaire

Dans le processus d'odométrie visuelle, une camera monoculaire calibrée est utilisée. La calibration de cette dernière et la détermination de la matrice des paramètres intrinsèques \mathbf{K} sont présentées dans la section 2.4.2. Ainsi, pour traquer la pose à l'instant (t_i) , trois images aux instants (t_i) , (t_{i-1}) et (t_{i-2}) sont utilisées. La pose à l'instant (t_i) est inconnue, et celles à (t_{i-1}) et (t_{i-2}) sont connues. Tout d'abord, des points d'intérêt \mathbf{x}_i dans les images sont extraits par le détecteur SURF [BAY et al., 2008]. Afin d'améliorer la robustesse des points d'intérêt extraits, et ainsi éviter toute redondance, une extraction avec une répartition uniforme des points sur les images est opérée. Les points d'intérêt extraits sont ensuite mis en correspondance en utilisant la distance calculée par SSD et en n'autorisant que des correspondances uniques.

$$D_{\text{SSD}}(p_{i1}, p_{i2}) = \sum (p_{i1} - p_{i2})^2 \quad (5.1)$$

Une étape de filtrage, en utilisant une contrainte épipolaire, est ensuite appliquée avec l'algorithme M-estimateur SAmple Consensus (MSAC) [TORR et ZISSERMAN, 2000] pour exclure les mises en correspondances aberrantes (*outliers*). Une étape de triangulation est opérée uniquement sur les mises en correspondance valides (*inliers*) et présentes dans les images précédentes aux instants (t_{i-2}) et (t_{i-1}) , et dans l'image courante à l'instant (t_i) . Cela permet d'estimer \mathbf{X}_n , i.e. les positions 3D des points d'intérêt extraits à l'instant (t_i) , et de les placer dans un repère relatif au centre optique de la caméra. Ainsi, la connaissance de ces points permet d'estimer les changements relatifs de pose du centre optique de la caméra. Une visualisation de cette approche est présentée sur la Figure 5.2.

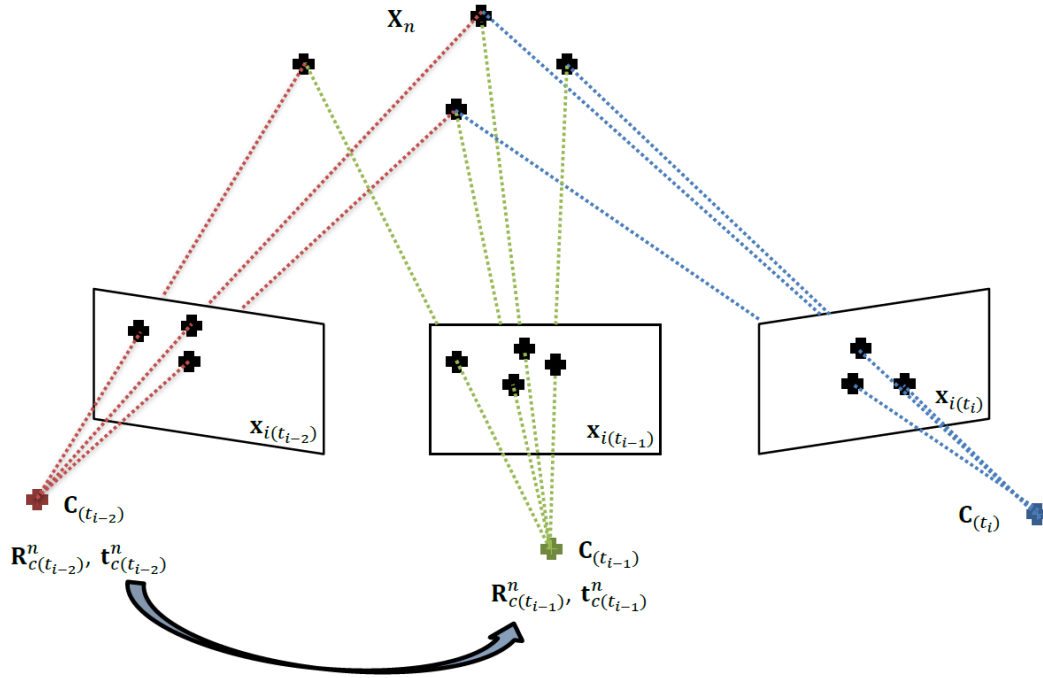


FIGURE 5.2 – Triangulation.

5.3 Estimation des longueurs de pas du piéton

5.3.1 Détection des instants de pas et estimation des longueurs de pas du piéton

Pour estimer les longueurs de pas du piéton, les signaux du gyroscope ω_{gyro} et de l'accéléromètre f_{acc} , acquis avec l'équipement tenu en main, sont analysés en suivant la procédure proposée dans [RENAUDIN et al., 2012; SUSI et al., 2013]. Une première classification des mouvements du piéton est opérée pour déterminer les phases de marche et les phases statiques. Les activités, basées sur des éléments décrivant le signal (i.e. la forme du signal, la fréquence et l'énergie suivantes), sont alors définies :

- *Static* : correspond aux phases d'immobilité de l'équipement mobile;
- *Walking* : correspond aux phases où le piéton marche.

Une détection d'extrêmes locaux et un seuillage sur l'énergie sont ensuite appliqués sur les signaux du gyroscope et de l'accéléromètre pour déterminer les événements t_{step_k} correspondant aux instants où le pied du piéton entre en contact avec le sol, au k^{eme} pas du piéton. La fréquence de pas f_k est alors calculée par *Transformée de Fourier à court termes* (STFT). Une deuxième classification des mouvements du piéton est opérée en analysant la variance des signaux inertiels pour déterminer le mode de portage de l'appareil selon [SUSI et al., 2013]. Bien que très divers, les mouvements de la main peuvent être séparés selon des classes définies par des caractéristiques du signal : Energie, contenu fréquentiel, évolution temporelle. Ces classes correspondent à différentes activités de la main du piéton portant la centrale inertielle.

- *Swinging* : Lors de la marche, le piéton a régulièrement les bras ballants le long du corps. Les bras oscillent alors suivant la fréquence de pas ou de foulée selon la vitesse. Les signaux de l'unité de mesures inertielles sont caractérisés par une répétition assez précise des oscillations (pics très visibles sur les fréquences de pas et de foulée) accompagnées d'une intensité relativement forte.
- *Texting* : Le piéton maintient sa main (avec l'équipement mobile) relativement fixe par rapport au tronc. Cette activité apparaît lorsque le piéton regarde l'écran de son équipement mobile, contenant l'unité de mesures inertielles, pour écrire/lire des messages, ou tout simplement naviguer sur des applications. Les intensités des signaux inertielles sont moindres

par rapport à l'activité *Swinging*.

- *Phoning* : Le piéton maintient son équipement mobile relativement proche de son oreille. Ce mode de portage apparaît notamment quand le piéton passe un appel ou écoute un message. Cette activité est semblable au mode de portage *Texting*.
- *Pocket* : L'équipement mobile est dans une poche ou un sac. Dans ce cas l'équipement mobile est relativement solidaire du piéton ou du sac dans lequel il est placé. L'énergie du signal est plus faible en moyenne que celle du mode *Swinging*.

Comme défini dans [RENAUDIN et al., 2012], un modèle générique est ensuite utilisé pour estimer les longueurs de pas $step_k$ du piéton. Il est basé sur la hauteur du piéton h_{ped} , sur la fréquence de pas f_k et sur un ensemble de trois paramètres $\{a, b, c\}$ déterminés sur 12 sujets.

$$step_k = h_{ped} \times (a \times f_k + b) + c \quad (5.2)$$

5.4 Estimation du facteur d'échelle

L'estimation du facteur d'échelle proposée dans ce chapitre est indépendante du processus d'estimation de la pose par odométrie visuelle monoculaire. Elle repose sur l'utilisation de l'estimation des longueurs de pas du piéton $step_k$, des instants de ses pas t_{step_k} et des instants de prise d'image par la caméra monoculaire. Une interpolation linéaire des longueurs de pas est faite pour déterminer l'amplitude de déplacement du piéton entre deux images, comme illustré sur la Figure 5.3. Par exemple, le facteur d'échelle de l'odométrie visuelle monoculaire utilisé au k^{eme} pas est alors calculée en utilisant une interpolation linéaire de la $k - 1^{eme}$ longueur de pas, i.e. la longueur de pas estimée entre les instants $t_{step_{k-1}}$ et t_{step_k} .

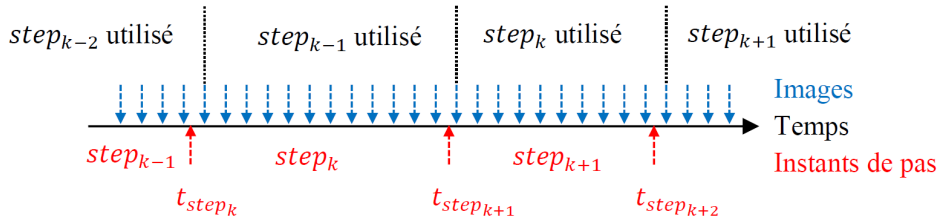


FIGURE 5.3 – Interpolation des longueurs de pas du piéton estimées.

Il est important de noter que l'estimation des longueurs de pas du piéton ne fournit que l'amplitude de ses déplacements dans le plan horizontal. En faisant l'hypothèse que les déplacements de l'appareil tenu en main par le piéton sont principalement sur le plan horizontal, les déplacements sur l'axe vertical ne sont pas pris en compte. Par conséquent, le facteur d'échelle est calculé en comparant les longueurs de pas du piéton et l'amplitude des déplacements de la caméra estimé par le processus d'odométrie visuelle monoculaire dans le plan horizontal. Ainsi, l'amplitude des déplacements du piéton dans le plan horizontal est définie comme suit :

$$D_{step}(t_i) = step_k \times \frac{\Delta t_i}{\Delta t_{step_k}} \quad (5.3)$$

où $D_{step}(t_i)$ est l'estimation de l'amplitude du déplacement du piéton sur le plan horizontal entre deux images consécutives aux temps (t_{i-1}) et (t_i) ; Δt_{step_k} est la différence de temps entre les instants de pas (k) et $(k + 1)$, et Δt_i est la différence de temps entre deux images consécutives aux instants (t_{i-1}) et (t_i) .

$$D_{VO}(t_i) = \|\mathbf{C}^n(t_i) - \mathbf{C}^n(t_{i-1})\| \quad (5.4)$$

où $D_{VO}(t_i)$ est l'amplitude du déplacement du centre optique de la caméra \mathbf{C}^n dans le plan horizontal, estimée par l'odométrie visuelle, entre les instants (t_{i-1}) et (t_i) . Enfin, le facteur d'échelle $s(t_i)$ est défini de la manière suivante :

$$s(t_i) = \frac{D_{step}(t_i)}{D_{VO}(t_i)} \quad (5.5)$$

Le facteur d'échelle est calculé à chaque estimation relative de la pose du centre optique de la caméra par le processus d'odométrie visuelle monoculaire, et est utilisé pour corriger l'amplitude du déplacement dans le plan horizontal. La position de la caméra sur le plan horizontal est relative à l'initialisation ou à la dernière estimation absolue de la pose dans un repère global.

5.5 Détermination de la position sur l'axe vertical

Dans le but d'avoir une estimation complète de la pose, plusieurs approches ont été étudiées pour estimer la position de l'équipement tenu en main sur l'axe vertical, et ainsi compléter l'estimation faite sur le plan horizontal. Les différentes approches testées sont présentées ci-dessous.

Double intégration de l'accélération et corrections ZUPT La première approche se base sur la mise en place d'une mise à zéro de la vitesse (ZUPT). Comme décrit dans l'état de l'art, à la différence des équipements montés sur le pied ou fixés sur le corps, les capteurs inertiels intégrés dans les équipements tenus en main ne permettent pas de mettre en place des corrections de type ZUPT pour limiter la dérive des estimations de la vitesse et de la position. En se basant sur les travaux présentés dans [ZHENG et al., 2016], une réinitialisation de la vitesse sur l'axe vertical est proposée pour les équipements tenus en main intégrant une unité de mesures inertielles. Cela permet de faire une estimation de la position sur l'axe vertical par double intégration de l'accélération.

Pression barométrique La seconde approche se base sur l'utilisation d'un baromètre intégré dans les équipements tenus en main par le piéton. En se basant sur les lois hydrostatiques, une variation de hauteur peut être définie en fonction d'une variation de pression :

$$\Delta h(t) = \left(\frac{R}{g}\right) * T(t) * \ln\left(\frac{P_0}{P(t)}\right) \quad (5.6)$$

Avec :

- R, la constante des gaz parfaits, avec $R = 287.05 \text{ J/Kg.K}$;
- g, l'accélération de la pesanteur à la surface de la Terre, avec $g \simeq 9.80665 \text{ m/s}^2$;
- $T(t)$, la température en Kelvin mesurée à l'instant (t), avec $T(\text{K}^\circ) = T(\text{C}^\circ) + 273,15$;
- $P(t)$, la pression en Pascal mesurée à l'instant (t) ;
- P_0 , la pression initiale.

En milieu extérieur et à une altitude donnée, les variations de pression traduisent un changement de conditions météorologique. Ce changement étant un processus lent (i.e. de l'ordre de plusieurs dizaines de minutes), l'hypothèse est faite que sur un cours intervalle de temps, la pression et la température restent constantes pour une hauteur définie. La pression étant également dépendante de l'altitude, si des variations de pression sont observées, cela traduit alors une variation d'altitude. Avec une position initiale sur l'axe vertical connue, les variations de hauteur dues aux mouvements du piéton peuvent alors être estimées. La hauteur dans un repère global peut ensuite être déterminée à chaque instant à partir de la connaissance de la hauteur initiale. Il est important de noter que l'estimation de la hauteur au moyen de la pression barométrique n'est exploitable qu'en milieu extérieur. En effet, des variations de pression sont observables lors des transitions du piéton entre le milieu extérieur et le milieu intérieur.

Utilisation du Modèle Numérique de Terrain La troisième approche se base sur le fait que le piéton est dans un contexte d'utilisation d'application d'aide à la mobilité sur son équipement tenu en main, i.e. qu'il est dans une position où il regarde la scène face à lui à travers un affichage en réalité augmentée et que le plan de l'écran de l'équipement tenu en main est approximativement parallèle à l'axe vertical. Cette position correspond au mode de portage dans lequel les mesures ont été acquises par chacun des piétons lors des différentes acquisitions en milieux urbains. Afin de proposer une estimation de la position de l'équipement sur l'axe vertical, la hauteur de chaque piéton h_{ped} ainsi que la hauteur du centre de l'écran h_{AR} sur lequel un affichage en réalité augmentée serait proposé, ont été mesurées. Ces mesures sont présentées dans le Tableau 5.1. Dans le cadre d'une visualisation d'information en réalité augmentée, l'hypothèse est faite que l'écart entre h_{ped} et h_{AR} reste le même pendant la marche pour un piéton donné.

Piéton	M1	M2	W1	W2	W3
Genre	Homme	Homme	Femme	Femme	Femme
h_{ped}	1.87 m	1.80 m	1.69 m	1.69 m	1.60 m
h_{AR}	1.67 m	1.60 m	1.55 m	1.50 m	1.35 m
Δh	0.20 m	0.20 m	0.14 m	0.19 m	0.25 m

TABLEAU 5.1 – Mesures de la hauteur du piéton et de la hauteur de l'équipement tenu en main.

Cela permet de définir $\Delta\hat{h}$, la différence moyenne entre la hauteur du piéton h_{ped} et la hauteur h_{AR} à laquelle se trouve le centre de l'écran de l'équipement qu'il tient en main, et sur lequel serait projeté des informations en réalité augmentée.

$$\Delta\hat{h} = \frac{1}{n} \sum_1^n (h_{ped} - h_{AR}) \quad (5.7)$$

Cette dernière est estimée expérimentalement sur cinq piétons comme étant de l'ordre de la vingtaine de centimètres. Ainsi, connaissant les dimensions de l'équipement du piéton, notamment Δh_{C-AR} , i.e. l'écart entre la position du centre optique de la caméra et le centre de l'écran, propre à chaque équipement grand public à disposition du piéton, la position du centre optique de la caméra sur l'axe vertical h_C peut être déterminé :

$$h_C = h_{ped} - \Delta\hat{h} + \Delta h_{C-AR} \quad (5.8)$$

Le centre optique de la caméra intégrée au matériel expérimental utilisé lors des différentes acquisitions est au même niveau que le centre de l'écran sur l'axe vertical dans le mode de portage définie précédemment, c'est à dire que $h_C = h_{AR}$ et que $\Delta h_{C-AR} = 0$. De ce fait, la position du centre optique de la caméra sur l'axe vertical est définie en fonction de la hauteur du piéton comme suit :

$$h_C = h_{ped} - \Delta\hat{h} \quad (5.9)$$

Avec la connaissance du contenu des Systèmes d'Information Géographique 3D, et plus particulièrement du Modèle Numérique de Terrain (MNT) de la zone d'acquisition, la position du centre optique de la caméra sur l'axe vertical peut être déterminée en environnement extérieur. Cela présente l'avantage de pouvoir estimer directement la pose de la caméra monoculaire dans une visualisation du contenu des SIG 3D, i.e. la dalle contenant les modèles 3D des bâtiments aux alentours de la position estimée, et ainsi améliorer la reprojection et la caractérisation des données via un affichage en réalité augmentée directement sur site.

5.6 Expérimentation : estimation de la localisation par odométrie visuelle monoculaire mise à l'échelle

5.6.1 Description des acquisitions

Afin d'évaluer la solution proposée dans ce chapitre, un parcours long de 0.7 km a été effectué en milieux urbains par trois piétons, avec en main l'équipement expérimental présenté dans la section 2.4. Ce parcours est en environnement extérieur, avec des passages en canyons urbains et des passages avec d'importantes réflexions spéculaires. Une visualisation de la trajectoire du parcours effectué lors des acquisitions est présentée sur la Figure 5.4. Le détail des différents piétons est présenté dans le Tableau 5.2. Durant toute l'acquisition, chaque piéton a tenu le matériel expérimental en main en mode *Texting*. Plus précisément, dans un mode de portage correspondant à l'observation de la scène face à eux, i.e. avec la caméra faisant face au monde, et un écran sur lequel serait proposé un affichage en réalité augmentée.

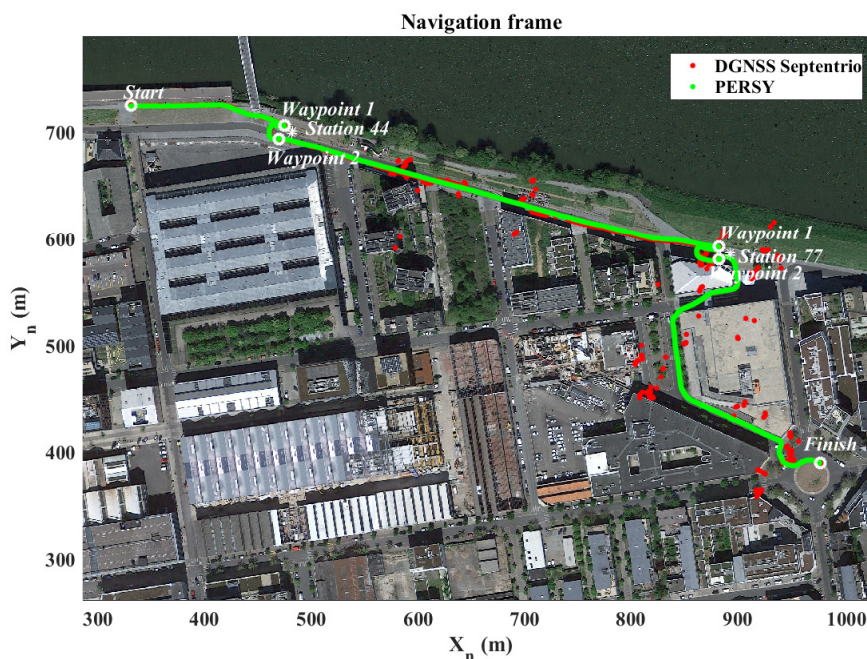


FIGURE 5.4 – Trajectoire de référence pour le parcours de 0.7 km lors de l'acquisition du piéton M2 et points de passages de référence à proximité des objets de référence sur le parcours (stations 44 et 77).

5.6.2 Définition de la référence terrain

Pour évaluer l'estimation de la pose de l'équipement tenu en main par le piéton lors de ses déplacements en milieux urbains, il est nécessaire d'établir une vérité terrain. Comme il est difficile d'avoir une référence sur les six degrés de liberté de l'équipement tenu en main, une évaluation de l'odométrie visuelle monoculaire mise à l'échelle est proposée en termes de distance à une trajectoire de référence sur le plan horizontal.

Récepteur GPS Haute Sensibilité La plateforme ULISS, présentée dans la sous-section 2.4.3, intègre un récepteur HS-GPS qui fournit des positions GPS à 5 Hz [ORTIZ et al., 2017]. Comme décrit dans le Tableau 5.2, avec cette plateforme, les positions de l'équipement tenu en main ne sont pas toujours disponibles, et lorsqu'elles le sont, la précision dépend du lieu où se trouve le piéton, i.e. des différents bâtiments induisant des multi-trajets, et du nombre de satellite disponible. Un écart type de l'erreur de positionnement GPS allant jusqu'à 30 mètres est observé pour les différentes

acquisitions. Cette solution n'est donc pas suffisamment fiable et précise pour être utilisée comme référence terrain.

Récepteur GNSS et traitement en mode différentiel Pour acquérir une trajectoire de référence plus fiable et plus précise, une antenne hélicoïdale attachée au poignet, comme illustré sur la Figure 5.5, et un récepteur GNSS Septentrio situé dans un sac à dos, ont été utilisés pour collecter les signaux des satellites GPS et GLONASS à 5 Hz lors des acquisitions. Avec ces dispositifs, les positions de l'antenne sont déterminées avec RTKLIB¹ à l'aide d'une solution GNSS en différentiel. Comme décrit dans le Tableau 5.2, dans certains espaces urbains, de manière identique au récepteur HS-GPS intégré dans la plateforme ULISS, les positions de l'antenne attachée au poignet ne sont pas disponibles tout le temps et ne sont pas précises lorsque le piéton est dans un canyon urbain. Ces résultats représentent évidemment la difficulté de la localisation en milieu urbain à l'aide de tels capteurs, et est loin de la précision nécessaire pour définir une référence terrain, ainsi que pour le fonctionnement des systèmes de réalité augmentée sur les équipements tenus en main. Cependant, pendant le post-traitement avec RTKLIB pour le positionnement DGNSS, un fichier contenant l'écart type de positionnement est généré. Il est ainsi possible d'identifier les moments où la valeur de l'écart type est très faible et les moments où elle atteint plusieurs mètres. Il est observé dans les fichiers générés et sur la Figure 5.4, que l'écart type de l'erreur de positionnement estimée était inférieur à un mètre lorsque les piétons marchent dans un environnement ouvert et que ce dernier pouvait atteindre 21 mètres lorsque les piétons marchaient dans le canyon urbain. En utilisant un récepteur GNSS et un traitement en différentiel, une trajectoire de référence n'est donc disponible qu'à ces moments précis de l'acquisition, ce qui n'est pas suffisant pour évaluer la précision de l'approche proposée sur l'ensemble de l'acquisition.



FIGURE 5.5 – Matériel expérimental du piéton, antenne hélicoïdale attachée au poignet et récepteur GNSS géodésique dans le sac à dos durant les acquisitions.

Système de référence piétonne PERSY monté sur le pied Afin d'estimer une trajectoire de référence continue sur l'ensemble des acquisitions, la plateforme PERSY [SCORNEC et al., 2017] a été montée sur le pied, du même côté que l'équipement du piéton tenu en main pendant les acquisitions, comme illustré sur la Figure 5.6. Les phases quasi statiques de l'accélération et du champ magnétique sont utilisées pour atténuer les erreurs des capteurs inertiels. Des mesures de phase GNSS sont également utilisées pour améliorer la précision de positionnement. Le traitement des données acquises avec PERSY permet de déterminer la position à 160 Hz et annonce avoir une erreur moyenne de positionnement de 0,3% de la distance parcourue [SCORNEC et al., 2017]. Comme

1. <http://www.rtklib.com/>

Piéton	W1	M1	M2
Genre	Femme	Homme	Homme
h_{ped}	1.69 m	1.87 m	1.80 m
Durée de l'Acquisition	550 s	486 s	574 s
Disponibilité du DGPS	86.3 %	78.9 %	54.8 %
Disponibilité du DGNSS	90.7 %	94.5 %	81.2 %

TABLEAU 5.2 – Informations sur les différents piétons et leur acquisition.

illustré sur la Figure 5.4, les positions estimées avec PERSY sont plus précises que celles calculées avec la solution DGNSS. Par conséquent, bien que la solution DGNSS soit absolue et plus précise en certains endroits de l'acquisition, une comparaison à la solution PERSY, qui est relative à l'initialisation, est préférée pour faire l'évaluation de la solution proposée dans ce chapitre. Les positions obtenues grâce à la solution DGNSS seront cependant affichées à titre de comparaison, et non d'évaluation.

Comme illustré sur la Figure 5.6, en plus de tout ces équipements, les piétons ont marché lors des acquisitions sur des points de référence marqués au sol d'un cercle blanc barré et ont fait un arrêt de plusieurs secondes sur ces derniers. Ces points de référence ont été choisis à proximité des objets d'intérêt, à savoir deux pour chaque objet de référence, afin que le piéton passe par des endroits où ces derniers sont visibles sur les enregistrements vidéo. Ces points de passages permettent d'évaluer la précision de l'estimation de la pose aux instants où les piéton ce situent sur ces points. Les emplacements exacts de ces points de référence ont été déterminé à l'aide de la solution DGNSS présentée dans la sous-section 4.5.2 et la sous-section 5.6.2. La position d'arrivée de l'acquisition a également été déterminée à l'aide de la solution DGNSS afin de pouvoir calculer l'erreur de positionnement en fin d'acquisition, et ainsi s'affranchir de l'erreur de positionnement de PERSY accumulée au cours de la marche.

5.6.3 Initialisation de la pose

Pour mettre en place l'odométrie visuelle monoculaire à l'échelle proposée dans ce chapitre, la pose initiale doit être définie dans le référentiel *Navigation* en début d'acquisition. Pour cela, la position de départ de l'acquisition a été déterminée à l'aide de la solution DGNSS présentée précédemment. Quant à l'orientation, elle a été déterminée au moyen de la solution proposée dans la section 6.2, i.e. l'intégration des signaux inertiels et magnétiques acquis par l'équipement tenu en main dans le filtre d'estimation des angles d'attitudes MAGYQ [RENAUDIN et COMBETTES, 2014].

Lors de l'acquisition en milieu urbain, il est prévu que le piéton marche sur des points de référence choisis à proximité d'objets de référence (i.e. le terminal des stations vélo). La reconnaissance de ces éléments permet d'estimer une pose plus précise que celle estimée par l'odométrie visuelle mise à l'échelle après plusieurs centaines de mètres parcourus. L'estimation de la pose basée sur la reconnaissance d'objet de référence est évaluée dans la section 4.7, et permet d'atteindre une précision de positionnement meilleure que 50 cm sur le plan horizontal. Ainsi, de la même manière que les corrections mises en place sur l'estimation de la trajectoire pour un processus de navigation à l'estime du piéton, présentée dans la section 6.3, l'estimation absolue de la pose de l'équipement tenu en main permet de corriger la pose estimée et de réinitialiser le processus d'odométrie visuelle monoculaire à l'échelle.

Il est important de noter qu'après la perte de détection d'un objet de référence, l'orientation de la trajectoire estimée avec le processus d'odométrie visuelle monoculaire est relative à cette estimation absolue de la pose. Une imprécision dans cette estimation entraîne donc une erreur de positionnement croissante au cours de la marche. La précision de l'estimation de la pose basée sur la reconnaissance d'objet de référence peut être déterminée par l'erreur de reprojection cal-

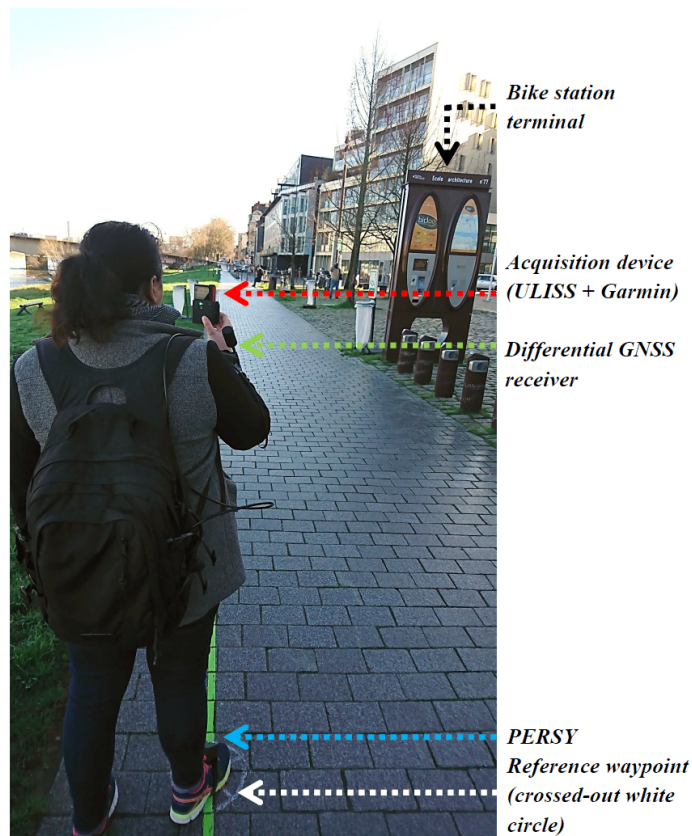


FIGURE 5.6 – Acquisition avec la plateforme expérimentale (ULISS + caméra) en main, l’antenne hélicoïdale attachée au poignet et le récepteur GNSS géodésique dans le sac à dos, la plateforme PERSY attachée au pied pour déterminer la position lors des acquisitions, et les points de référence au sol.

culée dans le processus. Généralement, la dernière vue avant la perte de détection d’un objet de référence est une vue rasante, correspondant au cas le plus dégradé pour estimer la pose, et donc à l’estimation la moins précise. Par conséquent, lorsque le piéton marque un arrêt de plusieurs secondes devant l’objet de référence, l’utilisation de la pose estimée à cet instant est préférée à l’utilisation de la dernière vue de l’objet de référence. Cela correspond au cas le plus précis de l’estimation de la pose basée sur la reconnaissance d’un objet de référence.

5.6.4 Résultats et analyses

Estimation des longueurs de pas du piéton

Comme présenté dans la section 5.3, les signaux de l’accéléromètre et du gyroscope acquis avec l’équipement tenu en main sont analysés. La Figure 5.7-(a) présente la détection des phases de marche et des phases statiques du piéton en opérant une détection d’extrêmes locaux et un seuillage sur l’énergie des signaux. L’analyse de la variance des signaux de l’accéléromètre et du gyroscope permet ensuite de faire une classification des activités lors des phases de marche, et plus particulièrement pour définir le mode de portage de l’équipement tenu en main (i.e. *Texting*, *Swinging* et *Phoning*). L’observation de la classification confirme bien que le piéton marche avec l’appareil tenu en main avec un portage en mode *Texting* (en vert), et que les moments où le piéton est statique (en rose) sur chacun des points de référence définis à proximité des objets de référence, sont correctement identifiées. La classification et l’analyse des signaux permet ensuite de déterminer la fréquence de pas du piéton, et d’estimer les longueurs de ses pas en appliquant le modèle précédemment défini. La Figure 5.7-(b) présente les longueurs de pas estimées pour le piéton *M1* lors de l’acquisition en milieux urbains grâce à l’analyse de sa fréquence de marche et la connaissance de sa hauteur.

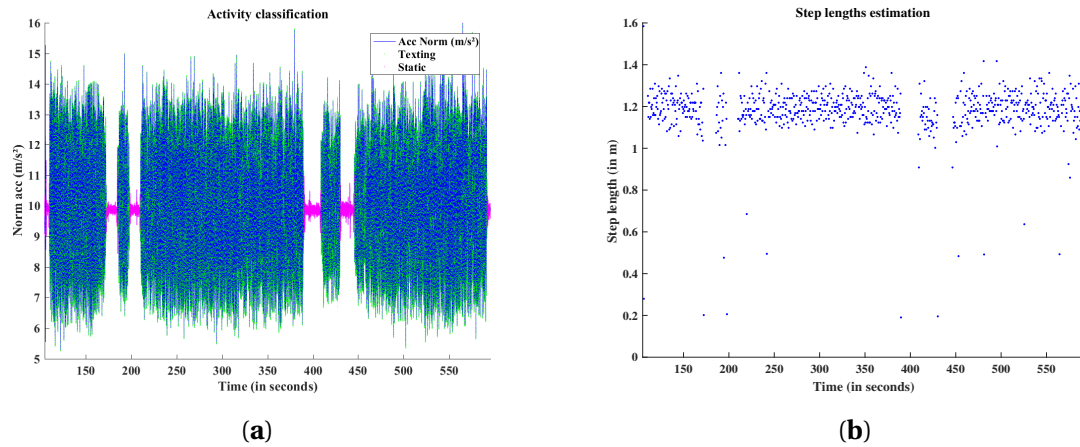


FIGURE 5.7 – (a) Classification de l'activité du piéton : Statique (en rose), Marche avec portage en mode *Texting* (en vert); (b) Estimation des longueurs de pas du piéton M1.

Odométrie visuelle mise à l'échelle

La Figure 5.8-(a) présente les trajectoires de référence, respectivement obtenues avec la solution avec PERSY (en bleu) et avec la solution DGNSS (en vert), et la trajectoire estimée avec la solution proposée dans ce chapitre (en rouge), pour les acquisitions des trois piétons différents. A partir de ces résultats, il est observé que les trajectoires estimées sont proches des trajectoires de référence. Cela signifie que la solution proposée, reposant sur l'estimation des longueurs de pas du piéton lors de ses déplacements en milieux urbains, permet de correctement mettre à l'échelle l'odométrie visuelle monoculaire.

Pour évaluer les performances de l'approche proposée, l'erreur de positionnement sur le plan horizontal est utilisée. Cette dernière correspond à la distance entre la position estimée et celles déterminées par chacune de solutions de référence. Les trajectoires estimées pour chacun des piétons sont présentées sur la Figure 5.8-(b).

Piéton	WI	M1	M2
Objet connu détecté	660 img (12 %)	768 img (15.8 %)	726 img (12.6 %)
% d'erreur de positionnement	< 5 m < 10 m	< 5 m < 10 m	< 5 m < 10 m
A la référence DGNSS (Partie 1)	91.97 % 99.73 %	67.73 % 81.34 %	59.79 % 63.95 %
A la référence PERSY (Partie 1)	92.95 % 99.19 %	48.69 % 87.42 %	61.28 % 63.93 %
A la référence DGNSS (Totale)	Référence fausse	Référence fausse	Référence fausse
A la référence PERSY (Totale)	53.95 % 58.33 %	43.71 % 77.34 %	47.21 % 57.31 %
Erreur moyenne de positionnement			
A la référence DGNSS (Partie 1)	3.58 m	5.26 m	7.82 m
A la référence PERSY (Partie 1)	6.60 m	5.14 m	7.74 m
A la référence DGNSS (Totale)	Référence fausse	Référence fausse	Référence fausse
A la référence PERSY (Totale)	9.39 m	5.75 m	9.99 m

TABLEAU 5.3 – Erreurs de positionnement.

Le Tableau 5.3 présente le pourcentage du trajet pour lequel l'erreur de positionnement de la solution proposée, par rapport aux références DGNS et PERSY, est maintenue en dessous d'un certain seuil. Les erreurs moyennes de positionnement, entre la solution proposée et les références utilisées, sont également présentées. Pour chaque piéton, le pourcentage et l'erreur moyenne de positionnement sont données pour la première partie du parcours, i.e. correspondent à la partie où le DGNS est précis, et pour l'ensemble de l'acquisition, i.e. comprenant des passages où le DGNS n'est pas disponible ou est très imprécis. Il est à noter que la deuxième moitié du parcours comprend également des passages où la vision est perturbée par d'importantes réflexions spéculaires.

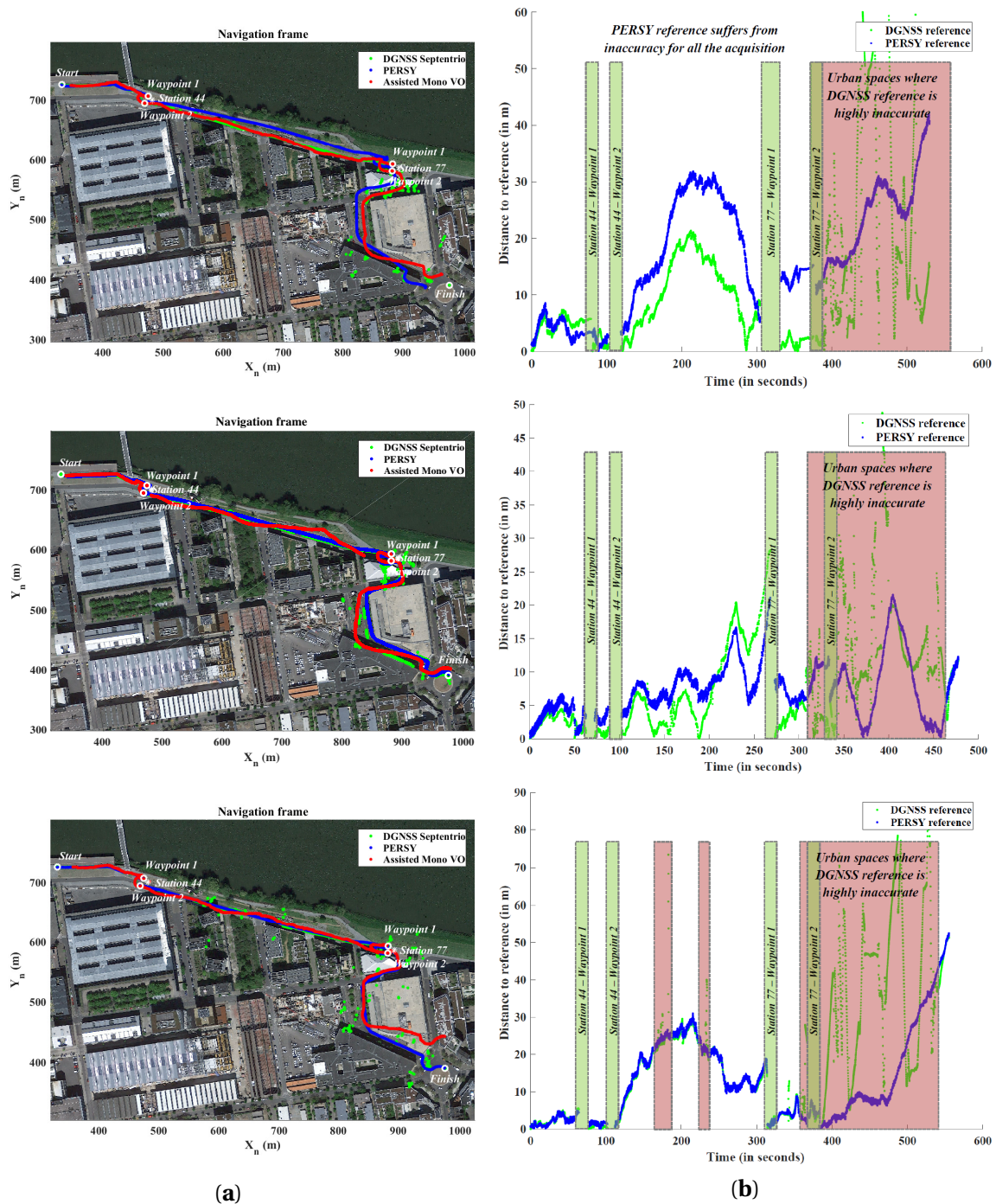


FIGURE 5.8 – (a) Trajectoires de référence au DGNS (en vert) et à PERSY (en bleu), trajectoire estimée avec l'odométrie visuelle mise à l'échelle (en rouge); (b) Erreur de positionnement de l'odométrie visuelle mise à l'échelle à la solution DGNS (en vert) et à la solution PERSY (en bleu). Les résultats sont présentés pour les piétons $W1$, $M1$ et $M2$ (du haut vers le bas).

En milieux urbains le piéton se retrouve fréquemment dans la situation où il doit éviter les autres piétons, les cyclistes, les voitures, les éléments de mobiliers urbain, et attendre aux passages pour piétons, l'obligeant à faire de petits déplacements ou à s'arrêter. De ce fait, la fréquence de marche du piéton et les longueurs de pas estimées varient de manière importante. Ceci est mis en évidence dans les acquisitions réalisées lorsque le piéton se déplace entre les deux points de référence d'une même station, se retrouvant ainsi dans la situation où il doit contourner des éléments du mobilier et éviter des piétons et des cyclistes. Une visualisation de la trajectoire estimée dans ces cas est proposée avec la Figure 5.9. Dans ce cas, les solutions courantes (i.e. positionnement GPS et navigation à l'estime) avec un appareil portatif n'ont pas bien fonctionné alors que la solution proposée dans ce chapitre a permis d'estimer une trajectoire lisse. Il est à noter que l'utilisation d'une longueur moyenne de pas [TOMAZIC et SKRJANC, 2015] aurait, dans une moindre mesure, également faussée l'échelle et l'estimation du déplacement.

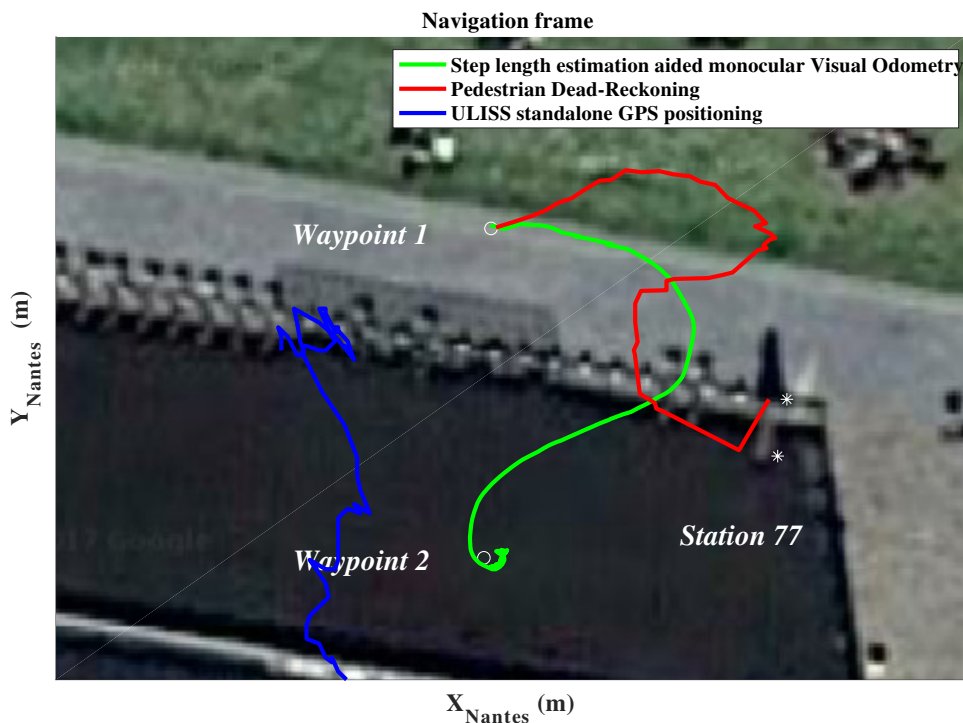


FIGURE 5.9 – Trajectoires estimées lors de petits déplacements du piéton avec le positionnement par GPS (en bleu), par navigation à l'estime du piéton (en rouge), par odométrie visuelle monoculaire à l'échelle (en vert).

Afin de compléter cette évaluation, la Figure 5.10 présente le pourcentage d'erreur de la solution en fonction de la distance parcourue par les différents piétons. Hormis pendant les premiers mètres parcourus, le pourcentage d'erreur de positionnement est maintenu en dessous des 10 %. Il est à noter que la correction effectuée, grâce à la reconnaissance d'un élément connu, permet d'atteindre un pourcentage d'erreur de quasiment 0 %. En comparaison des méthodes usuelles présentées dans section 3.2, la méthode proposée dans ce chapitre et les corrections permettent de maintenir une précision équivalente, voir meilleure. En effet, les solutions les plus précises de l'état de l'art permettent d'atteindre 5 % d'erreur sur quelques dizaines ou centaines de mètres parcourus, alors que la solution proposée présente ces performances sur quasiment un kilomètre.

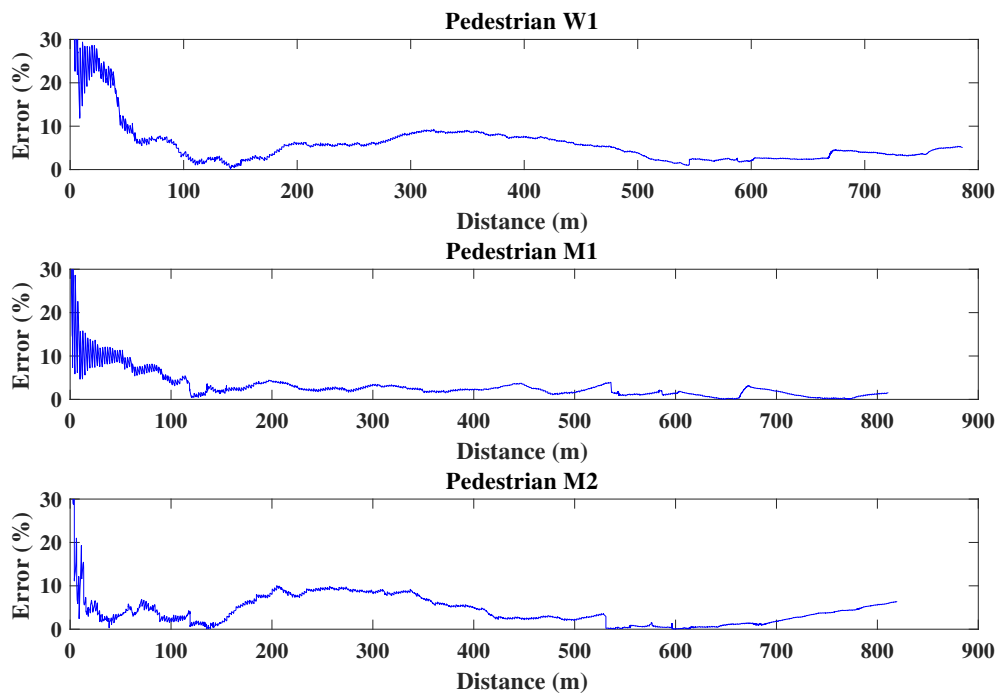


FIGURE 5.10 – Pourcentage d'erreur de positionnement en fonction de la distance parcourue pour les différents piétons.

Estimation de la position sur l'axe vertical

Double intégration de l'accélération et ZUPT Comme présenté dans la section 5.5 et dans [ZHENG et al., 2016], une détection des extrêmes locaux est opérée sur le signal de l'accéléromètre sur l'axe vertical. Cette détection est présentée sur la Figure 5.11-(a). Les minimas correspondant aux instants où le talon rentre en contact avec le sol, et les maximas correspondant aux instants où le pied est à plat sur le sol. La remise à zéro de la vitesse est ainsi opérée à ces instants spécifiques. En plus de cela, une détection des phases statiques du piéton est opérée sur le signal de l'accéléromètre. Cette détection est similaire à celle proposée pour l'estimation des longueurs de pas du piéton. Ces phases statiques sont présentée sur la Figure 5.11-(b). Une remise à zéro de la vitesse est également opérée lors de ces phases. Il est à noter que les résultats sont présentés pour le parcours de 3 km présenté dans la sous-section 6.6.1. L'acquisition étant plus longue, cela permet d'observer la dérive de l'estimation au cours de la marche sur un parcours long.

Ainsi, la vitesse sur l'axe vertical est calculée en intégrant une première fois le signal de l'accéléromètre sur sa composante verticale. La vitesse calculée, après application des ZUPT, est présentée sur la Figure 5.12. Les variations de la position sur l'axe vertical sont ensuite calculées en intégrant la vitesse. Comme définit dans la section 5.5, une hauteur initiale de l'équipement tenu en main est définit comme la hauteur du piéton à laquelle on soustrait une vingtaine de centimètres.

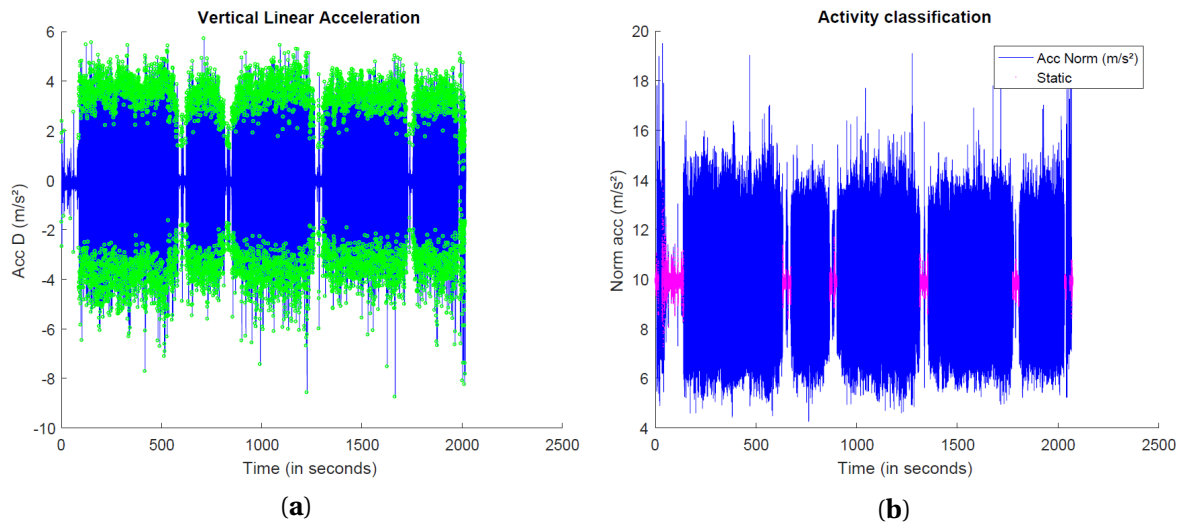


FIGURE 5.11 – (a) Détection des pics sur le signal de l'accéléromètre sur l'axe vertical; (b) Détection des phases statiques sur le signal de l'accéléromètre.

Cependant, la Figure 5.12 montre que l'amplitude des variations de la position sur l'axe vertical est de l'ordre de la centaine de mètres pour un parcours faisant une boucle de 3 kilomètres. Cette solution ne permet donc pas d'estimer la position sur l'axe vertical avec une précision suffisante pour permettre un affichage correct en réalité augmentée.

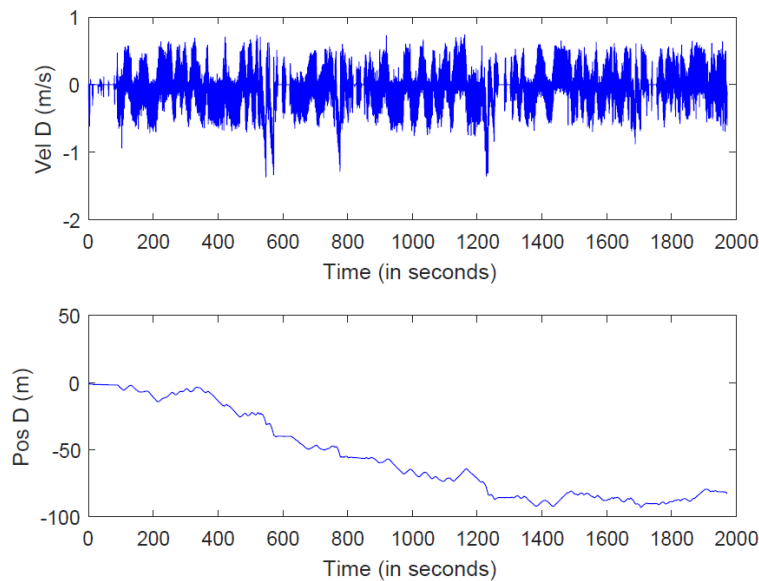


FIGURE 5.12 – Estimation de la vitesse avec application des ZUPT et estimation des variations de la position sur l'axe vertical.

Pression barométrique Comme présentée dans la section 5.5, la position sur l'axe vertical peut être estimée en milieux urbains grâce à la pression enregistrée par le baromètre intégré dans l'équipement tenu en main par le piéton. La mesure brute de la pression en sortie du baromètre est très bruitée, comme cela peut être observé sur la Figure 5.13-(a). Ainsi, pour être traitée plus facilement, et mieux refléter les changements de pression lors de la navigation du piéton en milieu urbain, une moyenne glissante est appliquée sur l'ensemble du signal. Il est à noter que les résultats sont également présentés pour le parcours de 3 km présenté dans la sous-section 6.6.1. L'acquisition étant plus longue, cela permet d'observer des changements de pression atmosphérique et l'impact que cela peut avoir sur l'estimation.

La Figure 5.13-(b) présente la mesure de la température durant une acquisition en milieu urbain. Il est facilement remarquable que la température mesurée est croissante et ne correspond pas à la température extérieure, mais certainement à la chaleur dissipée par l'équipement tenu en main par le piéton dans lequel est intégré le capteur de température. Du fait que la température mesurée n'est pas exploitable, dans des conditions classiques d'acquisition en milieu extérieur, l'hypothèse d'une température constante est faite pour une acquisition de courte durée. Ainsi, si la température est connue via une source d'information extérieure, cette dernière est utilisée comme constante. Autrement, si l'équipement est à température ambiante suite à une veille prolongée, la température mesurée en début d'acquisition est utilisée comme constante. Enfin, si aucun de ces cas ne se présente, une température arbitraire, mais réaliste, est fixée comme constante. Il est montré dans les calculs d'estimation de la variation de hauteur en fonction de la pression atmosphérique, que l'impact de la température et de ses variations ne sont que minimes.

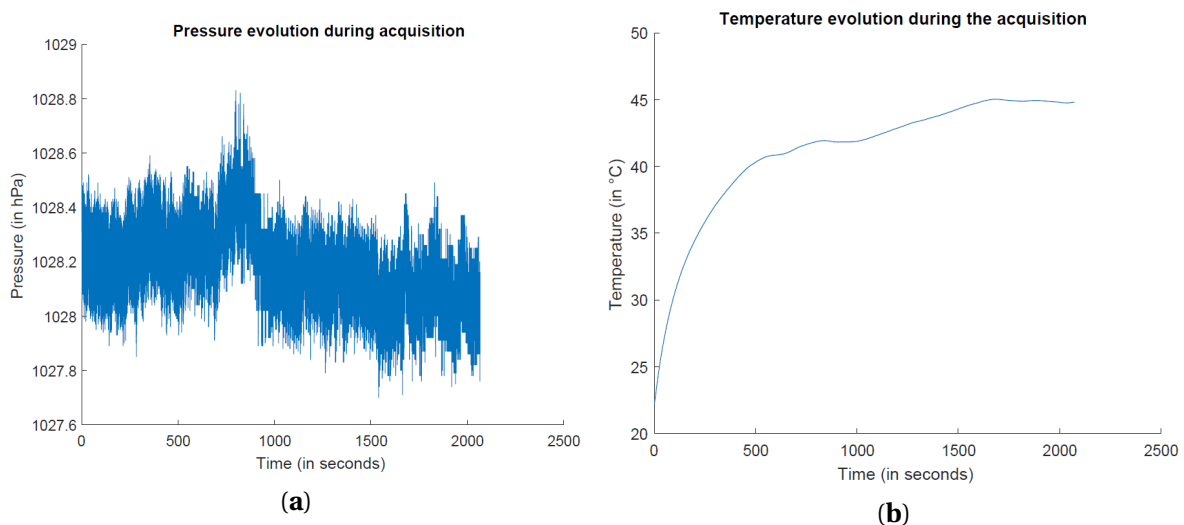


FIGURE 5.13 – (a) Mesures brutes de la pression durant l'acquisition; (b) Mesure de la température durant l'acquisition en milieu extérieur.

Ainsi, les variations de position sur l'axe vertical sont estimées et présentées sur la Figure 5.14 pour une acquisition de 40 minutes en milieu urbain. La hauteur initiale a été définie comme la hauteur du piéton moins une vingtaine de centimètres. Il peut être noté qu'une différence de 2.5 mètres est observée à la fin d'une boucle 3 kilomètres en environnement extérieur. Il est à prendre en compte que les acquisitions ont été effectuées dans la région Loire Atlantique. Cette dernière présentant des changements climatiques rapides, il se peut donc que la pression ait varié entre le début et la fin de l'acquisition, entraînant par conséquent des variations dans l'estimation de la hauteur. Malgré les quelques mètres d'erreur en fin de parcours, cette méthode est bien plus précise que la précédente pour estimer la position sur l'axe verticale. En effet, à la fin du parcours de 3 kilomètres, une erreur largement inférieure à 1% de la distance parcourue est atteinte. Cette méthode présente l'avantage d'être précise dans l'estimation de la variation de hauteur entre les instants de deux images consécutives, ou même sur de courte période.

Utilisation du Modèle Numérique de Terrain Comme présentée dans la section 5.5, en connaissant la hauteur du piéton, la position sur l'axe vertical peut être déterminée par rapport au Modèle Numérique de Terrain. Des visualisations des trajectoires de l'équipement tenu en main par le piéton dans le modèle 3D et dans *Google Earth* sont présentées sur la Figure 5.15. Bien que cette solution soit "pratique" pour compléter l'estimation de la pose, et ainsi permettre un affichage en réalité augmentée sur site, elle reste dépendante de la connaissance du Modèle Numérique de Terrain. En France, le Modèle Numérique de Terrain existe pour l'ensemble du territoire, cependant sa disposition constitue une première limite de la solution. De plus, de part la nature et l'architecture de l'environnement urbain, les milieux dans lesquels évoluent les piétons ne correspondent

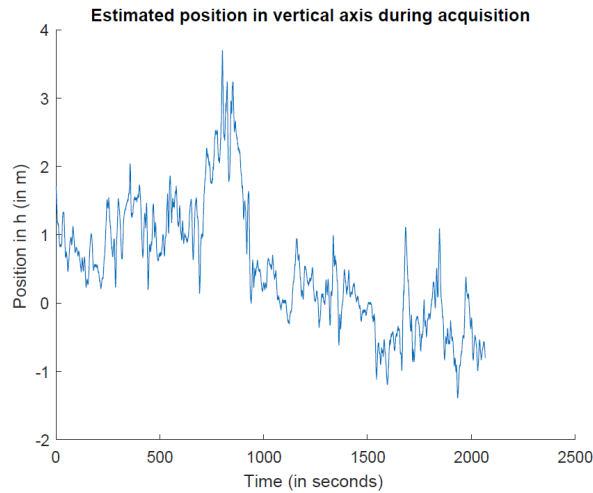
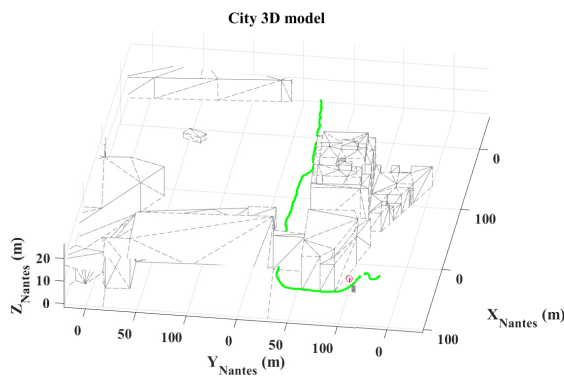
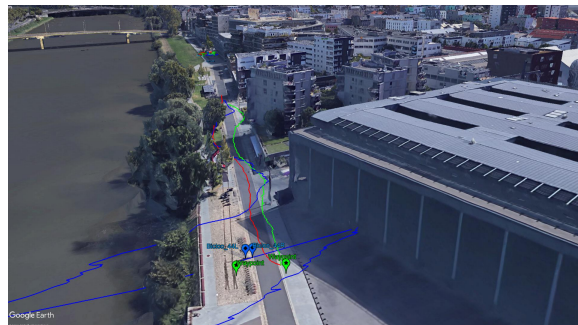
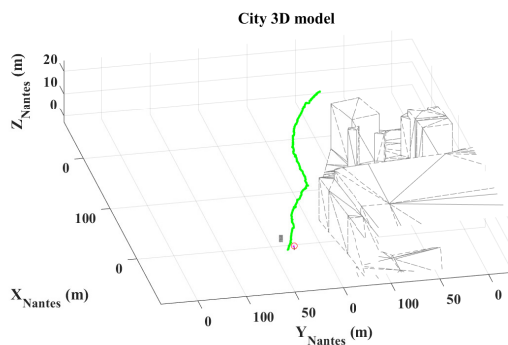


FIGURE 5.14 – Hauteur estimée durant l'acquisition.

pas forcément au Modèle Numérique de Terrain, introduisant de fait une erreur dans l'estimation de la hauteur de l'équipement tenu en main. Ainsi, la précision du Modèle Numérique de Terrain et les différences par rapport à l'architecture de l'environnement urbain constituent un seconde limite de la solution.



(a)

(b)

FIGURE 5.15 – (a) Visualisation dans le modèle 3D de la trajectoire estimée par odométrie visuelle mise à l'échelle; (b) Visualisation dans Google Earth de la trajectoire estimée par odométrie visuelle mise à l'échelle.

5.7 Conclusion

Dans le contexte de navigation piétonne en milieux urbains, la solution proposée dans ce chapitre permet une mise à l'échelle de l'odométrie visuelle monoculaire par l'estimation des longueurs de pas du piéton. Cela a pour objectif d'estimer en continu la pose de l'équipement tenu en main par le piéton lors de ces déplacements. La pose estimée n'est pas parfaite et plusieurs situations problématiques ont été identifiées. Il est apparu que la détermination des coordonnées des points du monde en 3D dans le processus d'odométrie visuelle monoculaire souffre d'imprécisions dans les zones ouvertes et dans les canyons urbains lorsque la caméra est pointée dans l'axe de la route. L'extraction des points d'intérêt, correspondant à des éléments qui se trouvent à plusieurs centaines de mètres de la caméra, apparaît ne pas être suffisamment précise pour permettre un suivi fiable de la pose. De plus, les passages avec des réflexions spéculaires (e.g. fenêtres et portes en verre) introduisent également des erreurs dans la détermination des coordonnées 3D des points d'intérêt. Ces éléments introduisent une erreur dans l'estimation de la pose, entraînant une erreur de positionnement croissante pendant la marche. Ces problèmes pourraient être atténués ou résolus à l'aide d'une méthode d'odométrie visuelle monoculaire plus sophistiquée telle que proposée dans [ENGEL et al., 2013].

Il est à noter que le sol n'est pas un plan horizontal parfait, et bien que le Modèle Numérique de Terrain de la zone d'acquisition soit connu, il n'est cependant pas suffisamment détaillé pour éviter les erreurs sur l'axe verticale correspondant à la topologie de l'environnement (e.g. trottoirs, esplanades, pontons, etc.). Tant que l'erreur de positionnement sur l'axe vertical n'excède pas quelques dizaines de centimètres, l'impact sera négligeable sur la qualité de la reprojection en réalité augmentée sur site.

Dans certaines situations, l'odométrie visuelle implémentée ne permet également pas d'estimer avec précision les changements d'orientation. Cela introduit un biais dans l'estimation de l'orientation, et entraîne une augmentation de l'erreur de positionnement pendant la marche.

Malgré tous les problèmes relevés, après analyse des résultats d'acquisition, il reste important de noter que l'objectif de résoudre le problème d'ambiguïté d'échelle de l'odométrie visuelle monoculaire est atteint. Il est observé que les erreurs de positionnement, majoritairement maintenues en dessous de la dizaine de mètres, étaient principalement dues à des erreurs d'estimation de l'orientation, que l'on retrouve également dans les processus de navigation à l'estime du piéton. L'approche proposée permet ainsi d'estimer directement l'échelle de l'odométrie visuelle monoculaire, et ce de manière continue, sans que le piéton n'ait à repasser au même endroit pour appliquer des techniques de *loop closure*. Cela ne correspondait pas à la logique de déplacement du piéton en milieu urbain. Cette solution peut également constituer une piste d'amélioration des algorithmes de SLAM pour déterminer l'échelle de la carte générée. La mise à profit de la reconnaissance d'objets de référence permet de réinitialiser l'estimation de la pose, et ainsi de corriger ponctuellement les dérives du processus d'odométrie visuelle monoculaire.

Chapitre 6

Fusion de données et estimation continue de la pose d'un équipement tenu en main

« La mesure d'une erreur est en même temps la mesure de la vérité correspondante. »

Claude de Saint-Martin

Sommaire

6.1 Introduction	70
6.2 Estimation des angles d'attitudes	70
6.3 Navigation à l'estime et corrections absolues	70
6.3.1 Estimation des longueurs de pas du piéton	71
6.3.2 Estimation relative de la position et corrections absolues	71
6.4 Estimation de la pose en statique avec couplage sur l'orientation	72
6.5 Estimation de la pose en dynamique	73
6.6 Expérimentation : corrections absolues d'un processus PDR	74
6.6.1 Description de l'expérimentation	74
6.6.2 Définition de la référence terrain	74
6.6.3 Résultats et analyses	76
6.7 Expérimentation : estimation continue de la pose	78
6.7.1 Description des acquisitions	78
6.7.2 Définition de la référence terrain	79
6.7.3 Résultats et analyses	79
6.8 Conclusion	80

6.1 Introduction

Ce chapitre présente les différentes approches proposées pour estimer en continu une pose qualifiée. Il est ainsi proposé de fusionner les données de l'unité de mesures inertielles et magnétiques avec les informations extraites grâce à la vision monoculaire. Le cœur de cette estimation repose sur la capacité à fusionner les données issues de ces capteurs, et sur l'utilisation d'informations opportunes pour corriger leur dérive. Les différentes méthodes d'hybridation qui ont été utilisées et/ou développées sont présentées dans les sections suivantes. Les résultats finaux de ces travaux de thèse y sont également présentés et analysés.

6.2 Estimation des angles d'attitudes

Pour estimer l'orientation \mathbf{R}_b^n du *Body* dans le référentiel *Navigation*, les données de l'unité de mesures inertielles et magnétiques sont fusionnées. Pour réaliser cette fusion, l'algorithme MAGYQ (Magnetic, Acceleration fields and Gyroscope Quaternion) est utilisé [RENAUDIN et COMBETTES, 2014]. Ce dernier est un filtre d'estimation d'attitude, basé sur le filtre de Kalman étendu, qui fusionne les signaux inertiels et magnétiques, et estime continuellement les erreurs en utilisant un paramétrage des angles sous forme de quaternion. Les modèles d'erreurs des capteurs définis dans la sous-section 2.4.3 sont ainsi implémentés dans les équations de mécanisation du filtre. Pour estimer les angles d'attitude de l'unité de mesures inertielles et magnétiques \mathbf{R}_b^n et atténuer les biais des capteurs, la mécanisation du filtre tire profit de phases opportunes sur les mesures de l'accéléromètre et du magnétomètre. Il est à noter que dans les équations suivantes, une valeur estimée est notée avec un chapeau ($\hat{\mathbf{q}}$) et une valeur conjuguée est notée avec une barre ($\bar{\mathbf{q}}$).

- La mise à jour *Quasi Static Acceleration* (QSA) : est réalisée lorsqu'une phase quasi statique dans le champ d'accélération \mathbf{f}_{ref}^n est observée;

$$(\delta \mathbf{z}_a)_{\mathbf{q}} = \left(\mathbf{f}_{ref}^n \right)_{\mathbf{q}} - \hat{\mathbf{q}}_b^n \otimes \left(\mathbf{y}_a^b - \hat{\mathbf{b}}_a \right)_{\mathbf{q}} \otimes \bar{\mathbf{q}}_b^n \quad (6.1)$$

- La mise à jour *Quasi Static Filed* (QSF) : est réalisée lorsqu'une phase quasi statique dans le champ magnétique \mathbf{m}_{ref}^n est observée, et ce malgré la présence de perturbations magnétiques.

$$(\delta \mathbf{z}_m)_{\mathbf{q}} = \left(\mathbf{m}_{ref}^n \right)_{\mathbf{q}} - \hat{\mathbf{q}}_b^n \otimes \left(\mathbf{y}_m^b \right)_{\mathbf{q}} \otimes \bar{\mathbf{q}}_b^n \quad (6.2)$$

Cette stratégie permet ainsi de corriger la dérive des capteurs et d'estimer les éléments de covariance associés aux angles d'attitude.

6.3 Navigation à l'estime et corrections absolues

Classiquement, les signaux inertiels sont intégrés selon une mécanisation "strap-down" pour calculer récursivement les positions du piéton (cf. sous-section 3.2.2). Ceci n'est possible que si la dérive des capteurs inertiels est fréquemment étalonnée ou réinitialisée, par exemple en utilisant une remise à zéro de la vitesse lors de la phase d'appui du pied sur le sol, connu sous le terme de "Zero velocity UPdaTes" (ZUPT) [FISCHER et al., 2013; SCORNEC et al., 2017]. Cependant, la correction ZUPT n'est pas facilement réalisable dans le contexte de la navigation piétonne avec un équipement tenu en main [ZHENG et al., 2016]. En effet, les moments où la vitesse est nulle se produisent rarement, contrairement au pied où les instants de vitesse nulle se produisent quand le pied est posé à plat sur le sol. De plus, lors de ses déplacements, le piéton est généralement représenté par la position de son centre de masse. Ceci est évidemment différent de l'emplacement de l'appareil tenu à la main, de fait que la main peut effectuer des mouvements libres même lorsque le centre de masse du piéton est statique. Par conséquent, au lieu de faire une double intégration de l'accélération mesurée, un modèle de longueur de pas est généralement adopté pour estimer les déplacements du piéton [LEPPÄKOSKI et al., 2012; RENAUDIN et al., 2012].

6.3.1 Estimation des longueurs de pas du piéton

Comme présenté dans la sous-section 5.3.1 et défini dans [RENAUDIN et al., 2012], un modèle générique est utilisé pour estimer les longueurs de pas $step_k$ du piéton. Ce dernier est basé sur la hauteur h_{ped} du piéton, sur la fréquence de pas f_k et sur un ensemble de trois paramètres $\{a, b, c\}$ déterminés sur 12 sujets.

$$step_k = h_{ped} \times (a \times f_k + b) + c \quad (6.3)$$

6.3.2 Estimation relative de la position et corrections absolues

Dans le processus de navigation à l'estime, la position \mathbf{p}_b^n , i.e. la position de l'équipement tenu en main par le piéton dans le référentiel *Navigation*, est une position relative par rapport à une position absolue de départ \mathbf{p}_0 . Une orientation initiale est également définie en comparant, respectivement les mesures de l'accéléromètre à la force gravitationnelle terrestre \mathbf{g} et les mesures du magnétomètre au champ géomagnétique terrestre \mathbf{B}_{Earth} . La trajectoire est ensuite calculée de manière itérative en appliquant, à chaque instant de pas, la longueur de pas estimée et le cap de marche (*yaw* ou *heading*).

$$\mathbf{p}_b^n(t_{step_{k+1}}) = \mathbf{p}_b^n(t_{step_k}) + step_k \times \theta_b^n(t_{step_k}) \quad (6.4)$$

Les différents caps de marche θ_b^n sont extraits des angles d'attitudes \mathbf{R}_b^n estimés grâce à la fusion des données de l'unité de mesures inertielles et magnétiques dans le filtre MAGYQ présenté dans la section 6.2. Une illustration du principe de navigation à l'estime du piéton est présenté sur la Figure 6.1.

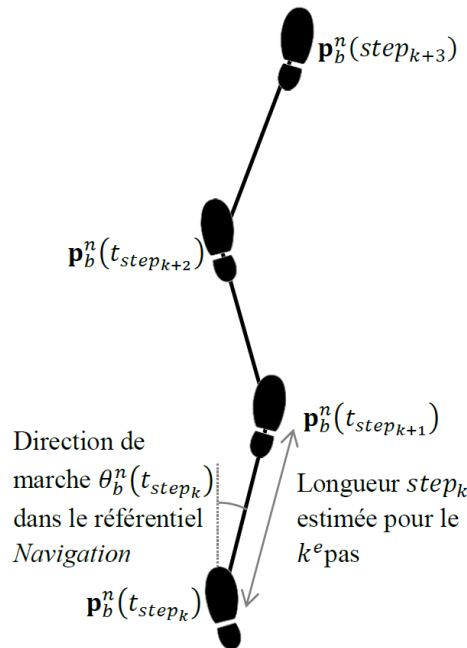


FIGURE 6.1 – Visualisation du principe de la navigation à l'estime du piéton.

De part la nature des capteurs, l'estimation des longueurs de pas et de l'estimation des caps de marche présentent une incertitude qui se traduit par une erreur de positionnement dans le processus de navigation à l'estime. Ainsi, lorsqu'un objet de référence est détecté, l'estimation absolue de la pose de la caméra dans le référentiel *Navigation* est utilisée pour corriger cette dérive. De plus, comme présenté dans la section 6.4, l'estimation absolue de la pose dans le référentiel *Navigation* sera fusionnée avec les orientations estimées au moyen des données inertielles et magnétiques.

6.4 Estimation de la pose en statique avec couplage sur l'orientation

La première méthode, décrite dans cette section, propose le couplage par moyenne pondérée de deux matrices de rotation. La première matrice de rotation est basée sur l'estimation de l'attitude au moyen des mesures inertielles et magnétiques présentée dans la section 6.2. La seconde matrice de rotation est basée sur l'estimation de la pose au moyen de la vision monoculaire et de la reconnaissance d'un objet de référence issu d'un SIG 3D présentée dans la section 4.5.

La fusion des matrices de rotation est effectuée en utilisant leurs expressions dans le domaine quaternion pour éviter un blocage de cardan, mais également pour simplifier les calculs en linéarisant les équations. Après avoir converti les deux matrices de rotation \mathbf{R}_{bIMMU}^n et $\mathbf{R}_{bVision}^n$ en quaternions, respectivement \mathbf{q}_{bIMMU}^n et $\mathbf{q}_{bVision}^n$, une moyenne pondérée est calculée avec la formulation suivante :

$$\mathbf{q}_{bFusion}^n = (\mathbf{q}_{bIMMU}^n)^{1-w} \otimes (\mathbf{q}_{bVision}^n)^w \quad (6.5)$$

Le facteur de pondération w , entre les estimations basées sur la vision monoculaire et celles basées sur les mesures de l'unité de mesures inertielles et magnétiques, est calculé comme suit :

$$w = \frac{\sigma_{IMMU}}{\sigma_{IMMU} + \sigma_{Vision}} \quad (6.6)$$

D'un côté, l'algorithme MAGYQ permet d'estimer l'incertitude angulaire σ_{IMMU} associé à l'estimation de l'orientation basée sur les mesures inertielles et magnétiques. De l'autre, les erreurs de reprojection θ_i pour l'ensemble des points reprojétés i , définies dans la sous-section 4.6.3, permettent de calculer une incertitude angulaire moyenne σ_{Vision} associée à l'estimation de la pose basée sur la vision monoculaire et la reconnaissance d'un objet de référence issu d'un SIG 3D, avec N le nombre de points géométriques utilisés.

$$\sigma_{Vision} = \frac{1}{N} \sum_{i=1}^N \theta_i \quad (6.7)$$

La nouvelle matrice de rotation $\mathbf{R}_{bFusion}^n$, obtenue grâce au calcul de $\mathbf{q}_{bFusion}^n$, est alors convertie en $\mathbf{R}_{oFusion}^c$, i.e. l'orientation du référentiel *Objet* par rapport à la caméra.

$$\mathbf{R}_{oFusion}^c = \mathbf{R}_b^c (\mathbf{R}_{bFusion}^n)^T \mathbf{R}_o^n \quad (6.8)$$

Une nouvelle pose $[\mathbf{R}_{oFusion}^c \mid \mathbf{t}_{oFusion}^c]$ est alors déterminée grâce à l'équation d'estimation de la pose basée sur la connaissance des mises en correspondance 2D-3D.

$$\tilde{\mathbf{x}}_i = \mathbf{K} [\mathbf{R}_{oFusion}^c \mid \mathbf{t}_{oFusion}^c] \tilde{\mathbf{X}}_o \quad (6.9)$$

Comme introduit dans la section 4.5, la nouvelle position de la caméra est alors déterminée dans le référentiel *Objet*.

$$\mathbf{t}_{oFusion}^c = -\mathbf{R}_{oFusion}^c \mathbf{C}_{oFusion}^o \quad (6.10)$$

Il ne reste plus qu'à reprendre les informations associées à l'objet de référence détecté ayant servi à l'estimation de l'orientation $\mathbf{R}_{bVision}^n$, pour replacer la position de l'équipement dans le référentiel *Navigation*.

$$\mathbf{C}_{oFusion}^n = [\mathbf{R}_o^n \mid \mathbf{t}_o^n] \mathbf{C}_{oFusion}^o \quad (6.11)$$

La pose $[\mathbf{R}_{bFusion}^n \mid \mathbf{t}_{bFusion}^n]$ peut ainsi être estimée de manière complète, que ce soit dans le référentiel *Objet* ou dans le référentiel *Navigation*.

$$\mathbf{t}_{bFusion}^n = -\mathbf{R}_{bFusion}^n \mathbf{C}_{oFusion}^n \quad (6.12)$$

6.5 Estimation de la pose en dynamique

La seconde méthode, décrite dans cette section, propose le couplage des données issues d'une part de l'unité de mesures inertielles et magnétiques, et d'autre part de la vision grâce à la reconnaissance d'objet de référence SIG 3D et à l'odométrie visuelle monoculaire mise à l'échelle par l'estimation des longueurs de pas du piéton. La Figure 6.2 présente une vue d'ensemble de l'approche proposée.

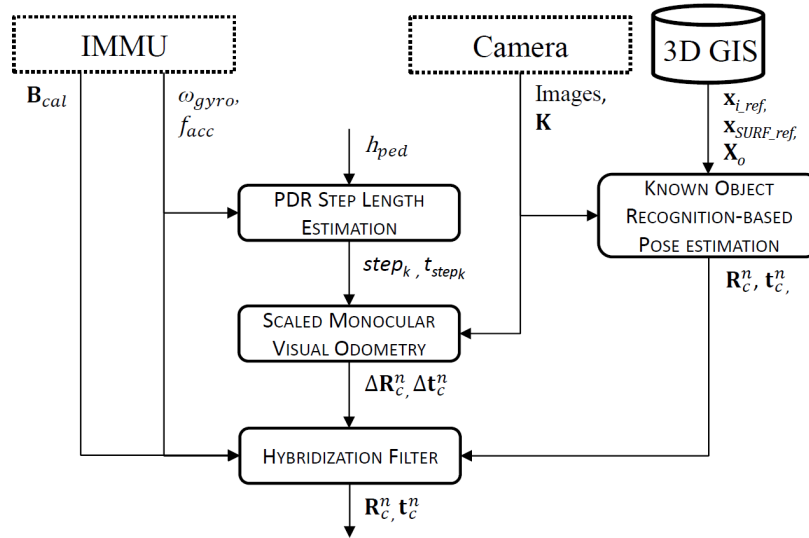


FIGURE 6.2 – Diagramme bloc de la solution de fusion entre les mesures inertielles et magnétiques, et les informations extraites de la vision monoculaire.

Dans l'algorithme MAGYQ, présenté dans la section 6.2, les deux mises à jour QSA et QSF sont utilisées dans le filtre de Kalman étendu pour estimer l'orientation R_b^n du *Body* dans le référentiel *Navigation* et atténuer les biais des capteurs. Grâce aux estimations absolues et relatives de la pose basée sur la vision monoculaire, deux mises à jour supplémentaires sont introduites dans la mécanisation du filtre :

- La mise à jour *Known Object Recognition based Pose Estimation* (KORPE) : est réalisée lorsqu'un objet de référence est détecté et qu'une pose absolue est estimée;
- La mise à jour *Assisted MONocular Visual Odometry* (AMOVO) : est réalisée lorsqu'une pose relative est estimée par odométrie visuelle.

Le vecteur d'état x comprend la position p_b^n et l'orientation R_b^n du *Body* dans le référentiel *Navigation* exprimée sous forme de quaternion q_b^n , et les sources d'erreur des capteurs (i.e. le biais du gyroscope sous forme quaternion b_{q_ω} et le biais de l'accéléromètre b_a).

$$x = [p_b^n \quad q_b^n \quad b_{q_\omega} \quad b_a] \quad (6.13)$$

Pour appliquer correctement la mise à jour visuelle dans le filtre d'hybridation, les informations basées sur la vision monoculaire (échantillonnées à $f_l = 10\text{Hz}$) sont fusionnées avec les mesures basées sur l'unité de mesures inertielles et magnétiques (échantillonnées à $f_h = 200\text{Hz}$) lorsque leur différence de temps Δ_{max} respecte la condition suivante :

$$0 < \Delta_{max} < \frac{1}{f_h} \quad (6.14)$$

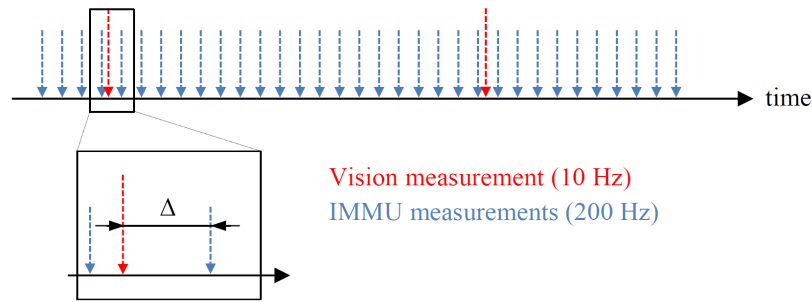


FIGURE 6.3 – Synchronisation entre les données vision et les données IMMU.

6.6 Expérimentation : corrections absolues d'un processus PDR

6.6.1 Description de l'expérimentation

Un parcours long de 2,9 km en milieux urbains a été effectuée par le piéton *MI*. Les informations sur le piéton *MI* et sur la durée de l'acquisition sont présentées dans le Tableau 6.1. Ce parcours est une boucle fermée ayant compris des passages en milieux couverts et en canyons urbains. Ces passages constituant des lieux où la réception des signaux satellitaires est fortement dégradé, et où le champ magnétique mesuré est perturbé par la présence d'éléments ferromagnétique. Durant toute l'acquisition, le piéton a tenu le dispositif expérimental, défini dans la section 2.4, en main dans un mode de portage *Texting*. Plus précisément, dans un mode de portage où la caméra est maintenue face au monde. Ce mode de portage est défini pour correspondre à l'utilisation d'un outil d'aide à la navigation piétonne avec un affichage en réalité augmentée sur la scène observée à travers l'écran de l'équipement.

En préliminaire de l'acquisition, l'étalonnage des magnétomètres doit être effectué dans un environnement avec un champ magnétique non perturbé afin de mesurer le champ magnétique terrestre [RENAUDIN et al., 2010; RENAUDIN et COMBETTES, 2014]. De plus, afin de suivre un nombre suffisant de satellites GNSS au début de la collecte de données, un emplacement avec un ciel ouvert est requis pour commencer l'acquisition. Comme présentée sur la Figure 6.4, un rond-point a ainsi été choisi comme emplacement de départ et d'arrivée de l'acquisition. Cet emplacement présente un ciel ouvert, nécessaire pour recevoir un nombre suffisant de signaux satellitaire, et est suffisamment éloigné des bâtiments et des éléments ferromagnétiques pour ne pas avoir de perturbations dans les mesures de champ magnétique.

Piéton	<i>MI</i>
Genre	Homme
h_{ped}	1.87 m
Durée de l'Acquisition	2109 s

TABLEAU 6.1 – Informations sur le piéton et sur l'acquisition.

6.6.2 Définition de la référence terrain

Pour évaluer l'estimation de la pose de l'équipement tenu en main par le piéton *MI* lors de ses déplacements en milieux urbains, il est nécessaire d'établir une vérité terrain. Comme il est difficile d'avoir une référence pour la pose d'un équipement tenu en main, une évaluation de l'estimation de la pose est proposée en termes de distance à des positions de référence sur le plan horizontal. Il est à noter que cette acquisition a été réalisée avant les autres acquisitions présentées dans cette thèse. La solution de référence avec la plateforme PERSY n'avait pas encore été intégrée au protocole expérimental.

Récepteur GPS Haute Sensibilité La plateforme ULISS, présentée dans la sous-section 2.4.3, intègre un récepteur HS-GPS qui fournit des positions GPS à 5 Hz [ORTIZ et al., 2017]. Comme décrit dans le Tableau 6.2, les positions de la plateforme tenue en main ne sont pas toujours disponibles, et lorsqu'elles le sont, la précision dépend du lieu où le piéton se trouve (i.e. canyon urbain, nombre de satellite disponible, etc.). Un écart type de l'erreur de positionnement GPS allant jusqu'à 30 mètres est observé pour les différentes acquisitions. Comme présentée sur la Figure 6.4, cette solution n'est donc pas suffisamment fiable et précise pour être utilisée comme référence terrain.

Récepteur GNSS et traitement en mode différentiel Pour acquérir une trajectoire de référence plus fiable et plus précise, une antenne hélicoïdale attachée au poignet, comme illustré sur la Figure 5.5, et un récepteur GNSS Septentrio situé dans un sac à dos ont été utilisés pour collecter les signaux des satellites GPS et GLONASS à 5 Hz lors des acquisitions. Avec ces dispositifs, les positions de l'antenne sont déterminées avec RTKLIB à l'aide d'une solution GNSS en différentiel. Comme décrit dans le Tableau 6.2, dans certains espaces urbains, de manière identique au récepteur HS-GPS intégré dans la plateforme ULISS, les positions de l'antenne attachée au poignet ne sont pas disponibles tout le temps et ne sont pas précises lorsque le piéton est dans un canyon urbain. Ces résultats représentent évidemment la difficulté de la localisation en milieu urbain à l'aide de tels capteurs, et est loin de la précision nécessaire pour définir une référence terrain, ainsi que pour le fonctionnement de systèmes de réalité augmentée sur les équipements tenus en main. Cependant, pendant le post-traitement avec RTKLIB pour le positionnement DGNSS, un fichier contenant l'écart type de positionnement est généré. Il est ainsi possible d'identifier les moments où la valeur de l'écart type est très faible et les moments où elle atteint plusieurs mètres. Il est observé dans les fichiers générés et sur la Figure 6.4, que l'écart type de l'erreur de positionnement estimée était inférieur à un mètre lorsque les piétons marchent dans un environnement ouvert et qu'il pouvait atteindre 21 mètres lorsque le piéton marchait dans un canyon urbain. En utilisant un récepteur GNSS et un traitement en différentiel, une trajectoire de référence n'est donc disponible qu'à ces moments précis de l'acquisition, ce qui n'est pas suffisant pour évaluer la précision de l'approche proposée.

Points de référence Comme défini dans la sous-section 4.7.2 et illustré sur la Figure 5.6, en plus de tout ces équipements, le piéton a marché lors des acquisitions sur des points de repère, marqués au sol d'un cercle blanc barré, et ont fait un arrêt de plusieurs secondes sur ces derniers. Ces points de référence ont été choisis à proximité des objets de référence, à savoir deux pour chaque station, afin que le piéton passe par des endroits où ces derniers seraient visibles sur les enregistrements vidéo, permettant ainsi de les reconnaître et d'estimer la pose. Ces points de référence permettent d'évaluer la précision de l'estimation de la pose aux instants où le piéton se situe sur ces derniers, en calculant la distance entre la position estimée et la position réelle. Les emplacements exacts de ces points de référence ont été déterminés à l'aide de la solution DGNSS présentée dans la sous-section 4.5.2 et la sous-section 4.7.2. La position de départ et d'arrivée de l'acquisition a également été déterminée à l'aide de la solution DGNSS afin de pouvoir calculer l'erreur de positionnement en fin d'acquisition.

Piéton	<i>MI</i>
Disponibilité du DGPS	63 %
Disponibilité du DGNSS	75 %
Objet de référence détecté	1316 images (6.2 %)

TABLEAU 6.2 – Résultats d'acquisition.

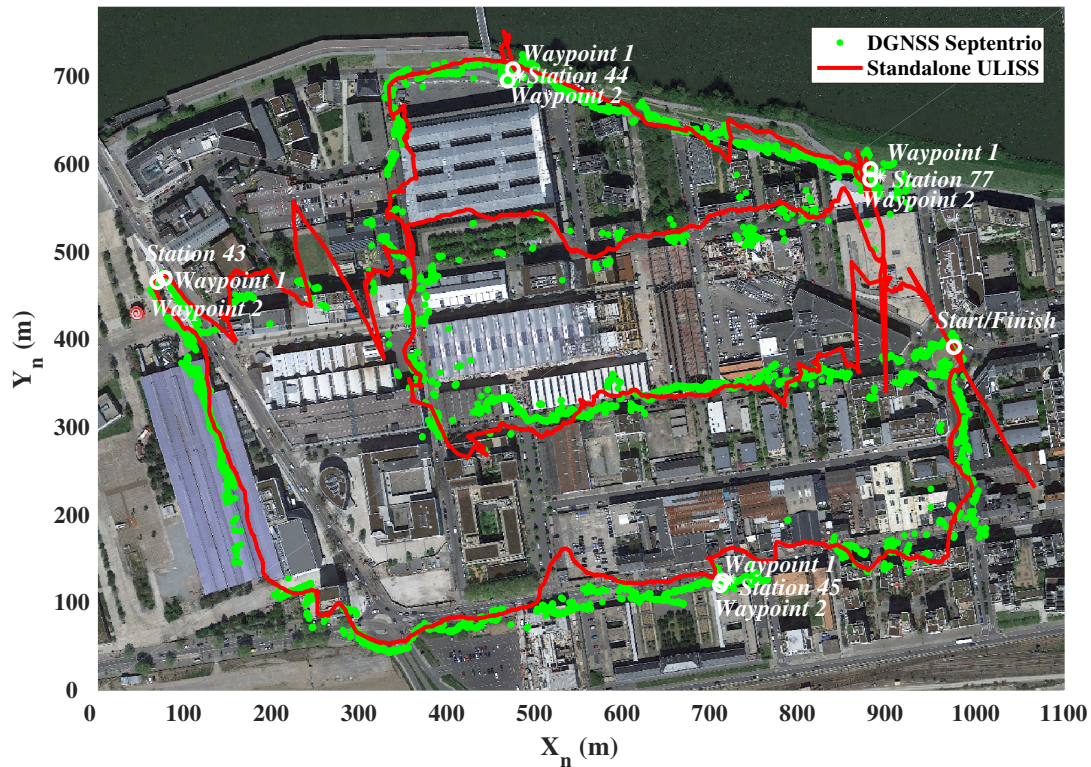


FIGURE 6.4 – Trajectoire de référence GPS standalone avec ULISS (en rouge) et trajectoire de référence DGNSS avec Septentrio (en vert) sur le parcours de 2.9 km en milieu urbain.

6.6.3 Résultats et analyses

Avec la détection de pas, la fréquence et les longueurs de pas définies dans la sous-section 6.3.2 et les caps extraits de l'estimation des angles d'attitude, une trajectoire relative est estimée. Tel que défini dans la sous-section 6.3.2, lorsque l'estimation de la pose, basée sur la reconnaissance d'objet de référence, est définie comme suffisamment précise, alors cette dernière est utilisée pour corriger la position estimée par le processus de navigation à l'estime. La Figure 6.5 permet de visualiser pour la marche en milieux urbains :

- La trajectoire de référence DGNSS (en vert) ;
- La trajectoire estimée par le processus de navigation à l'estime sans correction (en rouge) ;
- La trajectoire estimée par le processus de navigation à l'estime avec corrections par les estimations de poses absolues (en orange).

Le Tableau 6.3 présente la précision de localisation par rapport à la trajectoire de référence DGNSS pour la navigation à l'estime sans correction et pour la navigation à l'estime avec corrections basées vision.

Des estimations absolues de la pose sont réalisées à 0,9 km, 1,2 km, 1,9 km et 2,5 km de marche. Des corrections de l'estimation de la position du piéton sont alors opérées à ces instants. À la fin de la marche de 2,9 km, une erreur de positionnement d'environ 160 m est obtenue pour la localisation basée sur la navigation à l'estime du piéton. Cette erreur de positionnement est réduite à moins de 50 m pour la navigation à l'estime du piéton mise à jour par la reconnaissance d'objets de référence.

Bien que l'erreur de positionnement soit relative à la distance parcourue depuis le dernier positionnement absolu, une correction ponctuelle permet de limiter l'erreur de positionnement sur une longue marche piétonne en milieux urbains (cf. Figure 6.6).

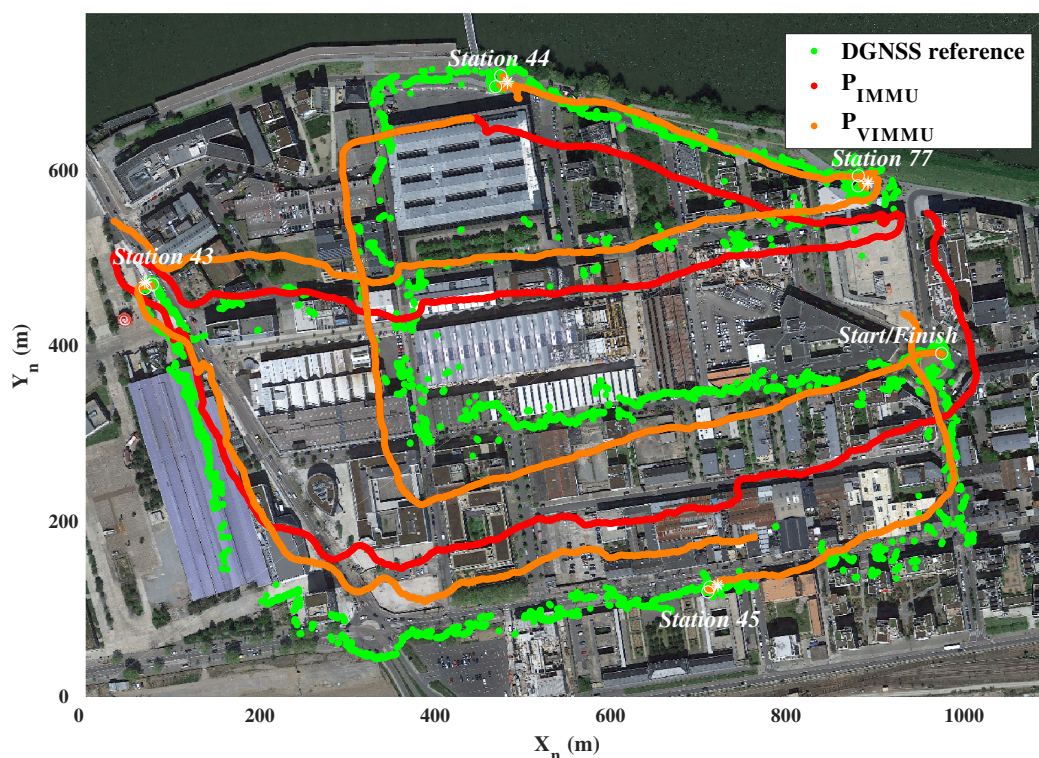


FIGURE 6.5 – Points de référence, trajectoire de référence DGNSS (en vert) et trajectoires obtenues pour la navigation à l'estime du piéton sans correction (en rouge) et avec corrections (en orange) sur le parcours de 2.9 km en milieux urbains.

Distance parcourue	0.5 km	0.9 km	1.0 km	1.2 km	1.9 km	2.0 km	2.5 km	2.9 km
Erreur PDR	40 m	60 m	40 m	45 m	60 m	30 m	100 m	160 m
Erreur PDR + Vision	40 m	< 0.4 m	5 m	< 0.4 m	< 0.4 m	10 m	< 0.4 m	30 m
Station détectée	X	44	X	77	43	X	45	X

TABLEAU 6.3 – Erreurs de positionnement.

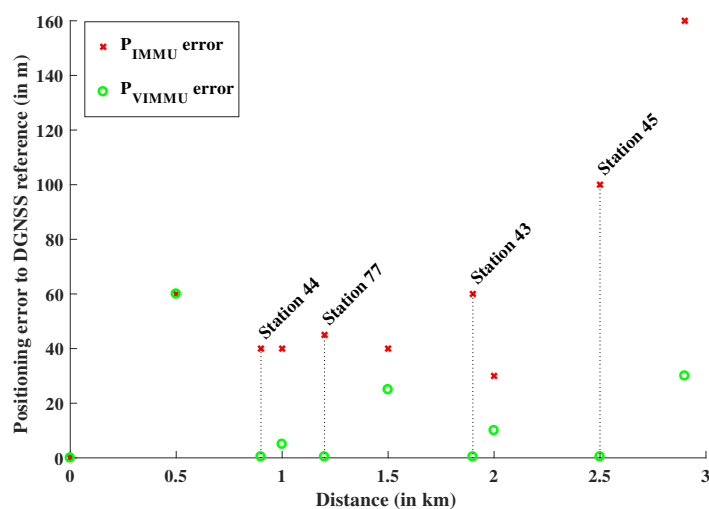


FIGURE 6.6 – Erreurs de positionnement du processus de navigation à l'estime du piéton et corrections appliquées.

6.7 Expérimentation : estimation continue de la pose

6.7.1 Description des acquisitions

Afin d'évaluer la solution d'estimation de la pose en dynamique proposée dans ce chapitre, un parcours long de 0.5 km a été effectué en milieu urbain par quatre piétons. Ces derniers ont tenus en main l'équipement expérimental présenté dans la section 2.4. Ce parcours comprend deux parties :

- Une première partie en environnement extérieur, avec des passages en canyons urbains et des passages avec des réflexions spéculaires;
- Une deuxième partie en environnement intérieur, avec le passage au rez de chaussée de l'École Nationale Supérieure d'Architecture de Nantes.

Une visualisation de la trajectoire du parcours effectué lors des acquisitions est présentée sur la Figure 6.7. Le détail des différents piétons est présenté dans le Tableau 6.4. Durant toute l'acquisition, chaque piéton a tenu le matériel expérimental en main en mode *Texting*. En précision de la définition donnée dans la sous-section 5.3.1, le mode *Texting* correspond, pour les acquisitions présentées dans cette thèse, à un mode de portage permettant l'observation de la scène face à eux, i.e. avec la caméra faisant face au monde, sur lequel serait proposé un affichage en réalité augmentée.

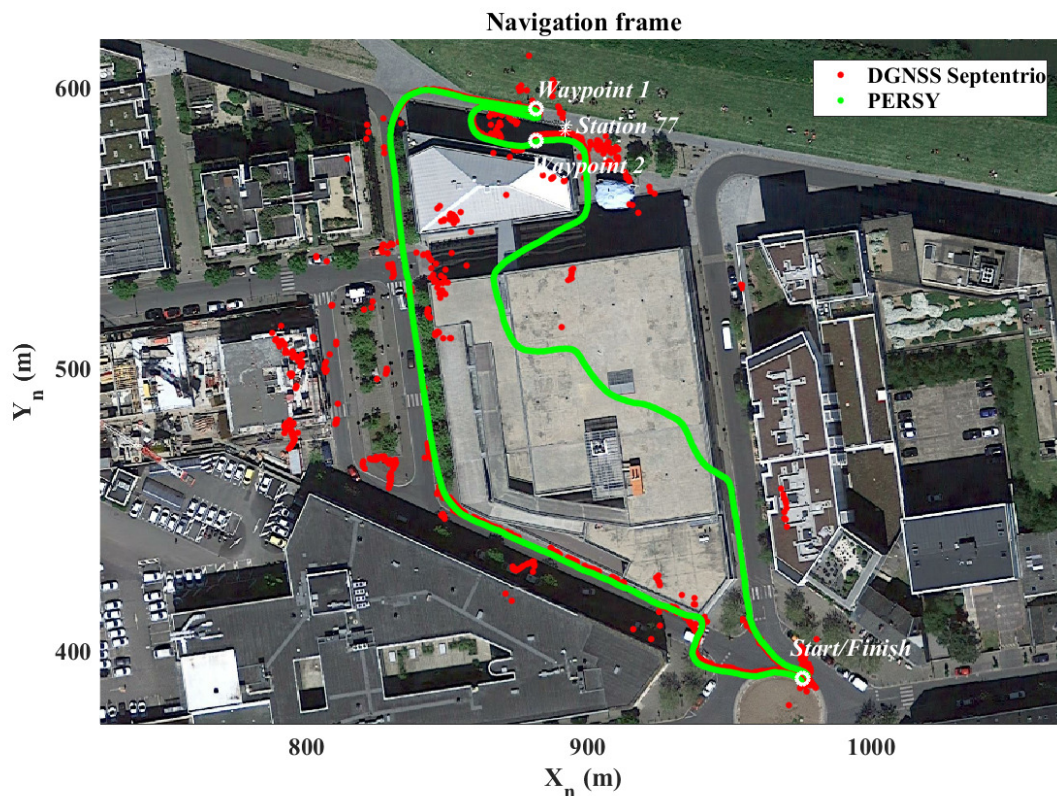


FIGURE 6.7 – Trajectoire de référence pour le parcours de 0.5 km lors de l'acquisition du piéton M2 et points de passages de référence à proximité d'un objet de référence sur le parcours (stations vélo 44).

6.7.2 Définition de la référence terrain

Pour évaluer l'estimation de la pose de l'équipement tenu en main par le piéton lors de ses déplacements en milieu extérieur et en milieu intérieur, il est nécessaire d'établir une vérité terrain. Pour cela le protocole et le matériel expérimental utilisé est identique à celui présenté dans la sous-section 5.6.2.

Piéton	W1	W2	M1	M2
Genre	Femme	Femme	Homme	Homme
h_{ped}	1.69 m	1.69 m	1.87 m	1.80 m
Durée de l'Acquisition	422 s	406 s	339 s	425 s
Disponibilité du DGPS	61.5 %	61.6 %	61.3 %	53.4 %
Disponibilité du DGNSS	63.0 %	68.8 %	62.8 %	71.8 %

TABLEAU 6.4 – Informations sur les différents piétons et leur acquisition.

6.7.3 Résultats et analyses

Pour évaluer le processus de fusion des données capteurs, les performances sont comparées à la trajectoire de référence PERSY et présentées en termes d'erreur de positionnement au *Waypoint 1* (i.e. juste avant la première détection de l'objet de référence qui se situe à peu près à la moitié du parcours) et au *Finish* (i.e. à la fin du parcours). Il est également présenté l'erreur moyenne de positionnement et l'erreur moyenne de positionnement par rapport à la distance parcourue, comme indiqué dans le tableau 6.5.

Piéton	W1	W2	M1	M2
Objet connu détecté	602 img (28.5%)	586 img (28.9%)	464 img (27.3%)	543 img (25.5%)
Erreur				
<i>Waypoint 1</i>	11.6 m	3.5 m	18.5 m	6.2 m
<i>Finish</i>	5.1 m	3.5 m	17.4 m	10.3 m
Erreur moyenne	9.3 m	10.1 m	18.1 m	8.6 m
% Erreur / Distance				
Partie 1 (extérieur)	11.2 %	11.3 %	18.1 %	12.1 %
Partie 2 (intérieur)	2.0 %	1.6 %	3.9 %	2.2 %
Parcours complet	6.1 %	6.6 %	11.0 %	7.2 %

TABLEAU 6.5 – Erreurs de positionnement.

La Figure 6.8 présente la trajectoire de référence DGNSS (en rouge), la trajectoire de référence PERSY (en vert) et la trajectoire estimée avec la solution proposée dans ce chapitre (en bleu) obtenue pour l'acquisition du piéton M2. A partir de ces résultats, on observe que la trajectoire estimée est proche de la trajectoire de référence PERSY. Cela signifie que le processus de fusion de capteurs proposé dans ce chapitre permet d'estimer la localisation d'un dispositif tenu en main, en espaces urbains et intérieurs, avec précision meilleure que les solutions GNSS et de navigation à l'estime du piéton avec les capteurs tenus en main.

L'estimation de la pose n'est cependant pas parfaite et plusieurs situations problématiques ont été identifiées. Il semble que la détermination des coordonnées des points du monde en 3D dans le processus d'odométrie visuelle monoculaire à l'échelle souffre d'imprécisions dans les canyons

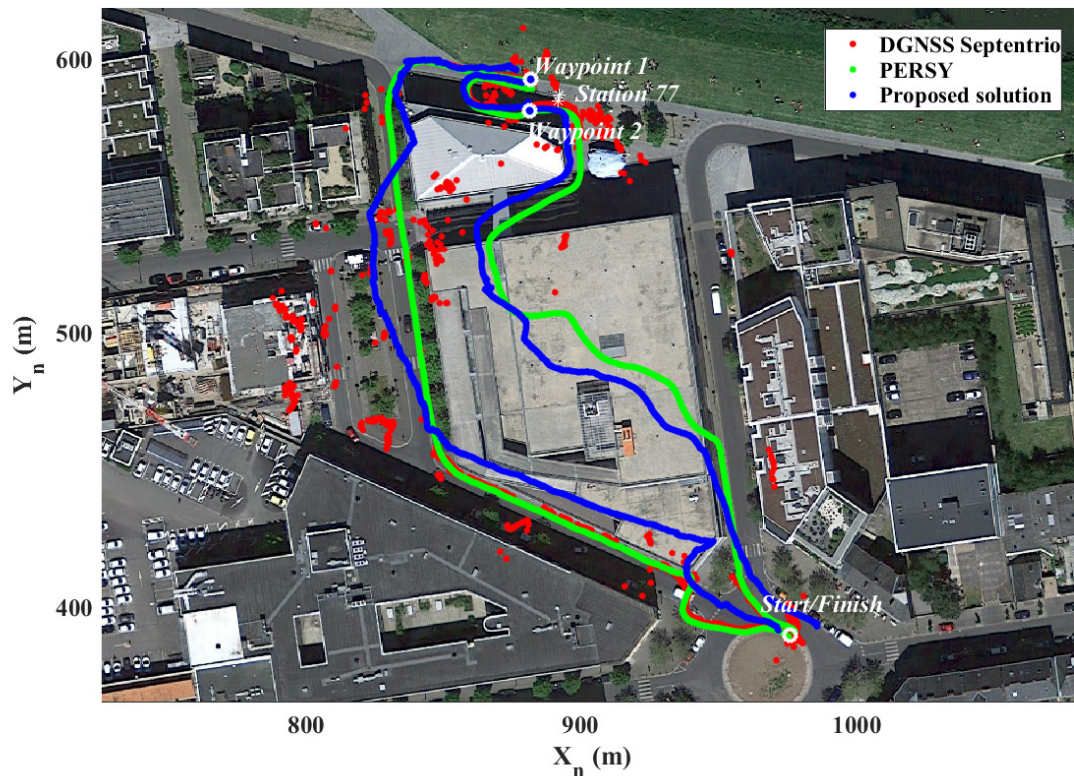


FIGURE 6.8 – Points de référence (en blanc), trajectoire de référence DGNS (en rouge), trajectoire de référence PERSY (en vert) et trajectoire estimée avec le processus de fusion des capteurs (en bleu) pour l’acquisition du piéton M2.

urbains et les zones ouvertes. L’extraction des points caractéristiques, correspondant à des éléments qui se trouvent à plusieurs centaines de mètres de la caméra, peut ne pas être suffisamment précise pour permettre un suivi relatif précis de la pose. Comme indiqué dans le tableau 6.5, les problèmes d’estimation sont mis en évidence en comparant l’erreur de positionnement en fonction de la distance parcourue pour la première section dans laquelle le piéton est en extérieur, et pour la deuxième section dans laquelle le piéton est principalement en intérieur.

Les passages urbains et intérieurs avec des réflexions spéculaires (e.g. fenêtres et portes en verre) introduisent également des erreurs dans la détermination des coordonnées 3D des points d’intérêt. Ces problèmes introduisent un biais dans le suivi de la pose (principalement dans l’estimation de l’orientation), ce qui entraîne une erreur de positionnement croissante pendant la marche, qui sera corrigée lors de la prochaine estimation absolue de la pose. Ces problèmes pourraient cependant être résolus à l’aide d’un processus d’odométrie visuelle monoculaire plus sophistiqué.

6.8 Conclusion

Dans le contexte de navigation piétonne en milieux urbains, l’approche proposée dans ce chapitre permet d’estimer en continu la pose d’un équipement tenu en main. Cette estimation est réalisée grâce à la fusion des données complémentaires issues des solutions présentées dans les chapitres précédents. Ainsi, l’estimation absolue de la pose basée sur la reconnaissance d’objets de référence, l’estimation relative de la pose basée sur l’odométrie visuelle monoculaire à l’échelle et l’estimation des angles d’attitude basée sur les données de l’unité de mesures inertielles et magnétiques ont été intégrées dans un filtre d’hybridation. D’une part, cela a permis de ponctuellement corriger un processus de navigation à l’estime et de réduire les erreurs de localisation dues

à la dérive des capteurs. D'autre part, cela a permis d'estimer en continu une pose qualifiée de l'équipement tenu en main.

Des acquisitions en milieux urbains, avec des passages en extérieur et en intérieur, ont mis en évidence les avantages et les difficultés de la solution d'hybridation proposée. En comparaison des solutions de l'état de l'art, la localisation du piéton avec été déterminée avec une meilleure précision et la reprojection d'informations en réalité augmentée sur l'écran de l'équipement mobile a pu être validée grâce à l'estimation qualifiée de la pose.

L'objectif de proposer une solution autonome pour estimer en continu la localisation et la pose d'un équipement tenu en main est atteint. Cette solution constitue en cela une innovation importante pour le développement d'outils d'aide à la navigation piétonne en milieux urbains.

Chapitre 7

Conclusion générale et perspectives

« Scio me nihil scire. »

Socrate

Sommaire

7.1 Conclusion et résumé des contributions	84
7.2 Perspectives	85

7.1 Conclusion et résumé des contributions

La localisation du piéton et l'estimation précise de la pose d'un équipement tenu en main sont des problématiques majeures dans les domaines de la navigation piétonne et de la réalité augmentée en milieux urbains. En effet, les environnements urbains constituent des zones présentant un grand nombre de perturbations pour les solutions classiquement utilisées (i.e. non ligne de vue des GNSS, perturbations magnétiques, occlusions, etc.). L'utilisation de capteurs bas coût, intégrés dans les équipements mobiles du piéton, ainsi que la volonté de développer des solutions autonomes, limitent d'autant plus les approches pour améliorer la localisation du piéton et proposer un affichage en réalité augmentée en milieux urbains.

Les travaux réalisés durant cette thèse ont ainsi eu pour objectifs de proposer des méthodes permettant d'améliorer la localisation en milieux urbains et l'estimation de la pose d'un équipement grand public tenu en main. Les principales contributions ont été d'intégrer les données 3D contenues dans les SIG, de prendre en compte les mouvements du piéton pour résoudre l'ambiguïté d'échelle de l'odométrie visuelle monoculaire et de fusionner les données complémentaires issues des capteurs visio-inertiels intégrés dans les équipements mobiles du piéton. Contrairement aux méthodes existantes de localisation et d'estimation de la pose, une attention particulière a été portée sur la précision des informations utilisées. En effet, de nombreux travaux, principalement dans le domaine de la réalité augmentée, utilisent des données erronées et/ou imprécises (e.g. positionnement GNSS, positionnement des modèles 3D) conduisant à des erreurs d'estimation et de reprojection. Ces erreurs sont corrigées par un alignement avec la scène réelle observée pour avoir une reprojection correcte, cependant les erreurs en amont ne sont pas corrigées. Ainsi, proposer une méthode d'estimation qualifiée de la pose, avec des corrections absolues grâce à des éléments de référence précis, permet de caractériser les données 3D contenues dans les SIG, de quantifier les erreurs et ainsi de pouvoir faire remonter les corrections nécessaires.

Pour estimer la pose, un premier travail d'identification d'éléments du mobilier des transports urbains a été effectué. Un second travail a été de mettre en place des mesures précises et des relevés GNSS en différentiel pour établir le modèle 3D d'un élément de référence, et de déterminer ses différentes positions à travers la ville avec une précision centimétrique. Les bases de données mises à disposition par la métropole de Nantes sont venues compléter cet ensemble de données. La détection et la reconnaissance de ces éléments de référence ont permis d'estimer la pose de manière absolue avec une erreur décimétrique en position et une erreur très inférieure au degré en orientation.

Pour compléter ces estimations ponctuelles, une solution d'odométrie visuelle monoculaire a été utilisée pour estimer la pose en continu. Dans le but de résoudre l'ambiguïté d'échelle de l'odométrie visuelle monoculaire, l'estimation dynamique des mouvements du piéton a été mise à profit, i.e. détermination des instants de pas et des longueurs de pas du piéton.

Dans les solutions existantes de navigation à l'estime du piéton (PDR) et d'estimation des angles d'attitude basée sur la fusion des données inertielles et magnétiques, les données complémentaires issues de la vision ont permis de ponctuellement corriger les erreurs de positionnement dus à la dérive des capteurs, et d'améliorer la précision de l'estimation des angles d'attitude réalisée avec le filtre MAGYQ. Cette meilleure estimation des angles d'attitude, et donc des caps de marche, a par conséquent permis d'améliorer la précision des solutions de navigation à l'estime du piéton.

Afin de tester les méthodes proposées, plusieurs acquisitions ont été réalisées en milieux urbains, par différents piétons, sur des parcours longs allant de 500 mètres à 3 kilomètres. Durant ces acquisitions, différents scénarios permettant de tester les limites des méthodes ont été mis en place (e.g. milieu urbain dense, milieu intérieur, passages avec réflexions spéculaires, passages avec vue dégagée, parcours longs, etc.). Les solutions proposées dans ces travaux de thèse ont ainsi pu être comparées aux solutions proposées dans l'état de l'art (i.e. GNSS, PDR). Il apparaît que dans les espaces urbains denses, une meilleure précision de localisation est atteinte en comparaison aux solutions précédemment évoquées.

L'estimation continue et qualifiée de la pose d'un équipement tenu en main par le piéton lors

de ses déplacements en milieux urbains permet également de proposer un affichage en réalité augmentée sur site. Après reprojection de modèles 3D contenus dans un SIG pour caractérisation, il apparaît que l'estimation de la pose est suffisamment précise pour permettre un affichage lorsqu'un objet de référence est détecté à une distance inférieure à la vingtaine de mètres, et le reste pendant plusieurs dizaines de mètres après la perte de détection d'objet de référence. En effet, les estimations étant relatives, elles continuent de souffrir d'une dérive au cours de la marche. De plus, les modèles 3D étant reprojétés sur des éléments de la scène à plusieurs dizaines de mètres, la stabilité de l'affichage en réalité augmentée constitue un point important à maîtriser avant de pouvoir proposer des outils d'aide à la navigation piétonne ou de permettre une caractérisation de contenu des SIG 3D par la foule.

7.2 Perspectives

Dans le contexte de navigation en milieux urbains et d'aide à la mobilité, plusieurs pistes d'améliorations sont envisagées. Une première voie d'amélioration serait d'augmenter le nombre d'objets de référence et d'étendre la couverture géographique, notamment en ajoutant d'autres éléments du mobilier des transports urbains à la liste des objets de référence. En réduisant la distance entre les objets de référence, cela aurait pour conséquences d'augmenter les chances d'en détecter, d'estimer des poses absolues et de corriger plus fréquemment les estimations relatives. Cela permettrait également de proposer plus souvent un affichage en réalité augmentée fiable. La Figure 7.1-(a) présente une visualisation de l'implantation des objets de référence (stations vélo) dans la ville de Nantes. Il est facilement remarquable que l'implantation des stations vélos est plus dense dans le centre ville de Nantes et dans les quartiers avoisinant. Il est important de noter que ces quartiers correspondent justement aux endroits pour lesquels la localisation par satellites est la plus imprécise, voir que la réception des signaux GNSS est impossible. La Figure 7.1-(b) présente quant à elle une visualisation de ce que serait l'implantation si la liste des objets de référence était étendue aux autres éléments du mobilier des transports urbains (stations de bus et de tram) dans la ville de Nantes.

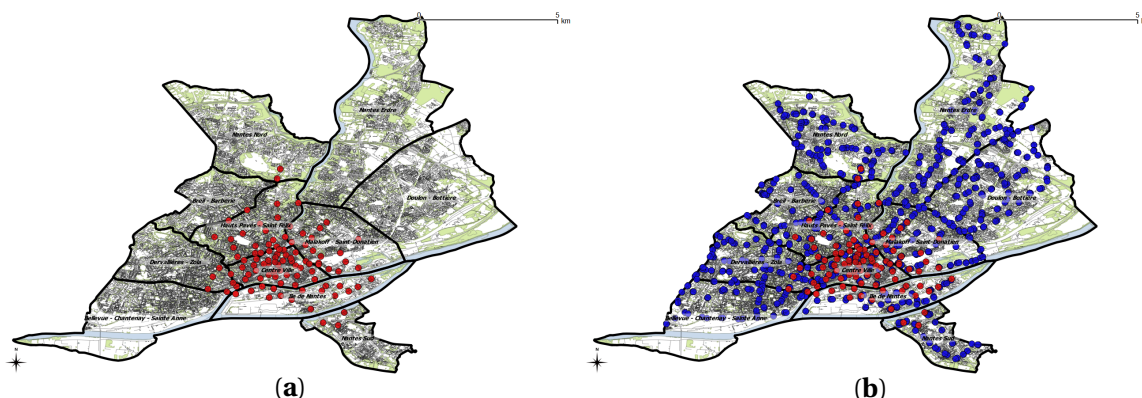


FIGURE 7.1 – (a) Implantation des stations vélo (en rouge) de la ville de Nantes; (b) Implantation des stations vélo (en rouge), des stations de bus et de tramway (en bleu) de la ville de Nantes.

Dans le cadre d'une extension à de nouveaux éléments du mobilier des transports urbains, i.e pouvant être plus complexes, une seconde voie d'amélioration serait d'estimer la pose en se basant sur l'ensemble du modèle, et plus seulement sur quelques points géométriques particuliers. La méthode du *Virtual Visual Servoing* [COMPORT et al., 2006; MARCHAND et al., 2016] constitue en cela une solution de choix pour effectuer l'estimation de la pose en utilisant directement le modèle 3D d'un objet.

Enfin, une troisième piste d'amélioration concerne la fusion des données visuelles, inertielles et magnétiques. Les solutions proposées dans cette thèse, bien que répondant aux objectifs fixés, peuvent être améliorées, simplifiées ou réorganisées. Ainsi, une nouvelle mécanisation du filtre

d'estimation de la pose, basée sur l'odométrie visuelle monoculaire mise à l'échelle par l'estimation des longueurs de pas du piéton, pourrait permettre d'avoir une meilleure estimation continue et qualifiée de la pose. Cette nouvelle mécanisation serait corrigée par les données inertielles et magnétiques (i.e. mesures des gyroscopes et phases statiques sur les mesures des accéléromètres et des magnétomètres), ainsi que par l'estimation de la pose basée sur la reconnaissance d'objets de référence. De manière similaire à ce qui a été proposé avec l'utilisation du filtre MAGYQ [RENAUDIN et COMBETTES, 2014], les mises à jour permettraient de corriger les erreurs et d'estimer les éléments de covariance, mais permettraient d'avoir une qualification sur la pose complète, et non plus sur les angles d'attitudes uniquement.

Bibliographie

- ALAHY, A., R. ORTIZ et P. VANDERGHEYNST. 2012, «Freak : Fast retina keypoint», dans *2012 IEEE Conference on Computer Vision and Pattern Recognition*, ISSN 1063-6919, p. 510–517, doi :10.1109/CVPR.2012.6247715. 21
- ANATI, R., D. SCARAMUZZA, K. G. DERPANIS et K. DANIILIDIS. 2012, «Robot localization using soft object detection», dans *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, ISSN 1050-4729, p. 4992–4999, doi :10.1109/ICRA.2012.6225216. 20, 21, 22
- ANTIGNY, N., M. SERVIERES et V. RENAUDIN. 2017, «Pedestrian track estimation with handheld monocular camera and inertial-magnetic sensor for urban augmented reality», dans *2017 International conference on Indoor Positioning and Indoor Navigation (IPIN)*. 22
- AQEL, M. O. A., M. H. MARHABAN, M. I. SARIPAN et N. ISMAIL. 2016, «Review of visual odometry : types, approaches, challenges, and applications», dans *SpringerPlus*. 23, 51
- ARTH, C., C. PIRCHHEIM, J. VENTURA, D. SCHMALSTIEG et V. LEPETIT. 2015, «Instant outdoor localization and SLAM initialization from 2.5D maps», *IEEE Transactions on Visualization and Computer Graphics*, vol. 21, n° 11, doi :10.1109/TVCG.2015.2459772, p. 1309–1318, ISSN 1077-2626. 26
- AZUMA, R., B. HOFF, H. NEELY et R. SARFATY. 1999, «A motion-stabilized outdoor augmented reality system», dans *Proceedings IEEE Virtual Reality (Cat. No. 99CB36316)*, ISSN 1087-8270, p. 252–259, doi :10.1109/VR.1999.756959. 23
- BAATZ, G., K. KÖSER, D. CHEN, R. GRZESZCZUK et M. POLLEFEYS. 2011, «Leveraging 3d city models for rotation invariant place-of-interest recognition», *International Journal of Computer Vision*, vol. 96, n° 3, p. 315–334, ISSN 1573-1405. 26
- BAY, H., T. T. A. ESS et L. V. GOOL. 2008, «Speeded-up robust features (SURF)», *Computer Vision and Image Understanding*, vol. 110, n° 3, doi :<http://dx.doi.org/10.1016/j.cviu.2007.09.014>, p. 346 – 359, ISSN 1077-3142. URL <http://www.sciencedirect.com/science/article/pii/S1077314207001555>, similarity Matching in Computer Vision and Multimedia. 21, 35, 51
- BILJECKI, F., H. LEDOUX et J. STOTER. 2016, «An improved lod specification for 3d building models», *Computers, Environment and Urban Systems*, vol. 59, doi :<https://doi.org/10.1016/j.compenvurbsys.2016.04.005>, p. 25 – 37, ISSN 0198-9715. URL <http://www.sciencedirect.com/science/article/pii/S0198971516300436>. xiii, 31
- BOAZ, B.-M., E. ELAZAR, L. HAREL et W. AYAL. 2011, «Improving accuracy of GNSS devices in urban canyons», dans *CCCG*. 11
- BOUGUET, J. Y. 2008, «Camera calibration toolbox for matlab», URL http://www.vision.caltech.edu/bouguetj/calib_doc/. 12
- BRADSKI, G. R. et A. KAEHLER. 2008, *Learning OpenCV, 1st Edition*, 1^{re} éd., O'Reilly Media, Inc., ISBN 9780596516130. 12

- CAPPELLE, C., M. E. B. E. NAJJAR, D. POMORSKI et F. CHARPILLET. 2007, «Localisation in urban environment using GPS and INS aided by monocular vision system and 3D geographical model», dans *2007 IEEE Intelligent Vehicles Symposium*, ISSN 1931-0587, p. 811–816, doi : 10.1109/IVS.2007.4290216. 21, 22, 26
- COMPORT, A. I., E. MARCHAND, M. PRESSIGOUT et F. CHAUMETTE. 2006, «Real-time markerless tracking for augmented reality : the virtual visual servoing framework», *IEEE Transactions on Visualization and Computer Graphics*, vol. 12, n° 4, doi :10.1109/TVCG.2006.78, p. 615–628, ISSN 1077-2626. 26, 85
- CORKE, P., J. LOBO et J. DIAS. 2007, «An introduction to inertial and visual sensing», . 24
- CUMMINS, M. et P. NEWMAN. 2011, «Appearance-only SLAM at large scale with FAB-MAP 2.0», *The International Journal of Robotics Research*, vol. 30, n° 9, p. 1100–1123. 20
- DALAL, N. et B. TRIGGS. 2005, «Histograms of oriented gradients for human detection», dans *Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 1 - Volume 01*, CVPR '05, IEEE Computer Society, Washington, DC, USA, ISBN 0-7695-2372-2, p. 886–893, doi :10.1109/CVPR.2005.177. URL <http://dx.doi.org/10.1109/CVPR.2005.177>. 34
- DAVID, P., D. DEMENTHON, R. DURAISWAMI et H. SAMET. 2003, «Simultaneous pose and correspondence determination using line features», dans *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, vol. 2, ISSN 1063-6919, p. II–424–II–431 vol.2, doi :10.1109/CVPR.2003.1211499. 25
- DAVISON, A. J., I. D. REID, N. D. MOLTON et O. STASSE. 2007, «MonoSLAM : Real-time single camera SLAM», *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, n° 6, p. 1052–1067. 24, 25
- DEMENTHON, D. F. et L. S. DAVIS. 1995, «Model-based object pose in 25 lines of code», *International journal of computer vision*, vol. 15, n° 1-2, p. 123–141. 25
- ENGEL, J., J. STURM et D. CREMERS. 2013, «Semi-dense visual odometry for a monocular camera», dans *Computer Vision (ICCV), 2013 IEEE International Conference on*, IEEE, p. 1449–1456. 67
- FEINER, S., B. MACINTYRE, T. HOLLERER et A. WEBSTER. 1997, «A touring machine : prototyping 3d mobile augmented reality systems for exploring the urban environment», dans *Digest of Papers. First International Symposium on Wearable Computers*, p. 74–81, doi :10.1109/ISWC.1997.629922. 23
- FISCHER, C., P. T. SUKUMAR et M. HAZAS. 2013, «Tutorial : Implementing a pedestrian tracker using inertial sensors», *IEEE pervasive computing*, vol. 12, n° 2, p. 17–27. 22, 70
- FRITZ, G., C. SEIFERT et L. PALETTA. 2006, «A mobile vision system for urban detection with informative local descriptors», dans *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, p. 30–30, doi :10.1109/ICVS.2006.5. 21
- GHINAMO, G., C. CORBI, P. LOVISOLO, A. LINGUA, I. AICARDI et N. GRASSO. 2015, *Accurate Positioning and Orientation Estimation in Urban Environment Based on 3D Models*, Springer International Publishing, Cham, ISBN 978-3-319-23222-5, p. 185–192, doi :10.1007/978-3-319-23222-5_23. URL http://dx.doi.org/10.1007/978-3-319-23222-5_23. 26
- GOODCHILD, M. F. 2007, «Citizens as sensors : The world of volunteered geography», *GeoJournal*, vol. 69, n° 4, doi :10.1007/s10708-007-9111-y, p. 211–221. 31

- HALLQUIST, A. et A. ZAKHOR. 2013, «Single view pose estimation of mobile devices in urban environments», dans *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, ISSN 1550-5790, p. 347–354, doi :10.1109/WACV.2013.6475039. 20, 21
- HARRIS, C. et M. STEPHENS. 1988, «A combined corner and edge detector», dans *In Proc. of Fourth Alvey Vision Conference*, p. 147–151. 21
- HARTLEY, R. et A. ZISSERMAN. 2003, *Multiple View Geometry in Computer Vision*, Cambridge University Press, ISBN 0521540518. 12, 13, 25, 37, 42
- HEISELE, B., I. RISKOV et C. MORGENSTERN. 2006, *Components for Object Detection and Identification*, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-540-68795-5, p. 225–237, doi :10.1007/11957959_12. URL https://doi.org/10.1007/11957959_12. 33
- HESCH, J. A., D. G. KOTTAS, S. L. BOWMAN et S. I. ROUMELIOTIS. 2014, «Consistency analysis and improvement of vision-aided inertial navigation», *IEEE Transactions on Robotics*, vol. 30, n° 1, doi :10.1109/TRO.2013.2277549, p. 158–176, ISSN 1552-3098. 22
- HETTIGE, B., H. HEWAMALAGE, C. RAJAPAKSHA, N. WAJIRASENA, A. PEMASIRI et I. PERERA. 2015, «Evaluation of feature-based object identification for augmented reality applications on mobile devices», dans *2015 IEEE 10th International Conference on Industrial and Information Systems (ICIIS)*, p. 170–175, doi :10.1109/ICIINFS.2015.7399005. 33
- HIDE, C., T. BOTTERILL et M. ANDREOTTI. 2010, «Low cost vision-aided imu for pedestrian navigation», dans *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, p. 1–7, doi :10.1109/UPINLBS.2010.5653658. 22
- HOLLERER, T., S. FEINER, T. TERAUCHI, G. RASHID et D. HALLAWAY. 1999, «Exploring mars : Developing indoor and outdoor user interfaces to a mobile augmented reality system», *Computers and Graphics*, vol. 23, p. 779–785. 23
- IM, T., D. COELHO, K. MUELLER et P. DE. 2018, «Smartphone based approximate localization using user highlighted texts from images», *Pervasive and Mobile Computing*, vol. 46, doi :<https://doi.org/10.1016/j.pmcj.2018.02.004>, p. 1 – 17, ISSN 1574-1192. URL <http://www.sciencedirect.com/science/article/pii/S1574119216304412>. 21
- JAIN, U., V. P. NAMBOODIRI et G. PANDEY. 2017, «Compact environment-invariant codes for robust visual place recognition», *CoRR*, vol. abs/1709.08103. URL <http://arxiv.org/abs/1709.08103>. 21
- JIANG, B., U. NEUMANN et S. YOU. 2004, «A robust hybrid tracking system for outdoor augmented reality», dans *IEEE Virtual Reality 2004*, ISSN 1087-8270, p. 3–275, doi :10.1109/VR.2004.1310049. 22, 26
- JIMENEZ, A. R., F. SECO, C. PRIETO et J. GUEVARA. 2009, «A comparison of pedestrian dead-reckoning algorithms using a low-cost MEMS IMU», dans *Intelligent Signal Processing, 2009. WISP 2009. IEEE International Symposium on*, IEEE, p. 37–42. 22
- KHAN, W. Z., Y. XIANG, M. Y. AALSALEM et Q. ARSHAD. 2013, «Mobile phone sensing systems : A survey», *IEEE Communications Surveys Tutorials*, vol. 15, n° 1, doi :10.1109/SURV.2012.031412.00077, p. 402–427. 2, 20
- KLEIN, G. et T. DRUMMOND. 2003, «Robust visual tracking for non-instrumental augmented reality», dans *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*, p. 113–122. 22

- KLEIN, G. et D. MURRAY. 2007, «Parallel tracking and mapping for small ar workspaces», dans *Mixed and Augmented Reality, 2007. ISMAR 2007. 6th IEEE and ACM International Symposium on*, IEEE, p. 225–234, doi :10.1109/ISMAR.2007.4538852. 24
- KNEIP, L., H. LI et Y. SEO. 2014a, *UPnP : An Optimal $O(n)$ Solution to the Absolute Pose Problem with Universal Applicability*, Springer International Publishing, Cham, ISBN 978-3-319-10590-1, p. 127–142, doi :10.1007/978-3-319-10590-1_9. URL http://dx.doi.org/10.1007/978-3-319-10590-1_9. 26
- KNEIP, L., H. LI et Y. SEO. 2014b, «Upnp : An optimal $o(n)$ solution to the absolute pose problem with universal applicability», dans *Computer Vision – ECCV 2014*, édité par D. Fleet, T. Pajdla, B. Schiele et T. Tuytelaars, Springer International Publishing, Cham, ISBN 978-3-319-10590-1, p. 127–142. 37
- KNOPP, J., J. SIVIC et T. PAJDLA. 2010, «Avoiding confusing features in place recognition», dans *Computer Vision – ECCV 2010*, édité par K. Daniilidis, P. Maragos et N. Paragios, Springer Berlin Heidelberg, Berlin, Heidelberg, ISBN 978-3-642-15549-9, p. 748–761. 21
- KNORR, S. B. et D. KURZ. 2016, «Leveraging the user's face for absolute scale estimation in hand-held monocular SLAM», dans *Mixed and Augmented Reality (ISMAR), 2016 IEEE International Symposium on*, IEEE, p. 11–17. 24
- KUMAR, R. S. M. 2014, «A survey on image feature descriptors», . 21
- KURZ, D., P. G. MEIER, A. PLOPSKI et G. KLINKER. 2014, «Absolute spatial context-aware visual feature descriptors for outdoor handheld camera localization overcoming visual repetitiveness in urban environments», dans *Computer Vision Theory and Applications (VISAPP), 2014 International Conference on*, vol. 2, p. 56–67. 26
- LATEGAHN, H., M. SCHREIBER, J. ZIEGLER et C. STILLER. 2013, «Urban localization with camera and inertial measurement unit», dans *Intelligent Vehicles Symposium (IV), 2013 IEEE*, IEEE, p. 719–724. 2, 20
- LEE, G. A., A. DUNSER, S. KIM et M. BILLINGHURST. 2012, «Cityviewar : A mobile outdoor ar application for city visualization», dans *2012 IEEE International Symposium on Mixed and Augmented Reality - Arts, Media, and Humanities (ISMAR-AMH)*, ISSN 2381-8360, p. 57–64, doi : 10.1109/ISMAR-AMH.2012.6483989. 23
- LEPPÄKOSKI, H., J. COLLIN et J. TAKALA. 2012, «Pedestrian navigation based on inertial sensors, indoor map, and wlan signals», dans *2012 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, ISSN 1520-6149, p. 1569–1572, doi :10.1109/ICASSP.2012.6288192. 24, 70
- LEUTENEGGER, S., M. CHLI et R. Y. SIEGWART. 2011, «Brisk : Binary robust invariant scalable keypoints», dans *2011 International Conference on Computer Vision*, ISSN 2380-7504, p. 2548–2555, doi :10.1109/ICCV.2011.6126542. 21
- LIANG, J. Z., N. CORSO, E. TURNER et A. ZAKHOR. 2015, *Multimodal Location Estimation of Videos and Images*, chap. Image-Based Positioning of Mobile Devices in Indoor Environments, Springer International Publishing, Cham, ISBN 978-3-319-09861-6, p. 85–99. 26
- LOWE, D. G. 1987, «Three-dimensional object recognition from single two-dimensional images», *Artif. Intell.*, vol. 31, n° 3, doi :10.1016/0004-3702(87)90070-1, p. 355–395, ISSN 0004-3702. URL [http://dx.doi.org/10.1016/0004-3702\(87\)90070-1](http://dx.doi.org/10.1016/0004-3702(87)90070-1). 26

- LOWE, D. G. 1999, «Object recognition from local scale-invariant features», dans *Proceedings of the International Conference on Computer Vision-Volume 2 - Volume 2*, ICCV '99, IEEE Computer Society, Washington, DC, USA, ISBN 0-7695-0164-8, p. 1150–. URL <http://dl.acm.org/citation.cfm?id=850924.851523>. 21
- LOWE, D. G. 2004, «Distinctive image features from scale-invariant keypoints», *Int. J. Comput. Vision*, vol. 60, n° 2, doi :10.1023/B:VISI.0000029664.99615.94, p. 91–110, ISSN 0920-5691. URL <https://doi.org/10.1023/B:VISI.0000029664.99615.94>. 21
- LYNEN, S., T. SATTLER, M. BOSSE, J. HESCH, M. POLLEFEYS et R. SIEGWART. 2015, «Get out of my lab : Large-scale, real-time visual-inertial localization», dans *Proceedings of Robotics : Science and Systems*, Rome, Italy, doi :10.15607/RSS.2015.XI.037. 20
- MAJDIK, A. L., D. VERDA, Y. ALBERS-SCHOENBERG et D. SCARAMUZZA. 2015, «Air-ground matching : Appearance-based GPS-denied urban localization of micro aerial vehicles», *J. Field Robot.*, vol. 32, n° 7, p. 1015–1039, ISSN 1556-4959. 21
- MARCHAND, E. et F. CHAUMETTE. 2002, «Virtual visual servoing : A framework for real-time augmented reality», dans *EUROGRAPHICS 2002 Conference Proceeding*, vol. 21(3), édité par G. Drettakis et H.-P. Seidel, Saarebrün, Germany, Germany, p. 289–298. URL <https://hal.inria.fr/inria-00352096>. 26
- MARCHAND, E., H. UCHIYAMA et F. SPINDLER. 2016, «Pose estimation for augmented reality : a hands-on survey», *IEEE Transactions on Visualization and Computer Graphics*, vol. PP, n° 99, doi :10.1109/TVCG.2015.2513408, p. 1–1, ISSN 1077-2626. 25, 37, 85
- MIKOLAJCZYK, K. et C. SCHMID. 2004, «Scale & affine invariant interest point detectors», *Int. J. Comput. Vision*, vol. 60, n° 1, doi :10.1023/B:VISI.0000027790.02288.f2, p. 63–86, ISSN 0920-5691. URL <https://doi.org/10.1023/B:VISI.0000027790.02288.f2>. 21
- MIKOLAJCZYK, K. et C. SCHMID. 2005, «A performance evaluation of local descriptors», *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, n° 10, doi :10.1109/TPAMI.2005.188, p. 1615–1630, ISSN 0162-8828. URL <https://doi.org/10.1109/TPAMI.2005.188>. 21
- MILFORD, M. J. et G. F. WYETH. 2012, «Seqslam : Visual route-based navigation for sunny summer days and stormy winter nights», dans *2012 IEEE International Conference on Robotics and Automation*, ISSN 1050-4729, p. 1643–1649, doi :10.1109/ICRA.2012.6224623. 21
- MIN, S., L. LEI, H. WEI et R. XIANG. 2012, «Interactive registration for augmented reality GIS», dans *2012 International Conference on Computer Vision in Remote Sensing*, p. 246–251, doi : 10.1109/CVRS.2012.6421269. 26
- MISHRA, A., K. ALAHARI et C. V. JAWAHAR. 2013, «Image retrieval using textual cues», dans *2013 IEEE International Conference on Computer Vision*, ISSN 1550-5499, p. 3040–3047, doi :10.1109/ICCV.2013.378. 21, 22
- MORENO-NOGUER, F., V. LEPETIT et P. FUA. 2007, «Accurate non-iterative O(n) solution to the PnP problem», dans *2007 IEEE 11th International Conference on Computer Vision*, ISSN 1550-5499, p. 1–8, doi :10.1109/ICCV.2007.4409116. 25, 37
- MORITA, H., M. HILD, J. MIURA et Y. SHIRAI. 2005, «View-based localization in outdoor environments based on support vector learning», dans *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, IEEE, p. 2965–2970. 20
- MOURIKIS, A. I. et S. I. ROUMELIOTIS. 2007, «A multi-state constraint kalman filter for vision-aided inertial navigation», dans *Proceedings 2007 IEEE International Conference on Robotics and Automation*, ISSN 1050-4729, p. 3565–3572, doi :10.1109/ROBOT.2007.364024. 22

- NASEER, T., M. RUHNKE, C. STACHNISS, L. SPINELLO et W. BURGARD. 2015, «Robust visual slam across seasons», dans *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, p. 2529–2535, doi :10.1109/IROS.2015.7353721. 21
- NISTÉR, D., O. NARODITSKY et J. BERGEN. 2004, «Visual odometry», dans *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1, Ieee, p. I–I. 23
- NÜTZI, G., S. WEISS, D. SCARAMUZZA et R. SIEGWART. 2011, «Fusion of IMU and vision for absolute scale estimation in monocular SLAM», *Journal of intelligent & robotic systems*, vol. 61, n° 1, p. 287–299. 24, 25
- ORTIZ, M., M. DE SOUSA et V. RENAUDIN. 2017, «A new PDR navigation device for challenging urban environments», *Journal of Sensors*, doi :10.1155/2017/4080479. URL <https://doi.org/10.1155/2017/4080479>. 7, 14, 56, 75
- REITMAYR, G. et T. W. DRUMMOND. 2006, «Going out : robust model-based tracking for outdoor augmented reality», dans *2006 IEEE/ACM International Symposium on Mixed and Augmented Reality*, p. 109–118, doi :10.1109/ISMAR.2006.297801. 26
- RENAUDIN, V., M. H. AFZAL et G. LACHAPPELLE. 2010, «Complete triaxis magnetometer calibration in the magnetic domain», *J. Sensors*, vol. 2010, p. 967 245 :1–967 245 :10. 16, 74
- RENAUDIN, V. et C. COMBETTES. 2014, «Magnetic, acceleration fields and gyroscope quaternion (MAGYQ)-based attitude estimation with smartphone sensors for indoor pedestrian navigation», *Sensors*, vol. 14, n° 12, doi :10.3390/s141222864, p. 22 864–22 890, ISSN 1424-8220. 58, 70, 74, 86
- RENAUDIN, V., M. SUSI et G. LACHAPPELLE. 2012, «Step length estimation using handheld inertial sensors», *Sensors*, vol. 12, n° 7, doi :10.3390/s120708507, p. 8507–8525, ISSN 1424-8220. URL <http://www.mdpi.com/1424-8220/12/7/8507>. 22, 52, 53, 70, 71
- RIBO, M., P. LANG, H. GANSTER, M. BRANDNER, C. STOCK et A. PINZ. 2002, «Hybrid tracking for outdoor augmented reality applications», *IEEE Computer Graphics and Applications*, vol. 22, n° 6, doi :10.1109/MCG.2002.1046629, p. 54–63, ISSN 0272-1716. 26
- ROSTEN, E. et T. DRUMMOND. 2006, «Machine learning for high-speed corner detection», dans *Proceedings of the 9th European Conference on Computer Vision - Volume Part I, ECCV'06*, Springer-Verlag, Berlin, Heidelberg, ISBN 3-540-33832-2, 978-3-540-33832-1, p. 430–443, doi : 10.1007/11744023_34. URL http://dx.doi.org/10.1007/11744023_34. 21
- RUBLEE, E., V. RABAUD, K. KONOLIGE et G. BRADSKI. 2011, «Orb : An efficient alternative to sift or surf», dans *2011 International Conference on Computer Vision*, ISSN 1550-5499, p. 2564–2571, doi :10.1109/ICCV.2011.6126544. 21
- RUOTSALAINEN, L., H. KUUSNIEMI et R. CHEN. 2010, «Overview of methods for visual-aided pedestrian navigation», dans *2010 Ubiquitous Positioning Indoor Navigation and Location Based Service*, p. 1–10, doi :10.1109/UPINLBS.2010.5653672. 20, 22
- SALARIAN, M. et R. ANSARI. 2015, «Efficient refinement of GPS-based localization in urban areas using visual information and sensor parameter», *CoRR*, vol. abs/1502.00319. 22
- SATOH, K., M. ANABUKI, H. YAMAMOTO et H. TAMURA. 2001, «A hybrid registration method for outdoor augmented reality», dans *Proceedings IEEE and ACM International Symposium on Augmented Reality*, p. 67–76, doi :10.1109/ISAR.2001.970516. 26
- SCARAMUZZA, D. et F. FRAUNDORFER. 2011, «Visual odometry [tutorial]», *IEEE robotics & automation magazine*, vol. 18, n° 4, p. 80–92. 23

- SCHAEFFER, C. 2013, «A comparison of keypoint descriptors in the context of pedestrian detection : Freak vs. surf vs. brisk», *Cité en*, vol. 12. 21
- SCHALL, G., D. WAGNER, G. REITMAYR, E. TAICHMANN, M. WIESER, D. SCHMALSTIEG et B. HOFMANN-WELLENHOF. 2009, «Global pose estimation using multi-sensor fusion for outdoor augmented reality», dans *2009 8th IEEE International Symposium on Mixed and Augmented Reality*, p. 153–162, doi :10.1109/ISMAR.2009.5336489. 26
- SCHOPS, T., J. ENGEL et D. CREMERS. 2014, «Semi-dense visual odometry for ar on a smartphone», dans *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, p. 145–150, doi :10.1109/ISMAR.2014.6948420. 25
- SCORNEC, J. L., M. ORTIZ et V. RENAUDIN. 2017, «Foot-mounted pedestrian navigation reference with tightly coupled gnss carrier phases, inertial and magnetic data», dans *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, p. 1–8, doi :10.1109/IPIN.2017.8115882. 57, 70
- SIVIC, J. et A. ZISSERMAN. 2003, «Video google : A text retrieval approach to object matching in videos», dans *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, IEEE, doi :10.1109/ICCV.2003.1238663. 20
- SMITH, R. 2007, «An overview of the tesseract OCR engine», dans *Ninth International Conference on Document Analysis and Recognition (ICDAR 2007)*, vol. 2, ISSN 1520-5363, p. 629–633, doi :10.1109/ICDAR.2007.4376991. 21, 34
- STRASDAT, H., J. MONTIEL et A. J. DAVISON. 2010, «Scale drift-aware large scale monocular SLAM», *Robotics : Science and Systems VI*, vol. 2. 24
- SUSI, M., V. RENAUDIN et G. LACHAPPELLE. 2013, «Motion mode recognition and step detection algorithms for mobile phone users», *Sensors*, vol. 13, n° 2, doi :10.3390/s130201539, p. 1539–1562, ISSN 1424-8220. URL <http://www.mdpi.com/1424-8220/13/2/1539>. 52
- SUTHERLAND, I. E. 1974, «Three-dimensional data input by tablet», *Proceedings of the IEEE*, vol. 62, n° 4, doi :10.1109/PROC.1974.9449, p. 453–461, ISSN 0018-9219. 25
- TAKACS, G., V. CHANDRASEKHAR, N. GELFAND, Y. XIONG, W.-C. CHEN, T. BISMPIGIANNIS, R. GRZESZCZUK, K. PULLI et B. GIROD. 2008, «Outdoors augmented reality on mobile phone using loxel-based visual feature organization», dans *Proceedings of the 1st ACM International Conference on Multimedia Information Retrieval, MIR '08*, ACM, New York, NY, USA, ISBN 978-1-60558-312-9, p. 427–434, doi:10.1145/1460096.1460165. URL <http://doi.acm.org/10.1145/1460096.1460165>. 21
- TAKETOMI, T., H. UCHIYAMA et S. IKEDA. 2017, «Visual SLAM algorithms : a survey from 2010 to 2016», *IPSJ Transactions on Computer Vision and Applications*, vol. 9, p. 1–11. 25, 51
- TOMAZIC, S. et I. SKRJANC. 2015, «Fusion of visual odometry and inertial navigation system on a smartphone», *Computers in Industry*, vol. 74, p. 119–134. 24, 62
- TORR, P. et A. ZISSERMAN. 2000, «Mlesac», *Computer Vision and Image Understanding*, vol. 78, n° 1, doi :10.1006/cviu.1999.0832, p. 138–156, ISSN 1077-3142. URL <http://dx.doi.org/10.1006/cviu.1999.0832>. 36, 51
- URBAN, S., J. LEITLOFF et S. HINZ. 2016, «Mlppn - A real-time maximum likelihood solution to the perspective-n-point problem», *CoRR*, vol. abs/1607.08112. URL <http://arxiv.org/abs/1607.08112>. 26, 37

- VALGREN, C. et A. LILIENTHAL. 2008, «Incremental spectral clustering and seasons : Appearance-based localization in outdoor environments», dans *Robotics and automation, 2008. ICRA 2008. IEEE international conference on*, IEEE, p. 1856–1861. 21
- VALGREN, C. et A. J. LILIENTHAL. 2010, «SIFT, SURF & seasons : Appearance-based long-term localization in outdoor environments», *Robotics and Autonomous Systems*, vol. 58, n° 2, p. 149–156. 21
- VENTURA, J., C. ARTH, G. REITMAYR et D. SCHMALSTIEG. 2014, «Global localization from monocular slam on a mobile phone», *IEEE transactions on visualization and computer graphics*, vol. 20, n° 4, p. 531–539. 20, 24
- VENTURA, J. et T. HOLLERER. 2012, «Wide-area scene mapping for mobile visual tracking», dans *2012 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, p. 3–12, doi : 10.1109/ISMAR.2012.6402531. 21
- WANG, R., M. SCHWÖRER et D. CREMERS. 2017, «Stereo DSO : Large-scale direct sparse visual odometry with stereo cameras», dans *International Conference on Computer Vision (ICCV), Venice, Italy*. 24
- WEISS, S. et R. SIEGWART. 2011, «Real-time metric state estimation for modular vision-inertial systems», dans *Robotics and Automation (ICRA), 2011 IEEE International Conference on*, IEEE, p. 4531–4537. 24
- WU, Y. 2016, «Image based camera localization : an overview», *CoRR*, vol. abs/1610.03660. URL <http://arxiv.org/abs/1610.03660>. 37
- YOU, S., U. NEUMANN et R. AZUMA. 1999, «Orientation tracking for outdoor augmented reality registration», *IEEE Computer Graphics and Applications*, vol. 19, n° 6, doi :10.1109/38.799738, p. 36–42, ISSN 0272-1716. 26
- YU, L., C. JOLY, G. BRESSON et F. MOUTARDE. 2016, «Improving robustness of monocular urban localization using augmented street view», dans *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*, IEEE, p. 513–519. 2, 21
- ZANDBERGEN, P. A. et S. J. BARBEAU. 2011, «Positional accuracy of assisted gps data from high-sensitivity gps-enabled mobile phones», *Journal of Navigation*, vol. 64, n° 3, doi :10.1017/S0373463311000051, p. 381–399. 2, 11, 20
- ZHANG, W. et J. KOSECKA. 2006, «Image based localization in urban environments», dans *3D Data Processing, Visualization, and Transmission, Third International Symposium on*, p. 33–40, doi : 10.1109/3DPVT.2006.80. 21
- ZHANG, Z. 2000, «A flexible new technique for camera calibration», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, n° 11, doi :10.1109/34.888718, p. 1330–1334, ISSN 0162-8828. 12
- ZHENG, L., Z. WU, W. ZHOU, S. WENG et H. ZHENG. 2016, «A smartphone based hand-held indoor positioning system», dans *Frontier Computing*, édité par J. C. Hung, N. Y. Yen et K.-C. Li, Springer Singapore, Singapore, ISBN 978-981-10-0539-8, p. 639–650. 54, 63, 70
- ZHENG, Y., Y. KUANG, S. SUGIMOTO, K. ÅSTRÖM et M. OKUTOMI. 2013, «Revisiting the pnp problem : A fast, general and optimal solution», dans *2013 IEEE International Conference on Computer Vision*, ISSN 1550-5499, p. 2344–2351, doi :10.1109/ICCV.2013.291. 26
- ZHOU, H., T. SATTLER et D. W. JACOBS. 2016, «Evaluating local features for day-night matching», dans *Computer Vision - ECCV 2016 Workshops - Amsterdam, The Netherlands, October 8-10 and 15-16, 2016, Proceedings, Part III*, p. 724–736, doi :10.1007/978-3-319-49409-8_60. URL https://doi.org/10.1007/978-3-319-49409-8_60. 21

Titre : Estimation continue de la pose d'un équipement tenu en main par fusion des données visio-inertielles pour les applications de navigation piétonne en milieu urbain

Mots clés : Navigation piétonne, fusion de données, Réalité augmentée

Résumé : Pour assister la navigation piétonne dans les espaces urbains et intérieurs, une estimation précise de la pose (i.e. la position 3D et l'orientation 3D) d'un équipement tenu en main constitue un point essentiel dans le développement d'outils d'aide à la mobilité (e.g. applications de réalité augmentée).

Dans l'hypothèse où le piéton n'est équipé que d'appareils grand public, l'estimation de la pose est limitée à l'utilisation de capteurs à faible coût intégrés dans ces derniers (i.e. un récepteur GNSS, une unité de mesure inertielle et magnétique et une caméra monoculaire).

De plus, les espaces urbains et intérieurs, comprenant des bâtiments proches et des éléments ferromagnétiques, constituent des zones difficiles pour la localisation et l'estimation de la pose lors de grands déplacements piétons.

Cependant, le développement récent et la mise à disposition d'informations contenues dans des Systèmes d'Information Géographiques 3D constituent une nouvelle source de données exploitable pour la localisation et l'estimation de la pose.

Pour relever ces défis, cette thèse propose différentes solutions pour améliorer la localisation et l'estimation de la pose des équipements tenus en main par le piéton lors de ses déplacements en espaces urbains et intérieurs.

Les solutions proposées intègrent l'estimation de l'attitude basée inertielle et magnétique, l'odométrie visuelle monoculaire mise à l'échelle grâce à l'estimation des déplacements du piéton, l'estimation absolue de la pose basée sur la reconnaissance d'objets SIG 3D parfaitement connus et la mise à jour en position de la navigation à l'estime du piéton.

Toutes ces solutions s'intègrent dans un processus de fusion permettant d'améliorer la précision de la localisation et d'estimer en continu une pose qualifiée de l'appareil tenu en main.

Cette qualification est nécessaire à la mise en place d'un affichage en réalité augmentée sur site.

Pour évaluer les solutions proposées, des données expérimentales ont été recueillies au cours de déplacements piétons dans un espace urbain avec des objets de référence et des passages intérieurs.

Titre : Continuous pose estimation of handheld device by fusion of visio-inertial data for pedestrian navigation applications in urban environments

Keywords : Pedestrian navigation, sensor fusion, augmented reality

Abstract : To support pedestrian navigation in urban and indoor spaces, an accurate pose estimate (i.e. 3D position and 3D orientation) of an equipment held in hand constitutes an essential point in the development of mobility assistance tools (e.g. Augmented Reality applications).

On the assumption that the pedestrian is only equipped with general public devices, the pose estimation is restricted to the use of low-cost sensors embedded in the latter (i.e. an Inertial and Magnetic Measurement Unit and a monocular camera).

In addition, urban and indoor spaces, comprising closely-spaced buildings and ferromagnetic elements, constitute challenging areas for localization and sensor pose estimation during large pedestrian displacements.

However, the recent development and provision of data contained in 3D Geographical Information System constitutes a new wealth of data usable for localization and pose estimation.

To address these challenges, this thesis proposes solutions to improve pedestrian localization and hand-held device pose estimation in urban and indoor spaces.

The proposed solutions integrate inertial and magnetic-based attitude estimation, monocular Visual Odometry with pedestrian motion estimation for scale estimation, 3D GIS known object recognition-based absolute pose estimation and Pedestrian Dead-Reckoning updates.

All these solutions are fused to improve accuracy and to continuously estimate a qualified pose of the handheld device. This qualification is required to validate an on-site augmented reality display.

To assess the proposed solutions, experimental data has been collected during pedestrian walks in an urban space with sparse known objects and indoors passages.