Vision robotique directe

Guillaume Caron

UNIVERSITÉ DE PICARDIE JULES VERNE

ÉCOLE DOCTORALE

SCIENCES, TECHNOLOGIES ET SANTÉ

Synthèse des travaux de recherche

pour obtenir le diplôme d'

HABILITATION A DIRIGER LES RECHERCHES

de l'Université de Picardie Jules Verne

Mention : Sciences pour l'ingénieur

Présentée par

Guillaume Caron

Vision robotique directe

parrainée par El Mustapha Mouaddib

préparée à l'UPJV au sein du laboratoire

Modélisation, Information et Systèmes

soutenue le 10 décembre 2019

Jury :

Rapporteurs :		
Marie-Odile Berger	-	DR, Inria Nancy Grand Est
Philippe Martinet	-	DR, Inria Sophia Antipolis
Olivier Balédent	-	MCU-HDR, Université de Picardie Jules Verne
Président :		
Christian Duriez	-	DR, Inria Lille-Nord-Europe
Examinateurs :		
Abderrahmane Kheddar	-	CNRS / Université de Montpellier
François Chaumette	-	DR, Inria Rennes-Bretagne Atlantique / IRISA
Claude Pégard	-	PR, Université de Picardie Jules Verne
Parrain :		
El Mustapha Mouaddib	-	PR, Université de Picardie Jules Verne

Table des matières

Préambule **1**

1 Synthèse d'activité **3**
1.1 Rapport d'activité . . . 4
1.1.1 Curriculum vitae . . . 4
1.1.2 Expériences d'enseignement . . . 6
1.1.3 Rayonnement . . . 8
1.1.4 Sollicitations d'expertise . . . 11
1.1.5 Contrats . . . 13
1.1.6 Résumé des activités de recherche . . . 15
1.1.7 Encadrement . . . 23
1.2 Production scientifique . . . 27
1.2.1 Articles de revues . . . 27
1.2.2 Articles dans les actes de conférences . . . 28
1.2.3 Autres publications . . . 31
1.2.4 Exposés invités . . . 31
1.2.5 Communications . . . 31
1.2.6 Logiciels et jeux de données . . . 33

2 Modélisation de la formation des images **35**
2.1 Formation géométrique des images . . . 35
2.1.1 Modèle de projection perspective . . . 35
2.1.2 Modèle de projection centrale unifié . . . 38
2.1.3 Modèle de projection centrale unifié à deux plans images . . . 40
2.2 Représentations des transformations géométriques . . . 42
2.3 Formation photométrique des images . . . 42

3 Etat de l'art en vision robotique directe **45**
3.1 Introduction . . . 46
3.2 Fondamentaux . . . 47
3.2.1 Fondements . . . 47
3.2.2 Méthodes d'optimisation usuelles . . . 48
3.2.3 Mise à jour des degrés de liberté . . . 51
3.2.4 Méthode d'optimisation efficace au second ordre . . . 55
3.2.5 Conclusion partielle . . . 56
3.3 Approches directes pures . . . 57
3.3.1 Quand des connaissances supplémentaires sont disponibles . . . 57
3.3.2 En faisant des hypothèses sur la scène ou le mouvement . . . 58
3.3.3 Localisation et cartographie simultanées . . . 70
3.3.4 Approches directes robustes . . . 71

3.4 Autres approches directes . . . 80
3.4.1 Critère de corrélation croisée . . . 80
3.4.2 Réseaux de neurones convolutionnels . . . 82
3.5 Approches directes étendues . . . 85
3.5.1 Distribution d'intensité . . . 85
3.5.2 Champs de descripteurs . . . 90
3.5.3 Approches basées noyaux . . . 92
3.5.4 Espace d'échelle . . . 95
3.6 Synthèse . . . 96

4 Suivi et asservissement visuels basés entropie photométrique 99
4.1 L'information mutuelle en suivi basé maquette virtuelle 3D . . . 100
4.1.1 Introduction . . . 100
4.1.2 Calcul de pose basé information mutuelle . . . 102
4.1.3 Matrices jacobienne et hessienne . . . 103
4.1.4 Résultats . . . 104
4.1.5 Conclusion partielle . . . 109
4.2 Exploration : Commande de caméra virtuelle basée entropie . . . 111
4.2.1 Introduction . . . 111
4.2.2 Exploration basée entropie photométrique . . . 112
4.2.3 Résultats . . . 113
4.2.4 Conclusion partielle . . . 116
4.3 Conclusion du chapitre . . . 117

5 Suivi visuel direct pur basé maquette virtuelle 3D 119
5.1 Calcul de pose direct pur basé maquette virtuelle 3D . . . 120
5.2 Recalage d'image perspective sur maquette 3D . . . 121
5.2.1 Introduction . . . 121
5.2.2 Etude de la fonction de coût . . . 125
5.2.3 Résultats . . . 126
5.2.4 Conclusion partielle . . . 129
5.3 Suivi visuel panoramique basé maquette virtuelle 3D . . . 129
5.3.1 Introduction . . . 129
5.3.2 Suivi visuel direct pur basé maquette virtuelle 3D . . . 131
5.3.3 Résultats . . . 131
5.3.4 Conclusion partielle . . . 133
5.4 Conclusion du chapitre . . . 134

6 Vision robotique directe basée mélange de potentiels 135
6.1 Le Mélange de Potentiels Photométriques . . . 136
6.1.1 Problématique . . . 136
6.1.2 Modélisation du mélange de potentiels photométriques . . . 137
6.2 Suivi visuel de plan adaptatif dans l'espace d'échelle . . . 138
6.2.1 Modélisation du problème . . . 138

6.2.2 Evaluation . . . 140
6.2.3 Conclusion partielle . . . 143
6.3 Asservissement visuel direct basé mélange de potentiels . . . 144
6.3.1 Définition du coût et de la loi de commande . . . 144
6.3.2 Matrice jacobienne relative à l'échelle . . . 145
6.3.3 Matrice jacobienne relative à la pose de la caméra . . . 145
6.3.4 Stratégie de mise à jour de l'étendue d'attraction . . . 147
6.3.5 Résultats . . . 148
6.3.6 Conclusion partielle . . . 152
6.4 Gyroscope visuel sphérique direct . . . 152
6.4.1 Vue d'ensemble . . . 152
6.4.2 Représentation d'image et mélange de potentiels sphériques . . . 153
6.4.3 Estimation directe de rotations . . . 155
6.4.4 Résultats . . . 157
6.4.5 Conclusion partielle . . . 161
6.5 Conclusion du chapitre . . . 162

7 Bilan et perspectives **165**
7.1 Bilan . . . 165
7.2 Projet de recherche . . . 167
7.2.1 Vue d'ensemble . . . 167
7.2.2 A court terme . . . 167
7.2.3 A moyen terme . . . 168
7.2.4 A long terme . . . 169
7.2.5 Positionnement du projet scientifique . . . 170

Bibliographie **173**

Préambule

La vision robotique directe désigne la vision artificielle comme organe de perception des robots en considérant le lien le plus étroit possible entre les pixels des images acquises par une caméra et l'état du robot qui l'embarque. Ce lien représente un fondement théorique, défini formellement, exploitable pour le suivi d'objet, la localisation de robot ou la commande automatique de ses déplacements.

Ces trois dernières problématiques générales sont très classiques en vision robotique. En effet, durant le dernier demi-siècle, elles ont été largement abordées, cependant massivement sous le paradigme des primitives visuelles géométriques. Concrètement, cela consiste à séparer le traitement d'images, qui détecte et met en correspondance ces primitives géométriques, de l'estimation ou de la commande. Cette division du problème apporte de nombreux avantages, notamment en terme de quantité de données à traiter dans une estimation de pose ou une loi de commande, en exploitant des lois purement géométriques. Cependant, les erreurs, aussi infimes soient-elles, faites à chaque étape s'accumulent.

Traiter le problème en un tout, c'est-à-dire estimer une pose ou commander un robot, directement à partir des intensités des pixels des images, apparaît donc intéressant pour atteindre une précision maximale. Le lien entre les intensités des pixels et les degrés de liberté de la caméra est, par contre, plus difficile à établir qu'avec des primitives géométriques, a fortiori pour de "grands" déplacements ou quand les conditions d'éclairage et/ou la scène changent, même partiellement. En effet, si l'estimation de transformations à deux degrés de liberté entre deux images directement à partir des intensités des pixels a été exprimée dès le début des années 1980, ce n'est que dans les vingt dernières années que cette approche a été étendue à plus de degrés de liberté puis à la commande automatique. Enfin, ce n'est qu'encore plus récemment que les recherches en vision directe contribuent à améliorer la robustesse et à étendre les domaines de convergences des estimateurs et des lois de commande. C'est dans ce cadre que s'inscrivent les travaux de recherche présentés dans ce mémoire.

Le mémoire se structure en sept chapitres.

Tout d'abord, le premier chapitre rapporte une synthèse des activités que j'ai menées pour ces recherches et l'enseignement, ainsi que du rayonnement qui en découle.

Ensuite, un premier groupe de deux chapitres rappelle, d'une part, des éléments de modélisation de la formation géométrique et photométrique des images (Chapitre 2) et rassemble, d'autre part, un état de l'art en vision robotique directe, ciblé, détaillé et reformulé sous un cadre théorique uniformisé (Chapitre 3).

En s'appuyant sur ce socle, un deuxième groupe de trois chapitres structure les contributions scientifiques, principalement réalisées dans le contexte des thèses de doctorat à l'encadrement desquelles j'ai participé. Cette structuration se fait selon la nature de l'information visuelle directe considérée, qu'elle s'appuie sur la théorie

de l'information (Chapitre 4), les intensités des images directement (Chapitre 5) ou, enfin, sur l'espace d'échelle (Chapitre 6). Ce dernier chapitre rassemble les contributions les plus fortes de mes travaux et ouvrant le plus de perspectives.

Enfin, le Chapitre 7 résume un bilan des contributions, sur les perspectives desquelles un projet scientifique est proposé pour les années à venir, selon plusieurs axes.

Chapitre 1

Synthèse d'activité

Sommaire

1.1 Rapport d'activité 4
1.1.1 Curriculum vitae 4
1.1.2 Expériences d'enseignement 6
1.1.3 Rayonnement 8
1.1.4 Sollicitations d'expertise 11
1.1.5 Contrats 13
1.1.6 Résumé des activités de recherche 15
1.1.6.1 Introduction 15
1.1.6.2 Activités de recherche relatives à la thèse de doctorat . 15
1.1.6.3 Activités de recherche postérieures à la thèse 17
1.1.6.4 Synthèse des activités de recherche postérieures à la thèse 22
1.1.7 Encadrement 23
1.2 Production scientifique 27
1.2.1 Articles de revues 27
1.2.2 Articles dans les actes de conférences 28
1.2.3 Autres publications 31
1.2.4 Exposés invités 31
1.2.5 Communications 31
1.2.6 Logiciels et jeux de données 33

1.1 Rapport d'activité

1.1.1 Curriculum vitae

Informations personnelles

Etat civil	34 ans, marié, un enfant
Adresse	Université de Picardie Jules Verne, 33 rue Saint-Leu, 80039 Amiens
Contact	+33 3 22 82 59 01 | guillaume.caron@u-picardie.fr
	http://mis.u-picardie.fr/∼g-caron

Cursus universitaire

Novembre 2010	Doctorat de l'**Université de Picardie Jules Verne**, Amiens Titre : "Estimation de pose et asservissement de robot par vision omnidirectionnelle" | Directeur : El Mustapha Mouaddib Mention Très Honorable.
Juillet 2007	Master STIC-Recherche de l'**Université de Reims, Champagne-Ardennes**, Reims Stage de Master : "Localisation de robot mobile par la vision en utilisant le plafond" Encadrants : El Mustapha Mouaddib et Ouiddad Labbani-Igbida Mention Très Bien. Résultat : 16,11/20 | classement : 1/17
Juillet 2005	Licence Informatique, **Université de Picardie Jules Verne** Mention Bien. Résultat : 15,85/20 | classement : 1/70

Parcours professionnel

Depuis Sep. 2011	Maître de conférences (CNU 61) de l'**Université de Picardie Jules Verne**, Amiens Labo. **Modélisation, Information et Systèmes, EA 4290** Thématique de recherche : Vision robotique directe
Oct. 2010 - Aoû. 2011	Ingénieur expert de recherche INRIA (postdoc), équipe Lagadic **IRISA - INRIA Rennes Bretagne Atlantique**, Rennes

Mobilité

Sep. 2019 - Août 2020	Délégation **CNRS** à l'**UMI 3218 JRL**, AIST, Tsukuba, Japon Collaboration : Abderrahmane Kheddar et Ryusuke Sagawa
Avr.-Mai 2013	Chercheur invité à l'**Université d'Osaka** au Japon Collaboration : Yasushi Yagi et Yasuhiro Mukaigawa (NAIST)
Nov. 2009 et Nov. 2010	Séjours de recherche dans l'équipe **INRIA Lagadic** de Rennes Collaboration : Eric Marchand

Distinctions

2016	Prime d'Encadrement Doctoral et de Recherche, **PEDR**
2015	Second **meilleur article** de la 12ième conférence internationale IEEE/RSJ URAI de la KROS
2011	**Qualification** "maître de conférences" en section 61 du CNU
2007-2010	**Allocation** de thèse de doctorat du MESR ("2nd tour" : sélection de 400 meilleurs dossiers)
2006-2007	**Bourse** de Master sur critères universitaires (MEN)

Sociétés savantes

Depuis 2013	Membre IAPR (via l'AFRIF)
Depuis 2007	Membre IEEE (Graduate Student Member jusqu'en 2011)

1.1.2 Expériences d'enseignement

Unités d'enseignement

Depuis Sep. 2018	Responsable du module (CM, TD, TP) de Vision avancée et réalité augmentée, **UPJV**, Amiens - étudiants de Master 2, responsable du module (CM, TD, TP) de Projet transversal, **UPJV**, Amiens - étudiants de Master 2, responsable du module (CM, TD, TP) de Vision pour la robotique, **UPJV**, Amiens - étudiants de Master 1, et intervenant du module (TP) Capteurs et instrumentation, **UPJV**, Amiens - étudiants de Licence 2,
Depuis Sep. 2015	Responsable du module (CM, TD, TP) de Mise en oeuvre de cellule robotisée, **UPJV**, Amiens - étudiants de Licence Professionnelle
Depuis Sep. 2011	Intervenant en Vision Industrielle (CM) **UPJV**, Nogent-sur-Oise (lycée Marie Curie) - étudiants de Licence Professionnelle
2013 - 2018	Intervenant en Vision et traitement d'images (CM, TD, TP), **UPJV**, Amiens - étudiants de Master 1
2013 - 2016	Responsable de l'Unité d'Enseignement d'Introduction à la robotique (CM, TD, TP), **UPJV**, Amiens - étudiants de Licence 3
2012 - 2016	Intervenant en Asservissement visuel (CM, TD, TP), **UPJV**, Amiens - étudiants de Master 2 et **Université de Bourgogne**, Le Creusot - Master International, jusqu'en 2014
2011 - 2018	Responsable des Unités d'Enseignement (CM, TD, TP) : Vision avancée (Robotique mobile, Localisation et navigation, jusqu'en 2014), **UPJV**, Amiens - étudiants de Master 2
Printemps 2018	Formation continue en photogrammétrie **UPJV**, Amiens - personnel ingénieur, chercheurs
Automne 2013	Formation continue en vision industrielle **UPJV**, Amiens - personnel industriel

D'autre part, je suis très impliqué dans le développement et le renouvellement des équipements pour les travaux pratiques et les projets de robotique et de vision en licence, licence professionnelle et master, avec l'achat, la mise en oeuvre, les mises à jour et évolutions de :

— 3 robots industriels au sein de cellules créées en interne (0 avant mon arrivée)

— 7 robots mobiles (3 anciens avant mon arrivée)

— 4 drones (aucun avant mon arrivée)

— 1 système de capture de mouvement 3D à 6 caméras IR (0 avant mon arrivée)

— 17 caméras industrielles et 1 scrutateur laser industriel (8 anciennes caméras avant mon arrivée)

et le montage des travaux pratiques associés à ces matériels.

Rayonnement pédagogique

20-21 Juin 2019 Membre du comité de programme des 2ièmes Journées Nationales de l'Enseignement de la Robotique (**GdR Robotique**), Toulouse.

18 Oct. 2018 Exposé aux premières Journées Nationales de l'Enseignement de la Robotique (**GdR Robotique**), Montpellier.

2017 - 2018 Porteur du projet d'innovation pédagogique **Projet transversal en équipe**.
Interne à l'UPJV (Budget : 10 K€), j'ai structuré une équipe pédagogique pour faire travailler ensemble les étudiants en formation initiale et ceux en alternance, combiner toutes les compétences acquises dans les unités d'enseignement du Master 2 "Vision Robotique" et établir une réalisation conséquente sur le thème de l'Industrie 4.0.

2014 Mon expérience d'enseignement en vision robotique m'a conduit à être invité à rédiger un article pour la revue **Techniques de l'ingénieur** [**D.2**].

1.1.3 Rayonnement

Responsabilités

Depuis Jan. 2018	Partie française du projet international franco-japonais PHC Sakura (MEAE) et STARS-EIC (Hauts-de-France) **FullScan**
Oct. 2015 - Aoû. 2019	Co-responsable de l'équipe de recherche Perception Robotique du laboratoire **MIS**
Déc. 2016 -	Partenaire UPJV et MT majeur du projet Interreg V-A **ADAPT**
2012 - 2019	Licence pro. "Robotique et Vision Industrielles" (RVI) de l'**UPJV**
2012 - 2016	Membre élu du conseil du laboratoire **MIS** (et aussi en 2009-2010)
2012 - 2016	Responsable de la commission des finances du laboratoire **MIS**
2011 - 2016	Organisateur des séminaires de l'équipe PR du **MIS**
2013 - 2015	Partenaire UPJV pour le projet Interreg IV-A **COALAS**

Animation scientifique

Aoû. 2018 -	Président du comité technique n^o 19 de l'**IAPR** "Computer Vision for Cultural Heritage Applications" (iapr.org/committees)
Jui. 2018 -	Co-animateur du thème "Traitement de l'image et du signal" du **GIS GRAISyHM**
Jan. 2018 -	Co-animateur de l'action "Vision guidée par les capteurs émergents" du **GdR ISIS** (Thème B - Image et vision)
Jan. 2018 -	Co-animateur de l'axe "Numérisation - virtualisation" de la **SFR "Numérique & patrimoine"**

Collaborations internationales

Laboratoire OMI - NAIST, Nara, Japon
Description : depuis 2013 (anciennement avec le Yagi lab de l'Univ. d'Osaka), nombreux séjours de recherche (dans les deux sens, de 1 semaine à 2 mois), y compris sur des supports de chercheurs invités ; travaux réguliers en commun, dont une publication d'article en revue internationale [**RI.3**] ; plusieurs dépôts de projets internationaux, dont un accepté (FullScan, PHC Sakura 2018 et 2019)

Instituto Tecnológico Superior de Misantla, Misantla Veracruz Mexique
Description : depuis 2014, accueil et co-encadrement de stagiaires de Master 1 et 2 et co-encadrement international de thèse depuis 2016 [**Doc.4**]

Laboratoire LTSS - Université Amar Telidji, Laghouat, Algérie
Description : depuis 2017, co-encadrement de co-tutelle internationale de thèse ([**Doc.5**]) et publication internationale ([**RI.1**])

Laboratoire CNRS/AIST UMI 3218/RL JRL, Tsukuba, Japon
Description : depuis 2019, accueil en délégation CNRS et recherche en interaction robotique

Laboratoire CVL - Univ. of Tokyo, Tokyo, Japon
Description : depuis 2017, organisation d'événements scientifiques (IEEE/CVF ICCV workshop on e-Heritage 2017 à Venise en Italie et prochaine édition en 2019 à Séoul en Corée), Professeur invité (Takeshi Oïshi) à l'UPJV en 2018 et travaux de recherche en visualisation 3D immersive

Laboratoire LIMU - Kyushu Univ., Fukuoka, Japon
Description : depuis 2018, recherche en interaction robotique et co-tutelle internationale de stage de Master (1,5 mois à l'UPJV - MIS, 3 mois au Japon)

Collaborations Nationales

Institut IRISA - INRIA/INSA/Univ. de Rennes 1, Rennes
Description : depuis 2011, postdoc et suite des travaux jusqu'en 2014 (2 publications internationales [**RI.6**, **CI.11**]) puis montage du projet Interreg VA ADAPT, travaux communs en cours dans le cadre de ce projet et au-delà ([**RI.2**])

Laboratoire IRSEEM - ESIGELEC, Rouen
Description : depuis 2013, montage et travaux en commun dans le cadre des projets Interreg IVA COALAS (publications internationales [**AI.3**, **EI.3**]), Interreg VA ADAPT et PHC Sakura FullScan

ERL CNRS VIBOT - UBFC, Le Creusot
Description : depuis 2017, montage et travaux en commun dans le cadre du projet PHC Sakura FullScan (publication d'article en revue internationale [**RI.3**]) et dépôt de projets

Evénements organisés

27 Oct. 2019	**IEEE/CVF ICCV** 2019 workshop on **e-Heritage** à Séoul en Corée du Sud (www.eheritage-ws.org)
26 Juin 2019	"Photogrammétrie : pratiques, chantiers et retours d'expérience", **SFR Numérique&Patrimoine** (Programme sur mis.u-picardie.fr/~sfr-np)
17 Juin 2019	"Jeunes chercheurs en traitement de signal/image", **GIS GRAI-SyHM** (www-lisic.univ-littoral.fr/~reflecto/GRAISyHM2019)
6 Juin 2019	"Capteurs visuels émergents : caméras événementielles", **GdR ISIS, GdR Robotique** (Programme sur www.gdr-isis.fr)
11 Oct. 2018	"Vision omnidirectionnelle : 25 ans de recherches", **GdR ISIS, GdR Robotique, UPJV, URN** (Programme sur www.gdr-isis.fr)
3 Juil. 2018	Membre du comité d'organisation de la "Journée Régionale des Doctorants en Automatique" en **région Hauts-de-France** à Amiens (mis.u-picardie.fr/ evenement/JRDA2018/)
20 Juin 2018	Deuxième Journée internationale "Vision 3D & robotique" du laboratoire **MIS**
29 Oct. 2017	**IEEE/CVF ICCV** 2017 workshop on **e-Heritage** à Venise en Italie (www.cvl.iis.u-tokyo.ac.jp/e-Heritage2017)
18 Nov. 2015	Journée **GdR ISIS** "Vision grand angle, multi-caméra et plénoptique" à Paris (Programme sur www.gdr-isis.fr)
15-19 Juin 2015	**ORASIS** : Congrès des jeunes chercheurs en vision par ordinateur à Amiens (orasis2015.sciencesconf.org)
16 Avr. 2014	Concours national "Faites de la science", phase régionale à l'**UPJV**
4 Juin 2012	Première Journée internationale "Vision 3D & robotique" du laboratoire **MIS**
24 Jan. 2012	Monument 3D, atelier de la conférence **RFIA** 2012
1er Avr. 2010	Première Journée des Jeunes Chercheurs du laboratoire **MIS** (JJCMIS 2010), qui se tient tous les ans depuis lors.

Diffusion et vulgarisation : principaux événements

7 Oct. 2016	Animateur du stand "e-Cathédrale au bout des doigts" **1er colloque du du cluster numérique d'Amiens Métropole**
24-26 Mai 2016	Responsable du stand du MIS au salon **Innorobo**, Paris
28-30 Avr. 2016	Instigateur et animateur du stand "e-Cathédrale au bout des doigts" aux journées **Connexions : rencontres du numérique** en Haut-de-France, Amiens
22 Mar. 2016	Conférencier invité au festival **Les Composites** à propos de réalité augmentée dynamique sur marionnette, Compiègne
19-20 Sep. 2015	Guide conférencier invité pour les **Journées Européennes du Patrimoine**, Cathédrale d'Amiens
16 Avr. 2015	Conférencier pour la réalité augmentée dynamique sur marionnette **Connexions : rencontres du numérique**, Amiens
14-15 Sep. 2013	Restitution intermédiaire du programme de recherche e-Cathédrale, **Journées Européennes du Patrimoine**, Logis du Roy à Amiens
Sep. 2012	Responsable du stand du MIS au festival **Science vs fiction** de l'APF, Puteaux

1.1.4 Sollicitations d'expertise

Jury de thèse de doctorat

- 7 mars 2019 : Examinateur de la thèse de Brice Renaudeau, Univ. de Limoges - XLIM
- 24 octobre 2014 : Examinateur de la thèse de Jean-Clément Devaux, Univ. d'Evry - IBISC

Comités de suivi de thèse de doctorat

- 2018, membre externe du comité de suivi de thèse de Nicolas Le Borgne (Insa Rennes - IRISA)
- 2017, membre externe du comité de suivi de thèse de Louis Lecrosnier (ESIGELEC - IRSEEM, Univ. de Rouen Normandie - LITIS)

Comités de sélection de postes de Maître de Conférences

- 2018, membre externe du comité de sélection du poste 61MCF4287 à Limoges (1ère phase sur dossier)
- 2015, membre externe du comité de sélection du poste 61MCF2501 à Nantes (finalement non réuni car mutation)
- 2014, membre local du comité de sélection du poste 61MCF813 à Amiens

Rapports sur des articles de recherche

Principales revues

1. *IEEE Robotics and Automation Letters, RA-L* depuis 2015.
2. *IEEE Transactions on Industrial Electronics, TIE* depuis 2016.
3. *Springer International Journal on Computer Vision, IJCV* en 2014 et 2016.
4. *IEEE Transactions on Systems, Man and Cybernetics : Systems, SMCA*, en 2016.
5. *OSA Journal of the Optical Society of America A, JOSA* en 2011 et 2016.
6. *IEEE Transactions on Robotics, T-RO* en 2014 et 2015.

Principales conférences

1. *IEEE Int. Conf. on Robotics and Automation, ICRA* depuis 2012.
2. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS* depuis 2011.
3. *IAPR Int. Conf. on Pattern Recognition Applications and Methods, ICPRAM* depuis 2019 (et membre du comité de programme).
4. *Congrès des jeunes chercheurs en vision par ordinateur, ORASIS* depuis 2017 (et membre du comité de programme).
5. *Conference on Robotics Innovation for Cultural Heritage, RICH* 2014 (et membre du comité scientifique).
6. *6th Int. Congress "Science and Technology for the Safeguard of Cultural Heritage in the Mediterranean Basin"* 2013 (et membre du comité scientifique et responsable d'une session "Cultural Heritage Identity").

Ateliers

1. Membre du comité scientifique de GeoUAV (*ISPRS Geospatial week* 2015).
2. Membre du comité de programme de "Perception pour le Véhicule Intelligent" (congrès *RFIA* 2014).
3. *Healthcare technology days, CareTECH* 2014.

Rapports sur des projets de recherche

1. Expert scientifique sur un dossier de thèse CIFRE (*ANRT*) en 2017.
2. Rapporteur sur des projets déposés à l'*Agence Nationale de la Recherche, ANR* en 2012 et depuis 2017.

1.1.5 Contrats

Contrats publics

01/01/2018 - 31/03/2021
Projet **FullScan**, financé par le PHC Sakura (Campus France) et co-financé par la région Hauts-de-France (STARS-EIC)
Financement total (France) : 28,5 K€ (part UPJV - MIS : 19,5 K€)
Rôle : Responsable du projet et recherche en numérisation nD
Partenaires : UPJV - MIS (Amiens, porteur), NAIST - OMI lab (Nara, Japon), UBFC - Le2I (Le Creusot), Esigelec - IRSEEM (Rouen)

01/12/2016 - 31/12/2020
Projet **ADAPT**, financé par le programme Interreg V A
Financement total FEDER : 5,9 M€, dont part UPJV - MIS : 335 K€ + 88 K€ de la région Hauts-de-France (post-doc)
Rôle : Responsable de module de travail, encadrement (thèse [**Doc.6**], postdoc [**PoD.3**], master) et recherche en commande robotique basée vision omnidirectionnelle
Partenaires majeurs : Esigelec - IRSEEM (Rouen, porteur), Insa - IRISA (Rennes), University of Kent (Cantebury, Angleterre), UPJV - MIS (Amiens), University College London - Aspire (Londres, Angleterre), Pôle Saint-Hélier (Rennes)

01/10/2015 - 01/10/2018
Projet **Transept**, financé par le CR Picardie
Financement total CR Picardie : 257 K€, part MIS : 139,5 K€
Rôle : Encadrement (postdoc [**PoD.2**], master) et recherche en recalage multi-modal 3D
Partenaires : UPJV - TRAME (Amiens, porteur), UPJV - MIS (Amiens), Amiens Métropole

01/01/2017 - 31/12/2017
Projet **DNI**, financé par la MESHS de Lille | Financement : 5 K€
Rôle : Recherche en reconstruction 3D à base de sémantique
Partenaires : UPJV - CERCLL (Amiens, porteur), UPJV - MIS (Amiens), University Debrecen (Hongrie), UPJV - EHGO (UMR Géo-graphie-cités, CNRS), Université de Lille - ALITHILA et IRCICA (Lille), Université d'Artois - Textes et cultures (Arras)

01/01/2015 - 31/12/2017
Projet **Athar 3D**, financé par le PHC Toubkal Franco-Marocain (Campus France) | Financement, part UPJV - MIS : 30 K€
Rôle : Encadrement (thèse [**Doc.3**]) et recherche en vision 3D
Partenaires : UPJV - MIS (Amiens, porteur français), IGN - LEOMI (Saint-Mandé), Université Mohammed V - LRIT (Rabat, Maroc), IAV Hassan II (Rabat, Maroc)

01/09/2013 - 30/06/2015 Projet **COALAS**, financé par le programme Interreg IV A
Financement total FEDER : 819 K€, part UPJV - MIS : 152 K€,

Rôle : Responsable d'activité, encadrement (postdoc [**PoD.1**], master) et recherche en commande de robot basée vision omnidirectionnelle

Partenaires majeurs : Esigelec - IRSEEM (Rouen, porteur), UPJV - MIS (Amiens), University of Essex (Colchester, Angleterre), University of Kent (Canterbury, Angleterre)

01/10/2012 - 01/10/2015 Projet **Assiduitas**, financé par le CR Picardie (appel structurant)
Financement total CR Picardie : 352 K€, part MIS : 302 K€

Rôle : Encadrement (thèse [**Doc.2**]) et recherche en commande de caméra virtuelle

Partenaires : UPJV - MIS (Amiens, porteur), UTC - Heudiasyc (Compiègne), ACAP (Amiens), Amiens Métropole

Contrats privés

2018 - 2019 Valorisation numérique pour la **Ville de Vitré** | Budget : 5 K€
Rôle : Acquisitions 3D et collaboration avec des archéologues

2017 Vision virtuelle pour **Thales Optronique SA** | Budget : 40 K€
Rôle : Gestion du projet et supervision de l'ingénieur de développement [**Ing.2**] recruté pour ce projet

2016 Robotique agro-alimentaire pour **Godé SAS** | Budget : 10 K€
Rôle : Gestion du projet et supervision de l'équipe de deux chercheurs et deux stagiaires

2015 - 2017 Réalité augmentée pour la compagnie **Le tas de sable - Chés panses vertes** | Budget : 5 K€
Rôle : Développement d'algorithme de tracking articulé et de vidéo-mapping dynamique et encadrement de stagiaire de Master

2014 Sécurité robotique pour la société **CEFF** | Budget : 3 K€
Rôle : Gestion du projet, développement d'un simulateur et réglages sur le terrain

2013 Vision industrielle pour la société **O2GAME** | Budget : 2,5 K€
Rôle : Développement d'étalonnage et de correction de distorsions

1.1.6 Résumé des activités de recherche

1.1.6.1 Introduction

Mes travaux de recherche traitent de la vision robotique directe. Il s'agit de la vision artificielle comme organe de perception des robots en considérant le lien le plus étroit possible entre les pixels des images acquises par une caméra et l'état du robot qui l'embarque ou qui est observé par cette caméra. Ce lien est exploité pour suivre des objets se déplaçant dans le champ de vue de la caméra, localiser le robot ou commander automatiquement ses déplacements pour lui permettre de naviguer. Mes travaux en vision robotique directe s'inscrivent dans trois axes :

Axe 1 Appariement et mouvement : appariement de primitives visuelles et estimation de la structure de l'environnement observé et/ou de mouvement (celui des primitives visuelles ou celui de la caméra elle-même)

Axe 2 Suivi basé modèle : suivi d'objet, de bâtiment, de scène dans les images d'une caméra en mouvement, connaissant une maquette 3D partielle de cet objet

Axe 3 Asservissement visuel : commande automatique de robot référencée capteur de vision

Ces trois axes structurent bien l'ensemble de mes travaux, que ce soit durant la thèse de doctorat (Partie 1.1.6.2) ou dans mes activités plus récentes (Partie 1.1.6.3).

Dans cette partie, les acronymes suivants distinguent le type de référence (bibliographie personnelle en page 27) :

- RI : article de revue internationale à comité de lecture
- RN : article de revue nationale à comité de lecture
- CI : article d'actes de conférences internationales à comité de lecture
- AI : article d'actes d'atelier international à comité de lecture
- CN : article d'actes de conférences internationales à comité de lecture
- D : publications diverses, sur invitation
- EI : communication (exposé) internationale
- EN : communication (exposé) nationale
- Inv : exposé invité
- PoD : post-doctorant encadré
- Doc : doctorant encadré
- Ing : ingénieur encadré

1.1.6.2 Activités de recherche relatives à la thèse de doctorat

Ce sont la vision et la stéréovision omnidirectionnelles qui ont été au coeur de mes travaux de thèse de doctorat. Cette perception visuelle omnidirectionnelle est obtenue avec une lentille grand angle fisheye ou avec un système catadioptrique, combinant un miroir et une ou plusieurs lentilles. Le thème principal de ma thèse de

doctorat a été de concevoir des approches de localisation et de navigation de robot mobile, en exploitant un système stéréo composé d'une seule caméra et de quatre miroirs, nommé FOO : Four On One. Le FOO acquiert quatre points de vue de la scène environnante en une seule image.

Axe 1 : Pour que la localisation soit possible, les images stéréoscopiques du FOO doivent être analysées. J'ai proposé une approche innovante travaillant avec les primitives géométriques que sont les droites verticales de l'environnement. Dans une seule acquisition du FOO, cette approche détecte les droites verticales de la scène, les apparie et les reconstruit dans l'espace [**CI.17**, **CN.10**], en utilisant les contraintes géométriques induites par le FOO.

Suivre des primitives géométriques dans les images du FOO au cours de son déplacement permet, de plus, d'estimer ce déplacement, tout en reconstruisant, partiellement, la scène. Pour ce faire, j'ai proposé le suivi photométrique de droites verticales et le suivi photométrique de zones planes de la scène en stéréovision omnidirectionnelle [**CI.12**], un seul plan apportant suffisamment d'informations pour estimer la pose relative entre deux poses du FOO. Plus particulièrement, le suivi photométrique consiste à considérer les intensités des pixels inclus dans la zone plane en entrée du problème d'optimisation estimant simultanément les paramètres du plan et le déplacement de la caméra, rendant ainsi le lien très étroit entre les intensités des pixels et l'état (la pose) du robot qui embarque le FOO.

Axe 2 : En se basant sur des travaux existants sur l'asservissement visuel virtuel (AVV), un formalisme puissant pour l'estimation de pose et le suivi, bénéficiant du large champ de connaissances de l'asservissement visuel, j'ai développé le suivi d'objet en utilisant sa maquette 3D filaire composée d'arêtes dans des séquences d'images acquises par le FOO [**RI.8**, **CI.16**, **EN.11**, **EN.7**].
Cette dernière méthode permet d'estimer la position et l'orientation de l'objet dans l'espace et je l'ai étendue à l'étalonnage stéréoscopique omnidirectionnel à base de points. Ce travail réalise le calcul simultané des paramètres de chaque caméra d'un banc stéréo, y compris le FOO, et leurs poses relatives [**CI.15**]. En collaboration avec un autre doctorant de l'équipe de recherche à laquelle j'appartenais, doctorant qui travaillait sur un système stéréo mixte, j'ai rendu ma méthode d'étalonnage hybride. Elle permet l'étalonnage de systèmes composés de caméras de différents types et modélisées différemment, comme une caméra fisheye et une caméra perspective, par exemple [**CI.13**, **CN.9**]. Ces derniers travaux ont mené à la réalisation d'un logiciel, partagé avec la communauté scientifique [**Log.1**] et à propos duquel un article synthétique a été publié [**D.3**].
De manière générale, travailler avec les images omnidirectionnelles, même si elles apportent un champ de vue omnidirectionnel, impose de prendre en compte les fortes distorsions de ces images. Elles conduisent à des problèmes de traitement d'images et de modélisation de primitives visuelles. J'ai alors proposé, pour la localisation d'un robot mobile par une caméra omnidirectionnelle externe et en utilisant une balise active embarquée, de formuler et comparer différentes représentations d'un point (plan image, sphère unitaire cartésienne ou polaire). Ceci a été fait dans le but de déterminer quelle représentation est la mieux adaptée pour l'estimation de

pose et a servi de travail préliminaire à l'asservissement visuel photométrique omnidirectionnel décrit ci-dessous.

Axe 3 : Tout comme le suivi photométrique de plans évoqué plus haut, considérer les intensités des pixels de l'image comme primitive d'un asservissement visuel permet d'éviter l'extraction de points ou de droites dans les images. L'asservissement visuel photométrique prenant une image directement en entrée de la loi de commande du robot, il n'y a même pas besoin de détection initiale (cf. suivi photométrique de plans). J'ai donc étendu l'asservissement visuel photométrique, existant en vision perspective, à la vision omnidirectionnelle en proposant, notamment, la représentation d'image et le calcul de gradients d'image adaptés aux distorsions des images omnidirectionnelles, conduisant aux meilleures performances de l'asservissement visuel d'un manipulateur industriel, comme d'un robot mobile [**RI.7**, **CI.14**, **EN.8**, **EN.9**, **EN.10**].

1.1.6.3 Activités de recherche postérieures à la thèse

Introduction

Les travaux de recherche que je mène depuis l'obtention du doctorat s'inscrivent toujours dans le thème de la vision directe mais s'étendent à d'autres types de caméra que la caméra omnidirectionnelle (ex : caméra sphérique polydioptrique). Cette vision directe s'entend, de plus, sur robot (ex : asservissement visuel direct à domaine de convergence étendu grâce à la modélisation de l'image par un mélange de gaussiennes photométriques) ou sans robot (ex : recalage précis pour la coloration photographique de maquette 3D haute définition) ainsi qu'en virtuel (ex : commande de caméra virtuelle pour l'assistance à la navigation dans les environnements 3D denses et riches en information).

Comme indiqué précédemment, l'ensemble de mes contributions se structure selon les trois axes évoqués (Partie 1.1.6.1). A cela s'ajoute un axe d'ouverture interdisciplinaire au sujet du patrimoine architectural à la fois à l'origine de certaines contributions (ex : commande de caméra virtuelle), comme contexte d'application d'autres contributions (ex : recalage précis) et source de réelles collaborations interdisciplinaires avec des géomètres, des informaticiens en science de l'éducation et des historiens de l'art.

Contributions dans les trois axes principaux de mes travaux

Axe 1 : La reconstruction 3D de points d'intérêt d'un objet observé par une caméra permet d'obtenir une estimation de la structure de cet objet. Plusieurs images de plusieurs points de vue différents d'un même objet sont nécessaires pour le reconstruire en 3D. Un système de vision catadioptrique à une seule caméra perspective et un ou plusieurs miroirs plans permet d'obtenir en une seule image plusieurs points de vue du même objet. Deux miroirs plans judicieusement agencés permettent, de plus, de couvrir toutes les faces d'un objet, aussi très utile pour une reconstruction

3D d'objet quand il n'est pas possible de faire prendre du recul à la caméra par rapport à cet objet. Ce type de système de vision existant dans la littérature mène à une reconstruction 3D de l'enveloppe de l'objet à partir de l'observation de sa silhouette dans diverses vues. Dans le cadre de la thèse de Noureddine Mohtaram, que j'ai co-encadré [**Doc.3**], nous nous sommes intéressés non pas à la silhouette de l'objet mais à l'ensemble de son apparence dans le but de pouvoir reconstruire plus de détails et ainsi obtenir une maquette 3D de l'objet plus fidèle à la réalité qu'en considérant les silhouettes. Pour ce faire, il faut d'abord détecter et mettre en correspondance des points image en prenant en compte les multiples réflexions induites par les deux miroirs. Pour cela, nous avons proposé une nouvelle approche à base de descripteur, nommée AMIFT [**CI.2**], qui donne de meilleurs résultats que les méthodes les plus performantes de l'état de l'art. Une fois la mise en correspondance faite, l'étalonnage géométrique du système mène à la reconstruction partielle de l'objet [**CN.2**].

Les points d'intérêt dans les images acquises, que ce soit d'un objet, comme évoqué ci-dessus, ou d'une scène, sont aussi utilisés pour estimer le mouvement de la caméra classique, notamment quand elle est embarquée sur un robot mobile. Les méthodes de type SLAM (Simultaneous Localization And Mapping) réalisent, de plus, simultanément, la reconstruction 3D de la scène et l'estimation du mouvement de la caméra. La robustesse du suivi de points d'intérêt dans les images acquises par une caméra mise en mouvement par un robot mobile est critique pour estimer ce mouvement. A ce sujet, je co-encadre Yassine Ahmine [**Doc.5**] qui, dans ses travaux de thèse, a proposé d'étendre la technique du suivi de points d'intérêt du Kanade-Lucas-Tracker dans l'espace d'échelle. L'espace d'échelle avait été, par le passé, étudié pour rendre robuste le suivi de points à des mouvements importants entre images successives d'un flux vidéo, en considérant plusieurs niveaux d'échelle (généralement entre 3 et 5 réductions de résolution) de l'image. La contribution du travail de Yassine est de considérer l'échelle comme un paramètre supplémentaire de la méthode de suivi, ne la limitant pas à quelques niveaux fixés arbitrairement, surpassant ainsi toutes les méthodes concurrentes de l'état de l'art [**RI.1**].

Une des limites majeures d'une caméra classique (caméra perspective) est qu'elle ne possède qu'un seul point de vue. Si déplacer la caméra permet d'avoir des images d'une même scène de différents points de vue et permet d'avoir une estimation de la structure de la scène, du mouvement de la caméra et/ou du mouvement d'un objet observé, cette estimation n'est faite qu'à un facteur d'échelle près. Deux caméras, ou plus, assemblées en un banc stéréoscopique étalonné, permettent de fixer l'échelle et la redondance d'informations permet d'avoir des estimations plus robustes aux perturbations, au prix d'un encombrement plus important du système de vision. La vision plénoptique permet de multiplier (énormément) les points de vue en une caméra aussi compacte qu'une caméra monoculaire classique grâce à l'ajout d'une matrice de micro-lentilles devant le capteur photosensible, certes au prix de points de vue très proches les uns des autres. Les caméras plénoptiques sont assez étudiées dans la littérature, notamment pour calculer des propriétés de matière à partir de l'observation de l'interaction entre un objet et une source lumineuse ou reconstruire

les profondeurs d'une scène. Néanmoins, peu de travaux existent en robotique et, avec Nathan Crombez, dont j'ai co-encadré un contrat postdoctoral [**PoD.2**], et en collaboration avec des chercheurs japonais du NAIST, je me suis intéressé à l'étude du calcul de pose d'objet plan dans les images non-conventionnelles acquises par ce type de caméra. Les contributions de cette étude portent sur la modélisation géométrique du problème, son estimation à partir de points d'intérêts appariés et la conclusion de l'apport de la vision plénoptique à l'estimation de plan en levant une limite théorique du cas à deux vues [**RI.3**, **Inv.3**].

Les estimations de structure et/ou de mouvement à base de primitives géométriques, comme les points d'intérêt évoqués ci-dessus, nécessitent une phase de détection et de mise en correspondance de ces primitives dans plusieurs images. Chacune de ces étapes souffre d'imprécisions qui s'accumulent et peuvent mener à des estimations parfois peu fiables. Je me suis donc intéressé à estimer le mouvement d'une caméra à partir des intensités de ses pixels directement. Il a fallu lever les limites des approches d'estimations photométriques qui, quoique très précises, souffrent d'un domaine de convergence restreint. Je me suis tout d'abord concentré sur le problème d'estimation de cap de caméra et, dans l'objectif d'avoir un domaine d'estimation global (c'est-à-dire sur 360°), j'ai, à nouveau, considéré la caméra catadioptrique omnidirectionnelle pour estimer le cap d'un robot mobile en mouvement. En collaboration avec Fabio Morbidi, également membre du laboratoire MIS, nous avons proposé d'exploiter la transformée de Fourier polaire pour transformer les images omnidirectionnelles dans le domaine fréquentiel où la correlation de deux images donne directement l'angle de rotation qui les sépare [**RI.5**, **EI.2**].
Ensuite, pour étendre l'estimation d'orientation aux deux autres angles de roulis et tangage en plus du cap, notamment intéressant pour la robotique aérienne, nous nous sommes orientés vers la vision sphérique polydioptrique dans l'objectif d'avoir un domaine d'estimation global pour les trois angles (c'est-à-dire toutes les rotations possibles). Toujours dans le but de calculer la rotation entre deux images nous avons proposé une nouvelle transformée d'image (sphérique), le mélange de potentiels photométriques qui, si elle ne permet pas encore d'obtenir directement les angles par un calcul de corrélation, permet de les calculer par une optimisation non-linéaire en quelques itérations [**CI.3**, **CN.1**]. Ce mélange de potentiels photométriques est un mélange de fonctions gaussiennes paramétrées par les coordonnées de chaque pixel sphérique, leurs intensités et par un paramètre d'extension commun à tous les pixels de l'image sphérique. Dans la terminologie statistique, le premier est l'équivalent de la moyenne d'une gaussienne, les secondes sont équivalentes aux pondérations du mélange de gaussiennes et le dernier est équivalent à l'écart-type d'une gaussienne. Le mélange de potentiels photométriques de chaque image permet d'appliquer la même méthodologie d'estimation que dans les méthodes photométriques mais dans un domaine de convergence bien plus étendu, a minima pour la rotation. Ce dernier travail étend à la vision sphérique le mélange de gaussiennes photométriques, proposé antérieurement, chronologiquement parlant, et décrit plus loin dans l'**Axe 3**.

Axe 2 : Dans le but d'éviter les étapes de traitement d'image de détection et mise en correspondance de primitives géométriques communes entre un objet ou une scène observée par une caméra pour en calculer la pose 3D, j'ai considéré des maquettes virtuelles 3D texturées, c'est-à-dire dont l'apparence photométrique est proche de celle de la scène réelle. Cette proximité d'apparence étant variable, il est difficile de comparer une image acquise par une caméra et une image de synthèse directement à partir de leur apparence photométrique.

C'est pourquoi, durant mon contrat postdoctoral, je me suis intéressé à l'information mutuelle, une mesure issue de la théorie de l'information, reformulée pour la donnée photométrique, et dérivée pour optimiser la pose de la caméra perspective observant la scène en maximisant l'information mutuelle [**RI.6**, **CN.8**]. Si l'information mutuelle est robuste à un réalisme minime de l'apparence de la maquette virtuelle, elle engendre un domaine de convergence parfois étroit du calcul de pose.

Dans le cas où l'apparence photométrique de la maquette 3D est plus réaliste, ce qui est le cas d'un nuage de points 3D colorés issus de scans laser et de photos plaquées sur ce nuage, le critère photométrique de base devient pertinent. Le critère photométrique, c'est-à-dire la différence pixel à pixel entre l'image acquise et l'image de synthèse, permet un calcul de pose précis de la caméra perspective dans un domaine de convergence moins restreint qu'avec l'information mutuelle. C'est ce que l'on a pu exprimer et montrer [**RN.2**, **AI.1**, **CN.7**] dans le cadre de la thèse de Nathan Crombez, que j'ai co-encadré [**Doc.1**]. Plus concrètement, la précision du calcul photométrique de pose est telle que l'on peut remplacer les couleurs du nuage de points 3D, pour donner une autre apparence réaliste (qualité photographique) à la maquette 3D.

En étendant le calcul photométrique de pose basé maquette 3D à la vision omnidirectionnelle embarquée sur robot mobile, nous avons proposé une nouvelle méthode de localisation de robot mobile basée vision et nuage de points 3D colorés [**CI.4**].

Ces méthodes de calcul photométrique de pose dépendent du réalisme, au moins photographique, des maquettes virtuelles considérées. Cependant, l'apparence globale de ces dernières en manque souvent quand on combine de nombreux scans, en particulier acquis à différents moments de la journée, voire des années différentes. Cela est dû au fait que les scanners laser du marché donnent une couleur à chaque point 3D grâce à une caméra qui balaie aussi la scène dans toute les directions. De plus, la cohérence de l'apparence dépend de l'étalonnage interne entre le télémètre laser et la caméra et la différence de résolution entre ces deux derniers. Idéalement, il faudrait pouvoir se passer de la caméra du scanner et être capable de mesurer la couleur du point d'impact laser directement. Pour ce faire, j'ai collaboré avec d'autres membres du laboratoire MIS pour proposer une méthode d'étalonnage de la quantité d'énergie du laser qui revient vers le scanner afin de pouvoir mesurer le niveau de vert du point d'impact laser [**RI.4**], c'est-à-dire l'une des trois composantes de couleur dans un codage "Rouge, Vert, Bleu".

L'ensemble de mes travaux sur cet axe a, plusieurs fois, fait l'objet d'exposés invités ces cinq dernières années [**Inv.10**, **Inv.8**, **Inv.5**, **Inv.2**], s'enrichissant ainsi au fil du temps.

Axe 3 : J'ai étendu les travaux d'asservissement visuel photométrique en vision omnidirectionnelle en l'adaptant mieux à la robotique mobile non-holonome que par le passé. J'ai donc proposé, en collaboration avec Youssef Alj, dont j'ai encadré le contrat postdoctoral [**PoD.1**], une nouvelle approche de commande référencée vision directe découplée, c'est-à-dire où le degré de liberté de translation du robot est contrôlé séparément de celui d'orientation [**CI.5**, **EI.3**, **AI.3**]. Ces derniers travaux ont permis au robot mobile de suivre des chemins visuels à forte courbure en réalisant des asservissements visuels successifs.

Mais la contribution majeure de mes travaux concerne la définition de la nouvelle primitive du mélange de gaussiennes basé image pour la commande de robot et de caméra virtuelle. C'est une contribution réalisée dans le cadre de la thèse de Nathan Crombez [**Doc.1**]. Il s'agit de considérer des fonctions gaussiennes issues de tous les pixels de chaque image en entrée de la loi de commande d'asservissement visuel. Chaque pixel donne une gaussienne, paramétrée par l'intensité du pixel qu'elle représente d'une nouvelle façon. Enfin, l'une des idées clé est d'ajouter un facteur d'extension commun à toutes les gaussiennes de l'image comme degré de liberté supplémentaire. Ces travaux ont permis d'étendre considérablement le domaine de convergence de l'asservissement visuel direct [**RI.2**, **CI.7**, **CN.3**].

Ensuite, nous avons étendu le concept du mélange de gaussiennes basé image comme primitive d'asservissement visuel au mélange de gaussiennes basées saillance visuelle pour l'exploration de scène et le cadrage pertinent d'objet [**CI.6**, **Inv.9**], dans le cadre de la thèse de Zaynab Habibi, que j'ai co-encadrée [**Doc.2**]. Dans cette même thèse, nous nous sommes orientés vers la maximisation du mélange de gaussiennes basées saillance pour la commande de caméra virtuelle, en combinant cette maximisation à d'autres critères de réalisme d'orientation et de mouvement de caméra [**CI.8**, **AI.2**, **CN.6**, **CN.4 EN.5**, **EN.2**, **EN.3**].

Le programme e-Cathédrale comme axe d'ouverture interdisciplinaire

Le programme de recherches e-Cathédrale[1] du laboratoire MIS porte sur la création de maquettes numériques d'édifices du patrimoine avec, comme fer de lance, la cathédrale d'Amiens. Ce programme interdisciplinaire va de la mesure in situ à la visualisation immersive en passant par l'étude architecturale.

L'ensemble des travaux de recherche évoqués précédemment a une intersection importante avec le programme e-Cathédrale, qu'il soit à l'origine de ces travaux [**CI.6**, **Inv.9**,**CI.8**, **AI.2**, **CN.6**, **CN.4 EN.5**, **EN.2**, **EN.3**, **RI.4**, **CN.2**, **CI.2**] ou qu'il en soit le contexte applicatif [**CN.7**, **RN.2**, **AI.1**].

En plus de ces travaux déjà évoqués, mon implication dans le programme e-Cathédrale a mené à des collaborations interdisciplinaires avec des chercheurs en informatique à propos de méthodes de conception générique de jeux sérieux pour l'éducation [**EI.4**] incluant des interfaces tangibles [**CI.10**], et s'adaptant à l'usager [**CI.9**].

1. mis.u-picardie.fr/e-cathedrale

Enfin, étant un acteur très actif du programme e-Cathédrale, je communique régulièrement sur l'état d'avancement du programme, de la méthodologie de numérisation jusqu'aux dispositifs de restitution immersive ou à distance, en passant par les traitements considérés [**EI.5**, **EN.4**, **Inv.7**, **Inv.6**, **Inv.4**, **RN.1**].

1.1.6.4 Synthèse des activités de recherche postérieures à la thèse

En résumé, mes travaux de recherche s'appuient sur un état de l'art qui avait montré qu'il était possible d'exploiter directement la primitive photométrique pour la commande de robot et l'estimation de mouvement dans l'espace, rapprochant grandement l'image acquise de la variation des degrés de liberté de la caméra qui l'acquiert, sans avoir besoin d'extraire et mettre en correspondance des primitives géométriques. Le coeur de mes travaux de recherche s'est attaché à identifier et repousser les limites de cette primitive photométrique, en terme de robustesse à des perturbations majeures, grâce à l'information mutuelle, d'une part, et en terme d'étendue du domaine de convergence, grâce aux mélanges de potentiels photométriques, d'autre part. Afin de montrer la puissance de cette dernière primitive directe, je m'attache aussi à la décliner pour différents types de caméra (perspective, sphérique) et pour diverses applications (commande de robot, estimation d'orientation dans l'espace). Mes travaux de recherche ont aussi permis d'introduire la navigation basée optimisation de saillance et d'entropie photométriques, notamment, directement calculées à partir des pixels de l'image, en environnement virtuel et réel.

1.1.7 Encadrement

Contrats postdoctoraux

PoD.1 "Asservissement visuel basé intensité pour une chaise roulante semi-autonome", **Youssef Alj** (recruté en postdoc à I3S-CNRS, Sophia en 2015 - 2016), mars 2014 - mai 2015 (taux d'encadrement : 100%) | publications en commun : [**CI.5**, **EI.3**]

PoD.2 "Géométrie des caméras plénoptiques", **Nathan Crombez** (recruté MCF CNU 61 en 2018 à l'UTBM), janvier 2016 - décembre 2016 (co-supervisé par El Mustapha Mouaddib, taux d'encadrement : 50%) | publication en commun (en tant que postdoc) : [**RI.3**]

PoD.3 "Vision grand angle pour fauteuil roulant intelligent", **Houssem Benseddik**, janvier 2018 - (co-supervisé par Fabio Morbidi, taux d'encadrement : 50%)

Thèses de doctorat

Doc.1 "Contributions aux asservissements visuels denses : nouvelle modélisation des images adaptée aux environnements virtuels et réels", **Nathan Crombez*** (recruté MCF CNU 61 en 2018 à l'UTBM), soutenue le **9 décembre 2015**. *Directeur de thèse* : Claude Pégard. Taux d'encadrement officiel : 70%. Financement : allocation ministérielle | publications en commun (en tant que doctorant) : [**RI.2**, **RN.2**, **CI.4**, **CI.7**, **AI.1**, **CN.7**, **CN.3**, **CN.7**]

Doc.2 "Vers l'assistance à l'exploration pertinente et réaliste d'environnements 3D très denses", **Zaynab Habibi*** (devenue ingénieur dans le privé), soutenue le **8 décembre 2015**. *Directeur de thèse* : El Mustapha Mouaddib. Taux d'encadrement officiel : 50%. Financement : projet régional structurant ASSIDUITAS | publications en commun : [**CI.6**, **CI.8**, **AI.2**, **CN.4**, **CN.6**, **Inv.9**]

Doc.3 "Vision multi-miroirs pour la reconstruction 3D", **Noureddine Mohtaram*** (devenu ingénieur dans le privé), co-tutelle avec l'université Mohammed V de Rabat (UM5, Maroc), soutenue le **19 septembre 2019**. *Directeur de thèse* : El Mustapha Mouaddib. *Co-encadrantes* : Amina Radgui (UM5) et Sanaa El Fkihi (UM5). Taux d'encadrement officiel : 30%. Financement : bourse d'excellence et projet international (PHC) Athar3D | publications en commun : [**CI.2**, **CN.2**]

Doc.4 "Navigation de drone basée vision", **Eder Rodriguez**, co-encadrement avec l'université Autonome de Tamaulipas (UAT, Mexique) depuis octobre 2016. *Directeur de thèse* : Claude Pégard. *Co-encadrant* : David Lara Alabazares (UAT). Financement : allocation du CONACYT (Mexique) pour effectuer sa thèse en France.

Doc.5 "Navigation autonome de robot mobile basée vision", **Yassine Ahmine***, co-encadrement avec l'université Amar Teloudji de Laghouat (UATL, Algérie) depuis novembre 2017. *Directrice de thèse* : Fatima Chouireb (UATL). *Co-encadrant* : El Mustapha Mouaddib. Financement : Bourse Profas B+ France-Algérie | publications en commun : [**RI.1**]

Doc.6 "Asservissement visuel omnidirectionnel de robot mobile", **Seif Eddine Guerbas***, depuis novembre 2017. *Directeur de thèse* : El Mustapha Mouaddib. Financement : projet international Interreg ADAPT.

Doc.7 "Vision catadioptrique adaptative", **Julien Ducrocq**°, depuis octobre 2019. *Directeur de thèse* : El Mustapha Mouaddib. Financement : allocation ministérielle.

* : Co-encadrement officiel accordé par le Conseil Scientifique de l'UPJV.

° : En attente d'autorisation officielle de co-encadrement par le Conseil Scientifique de l'UPJV.

Ingénieurs

Ing.1 "Accessibilité collaborative", Thibault Potin, février - juillet 2019. Taux d'encadrement : 100%. Financement : projet Interreg VA Adapt.

Ing.2 "Commande de caméra virtuelle", Nathan Sanchiz (devenu doctorant au MIS), janvier - juin 2017. Taux d'encadrement : 100%. Financement : contrat Thales OSA.

Stagiaires de Master

Masters 2

Sarah Delmas (ingénieur ENSTA-Bretagne), avril – sep. 2019
Support : projet Interreg V A Adapt
Sujet : Interfaçage de robot pour l'acquisition de données vision-inertiel synchronisées et asservissement visuel | Co-encadrant : Claude Pégard (PR, MIS – UPJV) | Taux d'encadrement : 80%

Julien Ducrocq (Master UPJV), mars – août 2019
Support : projet Interreg V A Adapt
Sujet : Asservissement visuel omnidirectionnel à large gamme de dynamique | Taux d'encadrement : 100%

Salah Cherigui (ingénieur CNAM), mai – décembre 2017
Support : fonds propres MIS
Sujet : Gestion des latences dans la commande déportée de drone
Co-encadrant : Claude Pégard (PR, MIS – UPJV) | Taux d'encadrement : 50%

Masters 2

Jordan Caracotte (Master UPJV), avril – septembre 2016
Support : projet régional Picardie-FEDER « Transept »
Sujet : Recalage de modèle CAO sur nuage de points 3D
Co-encadrant : El Mustapha Mouaddib (PR, MIS – UPJV) | Taux d'encadrement : 50%

Nancy Aguilar, janvier – mai 2016
Support : bourse PHILEAS
Sujet : Gyroscope visuel dense pour drone
Co-encadrant : David Lara (Enseignant-chercheur Univ. de Reynosa, Mexique) | Taux d'encadrement : 80%

Mounya Belghiti, avril – septembre 2015
Support : prestation, compagnie Chés Panses Vertes
Sujet : Réalité augmentée projetée | Taux d'encadrement : 100%

Jaouad Hajjami, juin – septembre 2015
Sujet : Gyroscope visuel dense | Taux d'encadrement : 100%

Eder Rodriguez, avril – juillet 2015
Support : bourse d'étude mexicaine (Univ. de Reynosa)
Sujet : Asservissement visuel de drone quadrotor
Co-encadrant : David Lara (Enseignant-chercheur Univ. de Reynosa, Mexique) | Taux d'encadrement : 80%

Noureddine Mohtaram, février – juillet 2013
Support : allocation de l'Univ. Mohammed V de Rabat au Maroc
Sujet : Mise en correspondance utilisant un système de vision basé sur deux miroirs plans
Co-encadrants : Amina Radgui (Enseignant-chercheur, Univ. Mohammed V), El Mustapha Mouaddib (PR, MIS – UPJV), Driss Aboutajdine (PR, Univ. Mohammed V)
Taux d'encadrement : 25%

Nicolas Cazy, avril – août 2013
Support : allocation du laboratoire MIS
Sujet : Commande embarquée d'un drone utilisant la vision
Co-encadrant : Abdelhamid Rabhi (MCF, MIS – UPJV) | Taux d'encadrement : 25%

Nathan Crombez, avril – août 2012
Support : allocation du laboratoire MIS
Sujet : Visualisation et navigation virtuelles de la maquette numérique de la cathédrale d'Amiens
Co-encadrant : El Mustapha Mouaddib (PR, MIS – UPJV) | Taux d'encadrement : 50%

Masters 1

Julien Ducrocq, avril – août 2018
Support : co-tutelle internationale (fonds propres : MIS - UPJV, et Kyushu University - LIMU)
Sujet : Réalité augmentée robotique
Co-encadrant : Hideaki Uchiyama (Associate Professor, Kyushu Univ., Japon) | Taux d'encadrement : 50%

Nancy Aguilar, mai – août 2015
Support : bourse d'étude mexicaine (Univ. de Reynosa)
Sujet : Etude et développements sur drone Parrot Bebop
Co-encadrant : David Lara (Enseignant-chercheur Univ. de Reynosa, Mexique) | Taux d'encadrement : 80%

Jaouad Hajjami, avril – août 2014
Support : projet Interreg IVA "COALAS : plateforme cognitive pour l'assistance aux personnes dépendantes"
Sujet : Système de vision stéréoscopique hybride pour robot mobile
Taux d'encadrement : 100%

Eder Rodriguez, octobre – novembre 2014
Support : bourse d'étude mexicaine (Univ. de Reynosa)
Sujet : Asservissement visuel de drone quadrotor
Co-encadrant : David Lara (Enseignant-chercheur Univ. de Reynosa, Mexique) | Taux d'encadrement : 80%

1.2 Production scientifique

1.2.1 Articles de revues

Revues internationales (8)

Synthèse : 2 articles sur les travaux de thèse [**RI.7**, **RI.8**] et 6 depuis lors, dont 1 en tant que postdoc [**RI.6**], puis 2 avec des doctorants encadrés [**RI.1**, **RI.2**] (en collaboration internationale et nationale), 1 avec un postdoc encadré [**RI.3**] (en collaboration internationale) et les autres avec des collègues enseignants-chercheurs de l'équipe de recherche à laquelle j'appartiens [**RI.4**, **RI.5**].

RI.1 "Adaptive Lucas-Kanade tracking", Y. Ahmine, G. Caron, E. Mouaddib, F. Chouireb, *Image and Vision Comput.* (Elsevier), Vol. 88, pp. 1-8, Août 2019

RI.2 "Visual Servoing with Photometric Gaussian Mixtures as Dense Features", N. Crombez, E. Mouaddib, G. Caron, F. Chaumette, *Transactions on Robotics* (IEEE), vol. 35, no. 1, pp. 49-63, Février 2019

RI.3 "Reliable planar object pose estimation in light-fields from best sub-aperture camera pairs", N. Crombez, G. Caron, T. Funatomi, Y. Mukaigawa, *Robotics and Automation Letters* (IEEE), Vol. 3, No. 4, pp. 3561 - 3568, Octobre 2018

RI.4 "Spherical target-based calibration of terrestrial laser scanner intensity. Application to colour information computation", E. Bretagne, P. Dassonvalle, G. Caron, *ISPRS Journal of Photogrammetry and Remote Sensing* (Elsevier), Vol. 144, pp. 14 - 27, Octobre 2018

RI.5 "Phase Correlation for Dense Visual Compass from Omnidirectional Camera-Robot Images", F. Morbidi et G. Caron, *Robotics and Automation Letters* (IEEE), Vol. 2, No. 2, pp. 688-695, Avril 2017

RI.6 "Direct model based visual tracking and pose estimation using mutual information", G. Caron, A. Dame et E. Marchand, *Image and Vision Computing*, Vol. 32, No. 1, pp. 54-63 (Elsevier), Janvier 2014

RI.7 "Photometric visual servoing for omnidirectional cameras", G. Caron, E. Marchand et E. Mouaddib, *Autonomous Robots*, Vol. 35, No. 2, pp. 177 - 193, (Springer), Octobre 2013

RI.8 "3D model based tracking for omnidirectional vision : A new spherical approach", G. Caron, E. Mouaddib et E. Marchand, *Robotics and Autonomous Systems*, Vol. 60, No. 8, pp. 1056 - 1068 (Elsevier), 2012

Revues nationales (2)

Synthèse : 1 article avec un doctorant encadré [**RN.2**] et 1 article avec des collègues enseignants-chercheurs de l'équipe de recherche à laquelle j'appartiens [**RN.1**].

RN.1 "Le patrimoine in silico. Exemple de la cathédrale d'Amiens.", E. Mouaddib, G. Caron, D. Groux, F. Morbidi, *In Situ - Revue des patrimoines* du Ministère de la Culture, vol. 39, 2019

RN.2 "Colorisation de nuages de points 3D par recalage dense d'images numériques", N. Crombez, G. Caron et E. Mouaddib, *Traitement du Signal*, vol. 31, No. 1-2, pp. 81 - 106 (Lavoisier), 2014

1.2.2 Articles dans les actes de conférences

Conférences internationales (17)

Synthèse : 5 articles avec des doctorants encadrés [**CI.2**, **CI.4**, **CI.6**, **CI.7**, **CI.8**], 1 article avec un postdoc encadré [**CI.5**], 4 articles avec des collègues enseignants-chercheurs de l'équipe de recherche à laquelle j'appartiens [**CI.1**, **CI.3**, **CI.9**, **CI.10**], 1 article en tant que postdoc [**CI.11**] et 6 articles sur les travaux de thèse [**CI.12**, **CI.13**, **CI.14**, **CI.15**, **CI.16**, **CI.17**].

CI.1 "Efficient reproduction of heritage buildings : a new way to exploit 3D point cloud slicing : An example with the Amiens Gothic Cathedral", D. Groux-Leclet, J. Lentremy, G. Caron, E. Mouaddib, *IEEE Digital Heritage Int. Congress (**DigitalHeritage**)*, pp.1-4, Octobre 2018, San Francisco, USA

CI.2 "AMIFT : Affine-mirror invariant feature transform", N. Mohtaram, A. Radgui, G. Caron, E. Mouaddib, *IEEE Int. Conf. on Image Processing (**ICIP**)*, pp. 893-897, Octobre 2018, Athènes, Grèce

CI.3 "Spherical Visual Gyroscope for Autonomous Robots using the Mixture of Photometric Potentials", G. Caron, F. Morbidi, *IEEE Int. Conf. on Robotics and Automation (**ICRA**)*, pp. 820-827, Mai 2018, Brisbane, Australie

CI.4 "Using dense point clouds as environment model for visual localization of mobile robot", N. Crombez, G. Caron et E. Mouaddib, **Best runner-up paper award** of the 12th *KROS-IEEE/RSJ Int. Conf. on Ubiquitous Robots and Ambient Intelli. (**URAI**)*, pp. 40-45, Octobre 2015, Goyang, Corée du Sud

CI.5 "Featureless omnidirectional vision-based control of non-holonomic mobile robot", Y. Alj et G. Caron, *KROS Int. Conf. on Ubiquitous Robots and Ambient Intelligence (**URAI**)*, pp. 95-100, Octobre 2015, Goyang, Corée du Sud

CI.6 "Good feature for framing : saliency-based Gaussian mixture", Z. Habibi, E. Mouaddib et G. Caron, *IEEE/RSJ Int. Conf. on Intelligent RObots and Systems (**IROS**)*, pp. 3682 - 3687, Octobre 2015, Hambourg, Allemagne

CI.7 "Photometric Gaussian Mixtures based Visual Servoing", N. Crombez, G. Caron et E. Mouaddib, *IEEE/RSJ Int. Conf. on Intelligent RObots and Systems (**IROS**)*, pp. 5486 - 5491, Octobre 2015, Hambourg, Allemagne

CI.8 "Assistive visual framing in 3D dense Points Cloud", Z. Habibi, G. Caron et E. Mouaddib, *IEEE Digital Heritage Int. Congress (**DigitalHeritage**)*, pp. 109-112, Octobre 2015, Grenade, Espagne

CI.9 "Toward the adaptive and context-aware Serious Game design", D. Leclet-Groux et G. Caron, *IEEE Int. Conf. on Advanced Learning Tech. (**ICALT**)*, poster, pp. 242-243, Juillet 2014, Athènes, Grèce

CI.10 "A Serious Game for 3D Cultural Heritage", D. Leclet-Groux, G. Caron, E. Mouaddib, A. Anghour, *IEEE Digital Heritage Int. Congress (**DigitalHeritage**)*, pp. 409 - 412, Octobre 2013, Marseille, France

CI.11 "Omnidirectional Visual Servoing using the Normalized Mutual Information", B. Delabarre, G. Caron, E. Marchand, *IFAC Symp. on Robot Control (**SYROCO**)*, Vol. 10, No. 1, pp. 102 - 107, Septembre 2012, Dubrovnik, Croatie

CI.12 "Tracking planes in omnidirectional stereovision", G. Caron, E. Marchand et E. Mouaddib, *IEEE International Conference on Robotics and Automation (**ICRA**)*, pp. 6306-6311, Mai 2011, Shanghai, Chine

CI.13 "Multiple camera types simultaneous stereo calibration", G. Caron et D. Eynard, *IEEE International Conference on Robotics and Automation (**ICRA**)*, pp. 2933 - 2938, Mai 2011, Shanghai, Chine

CI.14 "Omnidirectional Photometric Visual Servoing", G. Caron, E. et E. Mouaddib, *IEEE/RSJ International Conference on Intelligent RObots and Systems (**IROS**)*, pp. 6202 - 6207, Octobre 2010, Taipei, Taiwan

CI.15 "Single Viewpoint Stereoscopic Sensor Calibration", G. Caron, E. Marchand et E. Mouaddib, *International Symposium on Image/Video Communications (**ISIVC**)*, pp. 1 - 4, Septembre 2010, Rabat, Maroc

CI.16 "3D Model Based Pose Estimation For Omnidirectional Stereovision", G. Caron, E. Marchand et E. Mouaddib, *IEEE/RSJ International Conference on Intelligent RObots and Systems (**IROS**)*, pp. 5228 - 5233, Octobre 2009, St. Louis, Missouri, USA

CI.17 "Vertical Line Matching for Omnidirectional Stereovision Images", G. Caron et E. Mouaddib, *IEEE International Conference on Robotics and Automation (**ICRA**)*, pp. 2787 - 2792, Mai 2009, Kobe, Japon

Ateliers internationaux (4)

Synthèse : 2 articles avec des doctorants encadrés [**AI.1**, **AI.2**] et 2 articles en collaboration internationale [**AI.3**, **AI.4**], dont 1 en tant que postdoc [**AI.4**].

AI.1 "3D point cloud model colorization by dense registration of digital images", N. Crombez, G. Caron et E. Mouaddib, *3D Virtual Reconstruction and Visualization of Complex Architectures, **3DARCH**, ISPRS workshop*, Février 2015, Avila, Espagne

AI.2 "3D model automatic exploration : Smooth and Intelligent Virtual Camera Control", Z. Habibi, G. Caron et E. Mouaddib, ***e-Heritage**, workshop at ACCV 2014*, Novembre 2014, Singapour

AI.3 "COALAS : A EU Multidisciplinary Research Project for Assistive Robotics Neuro-rehabilitation", N. Ragot, G. Caron, M. Sakel et K. Sirlantzis, *Rehabilitation and Assistive Robotics, workshop at IEEE/RSJ **IROS***, Septembre 2014, Chicago, IL, USA

AI.4 "Evaluation of Model based Tracking with TrakMark Dataset", A. Petit, G. Caron, H. Uchiyama et E. Marchand, *TrakMark, workshop at IEEE Int. Symp. on Mixed and Augmented Reality (**ISMAR**)*, Octobre 2011, Basel, Suisse

Conférences nationales (10)

Synthèse : 5 articles avec des doctorants encadrés [**CN.2**, **CN.3**, **CN.4**, **CN.6**, **CN.7**], 1 article en collaboration internationale [**CN.5**], 1 article avec un collègue enseignant-chercheur [**CN.1**], 1 article en tant que postdoc [**CN.8**] et 2 articles sur les travaux de thèse [**CN.9**, **CN.10**].

CN.1 "Gyroscope visuel sphérique basé mélange de potentiels photométriques", G. Caron, F. Morbidi, ***RFIAP** : Reconnaissance de Formes, Image, Apprentissage et Perception* (AFRIF), Juin 2018, Marne-la-Vallée

CN.2 "Reconstruction 3D d'objet à partir d'une image catadioptrique multi-plan", N. Mohtaram, A. Radgui, G. Caron, E. Mouaddib, D. Aboutajdine, ***ORASIS** : Congrès des jeunes chercheurs en vision par ordinateur* (AFRIF), Juin 2017, Colleville-sur-mer

CN.3 "Asservissement visuel basé mélanges de gaussiennes photométriques", N. Crombez, G. Caron, E. Mouaddib, ***RFIA** : Reconnaissance de Formes et Intelligence Artificielle* (AFRIF), Juin 2016, Clermont-Ferrand

CN.4 "Mélange de gaussiennes basées saillance pour le cadrage visuel", Z. Habibi, E. Mouaddib, G. Caron, ***RFIA** : Reconnaissance de Formes et Intelligence Artificielle* (AFRIF), Juin 2016, Clermont-Ferrand

CN.5 "Estabilización Visual de un Quadrotor", E. Rodriguez, D. Lara, G. Romero Galvan, G. Caron, C. Pegard, ***SOMI** XXX : Congreso de Instrumentacion*, Octobre 2015, Durango, Mexique

CN.6 "Exploration réaliste et pertinente d'un nuage de points 3D dense et coloré", Z. Habibi, G. Caron et E. Mouaddib, ***ORASIS** : Congrès des jeunes chercheurs en vision par ordinateur*, (AFRIF), Juin 2015, Amiens

CN.7 "Colorisation photo-réaliste de nuages de points 3D", N. Crombez, G. Caron et E. Mouaddib, ***ORASIS** : Congrès des jeunes chercheurs en vision par ordinateur*, (AFRIF), Juin 2013, Cluny

CN.8 "L'information mutuelle pour l'estimation visuelle directe de pose", G. Caron, A. Dame et E. Marchand, ***RFIA** : Reconnaissance de Formes et l'Intelligence Artificielle*, (AFRIF), Janvier 2012, Lyon

CN.9 "Etalonnage simultané de systèmes stéréoscopiques hybrides", G. Caron et D. Eynard, ***ORASIS** : Congrès des jeunes chercheurs en vision par ordinateur*, (AFRIF), Juin 2011, Praz-sur-Arly

CN.10 "Mise en Correspondance de Droites Verticales dans les Images de Stéréovision Omnidirectionnelle", G. Caron et E. Mouaddib, ***ORASIS** : Congrès des jeunes chercheurs en vision par ordinateur* (AFRIF), Juin 2009, Trégastel

1.2.3 Autres publications

D.1 "3D model silhouette-based tracking in depth images for puppet suit dynamic video-mapping", G. Caron, M. Belghiti, A. Dessaux, arXiv :1810.03956, Octobre 2018

D.2 "Vision pour la robotique", G. Caron et E. Mouaddib, *Techniques de l'Ingénieur*, Septembre 2014

D.3 "Hybrid stereoscopic calibration", G. Caron et D. Eynard, *SPIE newsroom in Electronic Imaging & Signal Processing*, Juin 2011

1.2.4 Exposés invités

Inv.1 "Perception robotique : Commande et estimation basées vision(s)", *séminaire invité* au laboratoire XLIM de l'Université de Limoges, Mars 2019, France

Inv.2 "Les robots savent se repérer", *conférence flash invitée*, UFR Sciences de l'UPJV, Janvier 2019, Amiens, France

Inv.3 "Reliable planar object pose estimation in light-fields from best sub-aperture camera pairs", *séminaire invité* au NAIST, Novembre 2018, Nara, Japon

Inv.4 "Un élément de genèse de la SFR Numérique et Patrimoine en Hauts-de-France : le programme E-Cathédrale", *Journées SHS-Valo, MESHS de Lille*, Juin 2018, Lille, France

Inv.5 "Perception visuelle omnidirectionnelle en robotique mobile : exploration, localisation et navigation", *Journées Automatique et Automobile, GdR MACS*, Octobre 2017, Amiens, France

Inv.6 "e-Cathedral : On the digital archiving of the largest medieval Gothic church of France", *International multidisciplinary Workshop on Sensing, Reconstruction, and Recognition of Environment at NAIST*, Mars 2017, Nara, Japon

Inv.7 "e-Cathedral : On the digital archiving of the largest medieval Gothic church of France", *ACCV 2016 workshop on e-Heritage*, Novembre 2016, Taipei, Taïwan

Inv.8 "On the interest of virtual environments for actual cameras", séminaire anglophone à l'Optical Media Interface lab (NAIST), Novembre 2016, Nara, Japon

Inv.9 "Good Feature for Framing : Saliency-Based Gaussian Mixture", Z. Habibi, E. Mouaddib et G. Caron, *IROS 2015 workshop Int. Symp. on Attention and Cognitive Systems, ISACS*, Octobre 2015, Hambourg, Allemagne

Inv.10 "Real camera and virtual world interactions", séminaire anglophone pour le laboratoire LE2I et les Master VIBOT et Computer Vision, Décembre 2012, Université de Bourgogne, Le Creusot, France

1.2.5 Communications

Communications internationales

EI.1 "Omnidirectional visual perception in mobile robotics : toward assistive driving of powered wheelchair", G. Caron, *12th International Society of Physical and*

Rehabilitation Medicine World Congress (ISPRM), Juillet 2018, Paris, France

EI.2 "Phase Correlation for Dense Visual Compass from Omnidirectional Camera-Robot Images", F. Morbidi et G. Caron, *International Conference on Robotics and Automation* (IEEE), Mai 2017, Singapour

EI.3 "Omnidirectional photometric visual path following for wheelchair autonomous driving", Y. Alj, G. Caron et N. Ragot, *CareTECH : Healthcare Technology Days*, Décembre 2014, Rouen, France

EI.4 "From heritage building digitization to computerized education", G. Caron, D. Leclet-Groux, N. Crombez et E. Mouaddib, *6th Int. Cong. on Science and Technology for the safeguard of Cultural Heritage in the Mediterranean basin*, Octobre 2013, Athènes, Grèce

EI.5 "E-Cathedral : A multidisciplinary research program dedicated to the Cathedral of Amiens in France", G. Caron, E. Mouaddib, *1st Conference on "Robotics Innovation for Cultural Heritage", RICH*, Décembre 2012, Venise, Italie

Communications nationales

EN.1 "Estimation fiable de la pose d'un objet planaire à partir d'une sélection de paire d'images provenant d'une acquisition plénoptique", N. Crombez, G. Caron, T. Funatomi, Y. Mukaigawa, Journée *Co-conception : capteurs hybrides et algorithmes* (GDR ISIS, CNRS), Octobre 2019, Paris

EN.2 "Cadrage visuel assisté", Z. Habibi, E. Mouaddib, G. Caron, *présentation orale aux Journées de l'AFIG*, Novembre 2015, Lyon

EN.3 "Les mélanges de gaussiennes basées image pour la commande référencée vision", Z. Habibi, N. Crombez, G. Caron, E. Mouaddib, *présentation vidéo orale aux Journées Nationales de la Recherche en Robotique (JNRR)*, Octobre 2015, Saint-Valery-sur-Somme

EN.4 "e-Cathédrale : méthodologie de numérisation de la cathédrale d'Amiens et défis", G. Caron, N. Crombez, E. Mouaddib, *présentation orale aux journées de l'AFIG, Reims Image 2014*, Novembre 2014, Reims

EN.5 "Exploration automatique d'un environnement 3D : contrôle fluide et intelligent de caméra virtuelle", Z. Habibi, G. Caron, E. Mouaddib, *présentation affichée aux journées de l'AFIG, Reims Image 2014*, Novembre 2014, Reims

EN.6 "La stéréovision omnidirectionnelle compacte est-elle plénoptique ?", G. Caron et E. Mouaddib, Journée *Caméra 3D : de la modélisation à l'application* (GDR Robotique, CNRS), Juin 2013, UPMC, Paris

EN.7 "Suivi 3D et estimation de pose en stéréovision omnidirectionnelle", G. Caron, E. Marchand et E. Mouaddib, Journée *Suivi visuel 3D* (GDR ISIS, CNRS), Janvier 2011, UPMC, Paris

EN.8 "Asservissement visuel photométrique en vision omnidirectionnelle", G. Caron, E. Marchand et E. Mouaddib, poster à la *Journée des Jeunes Chercheurs en Robotique (JJCR)*, Novembre 2010, UPMC, Paris

EN.9 "Asservissement visuel en vision omnidirectionnelle", G. Caron, E. Mouaddib et E. Marchand, Journée de rencontre des chercheurs en image, vision et reconnaissance de forme des régions Nord-Pas-De-Calais et Picardie (Telecom Lille 1), Juin 2010, Lille

EN.10 "Asservissement visuel photométrique en vision omnidirectionnelle", G. Caron, E. Marchand, E. Mouaddib, Journée *Perception et Commande référencés capteurs non-conventionnels* (GDR Robotique, CNRS), Avril 2010, ENST, Paris

EN.11 "Estimation non-linéaire de pose d'objet 3D par stéréovision omnidirectionnelle", G. Caron, E. Marchand et E. Mouaddib, poster aux *Journées Nationales de la Recherche en Robotique (**JNRR** et **JJCR**)*, Novembre 2009, Neuvy/Barangeon

1.2.6 Logiciels et jeux de données

Log.1 HySCaS : Hybrid Stereoscopic Calibration Software
Etalonnage de bancs stéréo hybrides – Des bancs stéréo à différents modèles de caméras sont étalonnés par ce logiciel multi-plateforme et développé avec les librairies Qt, ViSP et OpenCV.
Implantation de **CI.13** : mis.u-picardie.fr/~g-caron/software.

Log.2 libPeR : librairie de l'équipe Perception Robotique
Rassemblement des développements logiciels associés aux travaux de l'équipe PR au sein d'une seule librairie interne écrite en langage C++.

Log.3 OVMIS : Omnidirectional Vision dataset of the MIS lab
Plusieurs jeux d'images omnidirectionnelles (hyper catadioptriques) permettant d'évaluer l'estimation de mouvement de bras manipulateur et de robot mobile. Créés pour [**RI.5**] (2,7 Go) : mis.u-picardie.fr/~g-caron/datasets.

Log.4 SVMIS : Spherical Vision dataset of the MIS lab
Plusieurs jeux d'images spheriques (fisheye polydioptiques) permettant d'évaluer l'estimation de mouvement de bras manipulateur et de drone. Créés pour [**CI.3**, **CN.1**] (1,6 Go) : mis.u-picardie.fr/~g-caron/datasets.

Log.5 LFMIS : Light Field dataset of the MIS lab
Plusieurs jeux d'images plénoptiques (grille de micro-lentilles) permettant d'évaluer l'estimation de mouvement d'objet plan.
Créés pour [**RI.3**] (2,2 Go) : mis.u-picardie.fr/~g-caron/datasets.

Log.6 AFAMIS : Template registration dataset of the MIS lab
110 000 images permettant d'évaluer le suivi de région d'image selon un modèle de mouvement de translation pure ou d'homographie. Créées pour [**RI.1**] (520 Mo) à partir des jeux de données MS-COCO et Yale faces :
mis.u-picardie.fr/~g-caron/datasets.

Chapitre 2

Modélisation de la formation des images

Sommaire

2.1 Formation géométrique des images 35
2.1.1 Modèle de projection perspective 35
2.1.2 Modèle de projection centrale unifié 38
2.1.3 Modèle de projection centrale unifié à deux plans images 40
2.2 Représentations des transformations géométriques 42
2.3 Formation photométrique des images 42

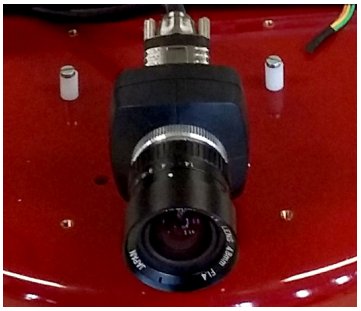
Les travaux de recherches rassemblés dans ce mémoire considèrent des types de caméras relativement variés. Ils nécessitent des modèles géométriques de projection adaptés à chaque type : le classique modèle de projection perspective pour les caméras usuelles, le modèle de projection centrale unifié pour les caméras panoramiques à point de vue unique et certaines caméras fisheyes, et une forme d'extension de ce dernier pour les caméras polydioptriques sphériques. Ces modèles sont décrits dans la partie 2.1, suivie de la partie 2.2 qui apporte quelques rappels sur les outils mathématiques pour représenter les transformations géométriques les plus communes dans l'espace.

D'autre part, ces caméras réalisent une image en collectant la luminance de la scène dans laquelle elles se trouvent. Dans l'état de l'art en vision robotique directe, certaines méthodes exploitent un modèle de la formation photométrique des images qui brièvement rappelé en partie 2.3.

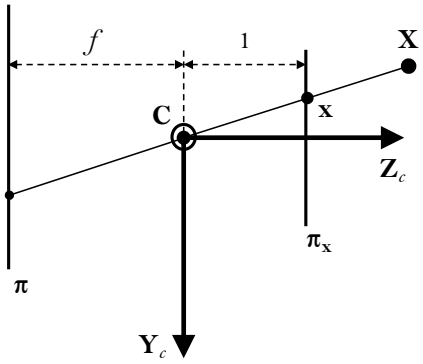
2.1 Formation géométrique des images

2.1.1 Modèle de projection perspective

Décrivant la formation géométrique d'une image (Figure 2.1(c)), observation d'une scène par un sténopé ou une caméra dont l'optique se rapporte à une "lentille mince" (Fig. 2.1(a)), le modèle de projection perspective, très connu, fait intervenir un centre de projection $\mathbf{C} \in \mathbb{R}^3$ et le plan de l'image $\boldsymbol{\pi}$ (Figure 2.1(b)). L'image étant un rectangle fini de $\boldsymbol{\pi}$, elle définit une section de la pyramide de sommet $\mathbf{C}$ et de hauteur infinie caractérisant le champ de vue de la caméra. Plus la distance



(a)



(b)

(c)

FIGURE 2.1 – Illustration de la projection perspective. (a) Caméra perspective. (b) Schéma de la projection perspective d'un point du monde 3D dans le plan image normalisé. (b) Exemple d'image acquise par une caméra caractérisable par le modèle de projection perspective à partir du triforium de la cathédrale d'Amiens : malgré l'exceptionnelle régularité de l'édifice, la largeur du vaisseau central dans l'image diminue du bas de l'image vers son centre alors qu'elle est sensiblement la même sur toute la longueur de l'édifice.

focale $f \in \mathbb{R}_+^*$ entre le centre optique et le plan image est grande, moins le champ de vue est ample, et inversement.

On définit le repère caméra $\mathcal{F}_c$ d'origine $\mathbf{C}$ et d'axes $\mathbf{X}_c \in \mathbb{R}^3$, tel que $||\mathbf{X}_c|| = 1$, et $\mathbf{Y}_c \in \mathbb{R}^3$, tel que $||\mathbf{Y}_c|| = 1$, parallèles au bords horizontal et, respectivement, vertical du rectangle image. Par convention usuelle, le sens de ces axes suit celui de l'organisation des pixels d'une image numérique acquise par une caméra, c'est-à-dire "vers la droite" pour $\mathbf{X}_c$ et "vers le bas" pour $\mathbf{Y}_c$. $\mathbf{Z}_c \in \mathbb{R}^3$, tel que $||\mathbf{Z}_c|| = 1$, est orthogonal aux précédents axes de telle sorte que $\mathcal{F}_c$ soit un repère direct, dont l'orientation est choisie telle que $\mathbf{Z}_c$ pointe vers l'avant de la caméra. Ce choix fait que les points tridimensionnels (3D) ${}^c\mathbf{X} = \begin{pmatrix} {}^cX & {}^cY & {}^cZ \end{pmatrix}^\mathsf{T} \in \mathbb{R}^3$ de la scène observée par la caméra, exprimés dans le repère caméra $\mathcal{F}_c$, aient leur troisième coordonnée positive.

Le modèle de projection perspective exprime un point image par l'intersection de la droite de vue $({}^c\mathbf{C}{}^c\mathbf{X})$, représentant le chemin suivi par un rayon lumineux, avec le plan image. Si, physiquement, le plan image est au-delà du centre optique par rapport au point 3D, tout plan de l'espace qui lui est parallèle peut former un plan image virtuel, identique au plan image réel, à une transformation de similitude près. Le plan image normalisé $\boldsymbol{\pi}_\mathbf{x}$, c'est-à-dire distant d'une unité de $\mathbf{C}$, du côté de ${}^c\mathbf{X}$, est généralement choisi en vision par ordinateur et aussi dans l'ensemble de ce document.

Par conséquent, le modèle de projection perspective exprime la projection d'un point 3D ${}^c\mathbf{X}$ dans $\boldsymbol{\pi}_\mathbf{x}$ en $\mathbf{x} = \begin{pmatrix} x & y \end{pmatrix}^\mathsf{T} \in \mathbb{R}^2$ par :

$$\begin{cases} x = \frac{{}^cX}{{}^cZ} \\ y = \frac{{}^cY}{{}^cZ} \end{cases} . \tag{2.1}$$

$\mathbf{x}$ est défini dans le plan image normalisé dont l'origine est sa propre intersection avec l'axe $\mathbf{Z}_c$. Il est à noter que ${}^c\mathbf{x} = \begin{pmatrix} x & y & 1 \end{pmatrix}^\mathsf{T} \in \mathbb{R}^3$ donne la direction de la droite de vue, exprimée dans $\mathcal{F}_c$ et passant par son origine. On le distingue de $\tilde{\mathbf{x}} = \begin{pmatrix} x & y & 1 \end{pmatrix}^\mathsf{T} \in \mathbb{P}^2$, la représentation homogène de $\mathbf{x}$ dans $\boldsymbol{\pi}_\mathbf{x}$. On écrit alors la fonction de projection perspective $pr()$:

$$\tilde{\mathbf{x}} = pr({}^c\mathbf{X}), \tag{2.2}$$

avec x et y exprimés comme en (2.1).

L'origine de l'image numérique acquise par une caméra étant en haut à gauche et son échantillonnage étant en pixels, le modèle de projection perspective considère une transformation supplémentaire, une transformation affine $\mathbf{K} \in \mathrm{Aff}(2)$, pour transformer $\boldsymbol{\pi}_\mathbf{x}$ vers le plan image numérique $\boldsymbol{\pi}_\mathbf{u}$. Cette transformation fait intervenir, généralement, quatre paramètres $\gamma_p = \{\alpha_u, \alpha_v, u_0, v_0\}$ dont $\alpha_u \in \mathbb{R}$ et $\alpha_v \in \mathbb{R}$ sont les facteurs d'échelle horizontal et, respectivement, vertical et $(u_0, v_0) \in \mathbb{R}^2$ sont les coordonnées du point principal, c'est-à-dire l'intersection de $\boldsymbol{\pi}_\mathbf{x}$ et $\mathbf{Z}_c$, exprimée dans l'image numérique. Ces paramètres, dits intrinsèques, caractérisent l'optique de la caméra, selon le modèle de projection perspective, et sont liés à la réalisation physique d'une image, notamment $\alpha_u = f/k_u$ et $\alpha_v = f/k_v$, avec $k_u \in \mathbb{R}_+^*$ et $k_v \in \mathbb{R}_+^*$ les dimensions d'une photodiode donnant un pixel dans l'image numérique [1]. Ainsi, le point $\tilde{\mathbf{u}} = \begin{pmatrix} u & v & 1 \end{pmatrix}^\mathsf{T} \in \mathbb{P}^2$ de l'image numérique est obtenu à partir de $\tilde{\mathbf{x}}$ par :

$$\tilde{\mathbf{u}} = \mathbf{K}\tilde{\mathbf{x}} \text{ avec } \mathbf{K} = \begin{pmatrix} \alpha_u & 0 & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} . \tag{2.3}$$

En rassemblant les deux étapes, on obtient la fonction de projection d'un point 3D ${}^c\mathbf{X}$ dans le plan image numérique :

$$\tilde{\mathbf{u}} = pr_{\gamma_p}({}^c\mathbf{X}) = \mathbf{K}pr({}^c\mathbf{X}). \tag{2.4}$$

1. L'expression du modèle physique du cheminement de la lumière vers le plan image réel et les développements montrant l'équivalence avec le modèle mathématique décrit dans cette sous-partie ont été volontairement omis par souci d'aller à l'essentiel sur ce sujet car très connus.

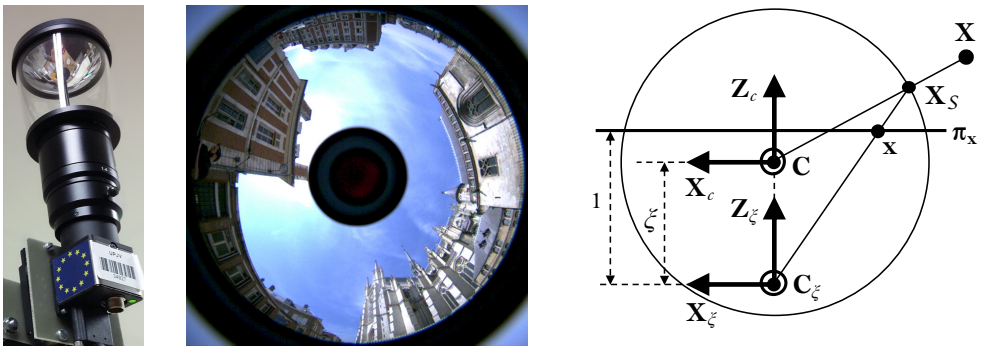
Remarque 1 (Enrichissement du modèle de projection perspective) *Le modèle de projection ci-dessus prend en compte le fait que les pixels puissent ne pas être parfaitement carrés mais rectangulaires. Il peut être simplifié au minimum avec un seul facteur d'échelle. Ce modèle peut aussi être enrichi pour prendre en compte : le fait que les pixels sont des parallélogrammes, les distorsions radiales ou tangentielles engendrées par l'optique utilisée, un défaut d'alignement de l'optique et de la matrice de photodiodes de la caméra, etc.* ◇

2.1.2 Modèle de projection centrale unifié

Le champ de vue des caméras pouvant suivre le modèle de projection perspective est limité par l'optique des objectifs existants, jusqu'aux alentours de 122° sans distorsion [2]. Au delà, les objectifs emploient des lentilles fisheye qui permettent d'obtenir des champs de vue de 180° et plus (jusqu'à 220°) ou des miroirs incurvés (Fig. 2.2(a)), au prix de fortes distorsions géométriques dans les images (Fig. 2.2(b)). Les associations de miroir hyperbolique de révolution et caméra perspective (caméra hypercatadioptrique), miroir parabolique de révolution avec caméra orthographique (caméra paracatadioptrique), pour les miroirs convexes, et l'association de miroir concave elliptique de révolution et de caméra perspective, engendrent une caméra panoramique à point de vue unique [Baker and Nayar, 1999] [3]. Pour ces caméras, le modèle de projection centrale unifié [Geyer and Daniilidis, 2000, Barreto and Araujo, 2001] est adapté. Il est aussi une bonne approximation de la projection réalisée par certaines caméras à objectif fisheye [Ying and Hu, 2004].

2. Distorsions inférieures à 1% selon `www.dxomark.com` au 1er mars 2017 pour l'objectif Sigma 12-24mm pour appareil photographique

3. Il existe aussi des combinaisons de réflexions convexe-concave qui respectent le point de vue unique.



(a) (b) (c)

FIGURE 2.2 – Vision panoramique. (a) Caméra hypercatadioptrique (Objectif V-Stone VS-C450MR-TK). (b) Image omnidirectionnelle obtenue avec un objectif catadioptrique (Place Saint-Michel, Amiens, 2015). (c) Schéma du modèle de projection centrale unifié.

Le modèle de projection centrale unifié (Fig. 2.2(c)) peut être vu comme une généralisation du modèle de projection perspective consistant à ajouter une étape préliminaire qui projette d'abord le point 3D ${}^c\mathbf{X}$ sur une sphère unitaire de centre $\mathbf{C}$ en $\mathbf{X}_\mathcal{S} \in \mathbb{R}^3$, tel que $||\mathbf{X}_\mathcal{S}|| = 1$:

$$\mathbf{X}_\mathcal{S} = \begin{pmatrix} X_\mathcal{S} \\ Y_\mathcal{S} \\ Z_\mathcal{S} \end{pmatrix} = pr_\mathcal{S}({}^c\mathbf{X}) \quad \text{avec} \quad \begin{cases} X_\mathcal{S} = \frac{{}^cX}{\rho} \\ Y_\mathcal{S} = \frac{{}^cY}{\rho} \\ Z_\mathcal{S} = \frac{{}^cZ}{\rho} \end{cases}, \ \rho = ||{}^c\mathbf{X}||, \tag{2.5}$$

avant de projeter ce dernier sur le plan image normalisé en $\mathbf{x}$, à l'aide d'un deuxième centre de projection $\mathbf{C}_\xi \in \mathbb{R}^3$, distant de $\xi \in \mathbb{R}_+$ du premier, selon l'axe $\mathbf{Z}_c$ [Barreto and Araujo, 2001] :

$$\mathbf{x} = pr\left(\mathbf{X}_\mathcal{S} + \begin{pmatrix} 0 & 0 & \xi \end{pmatrix}^\mathsf{T}\right) \quad \text{avec} \quad \begin{cases} x = \frac{X_\mathcal{S}}{Z_\mathcal{S}+\xi} \\ y = \frac{Y_\mathcal{S}}{Z_\mathcal{S}+\xi} \end{cases}. \tag{2.6}$$

$pr_\mathcal{S}()$ et $pr()$ se rassemblent en une seule fonction de projection pr_ξ du point 3D ${}^c\mathbf{X}$ en $\tilde{\mathbf{x}}$:

$$\tilde{\mathbf{x}} = pr_\xi({}^c\mathbf{X}) \quad \text{avec} \quad \begin{cases} x = \frac{{}^cX}{{}^cZ+\xi\rho} \\ y = \frac{{}^cY}{{}^cZ+\xi\rho} \end{cases}. \tag{2.7}$$

Pour finaliser le modèle de projection centrale unifié, la transformation du plan image normalisé au plan image numérique se fait de la même manière qu'avec le modèle de projection perspective (2.3). ξ rejoint l'ensemble des paramètres intrinsèques du modèle de projection centrale unifié $\gamma_u = \{\alpha_u, \alpha_v, u_0, v_0, \xi\}$. Ainsi, la fonction de projection d'un point 3D ${}^c\mathbf{X}$ dans le plan image numérique s'écrit :

$$\tilde{\mathbf{u}} = pr_{\gamma_u}({}^c\mathbf{X}) = \mathbf{K}pr_\xi({}^c\mathbf{X}). \tag{2.8}$$

La projection $pr_\xi()$ de la sphère vers le plan image est inversible, ce qui permet de retrouver un point sphérique, donc la droite de vue associée, à partir d'un point image :

$$\mathbf{X}_\mathcal{S} = pr_\xi^{-1}(\mathbf{x}) = \begin{pmatrix} \frac{\xi+\sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1}x \\ \frac{\xi+\sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1}y \\ \frac{\xi+\sqrt{1+(1-\xi^2)(x^2+y^2)}}{x^2+y^2+1} - \xi \end{pmatrix}. \tag{2.9}$$

Enfin, représenter les coordonnées de points sur la sphère par des coordonnées cartésiennes est redondant car, puisque la sphère est unitaire, $X_\mathcal{S}$, $Y_\mathcal{S}$ et $Z_\mathcal{S}$ ne sont pas linéairement indépendants ($||\mathbf{X}_\mathcal{S}|| = 1$). La représentation minimale d'un point sur la sphère se fait par les coordonnées sphériques d'azimuth ϕ et d'élévation θ ($\boldsymbol{\phi} = [\phi, \theta]^\top \in \mathbb{R}^2$), qui s'expriment à partir de $\mathbf{X}_\mathcal{S}$ par la fonction $c2s()$ (cartésien vers sphérique) :

$$\boldsymbol{\phi} = \begin{pmatrix} \phi \\ \theta \end{pmatrix} = c2s(\mathbf{X}_\mathcal{S}) = \begin{pmatrix} \arccos(Z_\mathcal{S}) \\ \arctan(Y_\mathcal{S}/X_\mathcal{S}) \end{pmatrix}. \tag{2.10}$$

Remarque 2 (Enrichissement du modèle de projection centrale unifié) *Tout comme le modèle de projection perspective, le modèle de projection centrale unifié peut être étendu en prenant en compte des paramètres intrinsèques supplémentaires, par exemple pour des distorsions radiales (Remarque 1).* ◇

Remarque 3 (Généralisation du modèle de projection perspective) *Ce modèle de projection unifié est valide pour toute caméra à point de vue unique, y compris les caméras perspectives. En effet, il suffit d'annuler ξ pour retrouver une projection perspective seule.* ◇

2.1.3 Modèle de projection centrale unifié à deux plans images

Pour atteindre un champ de vue de 360°, plusieurs objectifs ou plusieurs caméras sont nécessaires. Ce type de capteur visuel polydioptrique va de la combinaison de plusieurs caméras perspectives réparties sur la surface d'une sphère [Swaminathan and Nayar, 1999] (polycaméra) à la combinaison de deux lentilles fisheyes dos-à-dos, pour les plus compacts [Li, 2006]. Quelques produits sont disponibles sur le marché des professionnels de la photographie et du film 360 (ou de *réalité virtuelle*), pour le premier (ex : Insta360Pro), et de nombreux produits grand public (ex : Ricoh Theta, Samsung Gear 360, Garmin Virb 360, etc), pour le second.

Généralement, un modèle de projection caractérise chaque caméra, donc il y a autant de jeux de paramètres intrinsèques $\gamma_{m,j}$ que de caméras, m désignant le modèle de projection considéré et j, l'index de la caméra. A ceux-là s'ajoutent les paramètres extrinsèques qui expriment la pose $\mathbf{p}_{s,j} \in \mathfrak{se}(3)$, par abus de notation [Ma et al., 2004], de chaque caméra j, de repère $\mathcal{F}_{c_j}$, dans un repère commun associé au système polydioptrique, $\mathcal{F}_s$.

Cette modélisation de système polydioptrique est très similaire à celle des systèmes de vision stéréoscopique, ou multi-caméra en général. Cependant, les systèmes polydioptriques sphériques les plus compacts ont la spécificité d'être conçus pour que leurs objectifs aient des champs de vue complémentaires, réduisant ainsi leur nombre au minimum. Deux objectifs fisheye placés dos-à-dos suffisent à couvrir les 360° du champ de vue sphérique complet grâce à deux miroirs plans judicieusement positionnés entre les objectifs, reflétant ainsi la lumière vers une seule matrice photosensible [Li, 2006]. Pour encore plus de compacité, le fabricant de caméras Ricoh a remplacé, dans sa gamme Theta[4], les miroirs par deux prismes réorientant les rayons lumineux traversant les objectifs fisheye vers deux matrices photosensibles. Au delà de l'intérêt du très faible volume de cette dernière caméra polydioptrique sphérique (Fig. 2.3(a)), pour le grand public, comme en robotique, la proximité des deux objectifs fisheye rend aussi très proches leurs centres optiques. On peut alors faire l'approximation qu'ils sont confondus, en particulier quand les éléments de la scène observée sont suffisamment éloignés de la caméra sphérique[5] (distance de travail minimum des polycaméras [Swaminathan and Nayar, 1999]).

4. theta360.com
5. La gamme Theta de Ricoh est ici évoquée en particulier à titre d'exemple car elle possède, à

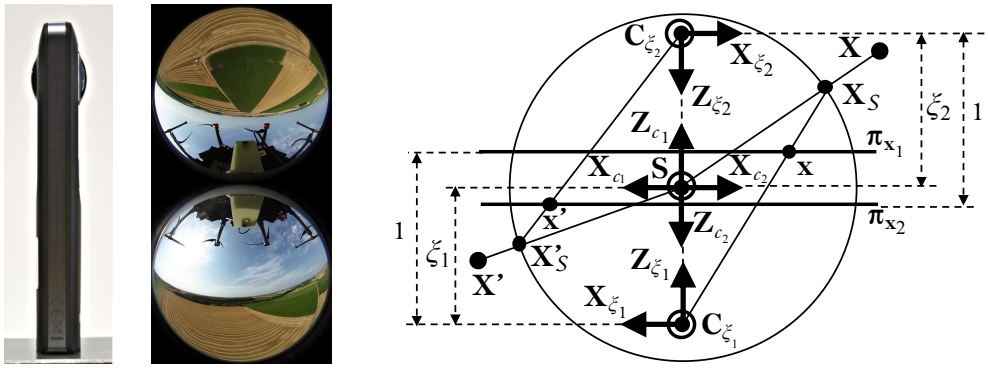


FIGURE 2.3 – Vision sphérique : (a) Caméra sphérique polydioptrique compacte (Ricoh Theta S, vue de profil avec les deux lentilles fisheye dans sa partie supérieure) acquérant des (b) images double-fisheye (Le champ à cailloux, Vaux-en-Amiénois, 2019) et dont la projection peut se représenter, sous hypothèses, par (c) une extension du modèle de projection centrale unifié à deux plans images.

Le modèle de projection de caméra sphérique polydioptrique compacte se restreint ainsi à une seule sphère dont chaque hémisphère est associé à l'une des deux images fisheye (Fig. 2.3(b)). Les lentilles, les matrices photosensibles et leur alignement pouvant être légèrement différents d'une caméra fisheye à l'autre, deux jeux de paramètres intrinsèques $\gamma_{u,j}, j \in \{1,2\}$ sont considérés. Cependant, puisque les deux caméras fisheye sont supposées partager la même origine, on peut fixer $\mathcal{F}_s = \mathcal{F}_{c_1}$ (Fig. 2.3(c)) et les paramètres extrinsèques, c'est-à-dire la pose de la deuxième caméra fisheye, relativement à la première, donc, se réduisent à l'orientation $\mathbf{r}_{s,2}$ (ou $\mathbf{r}_{1,2}$) $\in \mathfrak{so}(3)$, par abus de notation, représentation axe-angle de la matrice ${}^{c_2}\mathbf{R}_s = {}^{c_2}\mathbf{R}_{c_1} \in SO(3)$ [Caron and Morbidi, 2018]. On ré-exprime alors la projection d'un point sphérique ${}^{s}\mathbf{X}_{\mathcal{S}}$ de l'hémisphère associé à la caméra j dans le plan image normalisé de cette dernière (Eq. (2.6)) par :

$$\mathbf{x} = pr_j\left({}^{c_j}\mathbf{R}_s\,{}^{s}\mathbf{X}_{\mathcal{S}} + \begin{pmatrix}0 & 0 & \xi_j\end{pmatrix}^{\mathsf{T}}\right), \tag{2.11}$$

avec ${}^{c_1}\mathbf{R}_s = {}^{c_1}\mathbf{R}_{c_1} = \mathbf{I}_{3\times 3}$. En pratique, le signe de la troisième coordonnée de ${}^{c_j}\mathbf{X}_{\mathcal{S}} = {}^{c_j}\mathbf{R}_s\,{}^{s}\mathbf{X}_{\mathcal{S}}$ suffit à déterminer par laquelle des deux caméras il est visible.

l'heure actuelle, la plus faible distance entre ses lentilles fisheyes, par rapport à ses concurrentes. C'est donc la caméra sphérique pour laquelle l'approximation d'unicité de centre optique pour les deux objectifs fisheye est la plus tolérable.

2.2 Représentations des transformations géométriques

Dans la partie 2.1.3, écrire $\mathbf{r}_{1,2} \in \mathfrak{so}(3)$ et $\mathbf{p}_{s,j} \in \mathfrak{se}(3)$ sont des abus de notations. En effet, plus rigoureusement, $\mathbf{r}_{1,2} \in \mathbb{R}^3$ et $\mathbf{p}_{s,j} \in \mathbb{R}^6$ et ce sont leurs algèbres de Lie qui appartiennent, respectivement, à $\mathfrak{so}(3)$ et à $\mathfrak{se}(3)$ [Ma et al., 2004].

Néanmoins, cet abus de notation est fréquemment utilisé dans la littérature, tout comme dans ce mémoire, pour indiquer, avec une notation *succincte*, que $\mathbf{r}_{1,2} \in \mathbb{R}^3$ est la représentation axe-angle de la matrice de rotation ${}^{c_2}\mathbf{R}_{c_1} \in SO(3)$, groupe des rotations pures dans l'espace. Cette dernière s'exprime explicitement à partir de $\mathbf{r}_{1,2} = \theta\mathbf{w}$, tel que $\theta \in \mathbb{R}$ et $\mathbf{w} \in \mathbb{R}^3 : ||\mathbf{w}|| = 1$, par l'une des formules de Rodrigues [Ma et al., 2004] :

$$ {}^{c_2}\mathbf{R}_{c_1} = \cos\theta\, \mathbf{I}_{3\times 3} + (1 - \cos\theta)\mathbf{w}\mathbf{w}^\top + \sin\theta\mathbf{w}_\times. \tag{2.12} $$

Ces abus de notations permettent aussi d'indiquer, succinctement, que $\mathbf{p}_{s,j} \in \mathbb{R}^6$ est une représentation vectorielle de la matrice de transformation rigide ${}^{c_j}\mathbf{M}_s \in SE(3)$, groupe des transformations rigides dans l'espace. En considérant $\mathbf{p}_{s,j} = \left[{}^{c_j}\mathbf{t}_s^\top \;\middle|\; \mathbf{r}_{s,j}^\top\right]^\top$, tel que ${}^{c_j}\mathbf{t}_s \in \mathbb{R}^3$ et $\mathbf{r}_{s,j} \in \mathfrak{so}(3)$, par abus de notation, donc, on note ${}^{c_j}\mathbf{R}_s \in SO(3)$, la matrice de rotation exprimée à partir de $\mathbf{r}_{s,j}$ (Eq. (2.12)) et :

$$ {}^{c_j}\mathbf{M}_s = \begin{bmatrix} {}^{c_j}\mathbf{R}_s & {}^{c_j}\mathbf{t}_s \\ \mathbf{0}_{1\times 3} & 1 \end{bmatrix}. \tag{2.13} $$

Enfin, les relations ci-dessus sont inversibles, c'est à dire que l'on peut calculer explicitement $\mathbf{r}_{1,2}$ à partir de ${}^{c_2}\mathbf{R}_{c_1}$ ainsi que $\mathbf{p}_{s,j}$ à partir de ${}^{c_j}\mathbf{M}_s$ [Ma et al., 2004].

Remarque 4 (Généralité des expressions des transformations) *Ci-dessus, les repères indiqués pour les transformations ne sont que des exemples associés à la partie 2.1.3 et, bien entendu, les expressions de cette partie 2.2 sont valables pour tous repères entre lesquels une transformation rigide est possible.* ◇

2.3 Formation photométrique des images

Le modèle de Bergmann, Wang et Cremers [Bergmann et al., 2018] est ici repris pour mieux s'y référer dans le reste du document (Partie 3, notamment).

Tout point $\mathbf{X}$ d'une scène reflète la lumière d'une source dans toutes les directions. La quantité de lumière reflétée est la luminance énergétique $L(\mathbf{X}) \in \mathbb{R}_+$. Si cette luminance est reçue par un capteur en mouvement de la même façon, quelque soit son angle de vue, $\mathbf{X}$ est un point d'une surface lambertienne. $L(\mathbf{X})$ est captée par un photo-site d'une caméra. L'énergie totale reçue en $\mathbf{x}$ sur le plan image réel est l'éclairement énergétique, $E(\mathbf{x}) \in \mathbb{R}_+$. Au sein d'une même matrice de photo-sites (ou matrice photosensible) d'une caméra, selon la position de $\mathbf{x}$ sur le plan image, $E(\mathbf{x})$ peut varier, même pour $L(\mathbf{X})$ constante, particulièrement vers les bords de la matrice photosensible. Cet effet, le vignettage, a plusieurs sources comme le blocage partiel des rayons lumineux traversant l'objectif ou la géométrie

des lentilles [Kim and Pollefeys, 2008]. L'éclairement de $E(\mathbf{x})$ se calcule en multipliant $L(\mathbf{X})$, par un facteur de vignettage $V(\mathbf{x}) \in [0,1]$:

$$E(\mathbf{x}) = V(\mathbf{x})L(\mathbf{X}), \text{ tel que } \mathbf{x} = pr_{.}(\mathbf{X}), \tag{2.14}$$

$pr_{.}$ pouvant être la fonction de projection perspective $pr()$ (Eq. (2.2)) ou la fonction de projection centrale $pr_{\xi}()$ (Eq. (2.7)). $V(\mathbf{x})$ est une fonction non linéaire d'atténuation qui peut, pour certaines lentilles d'objectif de caméra, être approximée par un modèle polynomial radial [Goldmannan, 2010].

Pour obtenir une image, l'éclairement de la matrice photosensible est intégré pendant le temps d'exposition $\delta t_e \in \mathbb{R}_+$. En supposant l'éclairement constant pendant ce lapse de temps, on exprime l'éclairement accumulé par :

$$E_{acc}(\mathbf{x}) = \delta t_e E(\mathbf{x}). \tag{2.15}$$

La fonction $f : \mathbb{R} \to [\![0, 255]\!]$ de réponse de la caméra [Grossberg and Nayar, 2003], transforme ensuite $E_{acc}(\mathbf{x})$ en intensité dans l'image numérique (l'intervalle du codage des intensités est classiquement fait sur 8 bits mais si ce n'est pas le cas, f doit le prendre en compte). Alors, et à l'aide des équations (2.14) et (2.15), on exprime l'intensité $I(\mathbf{u})$ d'un pixel $\mathbf{u}$ à partir de la luminance $L(\mathbf{X})$ du point 3D $\mathbf{X}$ s'y projetant :

$$I(\mathbf{u}) = f\left(\delta t_e V(\mathbf{x})L(\mathbf{X})\right). \tag{2.16}$$

Chapitre 3

Etat de l'art en vision robotique directe

Sommaire

3.1 Introduction . . . 46
3.2 Fondamentaux . . . 47
3.2.1 Fondements . . . 47
3.2.2 Méthodes d'optimisation usuelles . . . 48
3.2.3 Mise à jour des degrés de liberté . . . 51
3.2.4 Méthode d'optimisation efficace au second ordre . . . 55
3.2.5 Conclusion partielle . . . 56
3.3 Approches directes pures . . . 57
3.3.1 Quand des connaissances supplémentaires sont disponibles . . . 57
3.3.2 En faisant des hypothèses sur la scène ou le mouvement . . . 58
3.3.2.1 Translation pure dans le plan image . . . 58
3.3.2.2 Mouvement affine dans l'image . . . 60
3.3.2.3 Mouvement projectif dans l'image . . . 61
3.3.2.4 Asservissement visuel photométrique . . . 68
3.3.3 Localisation et cartographie simultanées . . . 70
3.3.4 Approches directes robustes . . . 71
3.3.4.1 Modélisation statistique . . . 71
3.3.4.2 Modélisation de la variation de l'éclairage . . . 72
3.3.4.3 En présence de mouvement rapide . . . 78
3.3.4.4 Quand la résolution se dégrade . . . 79
3.4 Autres approches directes . . . 80
3.4.1 Critère de corrélation croisée . . . 80
3.4.2 Réseaux de neurones convolutionnels . . . 82
3.5 Approches directes étendues . . . 85
3.5.1 Distribution d'intensité . . . 85
3.5.1.1 Somme des variances conditionnelles . . . 85
3.5.1.2 Information mutuelle . . . 87
3.5.1.3 Ecart d'information normalisé . . . 89
3.5.1.4 Conclusion partielle . . . 89
3.5.2 Champs de descripteurs . . . 90
3.5.3 Approches basées noyaux . . . 92
3.5.3.1 Distributions d'intensité pondérée . . . 92
3.5.3.2 Noyaux photométriques . . . 93
3.5.3.3 Moments photométriques . . . 94
3.5.4 Espace d'échelle . . . 95
3.6 Synthèse . . . 96

3.1 Introduction

La vision robotique peut se définir comme étant l'ensemble des travaux traitants de la vision pour développer l'autonomie en robotique, à savoir le calcul de pose, la localisation, la cartographie, l'odométrie visuelle, le suivi et l'asservissement visuel. L'intersection avec la vision par ordinateur et la réalité augmentée est très importante et bon nombre d'approches de la première (ex : estimation de structure à partir du mouvement), comme de la seconde (ex : suivi d'objet marqué ou non), sont applicables dans un contexte robotique.

Ces trois domaines connexes ont contribué à l'estimation de géométrie, de mouvement et à la commande de système dynamique, largement en exploitant de l'information visuelle indirecte : les méthodes basées primitives géométriques. Une caméra ne mesurant pas directement des coordonnées de points, de droites ou d'autres éléments géométriques, le traitement d'image est nécessaire pour en extraire et alimenter l'estimateur ou la loi de commande. La vision robotique directe peut alors se définir en opposition à ces méthodes, c'est-à-dire qu'elle considère l'information directement mesurée par une caméra pour estimer des paramètres ou commander un robot : les intensités des pixels.

Primitives directes ou indirectes, dans les deux cas, ces problèmes d'estimation ou de commande s'appuient sur l'optimisation d'une fonction de coût $\mathscr{C}()$ calculée à partir de mesures visuelles $\mathcal{V}$ dépendant des degrés de liberté $\mathcal{D}$ du système :

$$\widehat{\mathcal{D}} = \arg\min_{\mathcal{D}} \mathscr{C}(\mathcal{D}, \mathcal{V}, \mathcal{P}, \mathcal{T}) \,. \tag{3.1}$$

Dans l'équation (3.1), $\mathcal{P}$ représente un ensemble de paramètres constants, généralement ceux de l'algorithme résolvant le problème, et $\mathcal{T}$ représente la référence ou la consigne (la tâche).

La suite de ce chapitre s'attache à présenter l'ensemble des méthodes de vision robotique directe à partir de l'écriture très générale du problème de l'équation (3.1).

Cet état de l'art se concentre sur les approches directes considérant les caméras passives et monoculaires, ou s'y apparentant, observant des objets ou des environnements rigides et non-articulés. D'autre part, le contexte robotique engendre naturellement une sélection des travaux de l'état de l'art étant temps-réel, ou proches de l'être. En s'appuyant sur les notions fondamentales du domaine ré-écrites sous un formalisme unique en partie 3.2, trois parties structurent l'état de l'art, aussi reformulé avec le même formalisme : d'abord les approches directes "pures", qui reposent sur la modélisation d'un coût pour un estimateur ou une loi de commande en considèrant les intensités des pixels des images directement (Partie 3.3) ; ensuite, une famille d'approches qui considère aussi les intensités des pixels directement, mais en modélisant leur comparaison par une vraisemblance, ou repoussant leur modélisation dans des approches basées apprentissage (Partie 3.4) ; enfin, les approches directes "étendues" reposant sur une transformée des intensités des pixels (Partie 3.5).

Ce chapitre se termine par une conclusion (Partie 3.6) synthétisant une structuration de tous les travaux rapportés et sur lesquels s'appuient les contributions des chapitres suivants.

3.2 Fondamentaux

3.2.1 Fondements

Les approches directes "pures" reposent sur l'hypothèse de conservation de l'intensité lumineuse, c'est-à-dire que l'observation d'un point 3D $\mathbf{X} \in \mathbb{R}^3$ de la scène par une caméra garde la même intensité $I(\mathbf{u}(\mathbf{p}, \mathbf{X}))$, quelle que soit la pose de vue $\mathbf{p} \in \mathfrak{se}(3)$ de la caméra et donc la position $\mathbf{u} \in \mathbb{R}^2$ du projeté de $\mathbf{X}$ dans l'image $I()$ [1] [Horn and Schunck, 1981]. Cette hypothèse, caractéristique des scènes lambertiennes, se réduit à une constance locale dans un intervalle de temps limité, pour des scènes plus réalistes, c'est-à-dire :

$$I(\mathbf{u}(\mathbf{p}, \mathbf{X}), t) = I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t + \delta t), \tag{3.2}$$

pour $\boldsymbol{\delta}\mathbf{p}$ et δt faibles et où $\oplus$ symbolise la composition des transformations rigides paramétrées par des éléments de $\mathfrak{se}(3)$. Cette dernière contrainte reste vague à cause de la difficulté à généraliser des tolérances sur $\boldsymbol{\delta}\mathbf{p}$ et δt car elles dépendent du contenu de la scène observée elle-même dans l'équation (3.2).

En s'appuyant sur l'équation (3.2), on peut formaliser le lien entre l'écart d'intensités entre deux images acquises à deux poses de caméras différentes et la pose relative $\boldsymbol{\delta}\mathbf{p}$ entre ces caméras (ou entre deux poses distinctes de la même caméra, soit son mouvement dans l'espace). Pour ce faire, on instancie la fonction de coût $\mathscr{C}()$ (Eq. (3.1)) en $\mathscr{C}_{SSD}()$, pour un ensemble $\mathcal{X}$ de points 3D de la scène :

$$\mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) = \frac{1}{2} \sum_{\mathbf{X} \in \mathcal{X}} \left(I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t + \delta t) - I(\mathbf{u}(\mathbf{p}, \mathbf{X}), t)\right)^2, \tag{3.3}$$

tel que $\mathcal{I} = \{I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t + \delta t), \forall \mathbf{X} \in \mathcal{X}\}$ et $\mathcal{I}^* = \{I(\mathbf{u}(\mathbf{p}, \mathbf{X}), t), \forall \mathbf{X} \in \mathcal{X}\}$. L'écriture de ces derniers ensembles d'intensités fait abstraction des potentielles limites du champ de vue de la caméra, faisant que certains points $\mathbf{X} \in \mathcal{X}$ peuvent être visibles d'un point de vue mais pas d'un autre, problème qui peut aussi apparaître à cause d'occultations.

L'équation (3.3) s'écrit aussi sous forme matricielle, en simplifiant les écritures en posant $I^*(\mathbf{u}(\mathbf{p}, \mathbf{X})) = I(\mathbf{u}(\mathbf{p}, \mathbf{X}), t)$ et $I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X})) = I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t + \delta t)$:

$$\begin{array}{rcl} \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) & = & \frac{1}{2} ||\mathbf{I}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathcal{X})) - \mathbf{I}^*(\mathbf{p}, \mathcal{X}))||^2 \\ & = & \frac{1}{2} ||\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*)||^2, \end{array} \tag{3.4}$$

où :

$$\begin{cases} \mathbf{I}^*(\mathbf{p}, \mathcal{X})) = \left[I^*(\mathbf{u}(\mathbf{p}, \mathbf{X}_1)), I^*(\mathbf{u}(\mathbf{p}, \mathbf{X}_2)), \ldots, I^*(\mathbf{u}(\mathbf{p}, \mathbf{X}_{|\mathcal{X}|}))\right]^\top \\ \mathbf{I}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathcal{X})) = \left[I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}_1)), I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}_2)), \ldots, I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}_{|\mathcal{X}|}))\right]^\top \end{cases}, \tag{3.5}$$

1. Une image acquise ou synthétisée ayant un support discret, l'accès à une intensité à des coordonnées réelles se fait par interpolation, négligée dans ces expressions.

et :

$$\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) = \mathbf{I}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathcal{X}) - \mathbf{I}^*(\mathbf{p}, \mathcal{X}) \tag{3.6}$$

L'équation (3.4), mène à l'expression du problème de calcul de pose relative (spécification de l'équation (3.1)) suivant :

$$\widehat{\boldsymbol{\delta}\mathbf{p}} = \arg\min_{\boldsymbol{\delta}\mathbf{p}} \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) . \tag{3.7}$$

Dès lors, résoudre l'équation (3.7) mène à la pose relative $\widehat{\boldsymbol{\delta}\mathbf{p}}$, optimale selon le critère de la fonction $\mathscr{C}_{SSD}()$.

3.2.2 Méthodes d'optimisation usuelles

Les méthodes de résolution de l'équation (3.7) sont généralement itératives, de type descente de gradient (convergence linéaire), Newton (convergence quadratique), Gauss-Newton (convergence quadratique, au plus), Levenberg-Marquardt (convergence quadratique, au plus) ou encore ESM (convergence quadratique, au moins), du fait de la non-linéarité entre les intensités des pixels et la pose relative $\boldsymbol{\delta}\mathbf{p}$ de la caméra. A chaque itération k, un incrément de pose $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \in \mathbb{R}^6$ est calculé, avec une expression différente, selon la méthode.

Descente de gradient : Il s'agit d'une approximation locale de la fonction de coût par un hyper-plan. La fonction est optimisée en faisant évoluer les degrés de liberté en direction opposée du gradient de la fonction de coût $\mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})$ (Eq. (3.6), dont les trois derniers paramètres ont été omis par soucis de compacité) :

$$\begin{aligned} \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} &= -\lambda \left. \frac{\partial \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}} \right|^{\top}_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \\ &= -\lambda \, \mathbf{J}^{\top}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} \mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(k)}) \end{aligned} \tag{3.8}$$

où $\lambda \in \mathbb{R}^+$ est un facteur d'échelle appliqué au gradient pour éviter de grands sauts pouvant engendrer la divergence de l'optimisation et en ayant posé, pour simplifier les expressions :

$$\begin{aligned} \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} = \left. \frac{\partial \mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}} \right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} &= \left. \frac{\partial \mathbf{I}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathcal{X})}{\partial \boldsymbol{\delta}\mathbf{p}} \right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \\ &= \left. \begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{u}} I \frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta}\mathbf{p}} \right)\Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{bmatrix} \right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} , \end{aligned} \tag{3.9}$$

car $\mathbf{I}^*(\mathbf{p}, \mathcal{X})$ (Eq. (3.5)) ne dépend pas de $\boldsymbol{\delta}\mathbf{p}$. $\nabla_{\mathbf{u}} I \in \mathbb{R}^{1\times 2}$ est le gradient spatial de l'image et $\partial\mathbf{u}/\partial\boldsymbol{\delta}\mathbf{p}$ est la matrice jacobienne géométrique de l'image. En vision perspective, $\nabla_{\mathbf{u}} I$ se calcule par convolution de l'image avec un filtre dérivatif vectoriel ou

matriciel régulier [Gonzalez and Woods, 2018] et la matrice jacobienne géométrique de l'image s'exprime par :

$$\begin{aligned}\frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta p}} &= \frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \boldsymbol{\delta p}} \\ &= \begin{bmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{bmatrix} \frac{\partial \mathbf{x}}{\partial \boldsymbol{\delta p}}\end{aligned} \tag{3.10}$$

où

$$\frac{\partial \mathbf{x}}{\partial \boldsymbol{\delta p}} = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -(1+x^2) & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix}, \tag{3.11}$$

en vision perspective [Chaumette and Hutchinson, 2006], grâce, notamment, aux équations (2.3) et (2.1) ou encore :

$$\frac{\partial \mathbf{x}}{\partial \boldsymbol{\delta p}} = \begin{pmatrix} -\frac{1+x^2(1-\xi(\Gamma+\xi))+y^2}{\rho(\Gamma+\xi)} & \frac{\xi xy}{\rho} & \frac{\Gamma x}{\rho} & xy & -\frac{(1+x^2)\Gamma-\xi y^2}{\Gamma+\xi} & y \\ \frac{\xi xy}{\rho} & -\frac{1+y^2(1-\xi(\Gamma+\xi))+x^2}{\rho(\Gamma+\xi)} & \frac{\Gamma y}{\rho} & \frac{(1+y^2)\Gamma-\xi x^2}{\Gamma+\xi} & -xy & -x \end{pmatrix}, \tag{3.12}$$

en projection centrale [Hadj-Abdelkader, 2006] ($\Gamma = \sqrt{1+(1-\xi^2)(x^2+y^2)}$).

Quelque soit le modèle de projection, la $\frac{\partial \mathbf{x}}{\partial \boldsymbol{\delta p}}$ est communément appelée la matrice d'interaction liée au point $\mathbf{x}$.

Newton : Il s'agit d'une approximation locale de la fonction de coût par une parabole (de dimension correspondant au nombre de degrés de liberté du problème). A partir du développement de Taylor-Young dans les espaces vectoriels normés au deuxième ordre de $\mathscr{C}_{SSD}(\boldsymbol{\delta p})$:

$$\begin{aligned}\mathscr{C}_{SSD}(\boldsymbol{\delta p}^{(k+1)}) \approx\ & \mathscr{C}_{SSD}(\boldsymbol{\delta p}^{(k)}) + \left.\frac{\partial \mathscr{C}_{SSD}(\boldsymbol{\delta p})}{\partial \boldsymbol{\delta p}}\right|_{\boldsymbol{\delta p}=\boldsymbol{\delta p}^{(k)}} \dot{\boldsymbol{\delta p}}^{(k)} \\ & + \frac{1}{2}\dot{\boldsymbol{\delta p}}^{(k)\top} \left.\frac{\partial^2 \mathscr{C}_{SSD}(\boldsymbol{\delta p})}{\partial \boldsymbol{\delta p}^2}\right|_{\boldsymbol{\delta p}=\boldsymbol{\delta p}^{(k)}} \dot{\boldsymbol{\delta p}}^{(k)},\end{aligned} \tag{3.13}$$

où la matrice hessienne de $\mathscr{C}_{SSD}$ s'exprime par :

$$\left.\frac{\partial^2 \mathscr{C}_{SSD}(\boldsymbol{\delta p})}{\partial \boldsymbol{\delta p}^2}\right|_{\boldsymbol{\delta p}=\boldsymbol{\delta p}^{(k)}} = \mathbf{J}_{\boldsymbol{\delta p}^{(k)}}^{\top}\mathbf{J}_{\boldsymbol{\delta p}^{(k)}} + \mathbf{H}_{\boldsymbol{\delta p}^{(k)}}, \tag{3.14}$$

avec :

$$\mathbf{H}_{\boldsymbol{\delta p}^{(k)}} = \sum_{\mathbf{X}\in\mathcal{X}} \left[\left.\frac{\partial^2 I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta p},\mathbf{X}))}{\partial \boldsymbol{\delta p}^2}\right|_{\boldsymbol{\delta p}=\boldsymbol{\delta p}^{(k)}} (I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta p},\mathbf{X})) - I^*(\mathbf{u}(\mathbf{p},\mathbf{X}))) \right]. \tag{3.15}$$

Puisque le but est d'atteindre $\boldsymbol{\delta}\mathbf{p}^{(k+1)} = \widehat{\boldsymbol{\delta}\mathbf{p}}$, la dérivée de $\mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p}^{(k+1)})$ par rapport à $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}$ est nulle en la solution, donc, en dérivant l'équation (3.13), on a :

$$\begin{aligned}\frac{\partial}{\partial \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}}\mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p}^{(k+1)}) &\approx \frac{\partial}{\partial \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}}\left[\mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p}^{(k)}) + \left.\frac{\partial \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}\right. \\ &\quad \left. + \frac{1}{2}\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)\top} \left.\frac{\partial^2 \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}^2}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}\right],\end{aligned} \tag{3.16}$$

qui donne :

$$\left.\frac{\partial \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} + \left.\frac{\partial^2 \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}^2}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \approx 0. \tag{3.17}$$

L'incrément $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}$ se calcule alors par :

$$\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \approx -\lambda \left.\frac{\partial^2 \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}^2}\right|^{-1}_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \left.\frac{\partial \mathscr{C}_{SSD}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}}, \tag{3.18}$$

ou, sous forme matricielle, et en détaillant légèrement l'expression :

$$\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} = -\lambda \left[\mathbf{J}^{\top}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} + \mathbf{H}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}\right]^{-1} \mathbf{J}^{\top}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} \mathbf{C}_{\mathbf{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(k)}). \tag{3.19}$$

L'insertion du $\lambda \in \mathbb{R}^+$ (Eq. (3.19)) a le même objectif que pour la descente de gradient (Eq. (3.8)). La formulation de l'algorithme de Newton pose le problème du calcul de la matrice hessienne, parfois couteux.

Gauss-Newton : C'est une approximation de la méthode de Newton prenant en compte la structure particulière des problèmes aux moindres carrés. On part d'un développement de Taylor-Young au premier ordre du vecteur $\mathbf{I}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathcal{X})$ (Eq. (3.5)), simplifié en $\mathbf{I}(\boldsymbol{\delta}\mathbf{p})$ pour alléger la notation :

$$\begin{aligned}\mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k+1)}) &\approx \mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)}) + \left.\frac{\partial \mathbf{I}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \\ &\approx \mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)}) + \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} \dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}\end{aligned} \tag{3.20}$$

Puisque le but est d'atteindre $\boldsymbol{\delta}\mathbf{p}^{(k+1)} = \widehat{\boldsymbol{\delta}\mathbf{p}}$, à convergence $\mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k+1)}) = \mathbf{I}^*$. De plus, si la matrice $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}$ (Eq. (3.20)) est de plein rang, la pseudo-inverse (de Moore-Penrose), notée $\cdot^{\dagger}$, peut lui être appliquée, ce qui mène à l'expression de $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}$ suivante :

$$\begin{aligned}\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} &= \lambda \mathbf{J}^{\dagger}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} \left(\mathbf{I}^* - \mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)})\right) \\ &= -\lambda \mathbf{J}^{\dagger}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} \mathbf{C}_{\mathbf{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(k)}).\end{aligned} \tag{3.21}$$

Levenberg-Marquardt : Il s'agit d'une extension de la méthode de Gauss-Newton, avec amortissement. En décomposant la pseudo-inverse de l'équation (3.21) :

$$\dot{\boldsymbol{\delta\mathbf{p}}}^{(k)} = -\lambda \left(\mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}^{\top} \mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}} \right)^{-1} \mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}^{\top} \mathbf{C}_{\mathbf{SSD}}(\boldsymbol{\delta\mathbf{p}}^{(k)}), \tag{3.22}$$

l'algorithme de Levenberg-Marquardt introduit une modification, pondérée par un scalaire $\mu \in \mathbb{R}^{+}$, de la diagonale du produit $\mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}^{\top} \mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}$ (Eq. (3.22)), afin de pouvoir en améliorer le conditionnement, tel que :

$$\dot{\boldsymbol{\delta\mathbf{p}}}^{(k)} = -\lambda \left(\mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}^{\top} \mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}} + \mu \operatorname{Diag}(\mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}^{\top} \mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}) \right)^{-1} \mathbf{J}_{\boldsymbol{\delta\mathbf{p}}^{(k)}}^{\top} \mathbf{C}_{\mathbf{SSD}}(\boldsymbol{\delta\mathbf{p}}^{(k)}). \tag{3.23}$$

Dans l'équation (3.23), $\mu = 0$ donne l'algorithme de Gauss-Newton (Eq. (3.21)). Toute autre valeur engendre un comportement entre l'algorithme de Gauss-Newton et celui de la méthode du gradient, de plus en plus proche de cette dernière à mesure que μ croît. En pratique, μ est fixé à une valeur initiale μ_0 à la première itération et il évolue d'une itération à l'autre, soit en diminuant si l'incrément $\dot{\boldsymbol{\delta\mathbf{p}}}^{(k)}$ a permis de faire décroître la fonction de coût (Eq. (3.4)), soit en augmentant dans le cas contraire, menant au re-calcul de $\dot{\boldsymbol{\delta\mathbf{p}}}^{(k)}$, jusqu'à ce que $\dot{\boldsymbol{\delta\mathbf{p}}}^{(k)}$ permette de faire décroître la fonction de coût.

3.2.3 Mise à jour des degrés de liberté

Une fois l'incrément $\dot{\boldsymbol{\delta\mathbf{p}}}^{(k)}$ calculé, la pose $\boldsymbol{\delta\mathbf{p}}^{(k)}$ courante de la caméra est mise à jour en $\boldsymbol{\delta\mathbf{p}}^{(k+1)}$ de plusieurs façons possibles.

Mise à jour explicite Deux premières approches explicites permettent de mettre à jour $\boldsymbol{\delta\mathbf{p}}^{(k)}$:

— l'approche additive [Lucas and Kanade, 1981] :

$$\boldsymbol{\delta\mathbf{p}}^{(k+1)} = \boldsymbol{\delta\mathbf{p}}^{(k)} + \dot{\boldsymbol{\delta\mathbf{p}}}^{(k)}. \tag{3.24}$$

— l'approche compositionnelle [Shum and Szeliski, 2000] : Soit l'application exponentielle de l'algèbre de Lie $\mathfrak{se}(3)$:

$$\exp_{\mathfrak{se}(3)} : \dot{\boldsymbol{\delta\mathbf{p}}}^{(k)} = [\boldsymbol{\upsilon}^{\top}, \boldsymbol{\omega}^{\top}]^{\top} \in \mathbb{R}^6 \longmapsto {}^{c^{(k)}}\mathbf{M}_{c^{(k+1)}} \in SE(3), \tag{3.25}$$

par abus de notation (cf. Remarque 5, page 52), dont le bloc rotation ${}^{c^{(k)}}\mathbf{R}_{c^{(k+1)}}$ et le vecteur translation ${}^{c^{(k)}}\mathbf{t}_{c^{(k+1)}}$ ont les expressions explicites suivantes :

$${}^{c^{(k)}}\mathbf{R}_{c^{(k+1)}} = \mathbf{I}_{3\times 3} + \operatorname{sinc}\theta \boldsymbol{\omega}_{\times} + \frac{1-\cos\theta}{\theta^2}\boldsymbol{\omega}_{\times}^2, \text{avec } \operatorname{sinc}\theta = \frac{\sin\theta}{\theta},\ \theta = \sqrt{\boldsymbol{\omega}^{\top}\boldsymbol{\omega}}, \tag{3.26}$$

et

$${}^{c^{(k)}}\mathbf{t}_{c^{(k+1)}} = \left[\mathbf{I}_{3\times 3} + \frac{1-\cos\theta}{\theta^2}\boldsymbol{\omega}_{\times} + \frac{1-\operatorname{sinc}\theta}{\theta^2}\boldsymbol{\omega}_{\times}^2 \right] \boldsymbol{\upsilon}, \tag{3.27}$$

correspondant, pour la rotation, aux formules de Rodrigues ($\boldsymbol{\upsilon}$: incrément en translation, $\boldsymbol{\omega}$: incrément en rotation).
${}^{c^{(k)}}\mathbf{M}_{c(k+1)}$ met à jour ${}^{c}\mathbf{M}_{c^*}^{(k)}$ par composition :

$$\begin{aligned} {}^{c}\mathbf{M}_{c^*}^{(k+1)} &= \left[{}^{c^{(k)}}\mathbf{M}_{c(k+1)}\right]^{-1} {}^{c}\mathbf{M}_{c^*}^{(k)} \\ &= {}^{c^{(k+1)}}\mathbf{M}_{c(k)}\, {}^{c}\mathbf{M}_{c^*}^{(k)}. \end{aligned} \tag{3.28}$$

On raisonne de façon similaire si $\boldsymbol{\delta}\mathbf{p}$ appartient à une autre algèbre de Lie (ex : $\exp_{\mathfrak{gl}(2)}$ si $\boldsymbol{\delta}\mathbf{p} \in \mathfrak{gl}(2)$, $\exp_{\mathfrak{sl}(3)}$ si $\boldsymbol{\delta}\mathbf{p} \in \mathfrak{sl}(3)$, etc). Au besoin, le résultat de la composition peut être transformé du groupe à l'algèbre de Lie par le logarithme du groupe [Ma et al., 2004].

Remarque 5 (Abus de notation d'algèbre de Lie) *L'équation (3.25) introduit un abus de notation en écrivant que* $\exp_{\mathfrak{se}(3)}$ *est une application de* $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \in \mathbb{R}^6$ *vers* ${}^{c^{(k)}}\mathbf{M}_{c(k+1)} \in SE(3)$. *En effet, en toute rigueur, en considérant* $\boldsymbol{\upsilon} \in \mathbb{R}^3$ *et* $\boldsymbol{\omega} \in \mathbb{R}^3$, *c'est l'algèbre de Lie de* $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} = [\boldsymbol{\upsilon}^\top, \boldsymbol{\omega}^\top]^\top$ *:*

$$\left(\begin{array}{c|c} \boldsymbol{\omega}_\times & \boldsymbol{\upsilon} \\ \hline 0 & 0 \end{array}\right),$$

qui appartient à $\mathfrak{se}(3)$ *et c'est sur cette dernière que* $\exp_{\mathfrak{se}(3)}$ *est applicable. Mais, comme évoqué dans la partie 2.2 pour d'autres opérations, l'abus de notation fait en équation (3.25) permet de simplifier les écritures.*

Enfin, pour indiquer que $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \in \mathbb{R}^6$ *a été contraint pour avoir la même signification qu'un vecteur de pose (relative), c'est-à-dire qu'une matrice de transformation rigide peut être obtenue directement à partir des six éléments de* $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}$ *(Eqs. 2.12 et 2.13), on pousse l'abus de notation jusqu'à écrire, dans ce cas précis :* $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \in \mathfrak{se}(3)$. ◇

Enfin, une fois les degrés de liberté de la transformation mis à jour, le nombre d'itérations k est incrémenté et une nouvelle mesure intermédiaire "virtuelle" est créée ($\mathcal{I}^{(k)}$, pendant la résolution itérative de la minimisation de $\mathscr{C}_{SSD}()$, Eq. (3.7), par exemple) pour calculer le nouvel incrément, quelque soit la méthode d'optimisation considérée parmi celles évoquées précédemment. La méthode de création de la nouvelle mesure intermédiaire dépendra de la nature des mesures $\mathcal{V}$ à recaler sur la référence $\mathcal{T}$ (synthèse d'image à partir d'une maquette virtuelle 3D ou transformation géométrique d'image, cf. Partie 3.3.1). Les approches additives et compositionnelles ont été montrées équivalentes si la transformation décrite par les degrés de liberté est inversible [Baker and Matthews, 2004].

Une troisième approche, dite compositionnelle inverse [Baker and Matthews, 2001] permute virtuellement les rôles de la mesure visuelle $\mathcal{I}$ et de la référence $\mathcal{I}^*$. Contrairement aux approches additives et compositionnelles pour lesquelles chaque itération k procède en deux temps, à

savoir le calcul de l'incrément et la mise à jour de la pose relative, l'approche compositionnelle inverse procède en trois temps. D'abord, le calcul de l'incrément se fait différemment, pour prendre en compte la permutation virtuelle de $\mathcal{I}$ et $\mathcal{I}^*$, en introduisant une autre pose relative $\boldsymbol{\delta}\mathbf{p}'$, celle de c^* par rapport à c, ce qui demande une ré-écriture du coût $\mathscr{C}_{SSD}()$ (Eq. (3.4)) en $\mathscr{C}_{SSD_{inv}}()$:

$$\begin{aligned}\mathscr{C}_{SSD_{inv}}(\boldsymbol{\delta}\mathbf{p},\mathcal{I},\{\mathbf{p},\mathcal{X}\},\mathcal{I}^*) &= \tfrac{1}{2}||\mathbf{I}^*(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p}',\mathcal{X})-\mathbf{I}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathcal{X})||^2 \\ &= \tfrac{1}{2}||\mathbf{C}_{\mathbf{SSD_{inv}}}(\{\boldsymbol{\delta}\mathbf{p},\boldsymbol{\delta}\mathbf{p}'\},\mathcal{I},\{\mathbf{p},\mathcal{X}\},\mathcal{I}^*)||^2.\end{aligned} \tag{3.29}$$

Dès lors, par exemple avec la méthode de Gauss-Newton, un incrément $\dot{\boldsymbol{\delta}\mathbf{p}}'^{(k)}$ est calculable à l'itération k pour minimiser $\mathscr{C}_{SSD_{inv}}$ (Eq. 3.29). La subtilité de l'approche compositionnelle inverse est de mettre à jour $\boldsymbol{\delta}\mathbf{p}$ à l'aide de $\dot{\boldsymbol{\delta}\mathbf{p}}'^{(k)}$, puisque la transformation de c^* vers c est l'inverse de la transformation de c vers c^*. On exploite l'exponentielle de l'algèbre de Lie au sein de laquelle $\dot{\boldsymbol{\delta}\mathbf{p}}'(k)$ est définie, par exemple $\mathfrak{se}(3)$ pour se comparer à l'équation (3.25), tel que :

$${}^{c^*(k)}\mathbf{M}_{c^*(k+1)} = \exp_{\mathfrak{se}(3)}\left(\dot{\boldsymbol{\delta}\mathbf{p}}'^{(k)}\right), \tag{3.30}$$

et :

$${}^{c(k)}\mathbf{M}_{c(k+1)} = \left[{}^{c^*(k)}\mathbf{M}_{c^*(k+1)}\right]^{-1}, \tag{3.31}$$

et l'équation (3.28) peut à nouveau être utilisée pour obtenir la nouvelle matrice ${}^{c}\mathbf{M}_{c^*}^{(k+1)}$ par composition.

L'approche compositionnelle inverse laisse donc $\boldsymbol{\delta}\mathbf{p}'$ constant, et même nul car :

$$\boldsymbol{\delta}\mathbf{p}'^{(k=0)} = \mathbf{0}, \tag{3.32}$$

tout en faisant évoluer $\boldsymbol{\delta}\mathbf{p}$ vers l'optimum au fil des itérations. L'approche compositionnelle inverse résout donc le problème suivant :

$$\widehat{\boldsymbol{\delta}\mathbf{p}} = \arg\min_{\boldsymbol{\delta}\mathbf{p}} \mathscr{C}_{SSD_{inv}}(\boldsymbol{\delta}\mathbf{p},\mathcal{I},\{\mathbf{p},\mathcal{X}\},\mathcal{I}^*). \tag{3.33}$$

L'approche compositionnelle inverse permet donc de résoudre le problème de recalage formulé à l'aide du coût $\mathscr{C}_{SSD_{inv}}$ (Eq. (3.33)), menant à la même solution $\widehat{\boldsymbol{\delta}\mathbf{p}}$ qu'en exploitant le coût $\mathscr{C}_{SSD}$ (Eq. (3.7)), sous réserve que la transformation considérée soit inversible [Baker and Matthews, 2004]. L'apport de l'approche compositionnelle inverse sur l'approche compositionnelle réside au niveau du calcul de la matrice jacobienne du développement de Taylor-Young au premier ordre de la $\mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}')$, par rapport à la nouvelle pose relative $\boldsymbol{\delta}\mathbf{p}'$:

$$\mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}'^{(k+1)}) \approx \mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}'^{(k)}) + \left.\frac{\partial\mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}')}{\partial\boldsymbol{\delta}\mathbf{p}'}\right|_{\boldsymbol{\delta}\mathbf{p}'=\boldsymbol{\delta}\mathbf{p}'^{(k)}} \dot{\boldsymbol{\delta}\mathbf{p}}'^{(k)}. \tag{3.34}$$

En suivant le même raisonnement que pour obtenir l'équation (3.21) selon la méthode de Gauss-Newton, on a :

$$\mathbf{I} \approx \mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}'^{(k)}) + \left.\frac{\partial\mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}')}{\partial\boldsymbol{\delta}\mathbf{p}'}\right|_{\boldsymbol{\delta}\mathbf{p}'=\boldsymbol{\delta}\mathbf{p}'^{(k)}} \dot{\boldsymbol{\delta}\mathbf{p}}'^{(k)}. \tag{3.35}$$

et :

$$\dot{\boldsymbol{\delta}\mathbf{p}}'^{(k)} = -\lambda \left.\frac{\partial \mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}')}{\partial \boldsymbol{\delta}\mathbf{p}'}\right|_{\boldsymbol{\delta}\mathbf{p}'=\boldsymbol{\delta}\mathbf{p}'^{(k)}} \left(\mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}'^{(k)}) - \mathbf{I}\right), \tag{3.36}$$

qui devient, puisque $\boldsymbol{\delta}\mathbf{p}'^{(k)} = \mathbf{0}, \forall k \in \mathbb{N}$ (Eq. (3.32)) :

$$\dot{\boldsymbol{\delta}\mathbf{p}}'^{(k)} = -\lambda \left.\frac{\partial \mathbf{I}^*(\boldsymbol{\delta}\mathbf{p}')}{\partial \boldsymbol{\delta}\mathbf{p}'}\right|_{\boldsymbol{\delta}\mathbf{p}'=\mathbf{0}} \mathbf{C}_{\mathbf{SSD_{inv}}}(\mathbf{0}). \tag{3.37}$$

La matrice jacobienne, ci-dessus, s'exprime, en détail :

$$\frac{\partial \mathbf{I}^*(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}', \mathcal{X})}{\partial \boldsymbol{\delta}\mathbf{p}'} = \begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{u}} I^* \frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta}\mathbf{p}'}\right)\Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{bmatrix} = \mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}, \tag{3.38}$$

avec $\mathbf{u}_i = \mathbf{u}(\mathbf{p}, \mathbf{X}_i), \forall \mathbf{X}_i \in \mathcal{X}$. Comme I^* est constante, $\nabla_{\mathbf{u}} I^*$ (Eq. (3.38)) peut n'être calculé qu'une seule fois (pour $k = 0$), et il en va de même pour $\frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta}\mathbf{p}'}$. $\mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}$ (Eq. (3.38)) n'est donc à calculer qu'une seule fois grâce à l'approche compositionnelle inverse contre un calcul à chaque itération de $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}}$ (Eq. (3.9)) dans les approches additive et compositionnelle. A l'itération suivante, seul le vecteur d'erreur sera à recalculer. L'approche compositionnelle inverse permet donc de réduire considérablement les temps de calcul de l'optimisation du même problème de recalage.

Pour finir sur la mise à jour explicite, il existe aussi une quatrième approche, dite additive inverse [Hager and Belhumeur, 1998], d'un coût calculatoire tout aussi faible que l'approche compositionnelle inverse, mais restreinte à un sous-ensemble des transformations admissibles par l'approche compositionnelle inverse, excluant les homographies et les rotations 3D [Baker and Matthews, 2004]. Cette approche n'est pas détaillée dans ce mémoire.

Choisir laquelle de ces quatre approches explicites utiliser dépend de la qualité des mesures visuelles $\mathcal{V}$ par rapport à celle de la référence $\mathcal{T}$. Si $\mathcal{T}$ souffre de plus de bruit que $\mathcal{V}$, privilégier l'approche additive ou l'approche compositionnelle donnera de meilleurs résultats qu'avec les autres approches du fait que les gradients spatiaux d'image seront calculés sur $\mathcal{V}$, plutôt que sur $\mathcal{T}$. A l'inverse, si $\mathcal{T}$ souffre d'un bruit inférieur ou similaire à $\mathcal{V}$, l'approche compositionnelle inverse donnera non seulement des résultats équivalents ou meilleurs que les approches additive ou compositionnelle, mais aussi avec un coût calculatoire plus faible.

Mise à jour implicite La mise à jour est implicite dans le cas de l'asservissement visuel basé image où le but est de déplacer physiquement la caméra vers $\mathcal{F}_{c^*}$, sans forcément calculer explicitement sa pose. Des lois de commandes cartésiennes peuvent s'écrire directement à partir des méthodes d'optimisation de la partie 3.2.2, en considérant $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \in \mathfrak{se}(3)$. Dans le cas régulier (voir [Siciliano et al., 2008] pour tous les cas) où le nombre de degrés de liberté dans l'espace de configuration du robot égale celui de son espace opérationnel, le vecteur de commande cinématique

cartésienne $\dot{\boldsymbol{\delta}}\mathbf{p}$ peut être transformé en vecteur de commande cinématique $\dot{\mathbf{q}}$, par :

$$\dot{\mathbf{q}} = {}^{e}\mathbf{J}_{e}(\mathbf{q})^{-1}\,{}^{e}\mathbf{V}_{c}\,\dot{\boldsymbol{\delta}}\mathbf{p}, \text{ tel que } {}^{e}\mathbf{J}_{e}(\mathbf{q}) = \frac{\partial \mathbf{p}}{\partial \mathbf{q}}, \tag{3.39}$$

où ${}^{e}\mathbf{V}_{c}$ est la matrice de changement de repère de torseur cinématique du repère caméra $\mathcal{F}_c$ au repère "effecteur" $\mathcal{F}_e$ du robot. $\dot{\mathbf{q}}$ est alors considéré comme consigne de vitesse par le contrôleur bas niveau du robot, pendant un pas de temps δ_t.

Un asservissement visuel réellement direct devrait exprimer le lien le plus étroit entre les intensités des pixels de la caméra et les entrées de commande du robot. Ainsi, en adaptant la méthode d'optimisation de Gauss-Newton (Eq. (3.21)), par exemple, à la commande de robot, on obtient :

$$\begin{aligned} \dot{\mathbf{q}} &= -\lambda \left[\frac{\partial \mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}}\,{}^{c}\mathbf{V}_{e}\,{}^{e}\mathbf{J}_{e}(\mathbf{q}) \right] \Big|^{\dagger}_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(t)}} \mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(t)}) \\ &= -\lambda\, \mathbf{J_{q}}(\boldsymbol{\delta}\mathbf{p}^{(t)})^{\dagger}\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(t)}). \end{aligned} \tag{3.40}$$

A noter que $\boldsymbol{\delta}\mathbf{p}^{(t)}$ (Eq. (3.40)) évolue au fil du temps et non des itérations puisque l'image courante $\mathcal{I}$ est mise à jour par une nouvelle acquisition d'image à chaque instant t avec un pas δ_t au cours du déplacement du robot.

Une autre raison de calculer directement $\dot{\mathbf{q}}$ à partir de la loi de commande de l'équation (3.40) plutôt que calculer $\dot{\boldsymbol{\delta}}\mathbf{p}$ puis le transformer en $\dot{\mathbf{q}}$ est que le résultat n'est pas toujours équivalent [Chaumette and Hutchinson, 2007]. Néanmoins, pour garder des expressions génériques aux contextes évoqués dans ce mémoire (calcul de pose, suivi, asservissement visuel, exploration d'environnement réel ou virtuel), les lois de commande présentées se concentreront sur la commande cartésienne.

3.2.4 Méthode d'optimisation efficace au second ordre

La méthode ESM [Malis, 2004], Efficient Second-order Minimization, étend la méthode de Gauss-Newton en deux aspects : au deuxième ordre et en combinant la mise à jour des degrés de liberté des approches compositionnelle et compositionnelle inverse. Soit le développement de Taylor-Young au deuxième ordre de $\mathbf{I}(\boldsymbol{\delta}\mathbf{p})$, évalué en $\widehat{\boldsymbol{\delta}\mathbf{p}}$:

$$\mathbf{I}^{*} \approx \mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)}) + \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}\dot{\boldsymbol{\delta}}\mathbf{p}^{(k)} + \frac{1}{2}\mathbf{H}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}\dot{\boldsymbol{\delta}}\mathbf{p}^{(k)}. \tag{3.41}$$

En considérant le développement de Taylor-Young au premier ordre de la matrice jacobienne $\mathbf{J}\,(\boldsymbol{\delta}\mathbf{p})$, évaluée en $\widehat{\boldsymbol{\delta}\mathbf{p}}$:

$$\mathbf{J}\left(\widehat{\boldsymbol{\delta}\mathbf{p}}\right) = \mathbf{J}\left(\boldsymbol{\delta}\mathbf{p}^{(k)}\right) + \left.\frac{\partial \mathbf{J}\,(\boldsymbol{\delta}\mathbf{p})}{\partial \boldsymbol{\delta}\mathbf{p}}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(k)}} \dot{\boldsymbol{\delta}}\mathbf{p}^{(k)}, \tag{3.42}$$

donnant :

$$\mathbf{J}^{*}_{\boldsymbol{\delta}\mathbf{p}} = \mathbf{J}\left(\boldsymbol{\delta}\mathbf{p}^{(k)}\right) + \mathbf{H}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}, \tag{3.43}$$

et en substituant l'équation (3.43) dans l'équation (3.41), on obtient :

$$\begin{aligned}\mathbf{I}^* &\approx \mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)}) + \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} + \frac{1}{2}\left(\mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}} - \mathbf{J}\left(\boldsymbol{\delta}\mathbf{p}^{(k)}\right)\right)\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \\ &\approx \mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)}) + \left(\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} + \frac{1}{2}\mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}} - \frac{1}{2}\mathbf{J}\left(\boldsymbol{\delta}\mathbf{p}^{(k)}\right)\right)\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} \\ &\approx \mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)}) + \frac{1}{2}\left(\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} + \mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}\right)\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}.\end{aligned} \tag{3.44}$$

En rassemblant $\mathbf{I}^*$ et $\mathbf{I}(\boldsymbol{\delta}\mathbf{p}^{(k)})$ du même côté et en substituant par $\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(k)})$, on obtient :

$$\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(k)}) \approx -\frac{1}{2}\left(\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} + \mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}\right)\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}, \tag{3.45}$$

qui est une approximation au second ordre de $\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p})$, sans calculer la matrice hessienne, pour $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}$ faible [Benhimane and Malis, 2004]. Seule $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}$ doit être recalculée à chaque itération tandis que $\mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}$ est constante, comme dans l'approche compositionnelle inverse de mise à jour des degrés de liberté du problème. Dès lors, l'incrément $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}$ s'obtient facilement en exploitant la pseudo-inverse :

$$\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} = -2\left(\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}} + \mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}\right)^{\dagger}\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(k)}). \tag{3.46}$$

3.2.5 Conclusion partielle

Quelles que soient les méthodes de résolution de la fonction de coût $\mathscr{C}_{SSD}$, on observe que la connaissance des coordonnées 3D des points de la scène observée à la pose $\mathbf{p}$ de référence est nécessaire pour résoudre l'équation (3.7). Les coordonnées des points 3D de la scène n'étant pas directement mesurées par une caméra, résoudre l'équation (3.7) va demander, soit d'introduire des connaissances supplémentaires (Partie 3.3.1), soit de modifier le problème (Eq. (3.7)) de deux manières :

— selon des hypothèses de contenu de la scène ou sur la nature du mouvement entre $\mathcal{I}$ et $\mathcal{I}^*$ (Partie 3.3.2).

— en basculant $\mathcal{X}$ dans l'ensemble des degrés de liberté du problème, c'est-à-dire estimer simultanément la scène et localiser la caméra (Partie 3.3.3).

3.3 Approches directes pures

3.3.1 Quand des connaissances supplémentaires sont disponibles

Les connaissances supplémentaires à introduire sont, notamment, une maquette 3D de la scène, pour connaître l'ensemble $\mathcal{X}$, et la pose $\mathbf{p}$ de caméra permettant de mesurer $\mathcal{I}^*$. Cependant, connaître $\mathbf{p}$ est un problème de calcul de pose en soit aussi, ce qui ne résout rien. Plusieurs solutions alternatives existent néanmoins.

Considérer une caméra de profondeur combinée à la caméra couleur (ou niveaux de gris) dans un système étalonné permet d'obtenir $\mathcal{X}$ exprimé dans le repère caméra ayant permis de mesurer $\mathcal{I}^*$. Dans ce cas $\mathbf{p} = \mathbf{0}$ et l'équation (3.4) peut être résolue [Steinbrücker et al., 2011]. La caméra de profondeur peut être temps-de-vol ou à lumière structurée [Giancola et al., 2018], ou encore plénoptique [Jeon et al., 2015]. Ce dernier type présente l'avantage de rester dans le cadre des caméras passives et peut s'apparenter au cas de la vision stéréoscopique qui permet aussi de résoudre l'équation (3.7), au moins à un facteur d'échelle près, en ajoutant la connaissance de rigidité entre les différents points de vue internes au système de vision [Comport et al., 2010]. Si la caméra de profondeur est utilisable en permanence comme la caméra passive, la connaissance d'une deuxième carte de profondeur à la deuxième pose d'acquisition peut aussi être prise en compte en modifiant l'équation (3.4) afin de considérer des données hétérogènes photométrique et géométrique [Kerl et al., 2013].

Considérer une maquette 3D "colorée" de la scène associe un ensemble d'intensités directement à $\mathcal{X}$ (une unique intensité associée à un unique point 3D $\mathbf{X} \in \mathcal{X}$). Cette maquette 3D colorée peut être créée à partir d'une caméra de profondeur combinée à la caméra couleur, une fois encore, en projetant chaque $\mathbf{X} \in \mathcal{X}$ dans $\mathcal{I}^*$ afin d'obtenir directement l'association intensité - point 3D [Lindner et al., 2007]. En restant dans le cadre des caméras passives monoculaires, la maquette 3D colorée de la scène peut être introduite comme connaissance supplémentaire, pour résoudre le problème de l'équation (3.7). La maquette s'obtient par une autre technique comme la modélisation par conception assistée par ordinateur ou, pour envisager un cadre applicatif plus réaliste, la lasergrammétrie, combinée à la prise de photographies [Héno and Chandelier, 2014], ou encore la photogrammétrie [Lindner, 2016], l'estimation de la structure d'environnement à partir du mouvement de caméra [Szeliski, 2011] ou la localisation et cartographie simultanées basées vision [Saputra et al., 2018]. Ces deux dernières approches produisent des nuages de points 3D colorés qui peuvent être considérés bruts ou simplifiés en maillage 3D texturé comme maquette de la scène pour résoudre l'équation (3.1) dont l'expression de la fonction de coût est une extension de l'équation (3.7), plus ou moins importante suivant les méthodes (voir Partie 3.5), prenant en compte le fait que les intensités de chaque point 3D $\mathbf{X} \in \mathcal{X}$ n'aient pas été acquises par la même caméra que celle qui acquiert $\mathcal{I}$.

3.3.2 En faisant des hypothèses sur la scène ou le mouvement

3.3.2.1 Translation pure dans le plan image

Quand l'objectif est de recaler une région $\mathcal{R}^*$ de $\mathcal{I}^*$ et la région $\mathcal{R}$ inconnue de $\mathcal{I}$ lui correspondant, la fonction de coût $\mathscr{C}_{SSD}()$ (Eq. (3.3)) se simplifie en [Lucas and Kanade, 1981] :

$$\mathscr{C}_{LK}(\boldsymbol{\delta u}, \mathcal{I}, \mathcal{R}^*, \mathcal{I}^*) = \frac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I(\mathbf{u}^* + \boldsymbol{\delta u}) - I^*(\mathbf{u}^*)\right)^2 \tag{3.47}$$

avec $\mathcal{I} = \{I(\mathbf{u}), \forall \mathbf{u} \in \mathcal{U}\}$, $\mathcal{I}^* = \{I^*(\mathbf{u}^*), \forall \mathbf{u}^* \in \mathcal{R}^*\}$ et $\mathcal{R}^* \subset \mathcal{U}$, tel que $\mathcal{U} = \{0, 1, \ldots, N_c - 1\} \times \{0, 1, \ldots, N_l - 1\}$, pour une image de définition $N_c \times N_l$ pixels. L'équation (3.47) s'écrit sous forme matricielle :

$$\begin{array}{rcl} \mathscr{C}_{LK}(\boldsymbol{\delta u}, \mathcal{I}, \mathcal{R}^*, \mathcal{I}^*) & = & \frac{1}{2}||\mathbf{I}(\boldsymbol{\delta u}) - \mathbf{I}^*||^2 \\ & = & \frac{1}{2}||\mathbf{C_{LK}}(\boldsymbol{\delta u})||^2 \end{array} \tag{3.48}$$

où $\mathbf{I}^* = \left[I^*(\mathbf{u}_1^*), I^*(\mathbf{u}_2^*), \ldots, I^*(\mathbf{u}_{|\mathcal{R}^*|}^*))\right]^\top$ et
$\mathbf{I} = \left[I(\mathbf{u}_1^* + \boldsymbol{\delta u}), I(\mathbf{u}_2^* + \boldsymbol{\delta u}), \ldots, I(\mathbf{u}_{|\mathcal{R}^*|}^* + \boldsymbol{\delta u}))\right]^\top, \forall \mathbf{u}_i^* \in \mathcal{R}^*$.

Les degrés de liberté du problème de minimisation de la fonction $\mathscr{C}_{LK}()$ sont directement définis dans le plan image par $\boldsymbol{\delta u} \in \mathbb{R}^2$. Par conséquent, l'écriture du coût $\mathscr{C}_{LK}()$ restreint, implicitement, le problème de sa minimisation à deux caméras c et c^* à projection orthographique et dont les repères $\mathscr{F}_c$ et $\mathscr{F}_{c^*}$ sont séparés d'une translation pure parallèle à leurs plans images, soit une pose relative $\boldsymbol{\delta p}$ à deux degrés de liberté. En pratique, pour la projection perspective et un $\boldsymbol{\delta u}$ assez faible, l'approximation faite par $\mathscr{C}_{LK}()$ (Eq. (3.47)) est suffisamment respectée pour le recalage de la région $\mathcal{I}^*(\mathcal{R}^*)$ sur l'image $\mathcal{I}$. C'est l'objet de l'article de Lucas et Kanade [Lucas and Kanade, 1981] qui peut être vu comme fondateur pour la vision robotique directe. Il propose un recalage direct en résolvant l'équation :

$$\widehat{\boldsymbol{\delta u}} = \arg\min_{\boldsymbol{\delta u}} \mathscr{C}_{LK}(\boldsymbol{\delta u}, \mathcal{I}, \mathcal{R}^*, \mathcal{I}^*)\,, \tag{3.49}$$

par une méthode d'optimisation de type Gauss-Newton :

$$\dot{\boldsymbol{\delta u}}^{(k)} = -\lambda \left.\frac{\partial \mathbf{C_{LK}}(\boldsymbol{\delta u})}{\partial \boldsymbol{\delta u}}\right|^{\dagger}_{\boldsymbol{\delta u} = \boldsymbol{\delta u}^{(k)}} \mathbf{C_{LK}}(\boldsymbol{\delta u}^{(k)}), \tag{3.50}$$

avec l'introduction du gradient spatial de l'image $\nabla_{\mathbf{u}} I$ pour exprimer la matrice jacobienne :

$$\frac{\partial \mathbf{C_{LK}}(\boldsymbol{\delta u})}{\partial \boldsymbol{\delta u}} = \begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{u}} I \, \frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta u}}\right)\Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{bmatrix} = \begin{bmatrix} \vdots \\ \nabla_{\mathbf{u}} I|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{bmatrix}, \tag{3.51}$$

car $\mathcal{I}^*$ et $\mathcal{R}^*$ sont constants et :

$$\forall \mathbf{u}^* \in \mathcal{R}^*, \mathbf{u} = \mathbf{u}^* + \boldsymbol{\delta}\mathbf{u} \Rightarrow \frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta}\mathbf{u}} = \mathbf{I}_{2\times 2}. \tag{3.52}$$

Enfin, à chaque itération, $\boldsymbol{\delta}\mathbf{u}$ est mis à jour par approche additive :

$$\boldsymbol{\delta}\mathbf{u}^{(k+1)} = \boldsymbol{\delta}\mathbf{u}^{(k)} + \dot{\boldsymbol{\delta}\mathbf{u}}^{(k)}. \tag{3.53}$$

Au lieu de calculer les deux paramètres du vecteur $\boldsymbol{\delta}\mathbf{u}$ pour chaque région d'intérêt, Lucas et Kanade proposent aussi, dans le même article [Lucas and Kanade, 1981], en considérant la pose relative $\boldsymbol{\delta}\mathbf{p}$ entre $\mathscr{F}_c$ et $\mathscr{F}_{c^*}$ connue, de minimiser l'écart des intensités en optimisant la coordonnée ${}^{c^*}Z$ de la région $\mathcal{R}^*$. Une unique coordonnée ${}^{c^*}Z$ étant considérée pour toute la région $\mathcal{R}^*$, hypothèse est faite d'une scène fronto-parallèle à la caméra c^* au sein de cette région. Dès lors, le problème de l'équation (3.49) devient :

$${}^{c^*}\widehat{Z} = \arg\min_{{}^{c^*}Z} \mathscr{C}_{LK_Z}\left({}^{c^*}Z, \mathcal{I}, \{\boldsymbol{\delta}\mathbf{p}, \mathcal{R}^*\}, \mathcal{I}^*\right), \tag{3.54}$$

avec :

$$\begin{array}{rcl} \mathscr{C}_{LK_Z}({}^{c^*}Z, \mathcal{I}, \{\boldsymbol{\delta}\mathbf{p}, \mathcal{R}^*\}, \mathcal{I}^*) & = & \frac{1}{2}||\mathbf{I}({}^{c^*}Z) - \mathbf{I}^*||^2 \\ & = & \frac{1}{2}||\mathbf{C}_{\mathbf{LK_Z}}({}^{c^*}Z)||^2 \end{array} \tag{3.55}$$

où $\mathbf{I}^* = \left[I^*(\mathbf{u}_1^*), I^*(\mathbf{u}_2^*), \ldots, I^*(\mathbf{u}_{|\mathcal{R}^*|}^*))\right]^\top$ et $\mathbf{I}({}^{c^*}Z) = \left[I(\mathbf{u}({}^c\mathbf{X}({}^{c^*}Z, \mathbf{u}_1^*, \boldsymbol{\delta}\mathbf{p}))), I(\mathbf{u}({}^c\mathbf{X}({}^{c^*}Z, \mathbf{u}_2^*, \boldsymbol{\delta}\mathbf{p}))), \ldots, I(\mathbf{u}({}^c\mathbf{X}({}^{c^*}Z, \mathbf{u}_{|\mathcal{R}^*|}^*, \boldsymbol{\delta}\mathbf{p})))\right]^\top$, $\forall \mathbf{u}_i^* \in \mathcal{R}^*$. En suivant le même raisonnement que pour minimiser $\mathscr{C}_{LK}$ (Eq. (3.48)), la minimisation de $\mathscr{C}_{LK_Z}$ (Eq. (3.55)) demande l'expression de la matrice jacobienne :

$$\frac{\partial \mathbf{C}_{\mathbf{LK_Z}}(Z)}{\partial Z} = \begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{u}} I \, \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \, \frac{\partial \mathbf{x}}{\partial {}^c\mathbf{X}} \, \frac{\partial {}^c\mathbf{X}}{\partial {}^{c^*}Z}\right)\Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{bmatrix}, \tag{3.56}$$

où, en vision perspective [Lucas and Kanade, 1981] :

$$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \begin{pmatrix} \alpha_u & 0 \\ 0 & \alpha_v \end{pmatrix}, \tag{3.57}$$

$$\frac{\partial \mathbf{x}}{\partial {}^c\mathbf{X}} = \begin{pmatrix} \frac{1}{{}^cZ} & 0 & -\frac{{}^cX}{{}^cZ^2} \\ 0 & \frac{1}{{}^cZ} & -\frac{{}^cY}{{}^cZ^2} \end{pmatrix}, \tag{3.58}$$

(dérivées des équations (2.3) et (2.1)) et

$$\frac{\partial \, {}^c\mathbf{X}}{\partial \, {}^{c^*}Z} = {}^c\mathbf{R}_{c^*} \frac{\partial \, {}^{c^*}\mathbf{X}}{\partial \, {}^{c^*}Z} = {}^c\mathbf{R}_{c^*} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}. \tag{3.59}$$

Enfin, à chaque itération, ${}^{c^*}Z$ est mis à jour par l'approche additive :

$${}^{c^*}Z^{(k+1)} = {}^{c^*}Z^{(k)} + {}^{c^*}\dot{Z}^{(k)}. \tag{3.60}$$

Que ce soit pour calculer $\boldsymbol{\delta u}$ (Eq. (3.49)) ou ${}^{c^*}Z$ (Eq. (3.54)), les gradients spatiaux de l'image $\nabla_{\mathbf{u}} I$ servent à résoudre le problème (Eq. (3.51) et (3.56)). Ce recalage est donc plus performant pour des régions aux contrastes variés plutôt qu'uniformes [Shi and Tomasi, 1994], au risque de divergence de l'optimisation pour une région dont les $\nabla_{\mathbf{u}} I$ ne sont pas assez variés.

3.3.2.2 Mouvement affine dans l'image

En se concentrant sur le calcul du $\boldsymbol{\delta u}$ (Eq. (3.49)) pour clarifier le propos, le modèle de mouvement d'une région $\mathcal{R}^*$ en $\mathcal{R}$ doit être étendu de la translation à deux dimensions ($\boldsymbol{\delta u}$) à une transformation plus générale, à plus de degrés de liberté pour que le recalage soit possible, même si la pose relative $\boldsymbol{\delta p}$ entre $\mathscr{F}_c$ et $\mathscr{F}_{c^*}$ n'est pas une translation pure parallèle aux plans images des deux caméras c et c^*. Ainsi, considérer que les deux régions sont liées par transformation affine [Shi and Tomasi, 1994] permet de considérer une pose relative $\boldsymbol{\delta p}$ plus libre. La fonction de coût $\mathscr{C}_{LK}$ (Eq. (3.47)) évolue donc en :

$$\mathscr{C}_{ST}(\{\mathbf{A}, \boldsymbol{\delta u}\}, \mathcal{I}, \mathcal{R}^*, \mathcal{I}^*) = \frac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I(\mathbf{A}\mathbf{u}^* + \boldsymbol{\delta u}) - I^*(\mathbf{u}^*)\right)^2, \tag{3.61}$$

où $\mathbf{A} \in \mathbb{R}^{2\times 2}$ tel que $\det(\mathbf{A}) \neq 0$.

Les quatre degrés de liberté supplémentaires de la transformation affine de $\mathscr{C}_{ST}()$ (Eq. (3.61)), par rapport à $\mathscr{C}_{LK}()$ (Eq. (3.47)), nécessitent :

— d'une part, qu'il y ait encore plus de contrastes variés dans la région d'intérêt qu'en translation pure : considérer une région de plus grandes dimensions peut apporter plus de variété [Hager and Belhumeur, 1998], mais augmente le risque de discontinuité de profondeur dans la scène pour la région considérée

— d'autre part, que le mouvement entre les deux régions à recaler soit suffisamment important pour réduire les ambiguïtés [Shi and Tomasi, 1994], c'est-à-dire que les six degrés de liberté de la transformation affine puissent être déterminés : le mouvement entre les régions doit cependant rester raisonnable pour que l'hypothèse de scène lambertienne (cf. Partie 3.2.1) ne soit pas violée.

Pour équilibrer les deux points ci-dessus, Shi et Tomasi [Shi and Tomasi, 1994] proposent d'estimer uniquement $\boldsymbol{\delta u}$ entre images successives d'une séquence de scène ou de caméra en mouvement et d'ajouter l'estimation de $\mathbf{A}$ seulement pour vérifier la cohérence du contenu de la région suivie dans l'image actuelle $\mathcal{I}(t_0 + n\delta t), n \in \mathbb{N}\backslash\{1\}$, par rapport à l'image d'origine $\mathcal{I}(t_0)$. Ce mécanisme permet de valider les régions à suivre dans une image, c'est-à-dire celles qui peuvent être suivies par l'algorithme, ce qui est aussi une façon de détecter les occultations. Ce principe, combiné à une première sélection des régions dans $\mathcal{I}(t_0)$ centrées autour de pixels dont le voisinage

est très varié mène au "célèbre tracker" KLT, des noms des principaux contributeurs Kanade, Lucas et Tomasi. Une autre raison de considérer des régions centrées autour de pixels dont le voisinage est très varié est que ces pixels représentent souvent un point géométrique dans l'espace 3D de la scène, donc indispensable pour une reconstruction de la structure 3D de la scène observée (cf. Partie 3.3.3).

C'est considérer la transformation affine qui fait apparaître la nécessité de générer une image intermédiaire $\mathcal{I}^{(k)}$ (cf. Partie 3.2.3) à chaque itération de l'optimisation de $\mathscr{C}_{ST}()$ (Eq. (3.61)). En effet, pour cette dernière, les axes horizontaux et verticaux de $\mathcal{I}$ ne sont plus contraints d'être parallèles et coplanaires à ceux de $\mathcal{I}^*$, contrairement à l'optimisation de $\mathscr{C}_{LK}()$ (Eq. (3.47)). Par conséquent, le calcul des gradients spatiaux de l'image doit se faire dans l'image $\mathcal{I}^{(k)}$. L'approche compositionnelle inverse présente, alors, un double intérêt pour résoudre $\mathscr{C}_{ST}()$ (Eq. (3.61)). D'abord, comme pour tous les types de transformation inversible, elle permet d'éviter la transformation géométrique d'image pour le calcul des gradients spatiaux de l'image, puisque ces derniers sont calculés dans $\mathcal{I}^*$, qui est fixe. Ensuite, l'approche compositionnelle inverse permet de réduire considérablement les temps de calcul, sans impact sur le résultat final, par rapport aux approches additives et compositionnelles (cf. Partie 3.2.3), ce qui est déjà vrai quand la transformation est une translation pure, mais encore plus quand le nombre de degrés de liberté est important.

3.3.2.3 Mouvement projectif dans l'image

Dans le cas où la région $\mathcal{R}^*$ à suivre dans l'image correspond à une zone plane de la scène, d'orientation quelconque, la transformation géométrique entre $\mathcal{R}^*$ et $\mathcal{R}$ est une homographie, tel que :

$$ {}^{c}\mathbf{H}_{c^*} = \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \in \mathbb{R}^{3\times 3} : {}^{c^*}\tilde{\mathbf{u}} \in \mathbb{P}^2 \longmapsto {}^{c}\tilde{\mathbf{u}} \in \mathbb{P}^2. \tag{3.62} $$

${}^{c}\mathbf{H}_{c^*}$ est une matrice 3×3 et c'est une transformation projective, donc définie à un facteur près, ce qui lui confère 8 degrés de liberté. Une contrainte supplémentaire est donc nécessaire pour l'estimer et on choisit généralement de fixer l'un de ses neufs éléments à 1 : $h_{33} = 1$, donc le vecteur des degrés de liberté est $\mathbf{h}_{\mathbb{P}^2} = [h_{11}, h_{12}, h_{13}, h_{21}, h_{22}, h_{23}, h_{31}, h_{32}]^\top \in \mathbb{R}^8$. Une possibilité alternative est de contraindre $\det(\mathbf{H}) = 1$, notamment pertinente pour de grands déplacements [Benhimane and Malis, 2004]. Cette contrainte fait appartenir $\mathbf{H}$ au groupe spécial linéaire $SL(3)$, dont l'algèbre de Lie $\mathfrak{sl}(3)$ possède l'application exponentielle $\exp_{\mathfrak{sl}(3)}$, valide localement :

$$ \exp_{\mathfrak{sl}(3)} : \mathbf{S}(\mathbf{h}_{\mathfrak{sl}(3)}) \in \mathfrak{sl}(3) \longmapsto \mathbf{H} \in SL(3), \tag{3.63} $$

où les éléments de $\mathbf{h}_{\mathfrak{sl}(3)} \in \mathbb{R}^8$ servent de poids dans la combinaison linéaire $\mathbf{S}$ de huit générateurs $\mathbf{G}_i$ qui forment une base de $\mathfrak{sl}(3)$ [Benhimane and Malis, 2004] :

$$ \mathbf{S}(\mathbf{h}_{\mathfrak{sl}(3)}) = \sum_{i=1}^{8} \mathbf{h}_{\mathfrak{sl}(3)_i} \mathbf{G}_i. \tag{3.64} $$

Ces générateurs $\mathbf{G}_i$ sont des matrices constantes à trace nulle et linéairement indépendantes [Mei et al., 2006] :

$$\mathbf{G}_1 = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_3 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_4 = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$
$$\mathbf{G}_5 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \mathbf{G}_6 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \mathbf{G}_7 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \mathbf{G}_8 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix}. \tag{3.65}$$

Contrairement à l'exponentielle de l'algèbre $\mathfrak{se}(3)$ (Eq. (3.25)) qui a une expression explicite, à l'aide des formules de Rodrigues, $\exp_{\mathfrak{sl}(3)}(\mathbf{S}(\mathbf{h}_{\mathfrak{sl}(3)}))$ est moins directe et de nombreux algorithmes existent pour la calculer par approximation de Taylor, décomposition en valeurs et vecteurs propres, etc [Golub and Van Loan, 1983]. Avec la matrice d'homographie, le coût $\mathscr{C}_{ST}()$ (Eq. (3.61)) peut être étendu de la transformation affine à la transformation projective en $\mathscr{C}_{BM}()$ [Benhimane and Malis, 2004, Benhimane and Malis, 2007] :

$$\begin{aligned} \mathscr{C}_{BM}(\mathbf{h}, \mathcal{I}, \{\mathcal{R}^*, {}^{\bar{c}}\mathbf{H}_{c^*}\}, \mathcal{I}^*) &= \tfrac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I({}^{c}\mathbf{H}_{\bar{c}}(\mathbf{h}) {}^{\bar{c}}\mathbf{H}_{c^*} \, \tilde{\mathbf{u}}^*) - I^*(\tilde{\mathbf{u}}^*) \right)^2 \\ &= \tfrac{1}{2} ||\mathbf{I}(\mathbf{h}) - \mathbf{I}^*||^2 \\ &= \tfrac{1}{2} ||\mathbf{C}_{\mathbf{BM}}(\mathbf{h})||^2, \end{aligned} \tag{3.66}$$

$\mathbf{h}$ étant $\mathbf{h}_{\mathbb{P}^2}$ ou $\mathbf{h}_{\mathfrak{sl}(3)}$ suivant la représentation choisie et où $\tilde{\mathbf{u}}^* = [u^*, v^*, 1]^\top$, en coordonnées homogènes, est l'équivalent de $\mathbf{u}^* = [u^*, v^*]^\top$, en coordonnées euclidiennes. Le paramètre ${}^{\bar{c}}\mathbf{H}_{c^*}$ indique une homographie *proche* de la solution et $\mathbf{h}$ représente donc l'incrément de transformation projective. L'accès à l'intensité $I(\tilde{\mathbf{u}})$, tel que :

$$\begin{gathered} \tilde{\mathbf{u}} \propto {}^{c}\mathbf{H}_{c^*} \, \tilde{\mathbf{u}}^* \\ \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \propto \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} u^* \\ v^* \\ 1 \end{bmatrix} \end{gathered} \tag{3.67}$$

se fait après passage de $\tilde{\mathbf{u}}$ en coordonnées euclidiennes :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} \dfrac{h_{11}u^* + h_{12}v^* + h_{13}}{h_{31}u^* + h_{32}v^* + h_{33}} \\ \dfrac{h_{21}u^* + h_{22}v^* + h_{23}}{h_{31}u^* + h_{32}v^* + h_{33}} \end{bmatrix}. \tag{3.68}$$

Minimiser $\mathscr{C}_{BM}()$ (Eq. (3.66)) pour $\mathbf{h} = \mathbf{h}_{\mathbb{P}^2}$ demande l'expression de la matrice jacobienne des intensités de l'image, courante pour les approches additive ou compositionnelle, désirée pour l'approche compositionnelle inverse, par rapport à $\mathbf{h}_{\mathbb{P}^2}$,

soit, pour l'approche compositionnelle inverse :

$$\left.\frac{\partial \mathbf{C_{BM}}(\mathbf{h}_{\mathbb{P}2})}{\partial \mathbf{h}_{\mathbb{P}2}}\right|_{\mathbf{h}_{\mathbb{P}2}=\mathbf{e}} = \left.\left[\begin{array}{c}\vdots\\ \left(\nabla_{\mathbf{u}} I^* \frac{\partial \mathbf{u}}{\partial \mathbf{h}_{\mathbb{P}2}}\right)\Big|_{\mathbf{u}=\mathbf{u}_i}\\ \vdots\end{array}\right]\right|_{\mathbf{h}_{\mathbb{P}2}=\mathbf{e}}, \tag{3.69}$$

$\mathbf{e}$ étant l'élément neutre de $\mathbf{h}_{\mathbb{P}2}$, c'est-à-dire correspondant à une homographie neutre ($^c\mathbf{H}_{c^*} = \mathbf{I_{3\times 3}}$), soit $\mathbf{e} = [1,0,0,0,1,0,0,0]$. A partir de l'expression générale :

$$\begin{aligned}\frac{\partial \mathbf{u}}{\partial \mathbf{h}} &= \begin{bmatrix}\frac{\partial u}{\partial h_{11}} & \frac{\partial u}{\partial h_{12}} & \cdots & \frac{\partial u}{\partial h_{32}}\\ \frac{\partial v}{\partial h_{11}} & \frac{\partial v}{\partial h_{12}} & \cdots & \frac{\partial v}{\partial h_{32}}\end{bmatrix}\\ &= \frac{1}{h_{31}u^*+h_{32}v^*+h_{33}}\begin{bmatrix}u^* & v^* & 1 & 0 & 0 & 0 & -u^*\alpha & -v^*\alpha\\ 0 & 0 & 0 & u^* & v^* & 1 & -u^*\beta & -v^*\beta\end{bmatrix},\end{aligned} \tag{3.70}$$

avec :

$$\begin{cases}\alpha = \frac{h_{11}u^*+h_{12}v^*+h_{13}}{h_{31}u^*+h_{32}v^*+h_{33}}\\ \beta = \frac{h_{21}u^*+h_{22}v^*+h_{23}}{h_{31}u^*+h_{32}v^*+h_{33}}\end{cases}, \tag{3.71}$$

on a :

$$\left.\frac{\partial \mathbf{u}}{\partial \mathbf{h}_{\mathbb{P}2}}\right|_{\mathbf{h}_{\mathbb{P}2}=\mathbf{e}} = \begin{bmatrix}u^* & v^* & 1 & 0 & 0 & 0 & -u^{*2} & -u^*v^*\\ 0 & 0 & 0 & u^* & v^* & 1 & -u^*v^* & -v^{*2}\end{bmatrix}, \tag{3.72}$$

car, pour rappel $h_{33} = 1$ et l'homographie neutre est telle que $^c\mathbf{H}_{c^*} = \mathbf{I_{3\times 3}}$.

La matrice jacobienne $\frac{\partial \mathbf{u}}{\partial \mathbf{h}_{\mathbb{P}2}}$ courante ne bénéficie pas de la simplification menant à l'équation (3.72) (sauf $h_{33} = 1$), puisque $\mathbf{h}_{\mathbb{P}2}$ est quelconque. Par conséquent, les approches additives, compositionnelles et la méthode d'optimisation ESM demandent plus de calcul par itération. Cependant, en considérant $\mathbf{h}_{\mathfrak{sl}(3)}$, on peut simplifier le calcul de la jacobienne courante en [Benhimane and Malis, 2004] :

$$\left.\frac{\partial \mathbf{C_{BM}}(\mathbf{h}_{\mathfrak{sl}(3)})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}}\right|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{h}^{(k)}_{\mathfrak{sl}(3)}} = \left.\left[\begin{array}{c}\vdots\\ \left(\nabla_{\mathbf{u}} I^{(k)}\big|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{h}^{(k)}_{\mathfrak{sl}(3)}} \left.\frac{\partial \mathbf{u}}{\partial \mathbf{h}_{\mathfrak{sl}(3)}}\right|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{e}}\right)\Big|_{\mathbf{u}=\mathbf{u}_i}\\ \vdots\end{array}\right], \tag{3.73}$$

permettant donc de ne recalculer que les gradients de l'image transformée, $\nabla_{\mathbf{u}} I^{(k)}$, à chaque itération.

Extension à la vision omnidirectionnelle Les méthodes de suivi décrites dans cette partie sont extensibles à la vision omnidirectionnelle en introduisant, tout d'abord, une transformation supplémentaire, celle du plan image à la sphère unitaire du modèle de projection centrale unifié (Partie 2.1.2) [Mei et al., 2006]. En effet, contrairement à la caméra perspective pour laquelle l'homographie $^c\mathbf{H}_{c^*}$ est valide dans le plan image, $^c\mathbf{H}_{c^*}$ est valide sur la sphère du modèle de projection unifié

mais pas directement dans le plan image, c'est-à-dire pour des points sphériques $\mathbf{X}_S$ (Eq. (2.5)) :

$$\begin{aligned} \mathbf{X}_S &\propto {}^c\mathbf{H}_{c^*}\, \mathbf{X}_S^* \\ \begin{bmatrix} X_S \\ Y_S \\ Z_S \end{bmatrix} &\propto \begin{bmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{bmatrix} \begin{bmatrix} X_S^* \\ Y_S^* \\ Z_S^* \end{bmatrix}. \end{aligned} \tag{3.74}$$

Il convient donc de reformuler le coût $\mathscr{C}_{BM}()$ (Eq. (3.66)) en [Mei et al., 2006] :

$$\begin{aligned} &\mathscr{C}_{M+_{\mathfrak{sl}(3)}}(\mathbf{h}_{\mathfrak{sl}(3)}, \mathcal{I}, \mathcal{P}_{M+_{\mathfrak{sl}(3)}}, \mathcal{I}^*) \\ &\quad = \tfrac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I(pr_{\gamma_u}({}^c\mathbf{H}_{\bar{c}}(\mathbf{h}_{\mathfrak{sl}(3)})\, {}^{\bar{c}}\mathbf{H}_{c^*}\, pr_{\gamma_u}^{-1}(\mathbf{u}^*))) - I^*(\mathbf{u}^*) \right)^2 \\ &\quad = \tfrac{1}{2} ||\mathbf{I_o}(\mathbf{h}_{\mathfrak{sl}(3)}) - \mathbf{I_o}^*||^2 \\ &\quad = \tfrac{1}{2} ||\mathbf{C}_{\mathbf{M}+_{\mathfrak{sl}(3)}}(\mathbf{h}_{\mathfrak{sl}(3)})||^2, \end{aligned} \tag{3.75}$$

où $\mathcal{P}_{M+_{\mathfrak{sl}(3)}} = \left\{ \mathcal{R}^*, {}^{\bar{c}}\mathbf{H}_{c^*}, \gamma_u \right\}$. Ci-dessus, $\mathbf{I_o}(\mathbf{h}_{\mathfrak{sl}(3)})$ indique que l'utilisation de $\mathbf{h}_{\mathfrak{sl}(3)}$ pour accéder aux intensités de l'image courante est différente de $\mathbf{I}(\mathbf{h})$ (Eq. (3.66))). L'indice $_\mathbf{o}$ est aussi utilisé pour les intensités désirées $\mathbf{I_o}^*$, même s'il n'y a pas de transformation de cette image, uniquement par souci de cohérence des écritures. Minimiser $\mathscr{C}_{M+}()$ demande l'expression de la matrice jacobienne suivante :

$$\frac{\partial \mathbf{C}_{\mathbf{M}+_{\mathfrak{sl}(3)}}(\mathbf{h}_{\mathfrak{sl}(3)})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}} = \begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{u}} I_o^{(k)} \frac{\partial \mathbf{u}}{\partial \mathbf{x}} \frac{\partial \mathbf{x}}{\partial \mathbf{X}_S} \frac{\partial \mathbf{X}_S}{\partial \mathbf{h}_{\mathfrak{sl}(3)}} \right)\Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{bmatrix}, \tag{3.76}$$

où $\nabla_{\mathbf{u}} I_o^{(k)}$ est le gradient spatial de l'image omnidirectionnelle, calculé classiquement, $\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$ est donné par l'équation (3.57) et :

$$\begin{aligned} \frac{\partial \mathbf{x}}{\partial \mathbf{X}_S} &= \begin{bmatrix} \frac{\partial x}{\partial X_S} & \frac{\partial x}{\partial Y_S} & \frac{\partial x}{\partial Z_S} \\ \frac{\partial y}{\partial X_S} & \frac{\partial y}{\partial Y_S} & \frac{\partial y}{\partial Z_S} \end{bmatrix} \\ &= \frac{1}{Z_S+\xi} \begin{pmatrix} 1 & 0 & -\frac{X_S}{Z_S+\xi} \\ 0 & 1 & -\frac{Y_S}{Z_S+\xi} \end{pmatrix} \end{aligned} \tag{3.77}$$

(dérivée de l'équation (2.6)).

La paramétrisation par groupe de Lie permet de simplifier l'évaluation de la dernière matrice jacobienne de l'équation (3.76) qui n'est à évaluer qu'une seule fois pour $\mathbf{h}_{\mathfrak{sl}(3)} = \mathbf{0}$ (menant à ${}^c\mathbf{H}_{\bar{c}} = \mathbf{I_{3\times 3}}$) [Mei et al., 2006] :

$$\left.\frac{\partial \mathbf{X}_S}{\partial \mathbf{h}_{\mathfrak{sl}(3)}}\right|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}} = \left.\frac{\partial \mathbf{X}_S}{\partial \mathrm{flat}(\mathbf{H})}\right|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}} \left.\frac{\partial \mathrm{flat}(\mathbf{H})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}}\right|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}} \tag{3.78}$$

avec :

$$\left.\frac{\partial \mathbf{X}_S}{\partial \mathrm{flat}(\mathbf{H})}\right|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}} = \begin{bmatrix} \mathbf{X}_S^{*\top} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_S^{*\top} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{X}_S^{*\top} \end{bmatrix} \tag{3.79}$$

et :

$$\left.\frac{\partial \text{flat}(\mathbf{H})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}}\right|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}} = \left[\text{flat}(\mathbf{G}_1)^\top \quad \text{flat}(\mathbf{G}_2)^\top \quad \dots \quad \text{flat}(\mathbf{G}_8)^\top\right], \tag{3.80}$$

où $\text{flat}(\mathbf{G}) = [G_{11}, G_{12}, G_{13}, G_{21}, \dots, G_{33}]$.

Enfin, le calcul de l'incrément $\dot{\mathbf{h}}_{\mathfrak{sl}(3)}$ par l'ESM se fait de la façon suivante [Mei et al., 2006] :

$$\dot{\mathbf{h}}^{(k)}_{\mathfrak{sl}(3)} = -\left[\begin{array}{c} \vdots \\ \left(\left(\frac{\nabla_{\mathbf{u}} I_o^{(k)} + \nabla_{\mathbf{u}} I_o^*}{2}\right)\left(\mathbf{J}_{\mathbf{u}\mathbf{X}_\mathbf{S}}\mathbf{J}_{\mathbf{X}_\mathbf{S}\mathbf{h}_{\mathfrak{sl}(3)}}\right)\Big|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}}\right)\Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{array}\right]^\dagger \mathbf{C}_{\mathbf{M}+_{\mathfrak{sl}(3)}}(\mathbf{h}^{(k)}_{\mathfrak{sl}(3)}), \tag{3.81}$$

où

$$\mathbf{J}_{\mathbf{u}\mathbf{X}_\mathbf{S}} = \left(\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{x}}{\partial \mathbf{X}_S}\right), \tag{3.82}$$

et

$$\mathbf{J}_{\mathbf{X}_\mathbf{S}\mathbf{h}_{\mathfrak{sl}(3)}} = \frac{\partial \mathbf{X}_S}{\partial \mathbf{h}_{\mathfrak{sl}(3)}}. \tag{3.83}$$

On remarque que puisque la fonction $pr_{\gamma_u}()$ de projection de la sphère du modèle unifié vers le plan image (et son inverse) est utilisée dans le suivi de plan par caméra omnidirectionnelle selon le coût $\mathscr{C}_{M+}()$ (Eq. (3.84)), il est nécessaire que la caméra soit étalonnée afin que ses paramètres intrinsèques γ_u soient connus. Pour se rapprocher du cas de la vision perspective dans lequel une homographie est directement valide dans le plan image, c'est-à-dire que le suivi peut se faire sans que la caméra ne soit étalonnée, il faut étendre la fonction de coût $\mathscr{C}_{M+}()$ en basculant les paramètres intrinsèques γ_u dans les degrés de liberté du problème. Cela permet de ne pas nécessiter la connaissance des paramètres intrinsèques γ_u de la caméra omnidirectionnelle a priori. $\mathscr{C}_{M+}()$ devient alors $\mathscr{C}_{SMM}()$ [Salazar-Garibay et al., 2009] :

$$\begin{aligned} &\mathscr{C}_{SMM}(\mathcal{D}_{SMM}, \mathcal{I}, \{\mathcal{R}^*, {}^{\bar{c}}\mathbf{H}_{c^*}\}, \mathcal{I}^*) \\ &\quad = \tfrac{1}{2}\sum_{\mathbf{u}^*\in\mathcal{R}^*}\left(I(pr_{\gamma_u}({}^{c}\mathbf{H}_{\bar{c}}(\mathbf{h}_{\mathfrak{sl}(3)})\,{}^{\bar{c}}\mathbf{H}_{c^*}\, pr^{-1}_{\gamma_u}(\mathbf{u}^*))) - I^*(\mathbf{u}^*)\right)^2 \\ &\quad = \tfrac{1}{2}||\mathbf{I_o}(\mathcal{D}_{SMM}) - \mathbf{I_o}^*||^2 \\ &\quad = \tfrac{1}{2}||\mathbf{C_{SMM}}(\mathcal{D}_{SMM})||^2, \end{aligned} \tag{3.84}$$

où $\mathcal{D}_{SMM} = \left\{\mathbf{h}_{\mathfrak{sl}(3)}, \gamma_u\right\}$. Minimiser $\mathscr{C}_{SMM}()$ ci-dessus passe donc par le calcul de l'incrément $\dot{\gamma_u}$ des paramètres intrinsèques de la caméra omnidirectionnelle en plus de $\dot{\mathbf{h}}_{\mathfrak{sl}(3)}$ (Eq. (3.114)) :

$$\begin{bmatrix} \dot{\mathbf{h}}^{(k)}_{\mathfrak{sl}(3)} \\ \dot{\gamma_u} \end{bmatrix} = -\left[\begin{array}{c} \vdots \\ \left(\left(\frac{\nabla_{\mathbf{u}} I_o{}^{(k)} + \nabla_{\mathbf{u}} I_o{}^*}{2}\right)\begin{bmatrix}\mathbf{J}_{\mathbf{u}\mathbf{h}_{\mathfrak{sl}(3)}} & \mathbf{J}_{\mathbf{u}\gamma_\mathbf{u}}\end{bmatrix}\Big|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}}\right)\Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{array}\right]^\dagger \mathbf{C_{SMM}}(\mathcal{D}_{SMM}), \tag{3.85}$$

avec $\mathbf{J}_{\mathbf{uh}_{\mathfrak{sl}(3)}} = \mathbf{J}_{\mathbf{uX_S}}\mathbf{J}_{\mathbf{X_S h}_{\mathfrak{sl}(3)}}$ et :

$$\mathbf{J}_{\mathbf{u}\gamma_\mathbf{u}} = \frac{\partial \mathbf{u}}{\partial \gamma_u} = \begin{pmatrix} \frac{\partial u}{\partial \alpha_u} & \frac{\partial u}{\partial \alpha_v} & \frac{\partial u}{\partial u_0} & \frac{\partial u}{\partial v_0} & \frac{\partial u}{\partial \xi} \\ \frac{\partial v}{\partial \alpha_u} & \frac{\partial v}{\partial \alpha_v} & \frac{\partial v}{\partial u_0} & \frac{\partial v}{\partial v_0} & \frac{\partial v}{\partial \xi} \end{pmatrix}, \tag{3.86}$$

où, en dérivant l'équation (2.8), on a :

$$\frac{\partial \mathbf{u}}{\partial \gamma_u} = \begin{pmatrix} x & 0 & 1 & 0 & -\frac{\alpha_u x}{Z+\xi} \\ 0 & y & 0 & 1 & -\frac{\alpha_v y}{Z+\xi} \end{pmatrix}. \tag{3.87}$$

Plusieurs régions planes peuvent être suivies en parallèle grâce aux méthodes présentées ci-dessus. Cependant, quand les plans sont rigidement liés entre eux dans la scène (ex : murs), on abandonne la représentation de la transformation sur SL(3), minimale pour un plan, au profit d'une représentation sur SE(3), pour la transformation rigide ${}^c\mathbf{M}_{c^*}$, à laquelle on ajoute les paramètres des plans suivis, à savoir un vecteur normal au plan ${}^{c^*}\mathbf{n}_j \in \mathbb{R}^3$, tel que $||{}^{c^*}\mathbf{n}_j|| = 1$, et la distance orthogonale ${}^{c^*}d_j \in \mathbb{R}$ du plan à l'origine du repère $\mathscr{F}_{c^*}$, pour chaque région $\mathcal{R}_j^*$, plane dans la scène, considérée [Mei et al., 2008]. En effet, la transformation entre deux régions $\mathcal{R}_j^*$ et $\mathcal{R}_j$ correspondantes reste une homographie mais en utilisant son expression spatiale [Hartley and Zisserman, 2004, Part. 13.1, p. 327] :

$$\mathbf{H}_j = {}^c\mathbf{R}_{c^*} - \frac{{}^c\mathbf{t}_{c^*}\,{}^{c^*}\mathbf{n}_j^{\mathsf{T}}}{{}^{c^*}d_j}, \text{ avec } {}^c\mathbf{t}_{c^*} \in \mathbb{R}^3 \text{ et } {}^c\mathbf{R}_{c^*} \in SO(3), \tag{3.88}$$

et où ${}^c\mathbf{R}_{c^*}$ et ${}^c\mathbf{t}_{c^*}$ sont le bloc rotation et le vecteur translation de ${}^c\mathbf{M}_{c^*} \in SE(3)$, respectivement. ${}^c\mathbf{M}_{c^*}$ est commune à tous les plans suivis, et représente donc une contrainte commune explicite, alors que ${}^{c^*}\mathbf{n}_j$ et ${}^{c^*}d_j$ sont propres à chaque plan.

En considérant à nouveau la caméra omnidirectionnelle étalonnée, $\mathscr{C}_{M+_{\mathfrak{sl}(3)}}()$ (Eq. (3.84)) se reformule sur SE(3), en prenant en compte p régions planes $\mathcal{R}_j^*, \forall j \in \{1, 2, \ldots, p\}$ [Mei et al., 2008] :

$$\begin{aligned} &\mathscr{C}_{M+_{\mathfrak{se}(3)}}(\mathcal{D}_{M+_{\mathfrak{se}(3)}}, \mathcal{I}, \mathcal{P}_{M+_{\mathfrak{se}(3)}}, \mathcal{I}^*) \\ &\quad = \tfrac{1}{2}\sum_{j=1}^{p}\sum_{\mathbf{u}^*\in\mathcal{R}_j^*}\left(I(pr_{\gamma_u}(\mathbf{H}_j(\boldsymbol{\delta}\mathbf{p}, {}^{c^*}\mathbf{n}_{\mathbf{d}j})\, pr_{\gamma_u}^{-1}(\mathbf{u}^*))) - I^*(\mathbf{u}^*)\right)^2 \\ &\quad = \tfrac{1}{2}||\mathbf{I}_\mathbf{o}(\mathcal{D}_{M+_{\mathfrak{se}(3)}}) - \mathbf{I_o}^*||^2 \\ &\quad = \tfrac{1}{2}||\mathbf{C}_{\mathbf{M}+_{\mathfrak{se}(3)}}(\mathcal{D}_{M+_{\mathfrak{se}(3)}})||^2, \end{aligned} \tag{3.89}$$

avec $\mathcal{D}_{M+_{\mathfrak{se}(3)}} = \left\{\boldsymbol{\delta}\mathbf{p}, \left\{{}^{c^*}\mathbf{n}_{\mathbf{d}j}\right\}_1^p\right\}$, où l'ensemble $\left\{{}^{c^*}\mathbf{n}_{\mathbf{d}j}\right\}_1^p$ est tel que chaque ${}^{c^*}\mathbf{n}_{\mathbf{d}j} = {}^{c^*}\mathbf{n}_j/{}^{c^*}d_j$ est associé à une région plane $\mathcal{R}_j^*$ de l'ensemble $\mathcal{P}_{M+_{\mathfrak{se}(3)}} = \left\{\mathcal{R}_j^*\right\}_1^p$, $\forall j \in \{1, 2, \ldots, p\}$. A noter que puisque $||{}^{c^*}\mathbf{n}_j|| = 1$, ${}^{c^*}\mathbf{n}_j$ a deux degrés de liberté. C'est pourquoi ${}^{c^*}\mathbf{n}_j$ et ${}^{c^*}d_j$ sont "rassemblés" en ${}^{c^*}\mathbf{n}_{\mathbf{d}j}$ (trois paramètres pour les trois degrés de liberté). Une fois ${}^{c^*}\mathbf{n}_{\mathbf{d}j}$ estimé, on calcule, trivialement :

$$\begin{cases} {}^{c^*}d_j = \dfrac{1}{||{}^{c^*}\mathbf{n}_{\mathbf{d}j}||} \\ {}^{c^*}\mathbf{n}_j = {}^{c^*}d_i\, {}^{c^*}\mathbf{n}_{\mathbf{d}j}. \end{cases} \tag{3.90}$$

En utilisant l'ESM, l'incrément des degrés de liberté se calcule par [Mei et al., 2008] :

$$\begin{bmatrix} {}^{c^*}\dot{\mathbf{n}}_{\mathbf{d}1}^{(k)} \\ {}^{c^*}\dot{\mathbf{n}}_{\mathbf{d}2}^{(k)} \\ \vdots \\ {}^{c^*}\dot{\mathbf{n}}_{\mathbf{d}p}^{(k)} \\ \dot{\boldsymbol{\delta}}\mathbf{p}^{(k)} \end{bmatrix} = - \begin{bmatrix} \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}\mathbf{n}_{\mathbf{d}1}} \\ \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}\mathbf{n}_{\mathbf{d}2}} \\ \vdots \\ \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}\mathbf{n}_{\mathbf{d}p}} \end{bmatrix}^{\dagger} \mathbf{C}_{\mathbf{M}+_{\mathfrak{se}(3)}}(\mathcal{D}^{(k)}_{M+_{\mathfrak{se}(3)}}), \tag{3.91}$$

avec

$$\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}\mathbf{n}_{\mathbf{d}j}} = \begin{bmatrix} \vdots \\ \left(\left(\frac{\nabla_{\mathbf{u}} {I_o}^{(k)} + \nabla_{\mathbf{u}} {I_o}^*}{2} \right) \mathbf{J}_{\mathbf{u}\mathbf{X}_\mathbf{S}} \begin{bmatrix} \mathbf{0}_{3\times 3(j-1)} & \mathbf{J}_{\mathbf{X}_\mathbf{S}\mathbf{n}_{\mathbf{d}j}} & \mathbf{0}_{3\times 3(p-j)} & \mathbf{J}_{\mathbf{X}_\mathbf{S}\boldsymbol{\delta}\mathbf{p}} \end{bmatrix} \Big|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}} \right) \Big|_{\mathbf{u}=\mathbf{u}_i} \\ \vdots \end{bmatrix}, \tag{3.92}$$

$\forall \mathbf{u}_i \in \mathcal{R}_j^*$ et avec $\mathbf{J}_{\mathbf{u}\mathbf{X}_\mathbf{S}}$ défini à l'équation (3.82).

Cependant, le calcul des gradients spatiaux de l'image omnidirectionnelle I_o doit être adapté à ses fortes distorsions, en les calculant directement sur sa représentation sphérique I_S [Demonceaux and Vasseur, 2009].

Les coûts $\mathscr{C}_{M+_{\mathfrak{sl}(3)}}()$ (Eq. (3.84)) et $\mathscr{C}_{M+_{\mathfrak{se}(3)}}()$ (Eq. (3.89)) se reformulent donc sur les images sphériques I_S et I_S^* en :

$$\begin{aligned} \mathscr{C}_{os_{\mathfrak{sl}(3)}}(\mathbf{h}_{\mathfrak{sl}(3)}, \mathcal{I}_S, \mathcal{P}_{C+_{\mathfrak{sl}(3)}}, \mathcal{I}_S^*) &= \tfrac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I_S({}^c\mathbf{H}_{\bar{c}}(\mathbf{h}_{\mathfrak{sl}(3)}) {}^{\bar{c}}\mathbf{H}_{c^*}\, \mathbf{X}_S^*)) - I_S^*(\mathbf{X}_S^*) \right)^2 \\ &= \tfrac{1}{2} ||\mathbf{I}_\mathbf{S}(\mathbf{h}_{\mathfrak{sl}(3)}) - \mathbf{I}_\mathbf{S}{}^*||^2 \\ &= \tfrac{1}{2} ||\mathbf{C}_{\mathbf{os}_{\mathfrak{sl}(3)}}(\mathbf{h}_{\mathfrak{sl}(3)})||^2, \end{aligned} \tag{3.93}$$

où $\mathcal{P}_{os_{\mathfrak{sl}(3)}} = \mathcal{P}_{M+_{\mathfrak{sl}(3)}} = \left\{ \mathcal{R}^*, {}^{\bar{c}}\mathbf{H}_{c^*}, \gamma_u \right\}$ ("os" pour "omnidirectionnel sphérique"), et, respectivement [Caron et al., 2011] :

$$\begin{aligned} \mathscr{C}_{C+_{\mathfrak{se}(3)}}(\mathcal{D}_{C+_{\mathfrak{se}(3)}}, \mathcal{I}_S, \mathcal{P}_{C+_{\mathfrak{se}(3)}}, \mathcal{I}_S^*) &= \tfrac{1}{2} \sum_{j=1}^{p} \sum_{\mathbf{u}^* \in \mathcal{R}_j^*} \left(I_S(\mathbf{H}_j(\boldsymbol{\delta}\mathbf{p}, {}^{c^*}\mathbf{n}_{\mathbf{d}j})\, \mathbf{X}_S^*) - I_S^*(\mathbf{X}_S^*) \right)^2 \\ &= \tfrac{1}{2} ||\mathbf{I}_\mathbf{S}(\mathcal{D}_{C+_{\mathfrak{se}(3)}}) - \mathbf{I}_\mathbf{S}{}^*||^2 \\ &= \tfrac{1}{2} ||\mathbf{C}_{\mathbf{C}+_{\mathfrak{se}(3)}}(\mathcal{D}_{C+_{\mathfrak{se}(3)}})||^2, \end{aligned} \tag{3.94}$$

où $\mathcal{D}_{C+_{\mathfrak{se}(3)}} = \mathcal{D}_{M+_{\mathfrak{se}(3)}} = \left\{ \boldsymbol{\delta}\mathbf{p}, \left\{ {}^{c^*}\mathbf{n}_{\mathbf{d}j} \right\}_1^p \right\}$ et $\mathcal{P}_{C+_{\mathfrak{se}(3)}} = \mathcal{P}_{M+_{\mathfrak{se}(3)}} = \left\{ \mathcal{R}_j^* \right\}_1^p$, $\forall j \in \{1, 2, \ldots, p\}$.

Dans les deux cas ($\mathscr{C}_{os_{\mathfrak{sl}(3)}}()$ et $\mathscr{C}_{C+_{\mathfrak{se}(3)}}()$), le calcul des matrices jacobiennes $\partial \mathbf{C}_{\mathbf{M}+_{\mathfrak{sl}(3)}}(\mathbf{h}_{\mathfrak{sl}(3)}) / \partial \mathbf{h}_{\mathfrak{sl}(3)}$ et $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}\mathbf{n}_{\mathbf{d}j}}$ change, en remplaçant le calcul des gradients spatiaux des images omnidirectionnelles ($\nabla_{\mathbf{u}} {I_o}^{(k)}$ et $\nabla_{\mathbf{u}} {I_o}^*$) par le calcul des gradients

spatiaux de l'image sphérique ($\nabla_{\mathbf{X}_S} {I_S}^{(k)}$ et $\nabla_{\mathbf{X}_S} {I_S}^*$) et en éliminant la jacobienne $\mathbf{J_{uX_S}}$ des expressions, puisque I_S s'échantillonne directement avec $\mathbf{X}_S$.

Le calcul des gradients spatiaux de l'image sphérique $\nabla_{\mathbf{X}_S} I_S$ signifie que l'on calcule la dérivée de I_S selon les trois axes cartésiens. Pour rappel, l'image sphérique est une surface paramétrée par deux angles d'azimuth θ et d'élévation ϕ ($\boldsymbol{\phi} = [\phi, \theta]^\top$, Partie 2.1.2). Les gradients de l'image sphérique sont donc d'abord calculés par rapport à ces paramètres, donnant $\nabla_{\boldsymbol{\phi}} I_S$, puis transformés par la jacobienne $\partial\boldsymbol{\phi}/\partial\mathbf{X}_S$, $\boldsymbol{\phi}$ s'écrivant en fonction de $\mathbf{X}_S$ (Eq. (2.10)). Pour calculer $\nabla_{\boldsymbol{\phi}} I_S$, les voisinages sphériques $\mathcal{A}_\phi$ et $\mathcal{A}_\theta$, de largeur L, linéaires en ϕ et θ, respectivement, sont déterminés par :

$$\begin{cases} \mathcal{A}_\phi = \left\{ \begin{pmatrix} \theta, & \phi + l\Delta_\phi \end{pmatrix}^\top, -\frac{L}{2} \leq l \leq \frac{L}{2}, l \neq 0 \right\} \\ \mathcal{A}_\theta = \left\{ \begin{pmatrix} \theta + l\Delta_\theta, & \phi \end{pmatrix}^\top, -\frac{L}{2} \leq l \leq \frac{L}{2}, l \neq 0 \right\} \end{cases}, \tag{3.95}$$

avec $\Delta_\phi = \Delta_\theta = \arccos(|\begin{pmatrix} 0 & 0 & 1 \end{pmatrix} . pr_\xi^{-1}(\begin{bmatrix} u_0 + 1 & v_0 & 1 \end{bmatrix}^\top)|)$, pour que le pas d'échantillonnage du voisinage corresponde à un pixel au centre de l'image.

Enfin, $\partial\boldsymbol{\phi}/\partial\mathbf{X}_S$ se calcule en dérivant la fonction $c2s()$ (Eq. (2.10)) :

$$\mathbf{J}_{\boldsymbol{\phi}\mathbf{X}_S} = \frac{\partial\boldsymbol{\phi}}{\partial\mathbf{X}_S} = \begin{bmatrix} \frac{\partial\phi}{\partial X_S} & \frac{\partial\phi}{\partial Y_S} & \frac{\partial\phi}{\partial Z_S} \\ \frac{\partial\theta}{\partial X_S} & \frac{\partial\theta}{\partial Y_S} & \frac{\partial\theta}{\partial Z_S} \end{bmatrix} = \begin{bmatrix} 0 & 0 & -\frac{1}{\sqrt{1-Z_S^2}} \\ -\frac{Y_S}{X_S^2(Y_S^2/X_S^2+1)} & \frac{1}{X_S(Y_S^2/X_S^2+1)} & 0 \end{bmatrix}, \tag{3.96}$$

et $\nabla_{\mathbf{X}_S} I_S$ s'obtient par :

$$\nabla_{\mathbf{X}_S} I_S = \nabla_{\boldsymbol{\phi}} I_S \mathbf{J}_{\boldsymbol{\phi}\mathbf{X}_S}. \tag{3.97}$$

3.3.2.4 Asservissement visuel photométrique

En asservissement visuel basé image à six degrés de liberté, c'est l'incrément de pose $\dot{\boldsymbol{\delta}\mathbf{p}}^{(t)} \in \mathfrak{se}(3)$, calculé à l'instant t, qui permet la mise à jour de l'image courante I en déplaçant physiquement la caméra. Ainsi, le coût "générique" $\mathscr{C}_{SSD}()$ (Eq. (3.4)) est reformulé en [Collewet et al., 2008] :

$$\begin{aligned} \mathscr{C}_{CMC}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{Z^*, \gamma_p\}, \mathcal{I}^*) &= \tfrac{1}{2}||\mathbf{I}(\boldsymbol{\delta}\mathbf{p}, Z^*) - \mathbf{I}^*)||^2 \\ &= \tfrac{1}{2}||\mathbf{C_{CMC}}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{Z^*, \gamma_p\}, \mathcal{I}^*)||^2, \end{aligned} \tag{3.98}$$

pour être minimisé par une loi de commande de type Levenberg-Marquardt, similaire à l'équation (3.23)) :

$$\mathbf{v_{CMC}} = \dot{\boldsymbol{\delta}\mathbf{p}}^{(t)} = -\lambda \left(\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(t)}}^\top \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(t)}} + \mu \operatorname{Diag}(\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(t)}}^\top \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(t)}}) \right)^{-1} \mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(t)}}^\top \mathbf{C_{CMC}}(\boldsymbol{\delta}\mathbf{p}^{(t)}), \tag{3.99}$$

$\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(t)}}$ étant exprimée comme en Eq. (3.9), et qui montre le meilleur comportement, malgré la considération du paramètre $Z = Z^*$, constant tout au long de l'asservissement visuel. Ce dernier point, bien que rendant une partie des matrices jacobiennes constante, représente une approximation importante, particulièrement quand l'asservissement visuel est à six degrés de liberté ou quand la scène n'est pas plane. Néanmoins, la redondance d'informations apportée par la considération de tous les

pixels de l'image dans la loi de commande permet une convergence précise malgré cette approximation, tout comme le fait que la scène observée ne soit pas réellement lambertienne ou en présence d'occultation partielle [Collewet et al., 2008].

L'asservissement visuel photométrique s'étend à l'asservissement visuel photométrique 3D en combinant une estimation simultanée de pose relative $\boldsymbol{\delta}\mathbf{p}$ et du Z^*, faite à chaque image acquise durant le déplacement de la caméra, à une loi de commande minimisant la pose relative [Silveira, 2014]. Pour l'estimation, la fonction de coût considérée est la suivante :

$$\begin{aligned} \mathscr{C}_S(\{\boldsymbol{\delta}\mathbf{p}, Z^*\}, \mathcal{I}, \gamma_p, \mathcal{I}^*) &= \tfrac{1}{2}||\mathbf{I}(\boldsymbol{\delta}\mathbf{p}, Z^*) - \mathbf{I}^*)||^2 \\ &= \tfrac{1}{2}||\mathbf{C_S}(\{\boldsymbol{\delta}\mathbf{p}, Z^*\}, \mathcal{I}, \gamma_p, \mathcal{I}^*)||^2, \end{aligned} \tag{3.100}$$

et le problème :

$$[\widehat{\boldsymbol{\delta}\mathbf{p}}, \widehat{Z}^*]^\top = \underset{\boldsymbol{\delta}\mathbf{p}, Z^*}{\arg\min}\, \mathscr{C}_S(\{\boldsymbol{\delta}\mathbf{p}, Z^*\}, \mathcal{I}, \gamma_p, \mathcal{I}^*), \tag{3.101}$$

est résolu par une méthode d'optimisation (cf. Partie 3.2.2) pour chaque image $\mathcal{I}$ acquise. Quant à la loi de commande, elle vise alors à minimiser $||\boldsymbol{\delta}\mathbf{p}||$ par :

$$\mathbf{v_S} = -\lambda \boldsymbol{\delta}\mathbf{p}. \tag{3.102}$$

Cet asservissement visuel photométrique 3D, appliqué aux scènes planes fronto-parallèles à la caméra en la pose désirée [Silveira, 2014] (un seul Z^* à estimer), présente une convergence deux fois plus courte (en terme de nombre d'images acquises au cours de l'asservissement visuel) qu'en considérant une ré-écriture de la loi de commande $\mathbf{v_{CMC}}$ (Eq. (3.99)) en deux phases, l'une d'estimation partielle de $\boldsymbol{\delta}\mathbf{p}$ (Z^* reste constant) à une seule itération, et l'autre de la même loi de commande $\mathbf{v_S}$ (Eq. (3.102)), au prix d'un besoin plus important en puissance de calcul et d'une condition initiale pour la pose relative $\boldsymbol{\delta}\mathbf{p}$ puisqu'elle est explicitement estimée. En pratique, cette approche d'asservissement visuel photométrique 3D est équivalente à un suivi plan, nécessitant une initialisation de ses paramètres, à partir duquel une pose relative est extraite par décomposition de l'homographie et utilisée dans la loi de commande $\mathbf{v_S}$ (Eq. (3.102)). Par conséquent, c'est uniquement la phase d'estimation qui est directe mais pas la loi de commande. L'asservissement visuel est donc 3D mais pas réellement direct.

Enfin, de la même manière que pour le suivi de plan(s), l'asservissement visuel photométrique s'étend à la vision omnidirectionnelle, soit en remplaçant le modèle de projection perspective par le modèle de projection stéréographique dans le calcul des jacobiennes, soit en travaillant à partir de l'image élevée sur la sphère du modèle de projection stéréographique et en adaptant le calcul des gradients spatiaux de l'image sur cette sphère. C'est cette dernière approche qui montre le meilleur comportement de l'asservissement visuel, encore amélioré en calculant directement les gradients spatiaux $\nabla_{\mathbf{X}_S} I_S$ de l'image sphérique, selon les trois axes cartésiens, plutôt qu'en les calculant par rapport aux coordonnées sphériques, suivi de leur transformation cartésienne (Eq. (3.97)) [Caron et al., 2013]. Pour calculer directement $\nabla_{\mathbf{X}_S} I_S$ sans passer par le calcul de $\nabla_{\boldsymbol{\phi}} I_S$, l'idée est la même que pour ce

dernier, sauf qu'au lieu de considérer des voisinages sphériques linéaires $\mathcal{A}_\phi$ et $\mathcal{A}_\theta$ (Eq. (3.95)), on considère des voisinages cartésiens $\mathcal{A}_{X_S}$, $\mathcal{A}_{Y_S}$, $\mathcal{A}_{Z_S}$, reprojetés sur la sphère avant d'être projetés dans l'image omnidirectionnelle pour accéder aux intensités. L'adaptation de l'asservissement visuel photométrique omnidirectionnel à la robotique non-holonome nécessite de découpler les degrés de liberté pour permettre de manoeuvrer *[Alj and Caron, 2015]*.

3.3.3 Localisation et cartographie simultanées

La cartographie consiste à reconstruire en 3D la scène, dans laquelle la caméra est déplacée, à partir des images acquises. Quand les poses de caméras de ces images sont inconnues, la cartographie et la localisation doivent se faire simultanément pour obtenir un résultat cohérent. Ainsi, comme évoqué précédemment (Partie 3.2.5), pour faire la localisation et cartographie simultanément à partir des intensités des pixels des images directement, on bascule $\mathcal{X}$, l'ensemble des points 3D de la scène observables, dans les degrés de liberté du problème d'optimisation :

$$\mathscr{C}_{SSD_{LC}}(\{\boldsymbol{\delta}\mathbf{p}, \mathcal{X}\}, \mathcal{I}, \mathbf{p}, \mathcal{I}^*) = \frac{1}{2} \sum_{\mathbf{X} \in \mathcal{X}} \left(I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t + \delta t) - I(\mathbf{u}(\mathbf{p}, \mathbf{X}), t)\right)^2 . \tag{3.103}$$

Ci-dessus, en général, $\mathbf{p} = \mathbf{0}_6$ et les points 3D $\mathcal{X}$ sont reconstruits dans le repère de la première caméra, tel que $\mathbf{u} = pr_{\gamma_m}(\mathbf{X})$ ($m = p$ en projection perspective, Eq. (2.4) et $m = u$ en projection centrale unifiée, Eq. (2.8)).

Bien entendu, minimiser le coût $\mathscr{C}_{SSD_{LC}}()$ (Eq. (3.103)) généralise le problème d'optimisation de l'équation (3.7) pour obtenir $\widehat{\mathcal{X}}$ en plus de $\widehat{\boldsymbol{\delta}\mathbf{p}}$. Les écritures et dimensions des matrices jacobiennes des méthodes d'optimisation (Partie 3.2.2) sont aussi modifiées et agrandies en conséquence.

Comme pour les autres approches directes pures, l'état de l'art s'est progressivement enrichi en commençant par peu de degrés de liberté de déplacement de la caméra et des scènes modestes en vision perspective [Matthies et al., 1988], pour atteindre des scènes très vastes et très riches en vision panoramique dont les mouvements sur les 6 degrés de liberté cartésiens sont permis [Caruso et al., 2015]. Les contributions intermédiaires majeures concernent la considération d'images clés, une représentation de l'environnement sous forme de voxels/octree plutôt qu'un nuage de points 3D non structuré et des méthodes d'optimisation adaptées à la haute non-linéarité du problème à optimiser, tout comme l'exploitation de la structure particulière (éparse) des matrices jacobiennes pour limiter les temps de calculs dans leurs inversions [Newcombe et al., 2011]. Le besoin de ressources calculatoires est longtemps resté un frein à cette famille de méthodes qui a donc développé des stratégies "semi-denses" de sélection des pixels disposant de suffisamment d'information pour mener à des points 3D fiables et visibles d'autres points de vue [Engel et al., 2013]. Cette méthodologie ayant rapidement trouvé ses limites, il s'est avéré plus efficace de diviser l'ensemble de la scène reconstruite en un graphe de poses de caméras à des images clés auxquelles sont associées des cartes de profondeurs semi-denses [Engel et al., 2014].

La précision de reconstruction s'accroit encore en considérant un modèle de formation photométrique des images (Partie 2.3), soit pour corriger les atténuations et perturbations des intensités de images avant d'appliquer l'un des algorithmes évoqués ci-dessus [Engel et al., 2014], soit pour les prendre directement en compte dans l'algorithme de localisation et cartographie simultanées, en ajoutant les paramètres du modèle de formation photométrique des images dans les degrés de liberté du problème [Bergmann et al., 2018] (cf. Partie 3.3.4.2).

3.3.4 Approches directes robustes

Même si les approches directes pures se montrent plutôt robustes face à la violation de l'hypothèse de base de scène lambertienne, grâce à la redondance d'informations dans l'image (ou dans une région de l'image), cette robustesse atteint ses limites quand cette violation est importante. C'est le cas des scènes aux matériaux réfléchissants (certains métaux et plastiques, vernis, etc) qui rendent l'intensité, supposée localement constante (Eq. (3.2)), variable. C'est aussi le cas en présence d'occultation, de flou de mouvement, ou de réponse variable au cours du temps et du déplacement de la matrice photosensible de la caméra. Enfin, comparer les pixels un à un perd de son sens quand de forts écarts de résolution existent entre les images (ou les régions) courante et désirée.

Trois familles d'approches directes pures robustes traitent ces problème en proposant :

— soit une modélisation statistique pour rejeter les aberrations quand une minorité de l'image est concernée,

— soit de modéliser les mécanismes physiques entrant en jeu (modèle d'illumination, modèle de flou) et en estimant leurs paramètres simultanément au suivi, à l'asservissement visuel ou à la localisation et cartographie simultanées,

— soit de ré-échantillonner l'image.

Ces trois familles d'approches directes pures robustes sont successivement présentées ci-après.

3.3.4.1 Modélisation statistique

Rendre robustes les approches directes par modélisation statistique consiste à réécrire la fonction de coût à l'aide d'une fonction robuste $\mathscr{R}()$ permettant de réduire la sensibilité aux aberrations. La fonction de coût générique $\mathscr{C}_{SSD}()$ (Eq. (3.3)) devient alors [Hager and Belhumeur, 1998] :

$$\mathscr{C}_{rSSD}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) = \frac{1}{2} \sum_{\mathbf{X} \in \mathcal{X}} \mathscr{R}\left(\left(I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t + \delta t) - I(\mathbf{u}(\mathbf{p}, \mathbf{X}), t) \right)^2 \right). \tag{3.104}$$

$\mathscr{R}(a)$ est un estimateur robuste [Huber, 1981] (M-Estimateur) à croissance subquadratique et monotone non décroissant en fonction de $|a|$ croissante. Le problème

d'estimation robuste (3.104) se converti en un problème de moindres carrés repondérés itératifs [Stewart, 1999] et $\mathscr{C}_{rSSD}()$ s'écrit alors, sous forme matricielle :

$$\begin{aligned}\mathscr{C}_{rSSD}(\boldsymbol{\delta}\mathbf{p},\mathcal{I},\{\mathbf{p},\mathcal{X}\},\mathcal{I}^*) &= \tfrac{1}{2}||\mathbf{W}\left(\mathbf{I}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathcal{X})-\mathbf{I}^*(\mathbf{p},\mathcal{X})\right)||^2 \\ &= \tfrac{1}{2}||\mathbf{W}\,\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p},\mathcal{I},\{\mathbf{p},\mathcal{X}\},\mathcal{I}^*)||^2,\end{aligned} \tag{3.105}$$

où $\mathbf{W}$ est une matrice diagonale de pondération. Chaque élément $w_i \in [0,1]$ de la diagonale de $\mathbf{W}$ traduit la confiance en la i-ème primitive (la i-ème intensité du vecteur d'erreur $\mathbf{I}-\mathbf{I}^*$). $\mathbf{W}$ est recalculée à chaque itération de la minimisation de $\mathscr{C}_{rSSD}()$ (3.105), donc l'incrément s'exprime, par exemple avec la méthode de Gauss-Newton :

$$\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} = -\lambda\left(\mathbf{W}\,\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}\right)^{\dagger}\mathbf{W}\,\mathbf{C_{SSD}}(\boldsymbol{\delta}\mathbf{p}^{(k)}). \tag{3.106}$$

Les poids sont calculés selon la fonction robuste $\mathscr{R}()$ choisie suivant un profil donné tel qu'une distribution gaussienne, par exemple (Tukey, Huber, Cauchy [Stewart, 1999]).

Considérer une pré-pondération de chaque intensité $I(\mathbf{u})$ de pixel $\mathbf{u}$ par l'écart-type $\sigma_{\mathcal{A}_\mathbf{u}}$ des intensités de son voisinage $\mathcal{A}_\mathbf{u}$, avant d'appliquer l'estimateur robuste permet de prendre en compte la corrélation spatiale des aberrations dans l'image [Mei et al., 2008]. Les résultats de suivi de plan en étendant le coût $\mathscr{C}_{M+_{\mathfrak{se}(3)}}()$ (3.89) à cette méthode de pondération donne des résultats de suivi par ESM plus stables, plus précis et pour un nombre moindre d'itérations que sans la prise en compte de la corrélation spatiale des aberrations [Mei et al., 2008].

3.3.4.2 Modélisation de la variation de l'éclairage

Rendre robuste les approches directes aux scènes à éclairage variable passe par le relâchement de l'hypothèse de conservation de l'intensité lumineuse (Eq. (3.2)) et donc, plus généralement, de scène lambertienne. On exprime alors une équation modélisant la variation de l'intensité lumineuse selon, par exemple, une hypothèse de changement global de l'intensité lumineuse par une transformation affine globale :

$$I(\mathbf{u}(\mathbf{p},\mathbf{X}),t) = \kappa(t+\delta t)I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathbf{X}),t+\delta t)+\iota(t+\delta t), \tag{3.107}$$

où $\kappa \in \mathbb{R}$ et $\iota \in \mathbb{R}$ sont des paramètres supplémentaires, photométriques cette fois-ci, s'ajoutant aux paramètres géométriques du mouvement $\boldsymbol{\delta}\mathbf{p}$.

Préparer les intensités $\mathcal{I}$ et $\mathcal{I}^*$, en les centrant et en les normalisant pour obtenir $\mathcal{I}_{ZN}$ et $\mathcal{I}^*_{ZN}$ (ZN : Zero-mean Normalized) à considérer directement dans les coûts de type SSD précédemment mentionnés, est une solution naïve mais efficace, comme le montre une partie des contributions de ce document, aussi bien en vision perspective *[Crombez et al., 2014]* qu'en omnidirectionnelle *[Crombez et al., 2015b]* (Partie 5.1), pour compenser les changements d'illumination au cours du suivi de l'asservissement visuel. Cependant, c'est au prix d'ajouter une étape préliminaire à la résolution directe du problème. D'autres approches, plus directes, considèrent les paramètres photométriques comme partie intégrante de l'expression de la fonction de coût à minimiser.

Pour encore plus de généralité, les changements d'illumination de la scène peuvent suivre un modèle à la fois global et local, tel que celui, empirique, de Phong [Phong, 1975] :

$$I_{Ph}(\mathbf{u}(\mathbf{p},\mathbf{X}),t) = k_s(\mathbf{X})\cos^{\eta}\alpha(\mathbf{p},t) + k_d(\mathbf{X})\cos\theta + k_a(\mathbf{X}), \tag{3.108}$$

qui est fait d'une composante spéculaire ($k_s(\mathbf{X})$), diffuse ($k_d(\mathbf{X})$) et ambiante ($k_a(\mathbf{X})$) et supposant une source de lumière ponctuelle (unique et statique, en Eq. (3.108), pour simplifier). $\alpha(\mathbf{p},t)$ est l'angle entre $\mathbf{r}$, c'est-à-dire la réflexion de la lumière provenant de la source sur la surface de la scène en $\mathbf{X}$, et la direction de vue $\mathbf{d}$, qui dépend de la pose $\mathbf{p}$ de la caméra, associée au pixel $\mathbf{u}$. θ est l'angle formé par la normale $\mathbf{n}$ à la surface en $\mathbf{X}$, et la direction $\mathbf{l}$ de la source lumineuse. η permet de régler la taille de la tâche spéculaire autour de $\mathbf{r}$ en $\mathbf{X}$. L'équation (3.108) fait clairement apparaître une dépendance de l'intensité à la pose $\mathbf{p}$ de la caméra.

L'asservissement visuel photométrique s'étend au modèle d'illumination de Phong [Collewet and Marchand, 2009] en ré-écrivant $\mathscr{C}_{CMC}()$ (Eq. (3.98)) en considérant des paramètres supplémentaires décrivant la scène, à savoir $\mathcal{N}$, $\mathcal{K}_s$, $\mathcal{K}_d$, $\mathcal{K}_a$ et $\mathcal{L}$, respectivement les normales à la surface de la scène en chaque point $\mathbf{X} \in \mathcal{X}$ observé, les trois ensembles de coefficients du modèle de Phong pour chaque $\mathbf{X} \in \mathcal{X}$ ($\mathcal{K}_a = \{k_a\}$ peut être considéré : la composante ambiante est commune à toute la scène) et, enfin, les positions des sources lumineuses dans la scène. Pour simplifier, en s'intéressant au cas particulier de la source de lumière directionnelle et co-localisée avec la caméra (anneau de LEDs autour de l'objectif de la caméra) le calcul de la géométrie 3D de la scène à chaque image durant le déplacement de la caméra n'est pas nécessaire, a fortiori quand la matrice d'interaction désirée est seule considérée dans la loi de commande [Collewet and Marchand, 2011]. Considérer la scène faite d'une unique matière, plane et fronto-parallèle à la caméra à la pose désirée simplifie encore le problème. Toutes ces hypothèses font que Z^*, $\mathcal{N} = \{\begin{pmatrix}0 & 0 & -1\end{pmatrix}^\top\}$, $\mathcal{L} = \{\mathbf{0}_{3\times 1}\}$ et $\mathcal{K} = \{k_s, k_d, k_a\}$ sont les mêmes pour tous les pixels de l'image et mènent au coût :

$$\begin{array}{rcl} \mathscr{C}_{CM_{Ph}}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \mathcal{P}_{CM_{Ph}}, \mathcal{I}^*) & = & \frac{1}{2}||\mathbf{I}(\boldsymbol{\delta}\mathbf{p}, \mathcal{P}_{CM_{Ph}}) - \mathbf{I}^*)||^2 \\ & = & \frac{1}{2}||\mathbf{C}_{\mathbf{CM_{Ph}}}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \mathcal{P}_{CM_{Ph}}, \mathcal{I}^*)||^2, \end{array} \tag{3.109}$$

avec $\mathcal{P}_{CM_{Ph}} = \{Z^*, \mathcal{N}, \mathcal{K}, \mathcal{L}, \gamma_p\}$. Enfin, les hypothèses mentionnées ci-dessus engendrent $\cos\theta = 1$ (Eq. 3.108), menant à l'expression suivante de la matrice d'interaction désirée associée à l'intensité de pixel, selon le modèle d'illumination de Phong [Collewet and Marchand, 2009] :

$$\mathbf{J}^*_{\mathbf{Ph}_{\boldsymbol{\delta}\mathbf{p}}} = \left[\begin{array}{c} \vdots \\ \frac{\eta k_s s^{\eta-1}}{||\tilde{\mathbf{x}}||}\begin{pmatrix} \frac{x}{\bar{Z}} & \frac{y}{\bar{Z}} & -\frac{x^2+y^2}{\bar{Z}} & y & -x & 0 \end{pmatrix}\Big|_{\mathbf{x}=\mathbf{x}_i(\mathbf{u}_i)} \\ \vdots \end{array}\right] - \mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}, \tag{3.110}$$

avec $s = \cos\alpha = \mathbf{r}^\top\mathbf{d} = -\begin{pmatrix}0 & 0 & -1\end{pmatrix}\tilde{\mathbf{x}}/||\tilde{\mathbf{x}}||$, $\bar{Z} = Z^*||\tilde{\mathbf{x}}||^2$ et $\mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}$ défini en équation (3.38). La seule différence avec [Collewet and Marchand, 2009] est que $\mathbf{J}^*_{\mathbf{Ph}_{\boldsymbol{\delta}\mathbf{p}}}$

(Eq. (3.110)) est définie en considérant le plan image numérique et non le plan image normalisé, par souci d'uniformisation avec le reste du document, ce qui se retrouve dans l'expression de $\mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}$ (Eq. (3.38) où I^* est dérivée par rapport à $\mathbf{u}$), mais qui est transparent dans la matrice à gauche du signe "$-$" dans l'équation (3.110)[2]. Les paramètres k_s et η sont réglés empiriquement. La différence d'expression des matrices jacobiennes $\mathbf{J}_{\mathbf{Ph}^*_{\boldsymbol{\delta}\mathbf{p}}}$ et $\mathbf{J}^*_{\boldsymbol{\delta}\mathbf{p}}$ permet de prendre en compte, directement dans la loi de commande, la variation d'intensité lumineuse perçue due au mouvement et donc, concrètement de compenser la différence d'apparence entre l'image courante et l'image désirée.

D'autres cas particuliers ont été étudiés avec le modèle d'illumination de Phong en asservissement visuel photométrique, comme la source lumineuse fixe dans la scène [Collewet and Marchand, 2011], par exemple. Ces travaux permettent de rendre l'asservissement visuel photométrique robuste aux variations de l'intensité lumineuse observée pendant l'asservissement visuel mais, à convergence, l'image acquise doit être la même que l'image désirée.

Pour relâcher les hypothèses faites sur les éléments et matières de la scène observée (uniformité, paramètres $\mathcal{K}, \mathcal{N}, \mathcal{L}$ connus), une alternative consiste à ré-écrire l'équation de conservation de l'intensité lumineuse (3.2) en un modèle de variation d'intensité affine semi-global [Silveira and Malis, 2007] :

$$I_{ai}(\mathbf{u}(\mathbf{p},\mathbf{X}),t) = \kappa(\mathbf{u})I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathbf{X}),t+\delta t)+\iota, \tag{3.111}$$

avec $\kappa(\mathbf{u}) \in \mathbb{R}$, un coefficient de variation d'intensité locale et $\iota \in \mathbb{R}$, le coefficient de variation d'intensité globale. L'idée consiste ensuite à dérouler les méthodes d'optimisation évoquées précédemment (Parties 3.2.2 et 3.2.4) en considérant les $\kappa(\mathbf{u})$ et ι comme des degrés de liberté supplémentaires, photométriques, cette fois-ci. Cependant, considérer un $\kappa(\mathbf{u})$ par pixel rendrait le problème de minimisation de la SSD (Eq. (3.7)), étendu pour prendre en compte la variation d'intensité formalisée à l'équation (3.111), insoluble à cause du nombre total de degrés de liberté du problème (les degrés de liberté géométriques, comme $\mathbf{p}$, autant de $\kappa(\mathbf{u})$ que de pixels dans l'image, et ι) supérieur au nombre de mesures (les pixels de l'image, soit $\mathcal{I}$). La solution consiste à diviser l'image en r régions contiguës $\mathcal{R}_j, \forall j \in \{1,2,\ldots,r\}$, bien moindres que des pixels dans l'image, et de considérer un κ_j par $\mathcal{R}_j$. Ainsi, l'équation (3.111) se ré-écrit en :

$$I_{ar}(\mathbf{u}(\mathbf{p},\mathbf{X}),t) = \kappa_j I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathbf{X}),t+\delta t)+\iota, \tag{3.112}$$

tel que $\mathbf{u} \in \mathcal{R}_j, \forall j \in \{1,2,\ldots,r\}$, $r < |\mathcal{U}|$ ($\mathcal{U}$ définit en équation 3.47).

En appliquant ce modèle affine de variation d'intensité lumineuse semi-globale au suivi direct de région $\mathcal{R}^*$, subdivisée en r sous-régions $\mathcal{R}_j, \forall j \in \{1,2,\ldots,r\}$,

2. La reprise des développements mathématiques originaux de l'asservissement visuel photométrique, exploitant le modèle d'illumination de Phong [Collewet and Marchand, 2009], pour exprimer $\frac{\partial I_{Ph}(\mathbf{u}(\mathbf{p},\mathbf{X}),t)}{\partial \mathbf{p}}$, insère uniquement le produit $\frac{\partial \mathbf{x}}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}} = \mathbf{I} \in \mathbb{R}^{2\times 2}$ au sein d'un produit matriciel pour exprimer $\frac{\partial \mathbf{d}}{\partial \mathbf{p}}$. L'impact étant neutre, la partie de l'équation (3.110) à gauche du signe "$-$" est identique dans les cas des plans images numérique et normalisé.

tel que $\mathcal{R}^* = \mathcal{R}_1 \cup \mathcal{R}_2 \cup \ldots \cup \mathcal{R}_r$, correspondant à un plan de la scène (formalisation avec l'algèbre de Lie $\mathfrak{sl}(3)$, cf. Partie 3.3.2), on exprime le coût $\mathscr{C}_{SM}()$ [Silveira and Malis, 2007] :

$$\begin{aligned}\mathscr{C}_{SM}(\left\{\mathbf{h}_{\mathfrak{sl}(3)}, \{\kappa_j\}_1^r, \iota\right\}, \mathcal{I}, \mathcal{R}^*, \mathcal{I}^*) &= \frac{1}{2}\sum_{j=1}^{r}\sum_{\mathbf{u}^*\in\mathcal{R}_j}(\kappa_j I({}^c\mathbf{H}_{c^*}(\mathbf{h})\,\tilde{\mathbf{u}}^*) + \iota - I^*(\tilde{\mathbf{u}}^*))^2 \\ &= \frac{1}{2}||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)}, \{\kappa_j\}_1^r, \iota) - \mathbf{I}^*||^2 \\ &= \frac{1}{2}||\mathbf{C_{SM}}(\mathbf{h}_{\mathfrak{sl}(3)}, \{\kappa_j\}_1^r, \iota)||^2,\end{aligned} \tag{3.113}$$

étendant ainsi le coût $\mathscr{C}_{BM}()$ (Eq. (3.66)), avec $r < |\mathcal{R}^*|$.

La minimisation de $\mathscr{C}_{SM}()$ par ESM étend celle de $\mathscr{C}_{BM}()$ (Eq. (3.66)), aux degrés de liberté photométriques additionnels $\{\kappa_j\}_1^r$ et ι, menant à l'expression suivante des incréments $\dot{\mathbf{h}}_{\mathfrak{sl}(3)}$, $\{\dot{\kappa}_j\}_1^r$ et $\dot{\iota}$:

$$\begin{bmatrix}\dot{\mathbf{h}}_{\mathfrak{sl}(3)}^{(k)} \\ \dot{\kappa}_1^{(k)} \\ \dot{\kappa}_2^{(k)} \\ \vdots \\ \dot{\kappa}_r^{(k)} \\ \dot{\iota}^{(k)}\end{bmatrix} = -\frac{1}{2}\mathbf{J_{SM}}^{\dagger}\mathbf{C_{SM}}(\mathbf{h}_{\mathfrak{sl}(3)}, \{\kappa_j\}_1^r, \iota), \tag{3.114}$$

avec :

$$\mathbf{J_{SM}} = \begin{bmatrix} \vdots & \mathbf{J_{I\kappa_1}}\big|_{\tilde{\mathbf{u}}^*=\tilde{\mathbf{u}}_1^*} & 0 & \ldots & 0 & \vdots \\ \vdots & 0 & \ddots & \vdots & \vdots & \vdots \\ \kappa_j^{(k)}\mathbf{J_{Ih_{\mathfrak{sl}(3)}}}\Big|_{\mathbf{u}=\mathbf{u}_i} & \vdots & 0 & \mathbf{J_{I\kappa_j}}\big|_{\tilde{\mathbf{u}}^*=\tilde{\mathbf{u}}_i^*} & 0 & \mathbf{J_{I\iota}} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \vdots & 0 & \ldots & 0 & 0 & \vdots \end{bmatrix}, \tag{3.115}$$

$\mathbf{J_{Ih_{\mathfrak{sl}(3)}}} = \left(\nabla_{\mathbf{u}} I^{(k)} + \nabla_{\mathbf{u}} I^*\right)\left(\mathbf{J_{uX_S}}\mathbf{J_{X_S h_{\mathfrak{sl}(3)}}}\right)\Big|_{\mathbf{h}_{\mathfrak{sl}(3)}=\mathbf{0}}$, $\mathbf{J_{I\kappa_j}} = I({}^c\mathbf{H}_{c^*}(\mathbf{h})\,\tilde{\mathbf{u}}^*) + I^*(\tilde{\mathbf{u}}^*)$ et $\mathbf{J_{I\iota}} = 2$. Les κ_j et ι sont mis à jour par approche additive. Le raisonnement en sous-régions permet, de plus, de traiter les sous/sur-expositions locales comme des occultations, en les éliminant du problème à convergence. Elles sont détectées par leur uniformité photométrique et leur élimination permet de rendre le suivi robuste aux réflexion spéculaires, pour peu qu'il reste suffisamment de sous-régions exploitables.

Cette dernière approche de suivi de plan apporte l'avantage de ne pas nécessiter de connaissance a priori sur les propriétés de matière de la scène, pas plus que le nombre, la couleur, l'intensité, le type ou la pose des sources lumineuses. De plus, grâce au modèle affine de variation d'illumination semi-global, le suivi de l'objet plan est possible, même quand les intensités de la région suivie sont différentes de celles de $\mathcal{I}^*$, à convergence. Le nombre r de sous-régions est cependant à régler judicieusement

pour obtenir le meilleur compris entre la robustesse aux changements d'illumination, le nombre de degrés de liberté et donc le conditionnement du problème ainsi que les temps de calcul.

De même, quand la définition des intensités désirées $\mathcal{I}^*$ peut être faite hors ligne, la variation d'intensité lumineuse peut se modéliser par une combinaison linéaire de $\mathcal{I}_j^*, \forall j \in \{1, 2, ..., b\}$ acquises à la même pose de caméra $\mathbf{p}^*$, mais avec des conditions d'illumination différentes :

$$I_{cl}(\mathbf{u}(\mathbf{p}, \mathbf{X}), t) = I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t + \delta t) + \mathbf{B}(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X})\boldsymbol{\tau}(t + \delta t), \tag{3.116}$$

où $\mathbf{B}(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}) \in \mathbb{R}^{1\times b}$ et $\tau(t) \in \mathbb{R}^{b\times 1}$. Concrètement, $\mathbf{B}(\mathbf{u})$ est le vecteur des b intensités acquises en $\mathbf{u}$ pour des illuminations différentes. $\boldsymbol{\tau}(t + \delta t)$ est partagé par tous les pixels considérés et forme les degrés de liberté photométriques ajoutés aux degrés de liberté géométriques, de façon similaire au problème de minimisation du coût $\mathscr{C}_{SM}()$ (Eq. (3.113)). On exprime alors le coût $\mathscr{C}_{HB}()$ pour le suivi de région $\mathcal{R}^*$ selon le modèle de mouvement affine [Hager and Belhumeur, 1998] (extension du coût $\mathscr{C}_{ST}$, Eq. (3.61)) :

$$\mathscr{C}_{HB}(\{\mathbf{A}, \boldsymbol{\delta}\mathbf{u}, \boldsymbol{\tau}\}, \mathcal{I}, \mathcal{R}^*, \mathcal{I}^*) = \frac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I(\mathbf{A}\mathbf{u}^* + \boldsymbol{\delta}\mathbf{u}) + \mathbf{B}(\mathbf{A}\mathbf{u}^* + \boldsymbol{\delta}\mathbf{u})\boldsymbol{\tau} - I^*(\mathbf{u}^*)\right)^2. \tag{3.117}$$

La minimisation de $\mathscr{C}_{HB}$ calcule simultanément les incréments de la transformation géométrique $\dot{\mathbf{A}}^{(k)}$ et $\dot{\boldsymbol{\delta}\mathbf{u}}^{(k)}$ et celui de la transformation photométrique $\dot{\boldsymbol{\tau}}^{(k)}$, utilisés pour mettre à jour les degrés de liberté du problème par approche additive [Hager and Belhumeur, 1998]. Si $b = 3$ est théoriquement le minimum, cette méthode est rendue encore plus robuste en remplaçant les intensités acquises pour différentes illuminations en $\mathbf{p}$, pour construire les vecteurs $\mathbf{B}(\mathbf{u})$, par les vecteurs associés aux $b - 2$ valeurs les plus fortes d'une décomposition en valeurs singulières (SVD) d'une base d'images d'un nombre bien supérieur à b, chacune ayant des illuminations différentes. Le premier élément de $\mathbf{B}(\mathbf{u})$ est l'intensité acquise en $\mathbf{p}$ aux conditions d'illuminations nominales ($j = 1$, par exemple) et le deuxième contient l'intensité moyenne de $\mathcal{I}_1^*(\mathcal{R}^*)$, pour pouvoir s'adapter à un changement global d'intensité. Pour les autres éléments de $\mathbf{B}(\mathbf{u})$, en pratique, considérer les valeurs des 4 vecteurs associés aux plus fortes valeurs singulières d'une SVD de 10 images permet de rendre robuste le suivi aux variations d'illumination globales et aux ombres [Hager and Belhumeur, 1998]. Vis-à-vis du coût $\mathscr{C}_{SM}()$ et de sa minimisation, $\mathscr{C}_{HB}()$ nécessite moins de degrés de liberté photométriques, certes au prix d'une phase d'apprentissage.

La plupart des modèles d'intensité variable rapportés ci-dessus (Eqs. (3.107), (3.111), (3.116)) sont écrits pour compenser la variation d'intensité entre $\mathcal{I}^*$ et $\mathcal{I}$ acquis, c'est-à-dire à partir des images seules mais pas de la scène observée. Le modèle d'illumination de Phong (Eq. (3.108)), quant à lui, fait apparaître clairement des paramètres dépendants de la scène, comme la direction de la source lumineuse, mais ses coefficients sont fixés empiriquement et identiques pour toute la scène. D'autre part, une partie des variations d'intensité des pixels

n'est pas imputable à un changement d'illumination de la scène mais à la caméra elle-même, ce à quoi les premiers modèles sont plus robuste que celui de Phong. Ainsi, plutôt que de chercher à optimiser une pose relative pour minimiser un écart entre intensités d'images, le modèle d'intensité variable, même le plus élémentaire (Eq. (3.107)) se ré-écrit en fonction de la luminance $L(.)$ du point 3D $\mathbf{X}$:

$$L(\mathbf{p}, \mathbf{X}, t) = \kappa(t + \delta t) L(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}, t + \delta t) + \iota(t + \delta t). \tag{3.118}$$

Même sans considérer la transformation photométrique affine [Engel et al., 2018], ce serait vraiment la scène qui serait considérée lambertienne [Bergmann et al., 2018] et non son image numérique, comme toutes les autres méthodes évoquées le font. On remonte à la luminance $L(\mathbf{p}, \mathbf{X})$ à partir de l'intensité mesurée en $I(\mathbf{u}(\mathbf{p}, \mathbf{X}))$ en inversant la fonction $f()$ (Eq. (2.16)) de formation photométrique de l'image (Partie 2.3) :

$$L(\mathbf{p}, \mathbf{X}) = \frac{f^{-1}\left(I(\mathbf{u}(\mathbf{p}, \mathbf{X}))\right)}{\delta t_e V(\mathbf{x}(\mathbf{u}))}. \tag{3.119}$$

Pour une image, un point 3D $\mathbf{X}$ étant associé à un unique point $\mathbf{u}$ du plan image numérique, selon les modèles de projection considérés en partie 2.1, plutôt que considérer $L(\mathbf{p}, \mathbf{X})$ directement dans un coût à minimiser, on peut considérer l'éclairement accumulé (Eq. (2.15)) en $\mathbf{x}$ du plan image réel en guise d'intensité image corrigée [Engel et al., 2018] :

$$I_{corr}(\mathbf{p}, \mathbf{X}) = \delta t_e \frac{f^{-1}\left(I(\mathbf{u}(\mathbf{p}, \mathbf{X}))\right)}{V(\mathbf{x}(\mathbf{u}))}, \tag{3.120}$$

pour reformuler tous les coûts précédents en les rendant robustes au vignettage, notamment, assurant, par exemple, un suivi de région d'image plus fiable, qu'en considérant directement $I()$, dans l'ensemble du champ de vue de la caméra, quelque soit le modèle de mouvement considéré. Le compromis que représente la considération de $I_{corr}()$ par rapport à $L()$ permet de conserver un traitement d'image simple pour le calcul de gradients, dans l'image, donc.

Dans ce qui précède, $I_{corr}()$ s'obtient à partir de $I()$ au prix d'un étalonnage photométrique de la caméra a priori [Engel et al., 2016], ou en ligne [Bergmann et al., 2018], pour connaître la fonction de réponse inverse de la caméra f^{-1} et la fonction de vignettage $V()$. Les variations de temps d'exposition au cours du temps pour adapter dynamiquement la caméra à la quantité variable de lumière dans la scène sont directement prises en compte par le facteur d'échelle $\kappa(t)$ (Eq. (3.118)) sur l'intensité (corrigée) [Engel et al., 2018].

Avec la prise en compte explicite de ces trois derniers facteurs de transformation de l'éclairement vers l'intensité de l'image numérique, ils sont de facto sortis du bruit, vis-à-vis duquel, la pondération que l'estimateur robuste (Eq. (3.104)) apporte à chaque intensité mesurée, permet de lutter. L'impact de la compensation du vignettage et de la fonction de réponse de la caméra sur la vision robotique directe a principalement été évalué en localisation et cartographie simultanée [Engel et al., 2018] mais peut clairement être bénéfique à l'ensemble des méthodes rapportées au début

de cette partie. En effet, ces compensations permettent de concentrer la robustesse de ces méthodes sur le reste des perturbations.

3.3.4.3 En présence de mouvement rapide

Un déplacement rapide de la caméra ou d'objet engendre du flou de mouvement dans l'image quand le temps d'exposition de la caméra δt_e est trop important, relativement à la vitesse de déplacement. En effet, l'équation exprimant l'éclairement accumulé sur un photo-site de la caméra (Eq. (2.15)), suppose, implicitement, que le point $\mathbf{x}$ du plan image reçoit la luminance du même point 3D $\mathbf{X}$, moyennant le vignettage, pendant l'accumulation. Pour prendre en compte un mouvement perceptible pendant δt_e, il convient de généraliser l'expression de $E_{acc}()$ (Eq. (2.15)) en :

$$E_{acc}(\mathbf{x}) = \int_0^{\delta t_e} V(\mathbf{x}) L(\mathbf{X}(t))\, dt, \tag{3.121}$$

tel que :

$$\forall\, t,\ \exists\, \mathbf{X}(t) \in \mathcal{X} \text{ tel que } \mathbf{x} = pr_{\cdot}({}^{c^{(t)}}\mathbf{M}_{c^{(0)}}\, {}^{c^{(0)}}\mathbf{M}_o\, {}^o\mathbf{X}), \tag{3.122}$$

${}^o\mathbf{X}$ étant l'expression du point $\mathbf{X} \in \mathcal{X}$ dans un repère fixe $\mathcal{F}_o$, quand la caméra est en mouvement (représenté par la transformation rigide ${}^{c^{(t)}}\mathbf{M}_{c^{(0)}}$, évoluant au cours du temps d'accumulation). La "dépendance" de $\mathbf{X}(t)$ au temps t indique que le point 3D se projetant en $\mathbf{x}$ n'est pas forcément toujours le même dans l'intervalle de temps δt_e. Pendant l'acquisition d'une image, ${}^{c^{(t)}}\mathbf{M}_{c^{(0)}}$ peut représenter un mouvement cohérent avec le mouvement de la caméra (ou de la scène) avant et après cette acquisition. Cette transformation peut donc être la même, à un facteur d'amplitude près. L'expression de l'accumulation de l'éclairement pourrait encore être généralisée au cas où les points 3D de la scène et la caméra sont en mouvement (différents) et au cas où tous les points de la scène ne suivraient pas le même mouvement (ex : scène rigide seulement par morceaux).

La prise en compte du flou de mouvement a été traitée pour améliorer le suivi d'objet plan basé intensités, donc sans étalonnage photométrique. En raisonnant dans l'image numérique et en considérant un modèle de mouvement projectif dans l'image, une intensité $I_{bh}()$ d'une image floue se modélise, à partir d'une version nette $I_{sha}()$ de cette même image, par [Mei and Reid, 2008] :

$$I_{bh}(\tilde{\mathbf{u}}, \mathbf{h}_b) = \int_0^1 I_{sha}(\exp_{\mathfrak{sl}(3)}(-t\, \mathbf{S}(\mathbf{h}_b))\, \tilde{\mathbf{u}})\, dt, \tag{3.123}$$

en considérant une normalisation du temps d'intégration et avec $\mathbf{h}_b \in \mathbb{R}^8$, suffisamment faible pour que $\exp_{\mathfrak{sl}(3)}()$, définie en équation (3.63), ait un sens.

Ensuite, en considérant que les intensités $\mathcal{I}^*$ de la région de référence $\mathcal{R}^*$ sont acquises sans flou de mouvement, la modélisation du flou de mouvement paramétrée par $\mathbf{h}_b$ (Eq. (3.123)) est appliquée aux intensités en $\mathcal{R}^*$ pour en générer une version floue, la plus semblable aux intensités $\mathcal{I}$, simultanément au calcul de l'homographie $\mathbf{h}_{\mathfrak{sl}(3)}$ optimale décrivant le mouvement de l'objet depuis le début

du suivi. Ainsi, la fonction de coût de type SSD $\mathscr{C}_{BM}()$ (Eq.(3.66)) s'étend en $\mathscr{C}_{MR_h}()$ [Mei and Reid, 2008] :

$$\mathscr{C}_{MR_h}(\{\mathbf{h}_{\mathfrak{sl}(3)}, \mathbf{h}_b\}, \mathcal{I}^{(t)}, \mathcal{R}^*, \mathcal{I}^*) = \frac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I({}^c\mathbf{H}_{c^*}(\mathbf{h}_{\mathfrak{sl}(3)})\, \tilde{\mathbf{u}}^*) - I^*_{bh}(\tilde{\mathbf{u}}^*, \mathbf{h}_b) \right)^2, \tag{3.124}$$

avec $I^*_{bh}()$, définie de façon identique à l'équation (3.123), en remplaçant I_{sha} par I^*.

La minimisation de $\mathscr{C}_{MR_h}()$, par approche compositionnelle [Mei and Reid, 2008], permet donc de prendre en compte un flou de mouvement incohérent avec les déplacements de la caméra autour de l'acquisition de l'image courante, au prix de huit degrés de liberté supplémentaires (ceux de $\mathbf{h}_b$) à estimer. Elle permet aussi de prendre en compte un flou de mouvement cohérent avec les déplacements mais le coût $\mathscr{C}_{MR_h}$ se simplifie aussi en remplaçant l'homographie $\mathbf{h}_b$ par un seul paramètre $a \in \mathbb{R}_+$ décrivant l'amplitude du mouvement, pendant l'acquisition de l'image, cohérent avec le mouvement $\mathbf{h}_{(t-\delta t_e,t)} \in \mathbb{R}^8$ effectué entre la région $\mathcal{R}^{(t-\delta t_e)}$ de l'image acquise précédente et $\mathcal{R}^{(t)}$, la courante à déterminer. $\mathscr{C}_{MR_h}$ devient alors $\mathscr{C}_{MR_a}$, dont les degrés de liberté sont $\mathbf{h}_{(t-\delta t_e,t)}$ et a, nécessitant comme paramètre supplémentaire ${}^{c^{(t-\delta t_e)}}\hat{\mathbf{H}}_{c^*}$, c'est-à-dire l'homographie optimale associée à l'image acquise précédente :

$$\begin{aligned} &\mathscr{C}_{MR_a}(\{\mathbf{h}_{(t-\delta t_e,t)}, a\}, \mathcal{I}, \{\mathcal{R}^*, {}^{c^{(t-\delta t_e)}}\hat{\mathbf{H}}_{c^*}\}, \mathcal{I}^*) \\ &= \tfrac{1}{2} \sum_{\mathbf{u}^* \in \mathcal{R}^*} \left(I(\exp_{\mathfrak{sl}(3)}(\mathbf{S}(\mathbf{h}_{(t-\delta t_e,t)}))\, {}^{c^{(t-\delta t_e)}}\hat{\mathbf{H}}_{c^*}\, \tilde{\mathbf{u}}^*) - I^*_{ba}(\tilde{\mathbf{u}}^*, \mathbf{h}_{(t-\delta t_e,t)}, a) \right)^2, \end{aligned} \tag{3.125}$$

avec $I^*_{ba}()$, définie en étendant l'expression de $I^*_{bh}()$ ($I_{bh}()$, Eq. (3.123)), tel que :

$$I_{ba}(\tilde{\mathbf{u}}, \mathbf{h}, a) = \int_0^1 I_{sha}\left(\exp_{\mathfrak{sl}(3)}(-t\, a\, \mathbf{S}(\mathbf{h}))\tilde{\mathbf{u}}\right)\, dt, \text{ tel que } \mathbf{h} \in \mathbb{R}^8, \tag{3.126}$$

où l'intensité nette I_{sha} est remplacée par I^* pour obtenir I^*_{ba}.

$\mathscr{C}_{MR_a}()$ se minimise par approche compositionnelle [Mei and Reid, 2008] ou ESM [Park et al., 2012], menant ainsi à une forme de généralisation de cette dernière méthode d'optimisation.

3.3.4.4 Quand la résolution se dégrade

Quand les changements d'échelle sont importants entre la consigne et la mesure, ce qui est le cas dans le suivi de plan quand l'objet s'éloigne énormément de la caméra alors que sa région de référence $\mathcal{R}^*$ a été acquise quand il était beaucoup plus proche de la caméra, ou quand son orientation devient presque perpendiculaire au plan image, la précision du suivi, en exploitant les coûts précédemment mentionnés, se dégrade, parfois au point de le faire échouer. L'échantillonnage spatial de l'éclairement pour former l'image numérique en est la cause principale et l'interpolation de l'image courante, pour obtenir les intensités $\mathcal{I}$ à comparer avec $\mathcal{I}^*$, ne suffit pas dans les cas extrêmes.

En reprenant l'idée de modifier $\mathcal{I}^*$ pour que ses intensités correspondent le plus possible à celles de $\mathcal{I}$, la dégradation de la résolution se modélise en prenant en compte le déplacement de la caméra (ou de l'objet), selon le modèle de transformation projective $\mathbf{H} \in SL(3)$, dans la formation de l'image numérique. La dégradation de résolution se modélise par une convolution, notée $*$, de l'image d'intensités $I_{hr}()$ avec un filtre gaussien affine, paramétré par l'approximation affine[3] $\mathbf{H}_A \in \mathbb{R}^{2\times 2}$ ($\det(\mathbf{H}_A) \neq 0$) de $\mathbf{H}$ [Ito et al., 2011] :

$$I_{rd}(\mathbf{u}, \mathbf{H}_A) = I_{hr}(\mathbf{u}) * \exp\left(-\frac{1}{2\sigma^2}\mathbf{u}^\top(\mathbf{H}_A^{-1}\mathbf{H}_A^{-\top} - \mathbf{I})^{-1}\mathbf{u}\right), \text{ avec } \sigma \in \mathbb{R}_+^*. \quad (3.127)$$

Le coût $\mathscr{C}_{BM}()$ (Eq. (3.66)), pour $\mathbf{h}_{\mathfrak{sl}(3)}$ s'étend en $\mathscr{C}_{IOD}()$, robuste à la dégradation de résolution [Ito et al., 2011] :

$$\mathscr{C}_{IOD}(\mathbf{h}_{\mathfrak{sl}(3)}, \mathcal{I}, \mathcal{P}_{IOD}, \mathcal{I}^*) = \frac{1}{2}\sum_{\mathbf{u}^*\in\mathcal{R}^*}\left(I({}^c\mathbf{H}_{\overline{c}}(\mathbf{h}_{\mathfrak{sl}(3)})\,{}^{\overline{c}}\mathbf{H}_{c^*}\,\tilde{\mathbf{u}}^*) - I^*_{rd}(\mathbf{u}^*, \mathbf{H}_A({}^{\overline{c}}\mathbf{H}_{c^*}))\right)^2, \quad (3.128)$$

avec $\mathcal{P}_{IOD} = \{\mathcal{R}^*, {}^{\overline{c}}\mathbf{H}_{c^*}\}$. Dans $\mathscr{C}_{IOD}()$, $\mathbf{H}_A$ approxime l'homographie proche de la solution et non l'homographie *totale* ${}^c\mathbf{H}_{\overline{c}}(\mathbf{h}_{\mathfrak{sl}(3)})\,{}^{\overline{c}}\mathbf{H}_{c^*}$ pour bénéficier des simplifications de l'algorithme ESM sur $\mathfrak{sl}(3)$ (Eq. (3.73)). En pratique, cette seconde approximation est acceptable puisque le $\hat{\mathbf{h}}_{\mathfrak{sl}(3)}$ optimal minimisant $\mathscr{C}_{IOD}()$ est faible [Ito et al., 2011].

3.4 Autres approches directes

3.4.1 Critère de corrélation croisée

Alors que les approches *pures* en vision robotique directe reposent sur l'expression d'un coût à minimiser basé SSD (Eq. (3.3)), la corrélation croisée normalisée (ZNCC : zero-mean normalized cross-correlation) représente une alternative. Contrairement à la SSD qui évalue la dissimilarité entre images, la ZNCC évalue la similarité entre images et est, par définition, adaptée aux changements globaux d'intensité dans l'image selon un modèle affine, là où il a fallu préparer les intensités ou étendre le coût à optimiser dans les approches basées SSD en intégrant un modèle de variation d'illumination (ex : Eq. (3.107), pour le modèle affine global). C'est aussi une approche directe puisqu'elle considère directement les intensités des pixels dans l'expression du critère générique $\mathscr{C}_{ZNCC}()$:

$$\begin{aligned}&\mathscr{C}_{ZNCC}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*)\\ &= \frac{\sum_{\mathbf{X}\in\mathcal{X}}\left(I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p}, \mathbf{X})) - m_{\mathcal{I}}\right)\left(I(\mathbf{u}(\mathbf{p}, \mathbf{X})) - m_{\mathcal{I}^*}\right)}{\sqrt{\sum_{\mathbf{X}\in\mathcal{X}}\left(I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p}, \mathbf{X})) - m_{\mathcal{I}}\right)^2}\sqrt{\sum_{\mathbf{X}\in\mathcal{X}}\left(I(\mathbf{u}(\mathbf{p}, \mathbf{X})) - m_{\mathcal{I}^*}\right)^2}},\end{aligned} \quad (3.129)$$

où $m_{\mathcal{I}}$, respectivement $m_{\mathcal{I}^*}$, est la moyenne des intensités $I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p}, \mathbf{X}))$, respectivement $I(\mathbf{u}(\mathbf{p}, \mathbf{X}))$, $\forall \mathbf{X} \in \mathcal{X}$.

3. L'approximation affine réalise l'extraction du bloc 2×2 supérieur gauche de la matrice d'homographie.

Le critère $\mathscr{C}_{ZNCC}()$ (Eq. (3.129)), par le centrage des intensités à comparer et leur normalisation, prend directement en compte une éventuelle transformation affine globale d'illumination. Une autre différence avec $\mathscr{C}_{SSD}()$ (Eq. (3.3)), c'est que $\mathscr{C}_{ZNCC}()$ est maximal à l'optimum, ce qui implique de mettre en place une maximisation :

$$\widehat{\boldsymbol{\delta}\mathbf{p}} = \arg\max_{\boldsymbol{\delta}\mathbf{p}} \mathscr{C}_{ZNCC}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) \,. \tag{3.130}$$

La résolution de l'équation précédente se fait par remontée de gradient ou la méthode de Newton [Scandaroli et al., 2012], par exemple, avec les mêmes caractéristiques que pour la minimisation (Partie 3.2.2). Les approches de Gauss-Newton et Levenberg-Marquardt ne sont théoriquement pas applicables pour traiter la corrélation croisée car, pour rappel, les approximations qu'elles font de la méthode de Newton sont permises par la nature du problème aux moindres carrés à minimiser.

Pour le suivi direct de plan, le critère $\mathscr{C}_{ZNCC}()$ (Eq. (3.129)) se reformule en :

$$\begin{aligned} \mathscr{C}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)}, \mathcal{I}, \mathcal{P}_{SMR}, \mathcal{I}^*) &= \frac{\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})^\top \mathbf{I}^*}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||\ ||\mathbf{I}^*||} \\ &= C_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)}), \end{aligned} \tag{3.131}$$

avec $\mathcal{P}_{SMR} = \{\mathcal{R}^*, {}^{\bar{c}}\mathbf{H}_{c^*}\}$. Pour calculer l'incrément $\dot{\mathbf{h}}_{\mathfrak{sl}(3)}$, la méthode de Newton nécessite l'expression de la matrice jacobienne de $C_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})$ [Scandaroli et al., 2012] :

$$\frac{\partial C_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}} = \left(\frac{\mathbf{I}^*}{||\mathbf{I}^*||} - C_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)}) \frac{\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)}||} \right)^\top \frac{\mathbf{J}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)}||}, \tag{3.132}$$

avec

$$\mathbf{J}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)}) = \frac{\partial \mathbf{C}_{\mathbf{BM}}(\mathbf{h}_{\mathfrak{sl}(3)})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}} - \frac{1}{|\mathcal{R}^*|} \begin{bmatrix} \vdots \\ \mathbf{1}_{1\times|\mathcal{R}|} \frac{\partial \mathbf{C}_{\mathbf{BM}}(\mathbf{h}_{\mathfrak{sl}(3)})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}} \\ \vdots \end{bmatrix}, \tag{3.133}$$

et de la matrice hessienne de $C_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})$, dont l'approximation à la solution suivante [Scandaroli et al., 2012] :

$$\begin{aligned} \frac{\partial^2 C_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})}{\partial \mathbf{h}_{\mathfrak{sl}(3)}^\top \partial \mathbf{h}_{\mathfrak{sl}(3)}} &\approx \frac{1}{\partial \mathbf{h}_{\mathfrak{sl}(3)}^\top \partial \mathbf{h}_{\mathfrak{sl}(3)}} \partial^2 \frac{\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})^\top \mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||\ ||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||} \\ &\approx -\frac{\mathbf{J}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})^\top \mathbf{J}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||^2} \\ &\quad + \frac{\mathbf{J}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})^\top}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||} \frac{\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)}) \mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})^\top}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||^2} \frac{\mathbf{J}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||}, \end{aligned} \tag{3.134}$$

permet de simplifier grandement l'expression et le calcul par rapport à la véritable matrice hessienne de $C_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)})$, tout en gardant sa propriété de matrice définie

négative, nécessaire pour ne pas perturber la maximisation. Grâce à l'approche compositionnelle inverse les avantages calculatoires du coût de type SSD sont conservés. Mais elle peut réduire le domaine de convergence alors l'approche compositionnelle lui est préférée [Scandaroli et al., 2012].

Pour étendre la robustesse du critère $\mathscr{C}_{SMR}()$ (Eq. (3.131)) aux changements d'illumination locaux, y compris aux réflexions spéculaires, et aux occultations partielles, on l'étend en $\mathscr{C}_{SMR_r}()$, en incorporant une pondération de la corrélation croisée [Scandaroli et al., 2012] :

$$\mathscr{C}_{SMR_r}(\mathbf{h}_{\mathfrak{sl}(3)}, \mathcal{I}, \mathcal{P}_{SMR}, \mathcal{I}^*) = \frac{\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})^\top \mathbf{W}\, \mathbf{I}^*}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||_{\mathbf{W}}\, ||\mathbf{I}^*||_{\mathbf{W}}}, \tag{3.135}$$

où la matrice de poids $\mathbf{W} \in \mathbb{R}_+^{*\,|\mathcal{R}^*|\times|\mathcal{R}^*|}$ est diagonale et $||\mathbf{I}||_{\mathbf{W}} = \sqrt{\mathbf{I}^\top \mathbf{W} \mathbf{I}}$.

Chaque élément w_i de la diagonale de $\mathbf{W}$ traduit la confiance en la similitude des intensités correspondantes de $\mathbf{I}$ et $\mathbf{I}^*$ et se calcule comme le produit des poids $w_j^e \in \mathbb{R}$ et $w_i^o \in \mathbb{R}$:

$$w_i = w_j^e(i) w_i^o, \tag{3.136}$$

$w_j^e(i)$ se calcule par sous-région $\mathcal{R}_j$ de la région $\mathcal{R}$ suivie pour la robustesse aux changements d'illumination locaux, sur une idée similaire à l'équation (3.112) modélisant les variations d'intensités par sous-régions. w_i^o se calcule par pixel, pour la robustesse aux occultations et renforcer la robustesse aux spécularités. Un même $w_j^e(i)$ est donc partagé par tous les pixels de coordonnées $\mathbf{u}_i \in \mathcal{R}_j$.

Brièvement, pour calculer chaque valeur $w_j^e(i)$, une classification k-means est appliquée sur les ZNCC des sous-régions $\mathcal{R}_j$ pour séparer la classe $\mathcal{G}^+$ des régions de confiance (les régions aux ZNCC proches de 1), de centroïde m_{w^e}, de la classe $\mathcal{G}^-$ des autres. Les régions $\mathcal{R}_j$ de la classe $\mathcal{G}^+$ voient leur poids $w_j^e(i) = 1$. Les autres régions voient leur poids calculé à partir d'une distribution de Huber [Stewart, 1999], fonction de la distance de leur ZNCC à m_{w^e}.

Quant aux poids w_i^o par pixel, ils sont aussi obtenus à partir d'une distribution de Huber, mais dont les paramètres sont issus du vecteur résiduel $\frac{\mathbf{I}^*}{||\mathbf{I}^*||} - \mathscr{C}_{SMR}(\mathbf{h}_{\mathfrak{sl}(3)}, \mathcal{I}, \mathcal{P}_{SMR}, \mathcal{I}^*) \frac{\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})}{||\mathbf{I}(\mathbf{h}_{\mathfrak{sl}(3)})||}$. Cette dernière pondération est une adaptation relativement directe de l'estimateur robuste introduit dans le coût $\mathscr{C}_{rSSD}()$ (Eq. (3.104)) mais dont la distribution a l'originalité d'être invariante aux changements d'illumination, plutôt que de considérer ces changements comme des aberrations.

3.4.2 Réseaux de neurones convolutionnels

Toutes les approches précédentes reposent sur une modélisation de l'évolution des degrés de liberté d'une transformation géométrique pour amener deux images, ou une ou plusieurs régions correspondantes dans deux images, à coïncider parfaitement. La variété des méthodes à elle-seule illustre l'effort de modélisation qui est toujours en cours pour résoudre ce problème le plus efficacement possible, dans les conditions les plus variées et ce, pour divers types de caméras. Cet effort est nécessaire pour

que les méthodes soient robustes et disposent d'un domaine de convergence étendu malgré l'importante non-linéarité des fonctions de coût.

Une alternative émergente consiste à basculer le modèle d'évolution de la transformation lui-même dans les degrés de liberté du problème. Elle repose sur des approches issues de l'intelligence artificielle, en particulier les réseaux de neurones convolutionnels (CNN : Convolutionnal Neural Networks) et les algorithmes d'apprentissage profond et par renforcement.

Ainsi, des algorithmes comme PoseNet [Kendall et al., 2015] et PoseCNN [Xiang et al., 2018] sont dédiés au calcul de pose d'objet en prenant, en entrée du CNN les pixels de l'image. Même si les taux d'identification d'une instance d'objet recherché dans une image atteignent des sommets avec cette famille de méthodes, elles ne sont pas vraiment directes, dans le sens où le CNN ne donne pas directement une pose aussi précise que les méthodes des parties précédentes. Un des problèmes est de combiner correctement les rotations et les translations de la pose pour leur apprentissage simultané dans la phase d'entraînement du CNN [Tekin et al., 2018]. Des méthodes raisonnant sur SE(3) existent pour du calcul de transformation rigide entre nuages de points 3D [Byravan and Fox, 2017] ou en stéréovision [Peretroukhin and Kelly, 2018] mais pas encore en vision monoculaire passive (ni sur SL(3)). Entraîner le CNN pour prédire l'apparence 2D de l'objet 3D est encore considéré comme la meilleure option pour réaliser l'apprentissage dans le plan, combinant ainsi, implicitement la rotation et la translation de la caméra dans l'espace [Tekin et al., 2018]. Cependant, si le CNN prédit une apparence (géométrique) dans l'image 2D, cela implique ensuite d'ajouter une étape géométrique (BB8 [Rad and Lepetit, 2017], SSD-6D [Kehl et al., 2017], [Tekin et al., 2018]), pour estimer la pose de l'objet. Le lien entre les intensités mesurées pour former l'image et les degrés de liberté du problème n'est donc toujours pas direct.

La pose relative entre deux images peut néanmoins être prédite suffisamment correctement pour commander un robot par asservissement visuel basé pose [Bateux et al., 2018]. Même si l'asservissement visuel ainsi formulé n'est pas direct, le calcul de pose relative entre la pose actuelle de la caméra et la pose correspondant à l'image désirée est, lui, bien la sortie du CNN ayant pris, en entrée, les pixels de l'image courante acquise par la caméra. On peut donc considérer que c'est une approche directe. Plus précisément, les poses relatives entre un jeu de 11000 images et la pose désirée, ainsi que les pixels de ces images, sont exploitées pour l'entraînement du CNN en 50 passes. Sans évaluation de la précision d'estimation de pose relative elle-même, la précision de positionnement d'un bras robotique à 6 degrés de liberté, dans la même scène que celle des images utilisées pour l'apprentissage, est submillimétrique [Bateux et al., 2018], tout comme l'asservissement visuel photométrique [Collewet and Marchand, 2011], par exemple. Les 11000 images nécessaires à l'apprentissage du CNN représentent une base bien supérieure à la seule image désirée nécessaire en entrée de la boucle de commande de l'asservissement visuel photométrique. Ces 11000 images sont cependant synthétisées hors ligne, à partir d'un modèle d'objet plan et texturé semblable à la scène réelle, et elles alimentent l'algorithme d'entraînement du CNN, exécuté hors ligne aussi.

Une variante consiste à entraîner le CNN à prédire la pose relative entre deux caméras, ayant chacune fait l'acquisition d'une image, à partir des pixels des deux images. En utilisant une base de 100000 images naturelles différentes pour texturer le modèle d'objet plan on peut générer aléatoirement deux poses de caméra observant ce plan, donc deux images. L'entraînement, en 50 passes, d'un CNN, en considérant les pixels de ces images associés à leur pose relative, a permis de prédire la pose relative entre une image courante et une image désirée. Cette prédiction est suffisamment correcte (pas d'évaluation de précision d'estimation de pose fournie [Bateux et al., 2018]) pour atteindre une précision centimétrique dans l'asservissement visuel basé pose exploitant la sortie de ce CNN [Bateux et al., 2018], dans une scène dont les images ne faisaient pas partie de la base d'apprentissage. Cela rapproche l'asservissement visuel basé pose exploitant les CNN d'un asservissement visuel classique pour lequel enregistrer une fois l'image désirée dans la scène suffit à exécuter l'asservissement. On peut donc imaginer que le même CNN peut être utilisé dans d'autres scènes (une seule scène est montrée dans [Bateux et al., 2018]).

Cette deuxième version de l'asservissement visuel basé pose exploitant les CNN n'est cependant pas non plus direct, surtout qu'il faut ensuite lancer un asservissement visuel photométrique pour atteindre la précision submillimétrique. Néanmoins, dans les deux versions, la convergence correcte de l'asservissement visuel indique que le modèle d'évolution de la transformation a bien été automatiquement *appris* par le CNN, certes au prix d'une phase d'apprentissage coûteuse en mémoire et en temps. Quoi qu'il en soit, cet asservissement visuel basé pose et CNN conduit à un taux de convergence plus élevé que l'asservissement visuel photométrique, pour la première méthode [Bateux et al., 2018] et une certaine robustesse à un écart initial importante dans l'image par rapport à l'image désirée.

Pour rendre direct l'asservissement visuel basé CNN, il faudrait que le CNN soit entraîné à directement produire les entrées de commande articulaire du robot, a minima les vitesses cartésiennes de la caméra/effecteur. La difficulté d'appliquer les méthodes orientées CNN évoquées ci-dessus en robotique est le besoin d'un nombre très important d'exemples pour entraîner correctement le CNN [Sünderhauf et al., 2018], ne serait-ce que de 10000 à 100000 exemples dans les méthodes rapportées ci-dessus [Bateux et al., 2018]. Il y a donc un fossé entre les approches basées CNN pour le calcul de pose basé modèle (et/ou le suivi de région dans l'image), où des millions d'exemples (synthétiques) sont possibles, et la commande de robot réel, où on ne peut faire cet "apprentissage" à base de millions d'exemples [Sünderhauf et al., 2018]. Pour résoudre ce problème, de récents travaux développent une adaptation d'un CNN entraîné en simulation vers un CNN pour le monde réel (approches sim-to-real) permettant de faire de l'asservissement visuel direct basé CNN [Zhang et al., 2019]. L'asservissement visuel direct basé CNN d'un bras à 7 degrés de liberté est mis en oeuvre pour atteindre une position 3D dans l'espace par rapport à un objet sur fond contrasté, néanmoins robuste à l'encombrement de ce fond.

3.5 Approches directes étendues

Les approches directes pures (Partie 3.3) s'attachent à exprimer le lien le plus direct possible entre les mesures faites par la caméra, c'est-à-dire les intensités des pixels de l'image, et les degrés de liberté du problème d'estimation de pose, de suivi ou de commande du déplacement de la caméra via un robot. Toute la complexité de la résolution du problème est donc concentrée dans la modélisation de la variation des intensités dans l'image et celles des degrés de liberté du problème.

Les approches directes étendues, quant à elles, considèrent une transformée des intensités de l'image, par exemple vers un espace d'échelle (Partie 3.5.4), en entrée de l'algorithme d'estimation ou de la loi de commande. Même si cette transformée peut apparaître comme une étape intermédiaire entre les intensités de l'image acquise et les degrés de liberté du problème, elle est généralement prise en compte dans la méthode de résolution. Il existe cependant quelques exceptions où les transformées des intensités de l'image des approches directes étendues considèrent, certes, directement tous les pixels de l'image (ou de la région d'intérêt) mais c'est leur résultat (ex : distribution d'intensité comme en partie 3.5.1) qui est mis directement en entrée de l'algorithme d'estimation ou de la loi de commande.

L'intérêt majeur de la transformation préliminaire des intensités de l'image dans les approches directes étendues est, selon les approches, de réduire considérablement la dimensionnalité du problème, de mieux maîtriser les trajectoires de la caméra dans l'espace, d'être robuste à des perturbations beaucoup plus importantes dans l'image que celles face auxquelles les approches directes peuvent lutter, comme la multimodalité, ou, enfin, d'étendre considérablement le domaine de convergence du suivi et de l'asservissement visuel directs.

3.5.1 Distribution d'intensité

Trois mesures de similitude entre images exploitant la théorie de l'information de Shannon [Shannon, 1948] appliquée aux intensités des images sont successivement présentées, avant d'être brièvement confrontées. Ces mesures ont l'avantage majeur de ne pas reposer directement sur les intensités des pixels des images mais sur une distribution des intensités de ces images, les rendant intrinsèquement robustes aux changements globaux d'illumination ainsi qu'à des perturbations arbitraires de l'espace des intensités (changements locaux en intensité, direction, spectre, voire modalité).

3.5.1.1 Somme des variances conditionnelles

L'adaptation du critère de la somme des variances conditionnelles (SVC) aux problèmes de suivi et d'asservissement visuel directs repose sur la motivation d'avoir un critère restant le plus proche possible des intensités des images acquises, tout en étant robuste aux variations non-linéaires d'intensité et à la multimodalité, comme pour le recalage d'images médicales [Pickering et al., 2009], contexte d'origine de cette mesure de similitude entre images.

La SVC repose sur le calcul de l'histogramme conjoint normalisé $\mathbf{P}_{\mathcal{I}\mathcal{I}^*}$ entre les intensités $\mathcal{I}$ de l'image courante et celles de l'image désirée $\mathcal{I}^*$, approximant la distribution d'intensité jointe. Plus précisément, $\mathbf{P}_{\mathcal{I}\mathcal{I}^*} \in \mathbb{R}_+^{N_i \times N_i}$, $N_i \in \mathbb{N}$ représentant le nombre d'intensités possibles dans les images considérées, c'est-à-dire, dans le cas général similaire à la partie 3.2.1, $\forall \mathbf{X} \in \mathcal{X}, I(\mathbf{u}(\mathbf{p}, \mathbf{X})) \in [\![0, N_i]\!]$ ($N_i = 255$ si les intensités de l'image sont codées sur 8 bits). Chaque élément $P_{\mathcal{I}\mathcal{I}^*}(i,j)$ de la matrice $\mathbf{P}_{\mathcal{I}\mathcal{I}^*}$ s'exprime alors, en introduisant dors-et-déjà les degrés de liberté $\boldsymbol{\delta}\mathbf{p}$ et paramètres $\{\mathbf{p}, \mathcal{X}\}$ du futur problème d'optimisation, par :

$$P_{\mathcal{I}\mathcal{I}^*}(i,j,\boldsymbol{\delta}\mathbf{p},\{\mathbf{p},\mathcal{X}\}) = \frac{1}{|\mathcal{X}|}\sum_{\mathbf{X}\in\mathcal{X}} d\left(I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathbf{X})) - i\right)\; d\left(I(\mathbf{u}(\mathbf{p},\mathbf{X})) - j\right), \tag{3.137}$$

avec la fonction $d : k \in \mathbb{Z} \longmapsto d(k)\mathbb{N}$, de type impulsion unitaire :

$$d(k) = \begin{cases} 1, & \text{si } k = 0 \\ 0, & \text{sinon} \end{cases}. \tag{3.138}$$

A partir des $P_{\mathcal{I}\mathcal{I}^*}(i,j)$, on exprime l'espérance conditionnelle par :

$$\mathscr{E}_{\mathcal{I}\mathcal{I}^*}(I \mid I^*,\boldsymbol{\delta}\mathbf{p},\{\mathbf{p},\mathcal{X}\}) = \frac{\sum_{i=0}^{N_i} i\, P_{\mathcal{I}\mathcal{I}^*}(i,I^*,\boldsymbol{\delta}\mathbf{p},\{\mathbf{p},\mathcal{X}\})}{\sum_{i=0}^{N_i} P_{\mathcal{I}\mathcal{I}^*}(i,I^*,\boldsymbol{\delta}\mathbf{p},\{\mathbf{p},\mathcal{X}\})}, \tag{3.139}$$

avec $I = I(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathbf{X}))$ et $I^* = I(\mathbf{u}(\mathbf{p},\mathbf{X}))$.

En appliquant cette modélisation de la SVC au suivi de région $\mathcal{R}^*$ plane dans l'image, considérant les homographies ${}^{\overline{c}}\mathbf{H}_{c^*} \in SL(3)$ et ${}^{c}\mathbf{H}_{\overline{c}}(\mathbf{h}) \in SL(3)$, $\mathbf{h} \in \mathbf{R}^8$ (cf. Partie 3.3.2.3), l'expression générique des $P_{\mathcal{I}\mathcal{I}^*}(i,j)$ (Eq. (3.137)) devient $P_{\mathcal{I}\mathcal{I}^*{}_{R+}}(i,j)$ [Richa et al., 2011] :

$$P_{\mathcal{I}\mathcal{I}^*{}_{R+}}(i,j,\mathbf{h},\{\mathcal{R}^*,{}^{\overline{c}}\mathbf{H}_{c^*}\}) = \frac{1}{|\mathcal{R}^*|}\sum_{\tilde{\mathbf{u}}^*\in\mathcal{R}^*} d\left(I({}^{c}\mathbf{H}_{\overline{c}}(\mathbf{h})\,{}^{\overline{c}}\mathbf{H}_{c^*}\,\tilde{\mathbf{u}}^*) - i\right)\; d\left(I^*(\mathbf{u}^*) - j\right), \tag{3.140}$$

et l'espérance conditionnelle :

$$\mathscr{E}_{\mathcal{I}\mathcal{I}^*{}_{R+}}(I \mid I^*,\mathbf{h},\{\mathcal{R}^*,{}^{\overline{c}}\mathbf{H}_{c^*}\}) = \frac{\sum_{i=0}^{N_i} i\, P_{\mathcal{I}\mathcal{I}^*{}_{R+}}(i,I^*,\mathbf{h},\{\mathcal{R}^*,{}^{\overline{c}}\mathbf{H}_{c^*}\})}{\sum_{i=0}^{N_i} P_{\mathcal{I}\mathcal{I}^*{}_{R+}}(i,I^*,\mathbf{h},\{\mathcal{R}^*,{}^{\overline{c}}\mathbf{H}_{c^*}\})}, \tag{3.141}$$

avec $I = I({}^{c}\mathbf{H}_{\overline{c}}(\mathbf{h})\,{}^{\overline{c}}\mathbf{H}_{c^*}\,\tilde{\mathbf{u}}^*)$ et $I^* = I^*(\mathbf{u}^*)$.

Enfin, l'expression de la SVC à optimiser est très similaire au coût de type SSD (Eq. 3.66), sauf qu'elle remplace les intensités désirées par l'espérance conditionnelle $\mathscr{E}_{\mathcal{I}\mathcal{I}^*{}_{R+}}()$ (Eq. (3.141)) [Richa et al., 2011] :

$$\begin{aligned} &\mathscr{C}_{R+}(\mathbf{h},\mathcal{I},\{\mathcal{R}^*,{}^{\overline{c}}\mathbf{H}_{c^*}\},\mathcal{I}^*) \\ &= \frac{1}{2}\sum_{\mathbf{u}^*\in\mathcal{R}^*}\left(I({}^{c}\mathbf{H}_{\overline{c}}(\mathbf{h})\,{}^{\overline{c}}\mathbf{H}_{c^*}\,\tilde{\mathbf{u}}^*) - \mathscr{E}_{\mathcal{I}\mathcal{I}^*{}_{R+}}(I \mid I^*,\mathbf{h},\{\mathcal{R}^*,{}^{\overline{c}}\mathbf{H}_{c^*}\})\right)^2, \end{aligned} \tag{3.142}$$

qui se résout par ESM (Partie 3.2.4), sans pouvoir bénéficier de la simplification (Eq. 3.73) permettant de ne calculer la partie géométrique de la matrice jacobienne courante qu'à la première itération puisque $\mathscr{C}_{\mathcal{I}\mathcal{I}^*{}_{R+}}()$, recalculé à chaque nouvelle image acquise, peut varier si les conditions d'illumination changent ou en présence d'occultation.

La SVC s'étend à l'asservissement visuel en reformulant le coût $\mathscr{C}_{R+}()$ par le remplacement du modèle de mouvement d'homographie par la transformation rigide dans l'espace du repère monde/objet au repère caméra [Delabarre and Marchand, 2012].

3.5.1.2 Information mutuelle

En vision directe, l'information mutuelle mesure le degré de similitude entre deux images en ne s'appuyant intrinsèquement que sur l'information partagée par ces images. Son expression et son calcul s'appuient sur les distributions d'intensité de chaque image et la distribution d'intensité jointe des deux images (Eq. (3.137)). Elles sont exploitées pour exprimer les entropies de Shannon jointe $H_{\mathcal{I}\mathcal{I}^*}(\delta\mathbf{p}, \{\mathbf{p}, \mathcal{X}\})$ et marginales $H_{\mathcal{I}}(\delta\mathbf{p}, \{\mathbf{p}, \mathcal{X}\})$ des deux images à comparer, tel que :

$$H_{\mathcal{I}\mathcal{I}^*}(\boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\}) = -\sum_{i=0}^{N_i}\sum_{j=0}^{N_i} P_{\mathcal{I}\mathcal{I}^*}(i, j, \boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\}) \log(P_{\mathcal{I}\mathcal{I}^*}(i, j, \boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\})), \tag{3.143}$$

et :

$$H_{\mathcal{I}}(\boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\}) = -\sum_{i=0}^{N_i} P_{\mathcal{I}}(i, \boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\}) \log(P_{\mathcal{I}}(i, \boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\})), \tag{3.144}$$

où les $P_{\mathcal{I}}()$ sont calculées de façon similaire aux $P_{\mathcal{I}\mathcal{I}^*}()$, mais pour une image.

L'information mutuelle s'exprime alors par :

$$MI(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) = H_{\mathcal{I}}(\boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\}) + H_{\mathcal{I}^*}(\{\mathbf{p}, \mathcal{X}\}) - H_{\mathcal{I}\mathcal{I}^*}(\boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\}), \tag{3.145}$$

qu'il convient de maximiser en optimisant les degrés de liberté $\boldsymbol{\delta}\mathbf{p}$ [Dame and Marchand, 2011] :

$$\widehat{\boldsymbol{\delta}\mathbf{p}} = \arg\max_{\boldsymbol{\delta}\mathbf{p}} MI(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*). \tag{3.146}$$

Pour s'assurer que la fonction de vraisemblance $MI()$ (Eq. (3.145)) soit suffisamment lisse pour limiter les minima locaux, différentiable sans approximation, tout en ayant un temps de calcul raisonnable, et, enfin, ait un domaine de convergence utile, le filtrage des images a priori et la quantification sont essentiels [Dame and Marchand, 2011]. Plus précisément, un filtrage gaussien sur les deux images permet d'améliorer le deuxième et le troisième point, sans changer la position de l'optimum de la fonction $MI()$. La quantification des images impacte les trois points. En particulier, la réduire empiriquement de $N_i = 255$ à

$N_i = 8$ [Dame and Marchand, 2011] permet de réduire le nombre de minima locaux et de gagner du temps dans le calcul des distributions d'intensité. Cependant, pour que ces dernières soient différentiables, elles sont remplacées par des fonctions B-splines, approximant des fonctions gaussiennes différentiables mais nécessitant un temps de calcul moins important. Ces fonctions B-splines sont donc directement intégrées dans l'expression des distributions d'intensité (Eq. (3.137)), en remplaçant la fonction $d()$ (Eq.(3.138)) par $d_{Bs} : k_{Bs} \in \mathbb{R} \longmapsto d_{Bs}(k_{Bs}) \in \mathbb{R}$ tel que [Dame and Marchand, 2011] :

$$d_{Bs}(k_{Bs}) = \begin{cases} (2+k_{Bs})^3, & \text{si } k_{Bs} \in [-2,-1] \\ 1+3(1+k_{Bs})+3(1+k_{Bs})^2-3(1+k_{Bs})^3, & \text{si } k_{Bs} \in [-1,0] \\ 1+3(1-k_{Bs})+3(1-k_{Bs})^2-3(1-k_{Bs})^3, & \text{si } k_{Bs} \in [0,1] \\ (2-k_{Bs})^3, & \text{si } k_{Bs} \in [1,2] \\ 0, & \text{sinon} \end{cases} . \tag{3.147}$$

Utiliser l'ordre 3 est suffisant pour avoir une différentiation lisse de la fonction B-spline. L'autre différence majeure entre les expressions des fonctions $d()$ et $d_{Bs}()$ (Eqs. (3.138) et (3.147)) réside dans leurs entrées respectives. En effet, pour la première, k est un entier alors que, pour la seconde, k_{Bs} est un réel. Cela permet de considérer des intensités de pixels des images non entières, afin de ne pas perdre d'information lors de la réduction de la quantification, qui devient alors une mise à l'échelle des intensités d'un intervalle $[\![0, N_I]\!] \subset \mathbb{N}$ initial, avec, en général, $N_I = 255$ à l'intervalle $[0, N_i] \subset \mathbb{R}$, avec $N_i = 8$, comme évoqué précédemment. En prenant cette mise à l'échelle en compte dans l'utilisation de la fonction $d_{Bs}()$, on étend l'expression des distributions d'intensité d'image en :

$$P_{\mathcal{I}_{Bs}}(i, \boldsymbol{\delta}\mathbf{p}, \{\mathbf{p}, \mathcal{X}\}) = \frac{1}{|\mathcal{X}|} \sum_{\mathbf{X} \in \mathcal{X}} d_{Bs}\left(\frac{N_i}{N_I} I(\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X})) - i\right). \tag{3.148}$$

On obtient $P_{\mathcal{I}\mathcal{I}^*{}_{Bs}}()$ en suivant le même raisonnement. En substituant ces nouvelles expressions dans celles des entropies, on rend, en cascade, la maximisation de l'information mutuelle $MI()$ (Eq. (3.146)) possible par la méthode de Newton, avec l'expression analytique des matrices jacobienne et hessienne, pour l'asservissement visuel [Dame and Marchand, 2011] et le suivi basé maquette virtuelle 3D texturée *[Caron et al., 2014]* en vision perspective, l'asservissement visuel omnidirectionnel *[Delabarre et al., 2012]* et le suivi d'un plan [Dame and Marchand, 2012], de plusieurs plans rigidement liés par une [Delabarre and Marchand, 2013] ou plusieurs [Fraissinet-Tachet et al., 2016] caméras perspectives.

Les extensions de l'asservissement visuel maximisant l'information mutuelle de la vision perspective à la vision omnidirectionnelle ou au suivi multiplan et multicaméra suivent la même méthodologie que l'adaptation de la minimisation des coûts basés SSD (photométriques) à ces cas. Le suivi basé maquette virtuelle 3D texturée maximisant l'information mutuelle entre image réelle et image de synthèse (rendu de la maquette 3D), est une contribution *[Caron et al., 2014]* détaillée en partie 4.1,

où les expressions des matrices jacobienne et hessienne de la méthode d'optimisation de Newton sont les mêmes que pour l'asservissement visuel.

A noter qu'une extension de l'information mutuelle, à savoir l'information mutuelle normalisée (NMI) [Studholme et al., 1999] :

$$NMI(\boldsymbol{\delta p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) = \frac{H_{\mathcal{I}^*}(\{\mathbf{p}, \mathcal{X}\}) + H_{\mathcal{I}}(\boldsymbol{\delta p}, \{\mathbf{p}, \mathcal{X}\})}{H_{\mathcal{I}\mathcal{I}^*}(\boldsymbol{\delta p}, \{\mathbf{p}, \mathcal{X}\})}, \tag{3.149}$$

a été exploitée pour l'asservissement visuel omnidirectionnel *[Delabarre et al., 2012]* mais n'a pas montré d'impact significatif par rapport à la version non normalisée, à part l'intérêt, seulement pratique dans ce contexte, de présenter une borne supérieure, contrairement à l'expression non normalisée.

3.5.1.3 Ecart d'information normalisé

Exprimé à partir des mêmes entropies marginales et jointe des images à comparer, l'écart d'information normalisé (NID : Normalised Information Distance), proposé plus récemment [Ming Li et al., 2004] que l'information mutuelle [Shannon, 1948] et que l'information mutuelle normalisée [Studholme et al., 1999], s'exprime par le coût suivant :

$$\mathscr{C}_{NID}(\boldsymbol{\delta p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*) = \frac{2H_{\mathcal{I}\mathcal{I}^*}(\boldsymbol{\delta p}, \{\mathbf{p}, \mathcal{X}\}) - H_{\mathcal{I}}(\boldsymbol{\delta p}, \{\mathbf{p}, \mathcal{X}\}) - H_{\mathcal{I}^*}(\{\mathbf{p}, \mathcal{X}\})}{H_{\mathcal{I}\mathcal{I}^*}(\boldsymbol{\delta p}, \{\mathbf{p}, \mathcal{X}\})}. \tag{3.150}$$

En exploitant les mêmes adaptations que pour l'information mutuelle, le coût $\mathscr{C}_{NID}()$ se minimise pour optimiser la pose d'une caméra [Pascoe et al., 2015] :

$$\widehat{\boldsymbol{\delta p}} = \arg\min_{\boldsymbol{\delta p}} \mathscr{C}_{NID}(\boldsymbol{\delta p}, \mathcal{I}, \{\mathbf{p}, \mathcal{X}\}, \mathcal{I}^*). \tag{3.151}$$

En particulier, ce coût a été exploité pour le calcul de pose de caméra basé maquette virtuelle 3D texturée [Pascoe et al., 2015] et la cartographie simultanée [Pascoe et al., 2017], ainsi que pour l'étalonnage extrinsèque direct caméra-scanner laser [Pascoe et al., 2015], par conséquent multimodal (comparaison des intensités d'images de la caméra avec les intensités, de nature totalement différente, des relevés lidar).

3.5.1.4 Conclusion partielle

Tout d'abord, dans les conditions idéales, le suivi de plan par SVC montre un meilleur taux de convergence pour un nombre d'itérations réduit par rapport à l'information mutuelle. La SVC se montre aussi plus robuste que l'ESM rendu robuste aux variations affines d'illumination. Le domaine de convergence de la SVC lui permet de continuer le suivi même quand le mouvement entre images successives devient trop important pour celui de l'information mutuelle. Cette dernière reste cependant intrinsèquement plus robuste aux réflexions spéculaires que les deux autres méthodes [Richa et al., 2011].

Ensuite, en résumé, l'asservissement visuel basé SVC montre une complexité, des performances et une robustesse s'intercalant entre l'asservissement visuel photométrique et l'asservissement visuel basé information mutuelle [Delabarre and Marchand, 2012].

Enfin, l'expression du coût $\mathscr{C}_{NID}()$ est théoriquement insensible à la quantité totale d'information des deux images. Par conséquent, il ne favorise pas les régions texturées de l'image au détriment de l'alignement global, contrairement à l'information mutuelle dans ce type de situation [Stewart and Newman, 2012].

3.5.2 Champs de descripteurs

Au lieu de considérer l'intensité des images acquises directement en entrée des algorithmes de suivi, de calcul ce pose ou de cartographie, l'idée générale de cette famille d'approches est de calculer un descripteur par pixel et d'utiliser cette information en entrée des algorithmes, pour une meilleure robustesse aux perturbations lumineuses et aux ambiguïtés.

Filtre dérivatif directionnel Rendre invariantes les intensités $\mathcal{I}$ et $\mathcal{I}^*$ des images à l'illumination, voire aux spécificités de capteurs multimodaux, permet d'envisager de ne comparer que l'information commune partagée par ces images. Les filtres dérivatifs représentent un moyen de transformation des intensités des images acquises vers un espace, dit d'*énergie* d'image [Burt, 1988], à considérer dans un critère à optimiser pour obtenir la transformation recherchée.

Afin de transformer une image vers cet espace invariant aux changements d'illumination et à la modalité, tout en ayant une discrimination de chaque pixel afin de limiter les minima locaux, des filtres dérivatifs directionnels sont appliqués selon quatre directions $\mathcal{O} = \{$'h', 'v', 'dd', 'dg'$\}$ (horizontal, vertical, diagonale du haut vers la droite et diagonale du haut vers la gauche) dans l'image, après un filtrage gaussien pour limiter les aberrations [Irani and Anandan, 1998]. Les quatre réponses I_d, tel que $d \in \mathcal{O}$, à ces filtres sont ensuite mises au carré menant ainsi aux quatre composantes "énergétiques" $\mathscr{I}_d$ par pixel, considérées dans un calcul de corrélation normalisée locale avec leurs équivalents I_d^* :

$$\begin{aligned}
&\mathscr{I}_d(\{\mathbf{A}, \boldsymbol{\delta}\mathbf{u}\}, \mathcal{I}_d, \{\mathbf{u}^*, L\}, \mathcal{I}_d^*) \\
&= \frac{\sum\limits_{\mathbf{u}_c^* \in \mathcal{U}_c(\mathbf{u}^*)} \left(I_d(\mathbf{A}\mathbf{u}_c^* + \boldsymbol{\delta}\mathbf{u})^2 - m_{I_d^2}\right)\left(I_d^*(\mathbf{u}_c^*)^2 - m_{I_d^{*2}}\right)}{\sqrt{\sum\limits_{\mathbf{u}_c^* \in \mathcal{U}_c(\mathbf{u}^*)} \left(I_d(\mathbf{A}\mathbf{u}_c^* + \boldsymbol{\delta}\mathbf{u})^2 - m_{I_d^2}\right)^2} \sqrt{\sum\limits_{\mathbf{u}_c^* \in \mathcal{U}_c(\mathbf{u})} \left(I_d(\mathbf{u}_c^*)^2 - m_{I_d^{*2}}\right)^2}},
\end{aligned} \tag{3.152}$$

tel que $\mathcal{U}_c(\mathbf{u}) = [\![u - L, u + L]\!] \times [\![v - L, v + L]\!]$, et $m_{I_d^2}$, respectivement $m_{I_d^{*2}}$, est la moyenne des $I_d()^2$, respectivement $I_d^*()^2$, sur la fenêtre carrée de côté $2L+1$. $\{\mathbf{A}, \boldsymbol{\delta}\mathbf{u}\}$ paramètrent la transformation géométrique affine (cf. Partie 3.3.2.2) à optimiser

pour maximiser le critère global [Irani and Anandan, 1998] :

$$\mathscr{S}_{IA}(\{\mathbf{A},\boldsymbol{\delta}\mathbf{u}\},\mathcal{I}_{\mathcal{O}},\{\mathcal{R}^*,L\},\mathcal{I}^*_{\mathcal{O}}) = \sum_{\mathbf{u}^*\in\mathcal{R}^*}\sum_{d\in\mathcal{O}} \mathscr{I}_d(\{\mathbf{A},\boldsymbol{\delta}\mathbf{u}\},\mathcal{I}_d,\{\mathbf{u}^*,L\},\mathcal{I}^*_d), \quad (3.153)$$

par la méthode de Newton, en considérant les énergies $\mathscr{I}_d()$ directement mesurées, ce qui mène à des expressions très simples des matrices jacobienne et hessienne de $\mathscr{S}_{IA}$ [Irani and Anandan, 1998].

Maximiser $\mathscr{S}_{IA}()$ avec $L = 3$ ou $L = 5$ permet de recaler deux images (la région $\mathcal{R}^*$ de l'équation (3.153) englobe toute l'image I^*) de modalités différentes (visible et infrarouge), dans une approche de multirésolution (cf. Partie 3.5.4) quand l'amplitude de la transformation affine est importante entre les deux images, à raison de 4 itérations par niveau de résolution, sauf pour le plus haut [Irani and Anandan, 1998]. Enfin, en pratique, après chaque itération de la boucle d'optimisation, les énergies courantes sont transformées avec l'incrément de transformation courant pour compenser l'impact des distorsions spatiales entre les images sur les énergies.

Autres approches Des variantes de la méthode détaillée ci-dessus considèrent un vecteur descripteur par pixel $\mathbf{u}$, dont chaque élément est la réponse en $\mathbf{u}$ d'un filtre appliqué l'image $I()$. Des méthodes reposent sur la somme des canaux de descripteurs en entrée de suivi de région [Antonakos et al., 2015], comme les gradients de l'image (2 canaux), les LBP [Ojala et al., 1996] (6 canaux), les histogrammes de gradients orientés [Dalal and Triggs, 2005], SIFT [Lowe, 1999] à 36 canaux. Ces descripteurs sont nativement robustes aux variations d'illumination et, de plus, les considérer dans un algorithme de suivi direct étendu apporte une certaine robustesse à la variabilité au sein d'une classe d'objets, plutôt que d'imposer de considérer un objet spécifique.

D'autres approches considèrent les canaux des descripteurs ensemble dans le coût à minimiser, comme une somme supplémentaire sur les canaux, imbriquée dans la somme sur les pixels. Ainsi, un jeu de quatre filtres dérivatifs gaussiens, dont les réponses sont seuillées, a été considéré [Crivellaro and Lepetit, 2014], engendrant donc une image à quatre canaux, ou encore des motifs binaires [Alismail et al., 2016], pour 8 canaux. Cela augmente la quantité d'information à traiter mais apporte de la robustesse au suivi de région plane et à la cartographie [Park et al., 2017] par rapport aux variations d'illumination, y compris en basse luminosité, tout en maintenant la précision à convergence.

Enfin, en augmentant les vecteurs descripteurs avec les intensités ou les couleurs des pixels de l'image, directement, on bénéfice du meilleur des deux types d'information, à savoir la capacité à traiter des environnements faiblement texturés ou, à l'opposé, très texturés, rendant polyvalent le calcul de pose basé maquette virtuelle 3D texturée d'objet [Zhong et al., 2018].

3.5.3 Approches basées noyaux

3.5.3.1 Distributions d'intensité pondérée

Les approches basées noyaux peuvent s'apparenter aux approches basées distribution d'intensité (Partie 3.5.1), mais en construisant différemment le lien entre une distribution d'intensité et les informations spatiales des degrés de liberté du mouvement d'une région d'image à suivre et/ou de la caméra. Ce lien s'exprime par une pondération de la fonction d'impulsion unitaire $d()$ (Eq. (3.138)), de l'expression de la distribution d'intensité, par la réponse d'une fonction noyau $K : \mathbb{R}_+ \longmapsto \mathbb{R}$ isotrope, convexe et monotone décroissante. On définit alors la distribution d'intensité pondérée $Q_{\mathcal{I}}$ de l'image courante, formulée pour le suivi en translation pure d'une région $\mathcal{R}^*$, de largeur L_u^* et de hauteur L_v^* dans le plan image numérique [Comaniciu et al., 2003], par :

$$Q_{\mathcal{I}}(i, \boldsymbol{\delta u}, \{\mathcal{R}^*\}) = \frac{\sum\limits_{\mathbf{u}^* \in \mathcal{R}^*} K(||\mathbf{L}^{-1}(\mathbf{u}^* + \boldsymbol{\delta u})||^2) d\left(I(\mathbf{u}^*) - i\right)}{\sum\limits_{\mathbf{u}^* \in \mathcal{R}^*} K(||\mathbf{L}^{-1}(\mathbf{u}^* + \boldsymbol{\delta u})||^2)}, \tag{3.154}$$

avec :

$$\mathbf{L} = \begin{pmatrix} L_u^*/2 & 0 & u_c^* \\ 0 & L_v^*/2 & v_c^* \\ 0 & 0 & 1 \end{pmatrix}, \tag{3.155}$$

mettant les coordonnées dans l'image à l'échelle et les centrant en $\mathbf{0}$ afin de permettre de définir la fonction $K()$ indépendamment de la taille de la région $\mathcal{R}^*$ dans l'image. u_c^* et v_c^* sont les coordonnées du centre de $\mathcal{R}^*$, la région de référence dont la distribution pondérée $Q_{\mathcal{I}^*}$ s'exprime de façon similaire à $Q_{\mathcal{I}}$ mais sans le $\boldsymbol{\delta u}$.

Selon un profil d'Epanechikov, la fonction noyau $K()$ s'exprime par [Comaniciu and Meer, 2002] :

$$K_E(k_n) = \begin{cases} 1 - k_n, & \text{si } k_n \in [0, 1] \\ 0, & \text{si } k_n > 1 \end{cases}. \tag{3.156}$$

Grâce à la pondération spatiale apportée par le noyau isotrope, on peut définir un coût lisse et dérivable (au moins par morceaux) entre les deux distributions d'intensité $Q_{\mathcal{I}}$ et $Q_{\mathcal{I}^*}$ directement, pour l'estimation du mouvement de la région $\mathcal{R}^*$ vers l'image courante [Hager et al., 2004] :

$$\begin{aligned} \mathscr{C}_{HDS}(\boldsymbol{\delta u}, \mathcal{I}, \{\mathcal{R}^*\}, \mathcal{I}^*) &= \left|\left| \begin{bmatrix} \vdots \\ \sqrt{Q_{\mathcal{I}}(i, \boldsymbol{\delta u}, \{\mathcal{R}^*\})} \\ \vdots \end{bmatrix} - \begin{bmatrix} \vdots \\ \sqrt{Q_{\mathcal{I}^*}(i, \{\mathcal{R}^*\})} \\ \vdots \end{bmatrix} \right|\right|^2 \\ &= ||\mathbf{Q}_{\mathcal{I}}(\boldsymbol{\delta u}) - \mathbf{Q}_{\mathcal{I}^*}||^2, \end{aligned} \tag{3.157}$$

conditions nécessaires pour mettre en place une minimisation du coût par descente de gradient ou la méthode de Newton [Hager et al., 2004].

Le coût $\mathscr{C}_{HDS}()$ s'étend du modèle de mouvement de la translation pure dans le plan image [Comaniciu et al., 2003] aux autres modèles de mouvements dans le plan image à plus de degrés de liberté [Hager et al., 2004], de façon similaire aux diverses extensions du coût direct pur $\mathscr{C}_{LK}()$ (Eq. (3.47)). Deux problèmes apparaissent cependant dans la résolution des problèmes d'optimisation prenant en compte plus de degrés de liberté que la translation pure dans le plan image pour les méthodes basées noyau. Il s'agit, d'une part, de la dynamique et l'équilibre de la distribution d'intensité (pondérée) et, d'autre part, de l'interaction entre cette distribution et la structure de l'image. Fondamentalement, le premier mène à un problème de sous-détermination où le nombre de contraintes apportées par les $Q_{\mathcal{I}}$ et $Q_{\mathcal{I}^*}$ est insuffisant pour calculer tous les degrés de liberté. Le second, quant à lui, est plutôt lié au problème d'ambiguïté perceptuelle pour lequel la distribution ne change pas, même si les degrés de liberté changent (plateau local de la fonction de coût $\mathscr{C}_{HDS}()$, notamment).

La solution consiste à considérer plusieurs noyaux, dont au moins certains sont anisotropes, pour apporter des informations complémentaires sur le mouvement [Hager et al., 2004]. Par exemple, en considérant un noyau supplémentaire de type triangulaire (largeur s) extrudé (longueur l) :

$$K_T(\mathbf{x}_L, \mathbf{n}) = \frac{4}{ls^2} \max\left(\frac{s}{2} - |\mathbf{x}_L \cdot \mathbf{n}|, 0\right), \tag{3.158}$$

avec $\mathbf{x}_L = \mathbf{L}^{-1}\mathbf{u}$. Dans les conditions idéales, ce noyau est le détecteur optimal de contour dans la distribution d'intensité pondérée, c'est-à-dire donnant la réponse maximale, localement.

En renommant $\mathbf{Q}_{\mathcal{I}}$ et $\mathbf{Q}_{\mathcal{I}^*}$ en $\mathbf{Q}_{\mathcal{I}_K}$ et $\mathbf{Q}_{\mathcal{I}^*_K}$, K désignant le noyau utilisé pour calculer les distributions, on écrit le coût multinoyau $\mathscr{C}_{HDS_m}()$ par extension du coût $\mathscr{C}_{HDS}()$ (Eq. (3.157)) en empilant autant de vecteurs des distributions d'intensité pondérées que de noyaux considérés :

$$\mathscr{C}_{HDS_m}(\boldsymbol{\delta}\mathbf{u}, \mathcal{I}, \{\mathcal{R}^*\}, \mathcal{I}^*) = \left\| \begin{bmatrix} \vdots \\ \mathbf{Q}_{\mathcal{I}_K}(\boldsymbol{\delta}\mathbf{u}) \\ \vdots \end{bmatrix} - \begin{bmatrix} \vdots \\ \mathbf{Q}_{\mathcal{I}^*_K}(\boldsymbol{\delta}\mathbf{u}) \\ \vdots \end{bmatrix} \right\|^2. \tag{3.159}$$

Dès lors, la difficulté est de sélectionner quelle collection de noyaux considérer pour que le coût apporte suffisamment d'informations indépendantes pour contrôler les degrés de liberté nécessaires au suivi. Par exemple, en combinant deux noyaux triangulaires extrudés de normales orthogonales, les deux translations horizontales sont contrôlées, et un troisième noyau conique (triangle de révolution) permet de contrôler l'échelle, menant à un suivi à trois degrés de liberté [Hager et al., 2004].

3.5.3.2 Noyaux photométriques

Le concept de fonction noyau reliée aux degrés de liberté du coût à optimiser pour suivre une région dans les images s'adapte au-delà du cadre des distributions

d'intensités en appliquant la fonction noyau directement à l'intensité de chaque pixel de l'image. En considérant tous les pixels, de coordonnées supposées continues, de l'image acquise par la caméra, on obtient [Kallem et al., 2007] :

$$\mathscr{S}(\boldsymbol{\delta}\mathbf{x}) = \int_{\boldsymbol{\pi}_u} K((\mathbf{K}^{-1}\mathbf{u} + \boldsymbol{\delta}\mathbf{x}))I(\mathbf{u})\,\mathrm{d}\mathbf{u}, \tag{3.160}$$

appliquant ainsi la fonction noyau sur les coordonnées normalisées des pixels, au sens du plan image normalisé. L'apport majeur est que $\mathscr{S}(\boldsymbol{\delta}\mathbf{x})$ est dérivable, quand $K()$ est lisse, même si l'image ne l'est pas à cause de ses discontinuités.

En considérant $\mathscr{S}^* = \mathscr{S}(\mathbf{0})$, calculé à partir des intensités $\mathcal{I}^*$ de l'image de référence, et le plan image de la caméra fronto-parallèle à la scène plane, située à 1 m, on exprime la loi de commande $\mathbf{v}_{\mathbf{K}+}$ à deux degrés de liberté, à partir de la méthode du gradient [Kallem et al., 2007] :

$$\mathbf{v}_{\mathbf{K}+} = \begin{bmatrix} v_X \\ v_Y \end{bmatrix} = \dot{\boldsymbol{\delta}\mathbf{x}}^{(t)} = -\lambda\,(\mathscr{S}(\boldsymbol{\delta}\mathbf{x}) - \mathscr{S}^*) \int_{\mathcal{U}} \left.\frac{\partial K((\mathbf{K}^{-1}\mathbf{u} + \boldsymbol{\delta}\mathbf{x}))}{\partial \boldsymbol{\delta}\mathbf{x}}\right|_{\boldsymbol{\delta}\mathbf{x}=\boldsymbol{\delta}\mathbf{x}^{(t)}} I(\mathbf{u})\,\mathrm{d}\mathbf{u}. \tag{3.161}$$

En pratique les deux degrés de liberté sont contrôlés séparément par deux noyaux, la méthode s'apparentant ainsi au suivi de région multi-noyau [Hager et al., 2004]. Ces noyaux sont gaussiens et leurs largeurs sont déterminées expérimentalement en cherchant le meilleur compromis entre l'étendue du domaine de convergence (largeur de gaussienne importante) et la précision à convergence (largeur de gaussienne réduite) [Kallem et al., 2007].

Le même raisonnement est suivi pour arriver au contrôle de la translation selon l'axe optique de la caméra ou de la rotation autour de cet axe, séparément, à ceci près que le noyau n'est pas appliqué sur les intensités de l'image acquise mais sur sa transformée de Fourier. En combinant les quatre lois de commande (une par degré de liberté), l'asservissement visuel peut corriger une pose relative appartenant au groupe $SE(2)$, augmenté de la translation selon la profondeur [Kallem et al., 2007].

3.5.3.3 Moments photométriques

Les moments photométriques s'apparentent à la fois au concept de noyaux photométriques et aux moments "classiques" découplant intrinsèquement les degrés de liberté de l'asservissement visuel [Chaumette, 2004]. L'apport des moments photométriques sur ces deux dernières approches est de pouvoir contrôler les degrés de liberté ensemble et à partir des intensités uniquement (sans passer par une transformée de Fourier pour certains degrés de liberté) sans traitement d'image préalable.

En considérant l'image du plan image normalisé, on définit le moment photométrique pondéré d'ordre $(o_x + o_y)$ par [Bakthavatchalam et al., 2018] :

$$\mathscr{M}_{o_x o_y}(t) = \iint_{\boldsymbol{\pi}_\mathbf{x}} x^{o_x} y^{o_y} w(\mathbf{x}) I(\mathbf{x}, t)\,\mathrm{d}x\,\mathrm{d}y, \tag{3.162}$$

dont l'expression est exploitée pour déterminer les moments photométriques et moments photométriques centrés d'ordre 0 à 3. Ces derniers sont combinés pour construire un vecteur d'autant de primitives visuelles que de degrés de liberté à contrôler, laissant envisager la possibilité d'une stabilité asymptotique globale de l'asservissement visuel, s'il n'y pas trop d'approximations faites [Chaumette and Hutchinson, 2006]. Concrètement, les moments photométriques, par leur construction, permettent un important découplage des degrés de liberté de l'asservissement visuel pour atteindre, dans les conditions nominales des trajectoires rectilignes ou très proches de l'être.

3.5.4 Espace d'échelle

Les approches multi-échelle permettent à la fois de réduire les temps de calcul et d'agrandir le domaine de convergence du suivi de région d'image. A partir de l'image $I^*()$, contenant la région de référence $\mathcal{R}^*$, et l'image $I()$, dans laquelle suivre cette région, deux pyramides de niveaux d'échelle sont générées en considérant $I_0^* = I^*$ et $I_0 = I$, comme étant leurs bases. En réduisant la taille de l'image I_e d'un facteur 2, pour obtenir l'image I_{e+1} d'échelle deux fois moindre que I_e, en cascade, l'ensemble $\{I_0, I_1, \ldots, I_{e_{max}}\}$ forme les sections parallèles à la base de la pyramide de niveaux d'échelle de I (même raisonnement pour I^*). Quand la pyramide est gaussienne, l'intensité de chaque pixel $I_e(\mathbf{u}_e)$ est obtenue à partir de l'image $I_{e-1}(\mathbf{u}_{e-1})$, convoluée avec un noyau gaussien $K_g()$ discret de dimension 5, tel que [Burt and Adelson, 1983] :

$$K_g(\mathbf{u}) = \mathbf{g}(u)\mathbf{g}(v)^\top, \tag{3.163}$$

avec :

$$\mathbf{g}(u) = \begin{cases} a_g & , \text{si} u = 0 \\ \frac{1}{4} & , \text{si} |u| = 1 \\ \frac{1}{4} - \frac{a_g}{2} & , \text{si} |u| = 2 \\ 0 & , \text{sinon} \end{cases}, \tag{3.164}$$

avec $a_g = 0,4$ pour que $\mathbf{g}()$ soit similaire à une gaussienne. L'expression de l'équation (3.164) assure que $\mathbf{g}()$ soit de norme unitaire, symétrique et d'égale contribution d'un niveau d'échelle e au suivant $(e+1)$. Effectuer ces convolutions en cascade d'un niveau d'échelle à l'autre avec un noyau constant est équivalent à appliquer un noyau de côté $2^{e+1}+1$ à chaque $I_0(\mathbf{u})$ pour directement déterminer $I_e(\frac{1}{2^e}\mathbf{u})$, pour un coût calculatoire bien moindre.

En considérant ces pyramides d'échelle, pour chaque nouvelle image dans laquelle suivre $\mathcal{R}^*$, le suivi est d'abord réalisé au niveau $e = e_{max}$ et, à convergence, l'état des degrés de liberté est conservé comme initialisation du suivi au niveau $e-1$ d'échelle, et ainsi de suite jusqu'à la base de la pyramide. Cette approche de suivi du niveau le plus grossier au plus fin [Bergen et al., 1992] exploite une discrétisation de l'espace d'échelle pour étendre le domaine de convergence du suivi direct. Le nombre de niveaux e_{max} dépend de la définition initiale de l'image. A titre d'exemple, pour

une définition classiquement utilisée en vision temps-réel tel que 640×480 pixels, e_{max} est généralement compris entre 3 et 5 [Bouguet, 2000].

D'autre part, le filtrage gaussien d'une pyramide multi-échelle augmente aussi la robustesse du suivi au bruit dans les images, aux occultations partielles et aux variations d'illumination.

La décomposition en ondelettes est une autre méthode de représentation multi-échelle des images, travaillant à la fois dans le domaine spatial, comme les pyramides gaussiennes, mais aussi dans le domaine fréquentiel. L'avantage est, notamment, d'engendrer l'équivalent d'un filtrage anisotrope, implicitement, contrairement aux pyramides gaussiennes pour lesquelles le filtrage est isotrope. De plus, le lien entre la variation des coefficients de la décomposition en ondelettes et la variation de la pose de la caméra qui acquiert les images s'exprime pour former l'expression analytique de la matrice jacobienne, utilisée dans la loi de commande de type Levenberg-Marquardt d'un asservissement visuel, directement dans l'espace d'échelle, minimisant l'écart entre les coefficients des décompositions en ondelettes de l'image courante et de l'image désirée [Duflot et al., 2018].

3.6 Synthèse

En conclusion de ce chapitre d'état de l'art, cette partie propose une représentation synthétique de l'ensemble des méthodes en les classant par type de coût ou de vraisemblance directe. Tout d'abord, le tableau 3.1 structure les approches directes pures et leurs quelques variantes, rendues robustes ou non. Ensuite, ce sont les méthodes directes étendues, c'est-à-dire reposant sur une transformée des intensités des pixels des images, qui sont rassemblées dans le tableau 3.2. Les références aux contributions rapportées et unifiées dans ce mémoire apparaissent en italique, dans les deux tableaux.

Dans les approches étendues, en plus des approches décrites dans les parties précédentes, des informations visuelles reposant sur deux espaces supplémentaires apparaissent dans le tableau 3.2 : l'espace fréquentiel et l'espace propre. Le premier repose sur la transformée de Fourier, polaire dans le cas des images panoramiques *[Morbidi and Caron, 2017]*, sphérique dans le cas images sphériques partielles ou complètes [Makadia and Daniilidis, 2003, Makadia et al., 2007]. Une fois la transformée appliquée, le cap relatif ou l'attitude relative de la caméra entre une pose courante et une pose de référence est déduit à partir du maximum de la corrélation des phases des transformées des deux images. Quant à raisonner dans l'espace propre, cela consiste à projeter l'image dans une base orthogonale pour exploiter l'information discriminante de l'image par rapport au mouvement de la caméra [Nayar et al., 1996] pour l'asservissement visuel [Deguchi, 2000] dont les performances sont meilleures (fluidité de mouvement, stabilité et domaine de convergence) quand ses matrices jacobiennes sont exprimées analytiquement [Marchand, 2019].

En résumé, on peut tout d'abord noter la variété des travaux menés par la

communauté scientifique en vision robotique directe, même en se concentrant sur la caméra passive, les environnements rigides et les méthodes applicables à la robotique, ou proches de l'être. Les méthodes vont de la considération directe des intensités des pixels d'une région d'image acquise pour suivre son déplacement à deux degrés de liberté, à la cartographie 3D et la localisation basée théorie de l'information appliquée aux intensités d'images panoramiques, à autant de degrés de liberté que de points 3D dans la scène, en plus des six de la pose de la caméra.

Malgré tous ces efforts, quelques champs de recherche restent à fouiller et d'autres à ouvrir, pour atteindre toujours plus de robustesse aux perturbations, voire les exploiter, et des domaines de convergence très vastes, voire globaux. C'est pourquoi certaines des contributions de ce mémoire s'inscrivent dans des catégories déjà existantes, pour décliner ou mettre à l'épreuve dans de nouveaux contextes des primitives directes pures et étendues, décrites dans les chapitres 4 et 5. Par contre, d'autres contributions sont plus fondamentales et permettent, notamment, de créer une nouvelle catégorie, celle des mélanges de potentiels photométriques. Elles sont décrites dans le chapitre 6.

TABLE 3.1 – Approches directes, partie 1 : intensités de pixels.

Information visuelle	application	références
photométrique	suivi	[Lucas and Kanade, 1981][Shi and Tomasi, 1994] [Baker and Matthews, 2001][Benhimane and Malis, 2004] [Baker and Matthews, 2004][Benhimane and Malis, 2007] [Mei et al., 2006][Mei et al., 2008] [Salazar-Garibay et al., 2009][Caron et al., 2011]
photométrique	asservissement visuel	[Collewet et al., 2008] [Caron et al., 2013][*Alj and Caron, 2015*]
photométrique	localisation et cartographie	[Matthies et al., 1988][Newcombe et al., 2011] [Engel et al., 2014][Caruso et al., 2015] [Engel et al., 2013]
photométrique+	suivi	[Hager and Belhumeur, 1998][Silveira and Malis, 2007] [Mei and Reid, 2008][Ito et al., 2011] [Park et al., 2012][*Crombez et al., 2014*] [*Crombez et al., 2015b*]
photométrique+	asservissement visuel	[Collewet and Marchand, 2011]
photométrique+	localisation et cartographie	[Engel et al., 2018][Bergmann et al., 2018]
corrélation	suivi	[Scandaroli et al., 2012]
réseaux de neurones	toutes	[Bateux et al., 2018][Tekin et al., 2018]

TABLE 3.2 – Approches directes, partie 2 : transformées des intensités.

Information visuelle	références
basée noyaux	[Kallem et al., 2007]
fréquentielle	[Makadia and Daniilidis, 2003][Makadia et al., 2007] *[Morbidi and Caron, 2017]*
ondelettes	[Duflot et al., 2018]
moments photométriques	[Bakthavatchalam et al., 2013][Bakthavatchalam et al., 2018] [Hadj-Abdelkader et al., 2018]
information mutuelle	[Dame and Marchand, 2011]*[Delabarre et al., 2012]* *[Caron et al., 2014]*[Dame and Marchand, 2012] [Delabarre and Marchand, 2013][Fraissinet-Tachet et al., 2016]
écart d'information normalisé	[Pascoe et al., 2015][Pascoe et al., 2017]
SVC	[Richa et al., 2011][Delabarre and Marchand, 2012]
espace propre	[Nayar et al., 1996][Deguchi, 2000][Marchand, 2019]
champs de descripteurs	[Irani and Anandan, 1998][Crivellaro and Lepetit, 2014] [Antonakos et al., 2015][Alismail et al., 2016] [Zhong et al., 2018][Park et al., 2017]
mélange de potentiels photométriques	*[Crombez et al., 2015a][Caron and Morbidi, 2018]*

Chapitre 4

Suivi et asservissement visuels basés entropie photométrique

Sommaire

4.1 L'information mutuelle en suivi basé maquette virtuelle 3D . 100
4.1.1 Introduction . . . 100
4.1.2 Calcul de pose basé information mutuelle . . . 102
4.1.3 Matrices jacobienne et hessienne . . . 103
4.1.4 Résultats . . . 104
4.1.5 Conclusion partielle . . . 109
4.2 Exploration : Commande de caméra virtuelle basée entropie 111
4.2.1 Introduction . . . 111
4.2.2 Exploration basée entropie photométrique . . . 112
4.2.3 Résultats . . . 113
4.2.4 Conclusion partielle . . . 116
4.3 Conclusion du chapitre . . . 117

Ce chapitre montre comment le lien formulé entre la variation de la distribution d'intensité d'une image et celle de la pose de la caméra (Partie 3.5.1), qui acquiert cette image, s'exploite dans le cadre des maquettes virtuelles 3D de bâtiments et de villes. Dans un premier temps, le critère d'information mutuelle est exploité pour le suivi visuel basé maquette virtuelle texturée (Partie 4.1) et, dans un deuxième temps, l'exploration visuelle de maquette virtuelle est abordée en exploitant l'entropie photométrique (Partie 4.2), avant de conclure le chapitre (Partie 4.3).

La méthode de suivi visuel basé maquette virtuelle *[Caron et al., 2014]* est le fruit de mes travaux à Inria Rennes-Bretagne Atlantique / IRISA, dans le cadre du projet ANR "CityVIP", en tant que postdoc. L'exploration de maquette virtuelle *[Habibi et al., 2015]* représente une partie des travaux réalisés dans le cadre de la thèse de Zaynab Habibi au laboratoire MIS de l'UPJV, que j'ai co-encadrée, dans le cadre du projet CR Picardie "Assiduitas", qui s'inscrit dans le programme e-Cathédrale du laboratoire.

4.1 L'information mutuelle en suivi basé maquette virtuelle 3D

4.1.1 Introduction

Les maquettes virtuelles 3D d'environnement apportent le potentiel d'éviter les phases d'exploration et de cartographie à tout robot mobile ou véhicule devant se localiser et se déplacer pour la première fois dans cet environnement. En effet, on peut imaginer qu'une seule maquette virtuelle soit partagée par tous les "usagers" de l'environnement qu'elle représente, même s'ils ont des capteurs de différentes caractéristiques, voire de différentes natures, si la maquette est suffisamment riche d'informations et précise.

Les maquettes virtuelles de villes, normalement créées par les géomètres (ex : IGN) à destination des humains (Fig. 4.1(a)), sont des candidates intéressantes et en constante amélioration. Se pose, cependant, le problème de correspondance des informations contenues dans la maquette virtuelle à celles apportées par les capteurs extéroceptifs du robot. En effet, les maquettes virtuelles de villes sont généralement issues de photogrammétrie et/ou de lasergrammétrie aérienne ou terrestre, c'est-à-dire avec des moyens de locomotion et des capteurs adaptés à la création de maquettes virtuelles fidèles à la réalité, mais différents de ceux des robots mobiles ou des véhicules, pour des raisons intrinsèques évidentes (ex : mobilité terrestre/aérienne) et pour des raisons de procédé et de coût (ex : appareils photographiques de photogrammétrie, stations de lasergrammétrie, etc). Ce problème de correspondance maquette/capteur est la contrepartie de l'absence de cartographie faite à l'aide des capteurs du robot en explorant l'environnement.

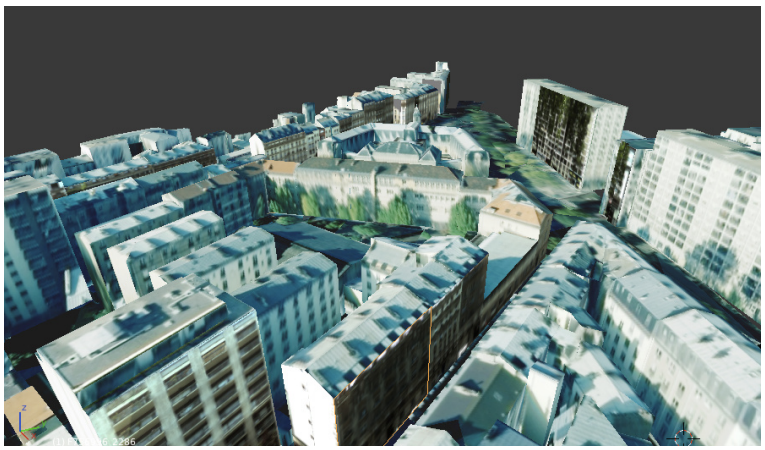
L'apparence visuelle des maquettes de villes, grâce aux textures (au sens de l'informatique graphique) plaquées sur les surfaces 3D maillées des bâtiments, est l'une des informations comparables aux mesures d'un capteur embarqué sur le robot : les images acquises par une caméra. Si la correspondance maquette/image est possible, les informations 3D de la maquette virtuelle permettront le calcul de pose précis et sans dérive [Royer et al., 2005] de la caméra et donc du robot qui l'embarque.

Pour rappel, le calcul de pose basé maquette 3D peut se faire avec divers types de primitives géométriques : les points [Lepetit and Fua, 2005], les droites [Jiang, 2006], leur combinaison [Rosten and Drummond, 2005] ou encore des modèles 3D filaires [Comport et al., 2006]. Les maquettes 3D en question sont toujours géométriques seules car formées uniquement des primitives géométriques à trouver dans l'image. L'information photométrique de l'image peut compléter l'information géométrique partagée par la maquette et l'image [Georgel et al., 2008, Pressigout and Marchand, 2007] pour améliorer la précision de l'estimation. Avec une maquette virtuelle texturée tel qu'évoqué plus haut, le calcul direct de pose de caméra est envisageable.

Cependant, exploiter l'information photométrique de la maquette virtuelle pour un calcul de pose direct demande un critère robuste aux imperfections de la maquette, qui est généralement précise mais pourvues d'approximations, d'aberrations

locales ou de manques (Fig. 4.1(c)), aux variations d'illumination entre les textures de la maquette et les images de la caméra, omniprésentes à l'extérieur, aux différences de résolution entre les textures de la maquette et les images de la caméra, a fortiori quand les textures sont issues d'images aériennes, et, enfin, aux occultations partielles. Dans l'état de l'art (Chapitre 3), l'information mutuelle est le critère ayant le potentiel de traiter un maximum de ces problèmes.

Par conséquent, l'asservissement visuel de caméra, embarquée sur l'effecteur d'un robot [Dame and Marchand, 2011], maximisant l'information mutuelle entre l'image courante, acquise par la caméra en mouvement, et l'image désirée, acquise par la même caméra à la pose désirée (Partie (3.5.1.2)), s'étend au problème de suivi visuel basé maquette 3D texturée. Spécifiquement, pour le calcul de pose d'une image (exemple en figure 4.1(b)), la caméra en mouvement est virtuelle, réalise une image de synthèse (Fig. 4.1(c)) et se déplace donc dans la maquette virtuelle (Fig. 4.1(a)), au fil des itérations de la maximisation de l'information mutuelle calculée entre image de synthèse et l'image (réelle). A convergence, la pose atteinte par la caméra virtuelle dans la maquette virtuelle, si elle est à l'échelle (voire géo-référencée), est la



(a)

(b)

(c)

FIGURE 4.1 – (a) La maquette virtuelle 3D texturée du XII^{ime} arrondissement de Paris, réalisée par l'IGN en 2010, (b) une image réelle acquise dans une rue et (c) la vue de synthèse correspondante.

même que celle de la caméra réelle, dont la pose est désormais calculée. Il s'agit d'un asservissement visuel virtuel [Marchand and Chaumette, 2002] direct pour lequel la caméra virtuelle doit suivre le même modèle de projection que la caméra réelle, avec les mêmes paramètres intrinsèques, dans le but que les images de synthèse aient la même géométrie que les images réelles.

Dans ce qui suit, la partie 4.1.2 adapte légèrement l'expression du critère d'information mutuelle photométrique au calcul de pose et apporte quelques précisions complémentaires à la partie 3.5.1.2 sur l'asservissement visuel basé information mutuelle. Ensuite, la partie 4.1.4 présente les résultats principaux avant la conclusion partielle (Partie 4.1.5).

4.1.2 Calcul de pose basé information mutuelle

Le problème de calcul de pose s'écrit de façon très similaire à celui de l'asservissement visuel [Dame and Marchand, 2011] présenté en partie 3.5.1.2, à ceci près que les intensités $\mathcal{I}_s$ de l'image courante sont issues de l'image de synthèse rendue à la pose de la caméra virtuelle dans la maquette virtuelle $\mathcal{M}$:

$$\widehat{\boldsymbol{\delta}\mathbf{p}} = \arg\max_{\boldsymbol{\delta}\mathbf{p}} MI(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}_s, \{\mathbf{p}, \mathcal{M}\}, \mathcal{I}^*). \tag{4.1}$$

La pose de la caméra virtuelle à convergence est directement celle de la caméra réelle $\widehat{\boldsymbol{\delta}\mathbf{p}}$ quand la caméra virtuelle a les mêmes paramètres intrinsèques γ_m que la caméra réelle, avec $m = p$ dans le présent travail.

Le problème de maximisation de l'information mutuelle (Eq. (4.1)) se résout itérativement à partir d'une pose $\boldsymbol{\delta}\mathbf{p}^{(0)}$, approchée de l'optimale, la pose optimale associée à l'image précédente acquise par la caméra réelle dans le contexte de suivi basé maquette virtuelle de cette partie. A cause des degrés de liberté couplés et des fortes non-linéarités de la fonction de vraisemblance $MI()$, la méthode de Newton (Partie 3.2.2) est préférée [Dame and Marchand, 2011] pour calculer les incréments de pose :

$$\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} = -\lambda \left[\mathbf{H}_{\mathbf{MI}_{\widehat{\boldsymbol{\delta}\mathbf{p}}}}\right]^{-1} \mathbf{J}_{\mathbf{MI}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}}^{\top}, \tag{4.2}$$

en suivant la même méthodologie que pour arriver à l'équation (3.19) pour le cas du coût $\mathscr{C}_{SSD}()$ (Eq. (3.3)), mais la nature différente du critère $MI()$ par rapport au coût $\mathscr{C}_{SSD}()$ mène à une équation différente. Une autre particularité est de considérer la matrice hessienne calculée pour la pose optimale dans le calcul de chaque $\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)}$. En plus de donner un meilleur comportement à convergence qu'avec la matrice hessienne désirée [Dame and Marchand, 2011], un gain non négligeable de temps de calcul est ici opéré car $\mathbf{H}_{\mathbf{MI}_{\widehat{\boldsymbol{\delta}\mathbf{p}}}}$ n'est calculé qu'une seule fois, à partir de l'image désirée seule, et non à chaque itération. La matrice jacobienne $\mathbf{J}_{\mathbf{MI}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}}$ est, quant à elle, calculée à chaque itération en prenant en compte les intensités $\mathcal{I}_s^{(k)}$ de l'image de synthèse rendue pour la pose courante $\boldsymbol{\delta}\mathbf{p}^{(k)}$.

4.1.3 Matrices jacobienne et hessienne

A partir des expressions de l'entropie marginale (Eq. (3.144)) et de l'entropie jointe (Eq.(3.143)), la fonction $MI()$ (Eq. (4.1)) se développe en :

$$MI(\boldsymbol{\delta \mathbf{p}}, \mathcal{I}_s, \{\mathbf{p}, \mathcal{M}\}, \mathcal{I}^*) = \sum_{i=0}^{N_i} \sum_{j=0}^{N_i} P_{\mathcal{I}_s\mathcal{I}^*}(i, j, \boldsymbol{\delta \mathbf{p}}) \log \left(\frac{P_{\mathcal{I}_s\mathcal{I}^*}(i, j, \boldsymbol{\delta \mathbf{p}})}{P_{\mathcal{I}_s}(i, \boldsymbol{\delta \mathbf{p}}) P_{\mathcal{I}^*}(j)} \right). \quad (4.3)$$

où les paramètres du problème ont été omis des distributions $P_{\mathcal{I}_s}$, $P_{\mathcal{I}^*}$ et $P_{\mathcal{I}_s\mathcal{I}^*}$ ainsi que la fonction de lissage des distributions (Eq. (3.148)), pour compacter les expressions.

A partir de l'équation (4.3) et des simplifications permises par la règle des dérivées en chaîne [Dowson and Bowden, 2006], les matrices d'interaction $\mathbf{J}_{\mathbf{MI}\boldsymbol{\delta \mathbf{p}}}$ et hessienne $\mathbf{H}_{\mathbf{MI}\boldsymbol{\delta \mathbf{p}}}$ s'expriment par :

$$\mathbf{J}_{\mathbf{MI}\boldsymbol{\delta \mathbf{p}}} = \sum_{i=0}^{N_i} \sum_{j=0}^{N_i} \mathbf{J}_{P_{\mathcal{I}_s\mathcal{I}^*}\boldsymbol{\delta \mathbf{p}}} \left(1 + \log \left(\frac{P_{\mathcal{I}_s\mathcal{I}^*}}{P_{\mathcal{I}^*}} \right) \right) \quad (4.4)$$

et

$$\mathbf{H}_{\mathbf{MI}\boldsymbol{\delta \mathbf{p}}} = \sum_{i=0}^{N_i} \sum_{j=0}^{N_i} \mathbf{J}^{\top}_{P_{\mathcal{I}_s\mathcal{I}^*}\boldsymbol{\delta \mathbf{p}}} \mathbf{J}_{P_{\mathcal{I}_s\mathcal{I}^*}\boldsymbol{\delta \mathbf{p}}} \left(\frac{1}{P_{\mathcal{I}_s\mathcal{I}^*}} - \frac{1}{P_{\mathcal{I}^*}} \right) + \mathbf{H}_{P_{\mathcal{I}_s\mathcal{I}^*}\boldsymbol{\delta \mathbf{p}}} \left(1 + \log \left(\frac{P_{\mathcal{I}_s\mathcal{I}^*}}{P_{\mathcal{I}^*}} \right) \right), \quad (4.5)$$

où indexes et degrés de liberté ont été omis par souci de compacité.

Pour exprimer les matrices jacobienne $\mathbf{J}_{P_{\mathcal{I}_s\mathcal{I}^*}\boldsymbol{\delta \mathbf{p}}}$ et hessienne $\mathbf{H}_{P_{\mathcal{I}_s\mathcal{I}^*}\boldsymbol{\delta \mathbf{p}}}$ de la distribution d'intensité jointe (Eqs. (4.4) et (4.5)), on repart de son expression générique (Eq. (3.137)) et on obtient :

$$\mathbf{J}_{P_{\mathcal{I}_s\mathcal{I}^*}(i,j,\boldsymbol{\delta \mathbf{p}})\boldsymbol{\delta \mathbf{p}}} = \frac{1}{|\mathcal{U}|} \sum_{\mathbf{u} \in \mathcal{U}} \mathbf{J}_{d_{Bs}\left(\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}}) - i\right)\boldsymbol{\delta \mathbf{p}}} d_{Bs}\left(\overline{I}^*(\mathbf{u}) - j \right) \quad (4.6)$$

$$\mathbf{H}_{P_{\mathcal{I}_s\mathcal{I}^*}(i,j,\boldsymbol{\delta \mathbf{p}})\boldsymbol{\delta \mathbf{p}}} = \frac{1}{|\mathcal{U}|} \sum_{\mathbf{u} \in \mathcal{U}} \mathbf{H}_{d_{Bs}\left(\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}}) - i\right)\boldsymbol{\delta \mathbf{p}}} d_{Bs}\left(\overline{I}^*(\mathbf{u}) - j \right), \quad (4.7)$$

où l'ensemble $\mathcal{U}$ des coordonnées des pixels de l'image est définit en équation (3.47), et avec :

$$\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}}) = \frac{N_i}{N_I} I_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}}), \quad (4.8)$$

les intensités mises à l'échelle en coordonnées non entières pour la fonction $d_{Bs}()$ basée B-spline (Eq. (3.147)) exploitée dans l'expression de la distribution d'intensité afin de la rendre dérivable :

$$\mathbf{J}_{d_{Bs}\left(\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}}) - i\right)\boldsymbol{\delta \mathbf{p}}} = -\frac{\partial d_{Bs}}{\partial i} \mathbf{J}_{\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}})\boldsymbol{\delta \mathbf{p}}} \quad (4.9)$$

$$\mathbf{H}_{d_{Bs}\left(\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}}) - i\right)\boldsymbol{\delta \mathbf{p}}} = \frac{\partial^2 d_{Bs}}{\partial i^2} \mathbf{J}^{\top}_{\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}})\boldsymbol{\delta \mathbf{p}}} \mathbf{J}_{\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}})\boldsymbol{\delta \mathbf{p}}} - \frac{\partial d_{Bs}}{\partial i} \mathbf{H}_{\overline{I}_s(\mathbf{u}, \boldsymbol{\delta \mathbf{p}})\boldsymbol{\delta \mathbf{p}}}. \quad (4.10)$$

Ci-dessus, les paramètres de $d_{Bs}()$ ont été omis pour compacter les expressions. $\mathbf{J}_{\overline{I}_s(\mathbf{u},\boldsymbol{\delta p})\boldsymbol{\delta p}}$ n'est autre que $\mathbf{J}_{\boldsymbol{\delta p}}$ (Eq. (3.9)), calculée à partir des intensités mises à l'échelle (Eq. (4.8)). De la même manière, $\mathbf{H}_{d_{Bs}(\overline{I}_s(\mathbf{u},\boldsymbol{\delta p})-i)\boldsymbol{\delta p}}$ est la matrice hessienne "photométrique" [Collewet and Marchand, 2011], calculée pour les intensités mises à l'échelle :

$$\mathbf{H}_{d_{Bs}(\overline{I}_s(\mathbf{u},\boldsymbol{\delta p})-i)\boldsymbol{\delta p}} = \frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta p}}^{\top} \nabla^2_{\mathbf{u}} \overline{I}_s \frac{\partial \mathbf{u}}{\partial \boldsymbol{\delta p}} + \nabla_u \overline{I}_s \, \mathbf{H}_u + \nabla_v \overline{I}_s \, \mathbf{H}_v, \tag{4.11}$$

où $\nabla_{\mathbf{u}} \overline{I}_s = (\nabla_u \overline{I}_s, \nabla_v \overline{I}_s)$ sont les gradients de l'image, $\nabla^2_{\mathbf{u}} \overline{I}$ sont les gradients des gradients de l'image et $\partial \mathbf{u}/\partial \boldsymbol{\delta p}$ est la matrice jacobienne géométrique de l'image (Eq. 3.11). $\mathbf{H}_u$ et $\mathbf{H}_v$ sont les matrices hessiennes des coordonnées d'un point par rapport à la variation de la pose de la caméra [Lapresté and Mezouar, 2004].

Ces dernières matrices jacobiennes et hessiennes dépendent à la fois des coordonnées du pixel $\mathbf{u}$ dans l'image et de la coordonnée ${}^cZ$ du point 3D de la scène qui se projette en $\mathbf{u}$. Dans le cas présent du calcul de pose basé maquette 3D texturée, les coordonnées ${}^cZ$ de tous les points de la maquette virtuelle observés par la caméra virtuelle sont connus lors de la réalisation de l'image de synthèse, pour laquelle l'algorithme classique de rendu est basé tampon de profondeur.

Durant le processus itératif d'optimisation, le déplacement de la caméra vers la pose optimale engendre un changement des coordonnées ${}^cZ$ de chaque point. La matrice hessienne $\mathbf{H}_{\mathbf{MI}_{\widehat{\boldsymbol{\delta p}}}}$ (Eq. (4.2)) à l'optimum étant considérée dans le calcul de l'incrément de pose à chaque itération, on fixe pour chaque pixel de l'image désirée la coordonnée ${}^{c*}Z = {}^{c(0)}Z$. Dans le suivi visuel basé maquette virtuelle, l'écart entre images successivement acquises par la caméra en mouvement dans l'environnement réel étant faible, cette approximation est suffisamment raisonnable pour que l'estimation de $\mathbf{H}_{\mathbf{MI}_{\widehat{\boldsymbol{\delta p}}}}$ soit précise, comme le montrent les résultats de la partie 4.1.4.

L'algorithme présenté en figure 4.2 résume le calcul de pose de caméra basé information mutuelle dans un processus de suivi visuel basé maquette virtuelle texturée.

4.1.4 Résultats

La méthode de suivi visuel basé maquette virtuelle 3D texturée par maximisation d'information mutuelle a été appliquée en considérant des maquettes d'échelle graduelle : boîte de sachets de thé (Fig. 4.3(a)), bâtiments d'habitation collective (Fig. 4.3(b)), plusieurs rues d'un arrondissement de Paris (Fig. 4.1). Les images de synthèse sont rendues à l'aide d'outils classiques de l'informatique graphique (OpenGL [1]/Ogre3D [2]) exploitant la puissance de calcul des cartes graphiques. Contrairement à l'image réelle, l'image de synthèse peut avoir des pixels "vides", c'est-à-dire dans lesquels aucun élément de la maquette virtuelle ne se projette (tous les pixels noirs de l'image en bas de la figure 4.3(a)). Ces pixels sont écartés du calcul des distributions d'intensité et de l'incrément de pose à chaque itération.

1. opengl.org
2. www.ogre3D.org

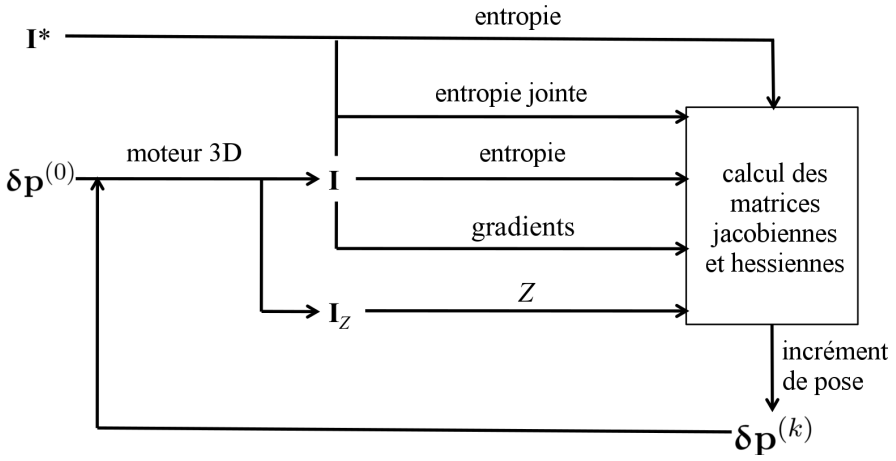
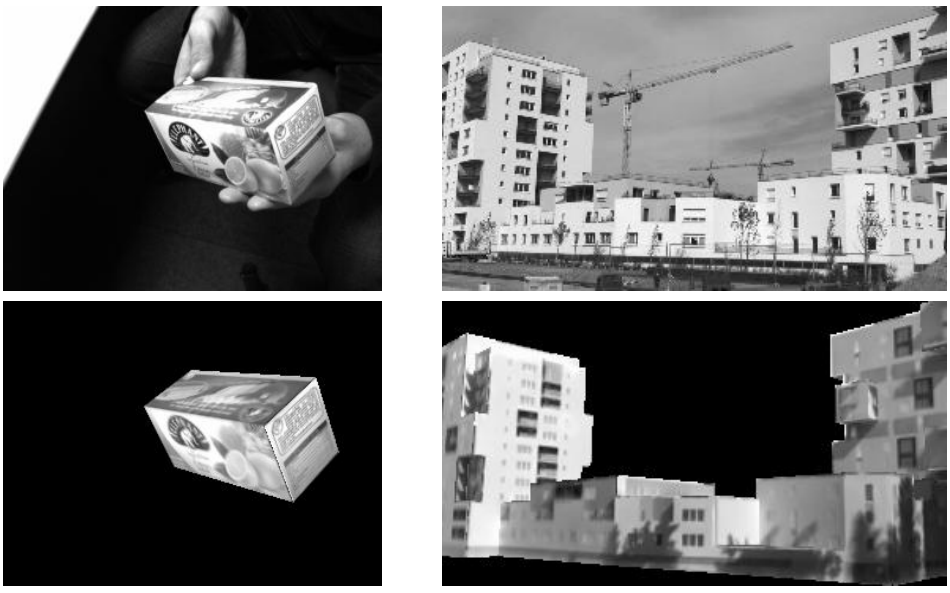


FIGURE 4.2 – Synoptique de l'algorithme d'estimation de pose basé information mutuelle. Le processus boucle jusqu'à ce que l'information partagée par les images courante (de synthèse) et désirée (réelle), représentées par les vecteurs d'intensités $\mathbf{I}$ et $\mathbf{I}^*$, soit stable. $\mathbf{I}_Z$ indique la sortie du moteur de rendu donnant les ${}^cZ$.

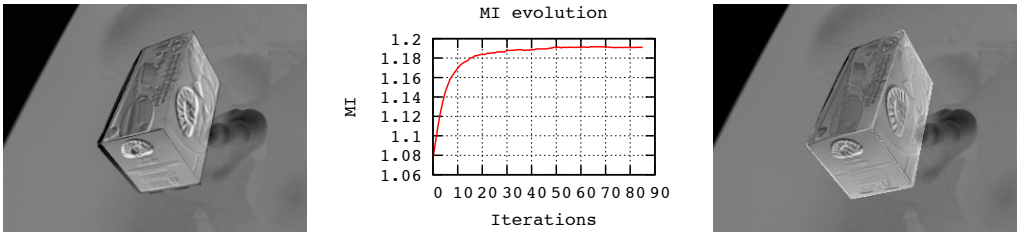


(a) (b)

FIGURE 4.3 – Exemples d'expérimentations de difficulté graduelle : (a) caméra fixe, petit objet (boîte de sachets de thés) en mouvement et (b) caméra tenue en main, grand bâtiment d'habitation collective. Dans les deux cas, l'image du haut est une image acquise par la caméra réelle et l'image du bas est l'image de synthèse pour la pose de caméra optimale selon le critère d'information mutuelle (poses de caméras virtuelle et réelle identiques).

Les pixels de l'image réelle aux mêmes coordonnées que les pixels vides de l'image de synthèse sont aussi exclus des calculs. Cette façon de faire est une approximation sauf à la pose optimale. Pour les deux types d'images, les pixels voisins des pixels ignorés sont aussi exclus du calcul de l'incrément de pose à chaque itération car on ne peut calculer les gradients des images aux coordonnées de ces pixels. Autrement dit, seuls les pixels exploitables pour l'ensemble des calculs sont considérés à chaque itération.

Validation du calcul de pose : L'expérience de la boîte de sachets de thé permet de valider le concept et de montrer le comportement du calcul de pose dans un cas simple qui présente une évolution lisse de la fonction $MI()$ qui atteint son maximum de façon logarithmique au fil des itérations (Fig. (4.4)) .



(a) (b) (c)

FIGURE 4.4 – Validation du calcul de pose sur un cas simple où la caméra observe une boîte dont la pose initiale est distante de 2,5 cm et 3,3° de l'optimale, ce que (a) la différence entre l'image de synthèse et l'image réelle montre bien (cette différence n'est pas exploitée par l'algorithme, elle sert juste d'illustration). (b) L'évolution de l'information mutuelle au fil des itérations traduit l'évolution itérative de la pose de la caméra virtuelle vers l'optimum, qui correspond bien à la pose de la caméra réelle, comme l'illustre (c) l'image de différence à convergence de la maximisation de la fonction $MI()$.

Evaluation de la précision du suivi : Avant les conditions réelles, la méthode de suivi basé maquette virtuelle a été évaluée en simulation à partir d'une séquence du jeu de données TrakMark [3] afin d'avoir une vérité terrain. Le suivi est un succès sur toute la séquence "Conference venue package 01" représentant une trajectoire d'environ 12 m pour un écart moyen à la vérité terrain de 15 mm en position et 0,15° en orientation, soit moitié moins que l'approche géométrique [Petit et al., 2012] par rapport à laquelle une comparaison a été faite.

Application à la localisation de véhicule embarquant une caméra : La méthode de suivi visuel basé maquette virtuelle texturée a été appliquée avec succès sur des séquences d'images dans lesquelles une boîte de sachets de thé est déplacée

3. trakmark.net

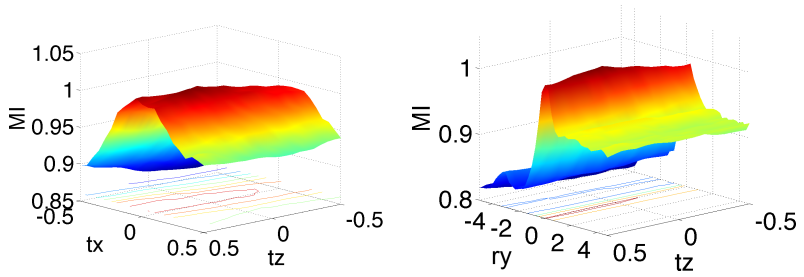
(a) (b)

FIGURE 4.5 – Exemples d'images de la séquence acquises dans le XII$^{\text{ime}}$ arrondissement de Paris avec occultations des bâtiments par des personnes et des véhicules.

par rapport à une caméra fixe, à l'intérieur (Fig. 4.3(a)), ou encore quand la caméra d'un smartphone, tenu en main pendant la marche, est pointée vers un bâtiment d'habitation collective, à l'extérieur, donc (Fig. 4.3(b)).

Mais le plus grand défi réside dans l'application à la localisation urbaine de véhicule embarquant une caméra (Fig. 4.5), exploitant ainsi une maquette virtuelle de ville dans un cas concret, comme évoqué en introduction (Partie 4.1(a)).

Une pose initiale est supposée fournie au début de la séquence de localisation d'environ 300 m par la méthode de suivi basé maquette virtuelle texturée, par exemple par GNSS à l'entrée de la ville, avant que la réception du signal ne soit perturbée par les bâtiments de la ville. Pour donner une indication sur la précision requise de la pose initiale dans ce contexte urbain, le domaine de convergence de la fonction $MI()$ est tracé en figure 4.6 pour deux paires de degrés de liberté. Ces tracés montrent que le maximum de la fonction $MI()$ peut être atteint d'une pose distante de plus de 70 cm de l'optimale, réduite à 50 cm en présence d'un écart de



(a) (b)

FIGURE 4.6 – Fonction de vraisemblance $MI()$ calculée entre une image réelle (la première de la séquence considérée en ville) et les images de synthèses en faisant varier deux degrés de liberté de la caméra virtuelle : translations horizontales (tx et tz) en mètre, d'une part (a), et translation longitudinale (tz) en mètre et cap (ry) en degrés, d'autre part (b).

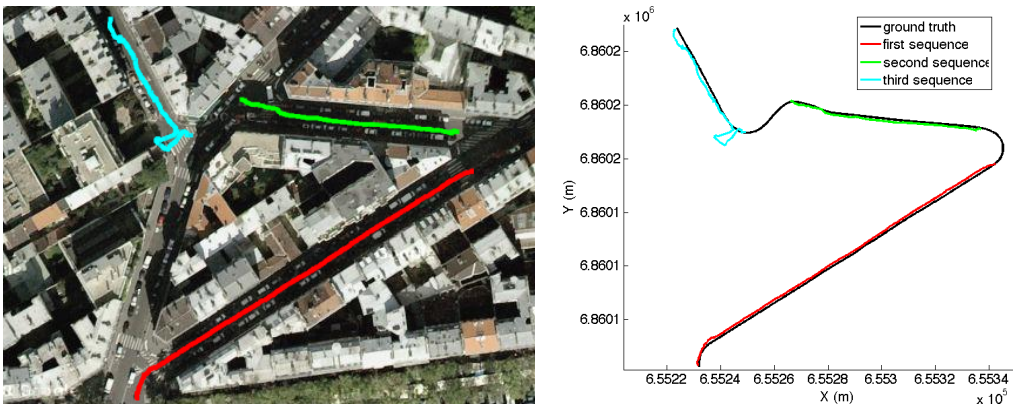
cap de 2,5°.

Cette dernière caractérisation expérimentale du domaine de convergence de la fonction $MI()$ pour le calcul de pose basé information mutuelle permet de déduire à quelle vitesse maximum le véhicule peut se déplacer sans mettre le suivi en difficulté. Ce calcul dépend de la fréquence d'acquisition de la caméra qui, pour 25 images par seconde, autorise une vitesse maximale de 63 Km/h, ce qui est au-dessus des limites autorisées en ville, le contexte d'application.

Une fois la première pose correctement initialisée, le suivi visuel basé maquette virtuelle est exécuté sur une séquence acquise en faisant rouler le véhicule dans trois rues du XII$^{\text{ime}}$ arrondissement de Paris. La trajectoire estimée est superposée sur une vue satellite de la ville pour une évaluation qualitative de la qualité d'estimation (Fig. 4.7(a)). La trajectoire calculée est bien alignée avec le centre des rues à sens unique, ce qui montre la stabilité et la précision des poses estimées, malgré les occultations des bâtiments par les voitures (Fig. 4.1(b) et 4.1(c)), les variations d'illumination, les vibrations de la camera ou encore le virage serré au début de la séquence (bas de la figure 4.7(a)).

Pour une comparaison quantitative, le véhicule embarquant la caméra pour faire l'acquisition de la séquence traitée précédemment, embarquait aussi d'autres capteurs (récepteur GNSS, centrale inertielle) produisant un ensemble de positions géoréférencées synchronisées avec les images (trajectoire noire en figure 4.7(b)). Comparée à cette autre estimation de la trajectoire du véhicule, les poses calculées par le suivi basé modèle 3D seul mènent à un écart moyen de 1,56 m (écart-type de 0,61 m, erreur maximale = 2,97 m, erreur minimale = 0,04 m). Rapporté à la distance totale parcourue de 286,58 m, l'erreur moyenne relative est de 0,54 %, sans dérive.

Contrairement à la simulation, la comparaison avec d'autres méthodes de suivi et calcul de pose basées maquette virtuelle n'a pas été possible car les primitives géométriques conventionnelles (ex : ASIFT [Morel and Yu, 2009], plus général que



(a) (b)

FIGURE 4.7 – Portions de trajectoire (rouge, vert et cyan) estimées par le suivi basé maquette virtuelle texturée, sans filtre.

SIFT) se mettent mal en correspondance à cause des différences entre les textures plaquées sur les façades et les images acquises par la caméra embarquée.

Ce sont d'ailleurs ces différences trop importantes entre la maquette virtuelle et la réalité qui empêchent le suivi basé maquette virtuelle d'estimer la trajectoire sur toute la longueur de la séquence (entre les trajectoires rouge et verte et entre les trajectoires verte et cyan en figure 4.7(a)). Comme évoqué en introduction (Partie 4.1.1), plusieurs facteurs entrent en jeu : le réalisme de la maquette, la résolution de la texture plaquée sur les façades et les occultations. La figure 4.8 montre les cas les plus difficiles où le suivi basé maquette virtuelle texturée échoue à cause des trop grandes approximations de la maquette par rapport à la réalité. Dans ces cas aussi, les approches basées primitives géométriques échouent : 23 % de fausses correspondances sur 26 correspondances obtenues par ASIFT, dont les correctes sont concentrées dans une petite région à droite des images (Fig. 4.8(a) et 4.8(b)) ; 51,5 % de fausses correspondances sur 64 correspondances ASIFT (Fig. 4.8(c) et 4.8(d)).

Dans ces secteurs, la maquette virtuelle est tout simplement inexploitable et basculer temporairement sur une autre technique, même moins précise que le calcul de pose basé information mutuelle, là où la maquette virtuelle est de qualité suffisante, est nécessaire le temps que le véhicule se déplace vers une zone où la maquette est de meilleure qualité.

Résultats en vidéo : https://youtu.be/emm-Eqo3220.

(a) réel (b) synthèse (c) réel (d) synthèse

FIGURE 4.8 – Deux cas où le suivi basé maquette virtuelle échoue. (a-b) la résolution de la texture de la maquette virtuelle est extrêmement basse, particulièrement à gauche. (c-d) une combinaison de problèmes d'occultation partielle, de placage erroné de texture et d'approximation trop importante de la géométrie de la scène réelle (le store est quasi perpendiculaire à la façade dans la réalité alors que son image est plaquée sur la façade verticale dans la maquette virtuelle).

4.1.5 Conclusion partielle

La prise en compte du critère d'information mutuelle dans le formalisme de calcul de pose a permis d'exploiter directement les intensités des pixels des images acquises par une caméra réelle dans un environnement réel à celles d'images de synthèses rendues par une caméra virtuelle observant une maquette virtuelle 3D texturée de cet environnement. Les expressions théoriques sont très proches de l'asservissement visuel basé information mutuelle et on confirme le même type de robustesse aux

occultations partielles (certaines approximations de la maquette virtuelle peuvent être considérées comme tel), aux variations d'illumination ainsi que, dans une mesure limitée, à la multi-modalité. En effet, comparer des images de synthèse avec des images réelles nécessite que la caméra virtuelle soit la plus fidèle possible à la caméra réelle mais tous les phénomènes physiques de la formation d'une image (cf. Chapitre 2, pour une évocation) ne sont pas pris en compte, loin de là.

Cependant, la simplification du mécanisme de formation des images permet d'atteindre des temps de rendu très réduits, laissant envisager de pouvoir faire tenir l'ensemble des itérations de la boucle d'optimisation pour atteindre la pose optimale dans le temps s'écoulant entre deux images acquises, ce qui n'était pas encore le cas dans le contexte applicatif visé dans ce travail au moment où il a été réalisé en 2011. En effet, l'ordre de grandeur était d'environ 4 secondes de traitement par image acquise, assez éloigné du temps-réel. Cependant, la carte graphique (GPU) était seulement utilisée classiquement pour le rendu des images de synthèse de la maquette virtuelle et une nouvelle implémentation du reste de l'algorithme adaptée aux traitements massivement parallèles des GPU combinée à leur évolution matérielle faisant toujours croître la fréquence des opérations élémentaires pourraient probablement suffire à atteindre des temps de calcul compatibles avec le temps-réel.

Une alternative consisterait à s'inspirer de travaux plus récents évitant le rendu d'image de synthèse à chaque itération de la boucle d'optimisation, gagnant ainsi un temps précieux. En faisant des rendus réguliers dans la maquette virtuelle hors ligne et en sélectionnant, en ligne, le rendu le plus proche de la pose actuelle de la caméra réelle avant de lancer la boucle d'optimisation, l'image des profondeurs associée à l'image de synthèse est exploitée pour la transformer progressivement vers l'image désirée [Ok et al., 2016].

Enfin, la qualité des maquettes virtuelles ne cesse de s'accroître, éliminant presque certains problèmes comme la faible résolution des textures ou les approximations faites sur la réalité, parfois si importantes que la maquette en devient inexploitable. Sans ces problèmes, d'autres critères, peut-être moins robustes mais de complexité moindre, que l'information mutuelle peuvent être envisagés, comme l'adaptation du coût SSD des approches directes pures (Partie 3.2.1) au suivi visuel basé maquette virtuelle texturée du chapitre 5.

4.2 Exploration : Commande de caméra virtuelle basée entropie

4.2.1 Introduction

L'amélioration permanente de la qualité et de la richesse des maquettes virtuelles 4.1.5, notamment de bâtiments du patrimoine historique [4], engendre un regain d'intérêt des usagers qui, cependant, ont besoin d'assistance pour les explorer [Freitag et al., 2018]. Une maquette virtuelle très détaillée et représentant un vaste environnement est un élément de motivation pour la découvrir mais la difficulté à pouvoir l'apprécier correctement reste un frein. Cette difficulté vient, en partie, des interfaces de visualisation et d'interaction actuelles que chacun veut pouvoir utiliser pour découvrir soi-même la maquette, quand l'interaction temps-réel est possible, sans pour autant savoir, de prime abord, ce qu'il est pertinent de regarder. Cela peut conduire à une lassitude rapidement suivie de l'arrêt de l'exploration. A l'inverse, les visites virtuelles scenarisées offrent une sélection de points de vue pertinents mais sont parfois perçues comme frustrantes car contraignant trop l'usager qui aimerait explorer une zone ou un détail absent de la visite virtuelle, conduisant aussi à la lassitude. La définition des visites virtuelles est, de plus, une étape préliminaire nécessaire dans ce dernier cas, demandant l'intervention d'un expert.

Dans le but de trouver un compromis entre l'exploration totalement libre d'une maquette virtuelle brute et la visite virtuelle scénarisée, des approches exploitant la théorie de l'information de Shannon, essentiellement des entropies visuelles géométriques calculées à partir de maillages polygonaux [Vázquez et al., 2001, Andujar et al., 2004, Polonsky et al., 2005, Bonaventure et al., 2011], définissent automatiquement des zones d'intérêt au sein de la maquette et calculent des chemins passant par ces zones. Cependant, aucune connaissance sur la pertinence de la visualisation lors du déplacement de la caméra virtuelle d'une zone pertinente à une autre n'est disponible.

D'autres approchent formulent le problème du déplacement de la caméra virtuelle dans ce cadre comme un problème d'asservissement visuel permettant de conserver un objet connu correctement cadré dans l'image, au sens cinématographique du terme [Courty and Marchand, 2001], permettant d'assurer la pertinence de la visualisation pendant le déplacement de la caméra virtuelle. Cependant, l'objet d'intérêt doit être préalablement défini.

Fusionner les deux approches en un asservissement visuel de caméra virtuelle basé entropie visuelle présente le potentiel d'atteindre automatiquement une zone pertinente de façon pertinente. Plus particulièrement, le critère de l'entropie photométrique est un bon candidat puisqu'il permet d'être générique aux représentations des maquettes virtuelles (maillage texturé ou nuage de points 3D), s'affranchissant ainsi de l'exclusivité des maillages des approches ayant déjà considéré l'entropie visuelle, mais géométrique, comme évoqué plus haut.

4. Programme e-Cathédrale du laboratoire MIS, mis.u-picardie.fr/e-cathedrale

4.2.2 Exploration basée entropie photométrique

Le principe consiste à commander automatiquement le déplacement de la caméra virtuelle en maximisant l'entropie photométrique de toute l'image et d'afficher les images de toutes les itérations de la boucle d'optimisation, chacune contenant de plus en plus d'information. Pour ce faire, une partie des développements mathématiques faits pour le suivi basé maquette virtuelle 3D texturée (Partie 4.1) est reprise, en particulier l'expression de l'entropie marginale (Eq. (3.144)) des intensités de l'image de synthèse $\mathcal{I}_s$:

$$H_{\mathcal{I}_s}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}_s, \{\mathbf{p}, \mathcal{M}\}) = -\sum_{i=0}^{N_i} P_{\mathcal{I}_s}(i, \boldsymbol{\delta}\mathbf{p}, \mathcal{I}_s, \{\mathbf{p}, \mathcal{M}\}) \log(P_{\mathcal{I}_s}(i, \boldsymbol{\delta}\mathbf{p}, \mathcal{I}_s, \{\mathbf{p}, \mathcal{M}\})), \tag{4.12}$$

avec les mêmes mises à l'échelle des intensités $\mathcal{I}_s$ auxquelles la fonction de lissage $d_{Bs}()$ est appliquée pour assurer la dérivabilité de la distribution d'intensité $P_{\mathcal{I}_s}()$.

La différence majeure avec toutes les méthodes décrites précédemment, c'est qu'il n'y a pas d'image désirée à laquelle comparer l'image courante pour optimiser les degrés de liberté de la caméra. Dès lors, le problème de maximisation de l'entropie photométrique s'écrit :

$$\widehat{\boldsymbol{\delta}\mathbf{p}} = \arg\max_{\boldsymbol{\delta}\mathbf{p}} H_{\mathcal{I}_s}(\boldsymbol{\delta}\mathbf{p}, \mathcal{I}_s, \{\mathbf{p}, \mathcal{M}\}). \tag{4.13}$$

En suivant la méthode d'optimisation de Newton pour résoudre l'équation (4.13), l'expression de l'incrément de pose est similaire à celle de l'information mutuelle (Eq. (4.2)) :

$$\dot{\boldsymbol{\delta}\mathbf{p}}^{(k)} = -\lambda \left[\mathbf{H}_{\mathbf{H}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}}\right]^{-1} \mathbf{J}_{\mathbf{H}_{\boldsymbol{\delta}\mathbf{p}^{(k)}}}^{\top}, \tag{4.14}$$

sauf que la matrice hessienne doit être calculée pour l'image courante, puisque l'image finale n'est pas connue, contrairement au cas du calcul de pose où l'on peut faire l'approximation que l'image désirée joue ce rôle (Eq. (4.2)). Dès lors, on exprime :

$$\mathbf{J}_{\mathbf{H}_{\boldsymbol{\delta}\mathbf{p}}} = \sum_{i=0}^{N_i} \mathbf{J}_{P_{\mathcal{I}_s}\boldsymbol{\delta}\mathbf{p}} \left(1 + \log\left(P_{\mathcal{I}_s}\right)\right) \tag{4.15}$$

et

$$\mathbf{H}_{\mathbf{H}_{\boldsymbol{\delta}\mathbf{p}}} = \sum_{i=0}^{N_i} \mathbf{J}_{P_{\mathcal{I}_s}\boldsymbol{\delta}\mathbf{p}}^{\top} \mathbf{J}_{P_{\mathcal{I}_s}\boldsymbol{\delta}\mathbf{p}} \frac{1}{P_{\mathcal{I}_s}} + \mathbf{H}_{P_{\mathcal{I}_s}\boldsymbol{\delta}\mathbf{p}} \left(1 + \log\left(P_{\mathcal{I}_s}\right)\right), \tag{4.16}$$

avec :

$$\mathbf{J}_{P_{\mathcal{I}_s}(i,\boldsymbol{\delta}\mathbf{p})\boldsymbol{\delta}\mathbf{p}} = \frac{1}{|\mathcal{U}|} \sum_{\mathbf{u}\in\mathcal{U}} \mathbf{J}_{d_{Bs}\left(\overline{I}_s(\mathbf{u},\boldsymbol{\delta}\mathbf{p})-i\right)\boldsymbol{\delta}\mathbf{p}} \tag{4.17}$$

$$\mathbf{H}_{P_{\mathcal{I}_s}(i,\boldsymbol{\delta}\mathbf{p})\boldsymbol{\delta}\mathbf{p}} = \frac{1}{|\mathcal{U}|} \sum_{\mathbf{u}\in\mathcal{U}} \mathbf{H}_{d_{Bs}\left(\overline{I}_s(\mathbf{u},\boldsymbol{\delta}\mathbf{p})-i\right)\boldsymbol{\delta}\mathbf{p}}, \tag{4.18}$$

les matrices jacobiennes $\mathbf{J}_{d_{Bs}\left(\overline{I}_s(\mathbf{u},\boldsymbol{\delta}\mathbf{p})-i\right)\boldsymbol{\delta}\mathbf{p}}$, respectivement hessiennes $\mathbf{H}_{d_{Bs}\left(\overline{I}_s(\mathbf{u},\boldsymbol{\delta}\mathbf{p})-i\right)\boldsymbol{\delta}\mathbf{p}}$, étant définies en équations (4.9) et, respectivement, (4.10).

4.2.3 Résultats

L'exploration maximisant l'entropie photométrique est validée sur des nuages de points 3D denses, où chaque point 3D a une couleur RGB, c'est-à-dire codée sur les trois canaux Rouge, Vert et Bleu. Les trois nuages de points 3D sur lesquels la méthode est appliquée sont obtenus par scanner laser (Faro Focus 3D), équipé d'une caméra qui balaye l'environnement dans un second temps afin de pouvoir affecter une couleur à chaque point 3D. Quand l'environnement est trop grand ou complexe pour être numérisé par un seul scan, le scanner est déplacé pour réaliser plusieurs scans, qui sont ensuite assemblés pour former le nuage de points 3D final qui vient en entrée de l'algorithme d'exploration. Aucun pré-traitement du nuage de points 3D n'est réalisé. Avec ou sans assemblage, le nuage de points 3D est dense mais cette densité n'est que rarement constante.

Les images de synthèse sont rendues avec les mêmes outils que ceux rapportés en partie 4.1.4. Comme dans cette dernière, seuls les pixels exploitables pour l'ensemble des calculs sont considérés à chaque itération : les pixels de l'image de synthèse n'ayant pas accueilli la projection d'un point 3D du nuage et leurs voisins sont ignorés du calcul de l'incrément de pose. Dans le cas du nuage de points 3D, cette sélection des pixels à considérer dans les calculs peut conduire à un ensemble vide de pixels quand la densité du nuage de points 3D est trop faible par rapport à la résolution de la caméra. Cependant, la haute densité des nuages de points 3D considérés dans cette partie limite grandement ce problème.

Validation de la maximisation d'entropie photométrique : L'exploration maximisant l'entropie photométrique est tout d'abord validée sur le nuage de points 3D d'une salle de travail aux murs blancs sur lesquels des posters sont affichés. L'algorithme d'optimisation est exécuté à partir d'une pose manuellement fixée de telle sorte que la caméra pointe une zone de la maquette virtuelle faible en information (image de gauche en figure 4.9). L'image de synthèse rendue à chaque itération de la boucle d'optimisation de l'équation (4.13) est affichée et un sous ensemble des 30 itérations requises pour converger au maximum local d'entropie photométrique est reporté en figure 4.9. Une évaluation qualitative visuelle permet de constater que le contenu de l'image finale est plus varié que celui de l'image initiale avec un accroissement progressif de cette variété au fil du déplacement de la caméra virtuelle.

FIGURE 4.9 – Extrait de la série d'images obtenues durant le déplacement progressif de la caméra virtuelle par maximisation de l'entropie photométrique de ces images au sein d'un nuage de points 3D d'une salle de travail aux murs plats et lisses.

Analyse succincte de la maximisation d'entropie photométrique : De difficulté graduelle, une maquette virtuelle, sous la forme d'un nuage de points 3D, de l'intérieur de la cathédrale d'Amiens, bien moins plate et beaucoup plus riche, est considérée dans une seconde validation. La pose initiale de la caméra virtuelle est fixée manuellement à "l'extérieur" de la maquette virtuelle, de telle sorte qu'elle ne soit que partiellement visible, dans un peu plus de la moitié du champ de vue (Fig. 4.10, image en haut à gauche : les parties blanches des images sont "vides"). Qualitativement, la maximisation de l'entropie photométrique permet de remplir progressivement le champ de vue de la caméra virtuelle avec un accroissement progressif de la quantité d'information présente dans l'image au fil des itérations.

L'accroissement de la quantité d'information dans l'image à chaque itération de la maximisation de l'entropie photométrique est clair sur le tracé de l'évolution de l'entropie photométrique au fil des itérations (Fig. 4.11).

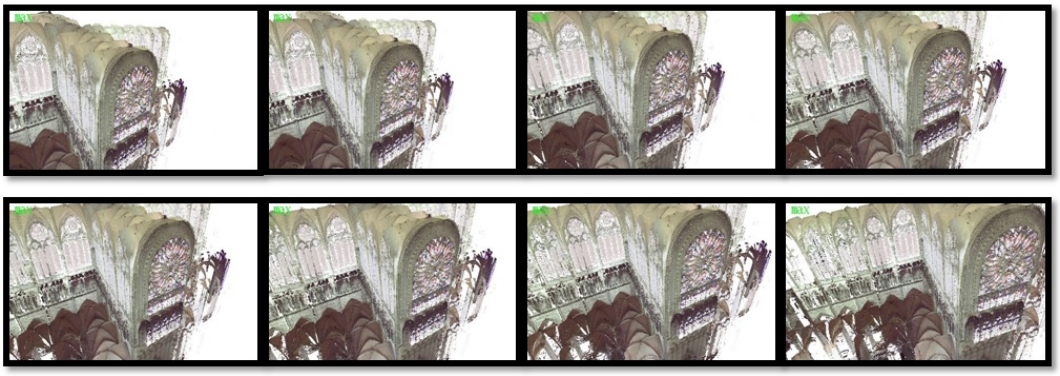


FIGURE 4.10 – Extrait de la série d'images obtenues durant le déplacement progressif de la caméra virtuelle par maximisation de l'entropie photométrique de ces images au sein d'un nuage de points 3D de l'intérieur de la cathédrale d'Amiens obtenu à partir de nombreuses positions de mesures.

Application à l'exploration de maquette virtuelle : L'évolution de l'entropie au fil des itérations (Fig. 4.11), avec un gain (λ, Eq. (4.14)) constant, illustre la non-linéarité du critère d'entropie photométrique vis-à-vis des six degrés de liberté de la caméra. En pratique, cette non-linéarité peut engendrer une apparence de mouvement lent dans les zones où l'information varie peu, rapide dans les zones où l'information varie plus. Le passage d'un comportement à l'autre est visible par les discontinuités de l'évolution de l'entropie photométrique (Fig. 4.11, itérations 13, 54, 65 et 87). Ces saccades du mouvement apparent dans l'image sont à éviter pour le confort de l'usager. Pour ce faire, la maximisation de l'entropie photométrie est incorporée dans un schéma de commande hiérarchique, contraignant l'amplitude du flot optique, qui est facilement prédit grâce à l'incrément de pose courant $\dot{\boldsymbol{\delta}}\mathbf{p}^{(k)}$ (Eq. (4.14)) et la connaissance de la maquette virtuelle $\mathcal{M}$. Pour parachever le

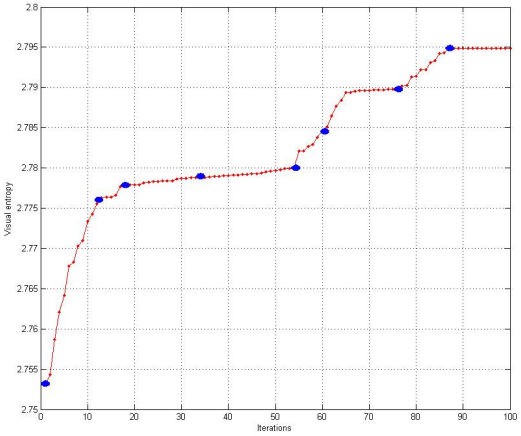


FIGURE 4.11 – Evolution de l'entropie photométrique de l'image de synthèse, au fil des itérations de se maximisation. Les gros points bleus sur la courbe d'évolution sont associés aux images de la figure 4.10.

réalisme et le confort visuel du mouvement de la caméra virtuelle, deux contraintes classiques sont ajoutées en amont et étendent le calcul de l'incrément de pose $\boldsymbol{\delta}\dot{\mathbf{p}}^{(k)}$ (Eq. (4.14)) pour former une loi de commande hybride de caméra virtuelle : la verticale de la pose de vue et l'évitement d'obstacles *[Habibi et al., 2015]*[5].

La figure 4.12 présente une série d'images extraites à différentes itérations d'exécution de la loi de commande virtuelle hybride-hiérarchique sur un nuage de points 3D de l'extérieur du palais de justice d'Amiens. A partir d'une pose de caméra où près de 30% du champ de vue ne montre rien, l'algorithme déplace automatiquement la caméra pour remplir le cadre avec un maximum d'information dans l'image finale à convergence.

Remarque 6 (A propos de la densité du nuage de points 3D) *Le mécanisme d'évitement d'obstacles, qui est en fait une contrainte de maintient de la caméra au-delà d'une distance des points 3D du nuage, évoqué dans le dernier paragraphe, peut aussi être vu comme un moyen indirect d'éviter que l'image de synthèse n'accueille trop peu de points 3D pour pouvoir calculer l'entropie photométrique, et donc l'incrément (comme évoqué au début de cette partie). Une autre alternative pour lutter contre ce problème potentiel consiste à considérer une caméra virtuelle de basse définition, qui limiterait les manques dans l'image, pour le calcul de l'incrément de pose, tout en affichant une image de définition supérieure pour l'usager. Cela aurait aussi pour effet bénéfique de réduire les temps de calcul, puisque moins de pixels seraient considérés. Une étude reste cependant à mener à propos de l'impact de la définition de l'image sur le comportement de la maximisation de*

5. Ces contraintes et le schéma de commande hiérarchique ne sont pas détaillés ici car reposant sur des critères géométriques, non directs.

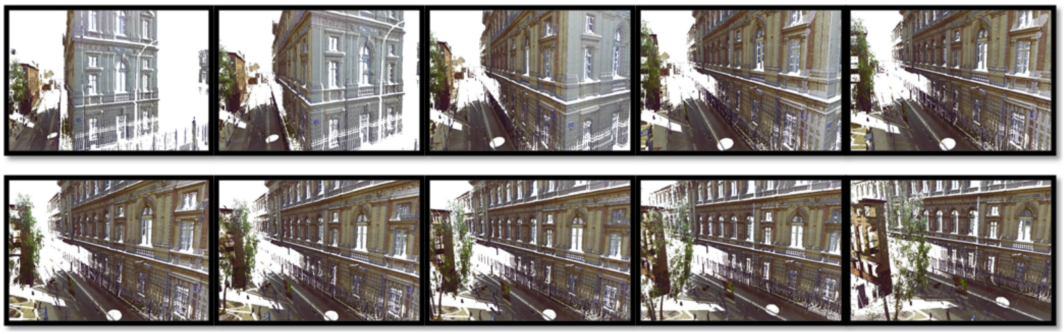


FIGURE 4.12 – Extrait de la série d'images obtenues durant le déplacement de la caméra virtuelle par le schéma de commande hybride-hiérarchique maximisant l'entropie photométrique tout en gardant la verticale, en évitant les obstacles et en ayant un mouvement apparent fluide et quasi constant. La maquette virtuelle est un nuage de points 3D de la rue Lesueur, à Amiens, qui longe le palais de justice.

l'entropie photométrique car il faut qu'il y ait suffisamment de pixels considérés pour que les distributions d'intensité aient un sens. Néanmoins, selon le contenu de l'image, le seuil peut être très bas puisqu'en se référant aux résultats de suivi visuel de région plane basé information mutuelle [Dame and Marchand, 2010], exploitant donc l'entropie photométrique, on peut déduire que des régions de l'ordre de 80×65 *pixels peuvent être suivies avec succès. Dans les expérimentations menées pour l'exploration basée entropie photométrique, ces problématiques n'ont jamais été rencontrées, même si aucun soin particulier n'a été mis en place pour les éviter.* ◇

4.2.4 Conclusion partielle

L'incorporation de l'expression de l'entropie photométrique de toute l'image dans un problème d'optimisation des degrés de liberté de caméra virtuelle au sein d'une maquette virtuelle, maximisant cette entropie, a permis de déplacer automatiquement cette caméra vers un maximum local d'information visuelle. A chaque itération de la boucle d'optimisation, les images synthétisées sont de plus en plus riches, en terme de quantité d'information, au sens de l'entropie, rendant pertinente la façon dont le point de vue le plus pertinent, localement, est rejoint.

Cette méthode, s'appuyant sur les développements théoriques faits pour le calcul de pose basé information mutuelle (Partie 4.1), peut s'appliquer à toute maquette virtuelle pourvue de texture, au sens de l'informatique graphique (une couleur par point, pour les nuages de points 3D), sans calcul a priori, ni sans connaître la maquette virtuelle a priori, non plus (pas de sémantique et seule une pose initiale qui permette de voir une partie de la maquette est requise). Cette caractéristique reste rare dans l'état de l'art car la plupart des méthodes de caractérisation de pertinence de point de vue reposent sur des critères géométriques explicites tel

que le nombre de triangles visibles, la surface projetée dans l'image, des caractéristiques sur la silhouette de la maquette virtuelle dans l'image, etc, certes la plupart du temps combinés dans des critères inspirés de la théorie de l'information [Bonaventura et al., 2018].

Le problème de temps de calcul évoqué pour le calcul de pose basé information mutuelle, au moment où son développement avait été fait (Partie (4.1.5)), est éliminé dans l'exploration maximisant l'entropie photométrique puisque les rendus d'images de synthèse faits à chaque itération sont autant de résultats à afficher à l'usager. Cela compense le fait qu'il faille calculer la matrice hessienne à chaque itération.

Au sein du contexte ayant motivé le développement de cette méthode d'exploration, les perspectives sont d'intégrer des entrées utilisateur afin de rendre la navigation interactive.

4.3 Conclusion du chapitre

Ce chapitre a montré comment l'asservissement visuel basé information mutuelle s'étend au calcul de pose et au suivi visuels de maquette virtuelle. Les résultats expérimentaux montrent la validation de la méthode et confirment la robustesse du critère d'information mutuelle aux changements d'illumination, occultations partielles et à la nature différente, certes légèrement dans le cas présent, des images comparées (réelle et synthétique). L'information mutuelle montre aussi une certaine robustesse aux écarts de résolution entre les images acquises et les textures du maillage 3D texturé formant la maquette virtuelle, tant que ces écarts restent raisonnables, c'est-à-dire qu'il y a suffisamment d'informations comparables partagées par les images. Quand la résolution de la texture de la maquette virtuelle est trop basse ou quand elle présente d'autres écarts trop importants à la réalité (géométrie trop approximée, texture erronée, etc) le suivi basé maquette virtuelle seul atteint ses limites, mais elles sont largement supérieures à celles des classiques descripteurs ponctuels.

D'autre part, ce chapitre a montré une ré-utilisation des développements théoriques de l'asservissement et du suivi visuels exploitant l'information mutuelle, dans un contexte, en apparence éloigné de la robotique, d'exploration de maquette virtuelle dense et détaillée. Sans information désirée connue a priori, la maximisation de l'entropie photométrique calculée dans les images de synthèse en optimisant les degrés de liberté de la caméra permet d'atteindre la zone la plus riche, localement, à partir de n'importe quelle pose initiale de caméra (qui permette tout de même de percevoir une partie de la maquette virtuelle). Les résultats montrent le comportement de la méthode et des extraits de séquences d'images synthétisées au fil des itérations de la maximisation de l'entropie photométrique dans divers environnements. Ils laissent, de plus, entrevoir l'intérêt d'intégrer cette méthode dans un système interactif d'exploration de maquette virtuelle, voire comme élément d'un système d'exploration d'environnement par un robot mobile, nécessitant de prendre en compte la modélisation cinématique de ce dernier.

Chapitre 5

Suivi visuel direct pur basé maquette virtuelle 3D

Sommaire

5.1 Calcul de pose direct pur basé maquette virtuelle 3D 120
5.2 Recalage d'image perspective sur maquette 3D 121
5.2.1 Introduction 121
5.2.2 Etude de la fonction de coût 125
5.2.3 Résultats 126
5.2.4 Conclusion partielle 129
5.3 Suivi visuel panoramique basé maquette virtuelle 3D 129
5.3.1 Introduction 129
5.3.2 Suivi visuel direct pur basé maquette virtuelle 3D 131
5.3.3 Résultats 131
5.3.4 Conclusion partielle 133
5.4 Conclusion du chapitre 134

Ce chapitre propose un retour au coût direct pur pour le suivi visuel basé maquette virtuelle. En effet, si, pour ce dernier problème, le critère direct étendu de l'information mutuelle, prouvé très robuste, a montré des limites, c'est essentiellement dû au niveau de qualité des maquettes (Partie 4.1.5). Quand cette qualité est suffisamment bonne, le coût élémentaire de type SSD (Partie 3.2.1) retrouve de son intérêt car, dans les conditions idéales, il offre au minimum les mêmes qualités que l'information mutuelle.

C'est ce que l'extension du coût direct pur au calcul de pose et au suivi visuels basés maquette virtuelle (Partie 5.1) montre, d'une part dans un contexte de recalage de photographies sur nuage de points 3D dense (Partie 5.2) et, d'autre part dans le suivi visuel basé maquette virtuelle en vision panoramique centrale (Partie 5.3), avant de conclure le chapitre. Le premier se fait dans le but d'améliorer encore l'apparence de la maquette virtuelle *[Crombez et al., 2014]* et, le deuxième, pour la localisation de robot mobile en mouvement *[Crombez et al., 2015b]*.

Ces travaux ont été réalisés dans le cadre de la thèse de Nathan Crombez au laboratoire MIS de l'UPJV, que j'ai co-encadrée, sur allocation ministérielle, et qui s'inscrit dans le cadre du programme e-Cathédrale du laboratoire MIS.

5.1 Calcul de pose direct pur basé maquette virtuelle 3D

Le calcul de pose direct pur basé maquette virtuelle se formule de façon similaire au calcul de pose exploitant l'information mutuelle (Partie 4.1), en ce qui concerne le fait que les intensités $\mathcal{I}^*$ de l'image désirée sont issues d'une image acquise par une caméra réelle et sont comparées aux intensités $\mathcal{I}_v^{(k)}$ de chaque image de synthèse rendue à partir de la maquette virtuelle à chaque itération k de l'optimisation du coût de type SSD. La formulation et la résolution du problème sont aussi très proches de l'asservissement visuel photométrique (Partie 3.3.2.4), sauf que, dans le cas présent, les profondeurs des points 3D de la scène observée sont connus grâce à la maquette virtuelle, évitant ainsi une approximation. D'autre part, la pose de la caméra est aussi explicitement mise à jour.

Si la maquette virtuelle est un nuage de points 3D dense dont chaque point possède une couleur, ou une intensité de nature similaire à celles des pixels des images numériques (cf. Partie 2.3), optimiser la pose de la caméra virtuelle en minimisant l'écart entre les intensités de l'image réelle et celles des images de synthèse est exactement le problème exprimé dans les fondements des approches directes pures (Partie 3.7), à la nature des intensités courantes près ($\mathcal{I}_v$ à la place de $\mathcal{I}$). Il se résout itérativement à l'aide de l'algorithme de Levenberg-Marquardt pour lequel l'incrément de pose à chaque itération est calculé comme en équation (3.23).

Une variante mineure, mais qui a un impact non négligeable, comme le montre la Partie 5.2 est cependant introduite dans cette partie. Elle consiste à préparer les intensités des images entre lesquelles évaluer l'écart afin de rendre le critère de comparaison SSD robuste aux changements d'illumination affines entre l'apparence de la maquette virtuelle et les images acquises. Pour ce faire, avant de les utiliser en entrée de l'algorithme d'optimisation, les intensités $\mathcal{I}_v^{(k)}$ et $\mathcal{I}^*$ sont centrées en zéro et mises à l'échelle pour que la moyenne de leurs valeurs absolues soit égale à 1 :

$$\forall \mathbf{u} \in \mathcal{U}, \widetilde{I}_v(\mathbf{u}) = \frac{I_v(\mathbf{u}) - m_{\mathcal{I}_v^{(k)}}}{\frac{1}{|\mathcal{U}|}\sum_{i=1}^{|\mathcal{U}|} |I_v(\mathbf{u}_i) - m_{\mathcal{I}_v^{(k)}}|}, \tag{5.1}$$

avec

$$m_{\mathcal{I}_v^{(k)}} = \frac{1}{|\mathcal{U}|}\sum_{\mathbf{u}\in\mathcal{U}} I_v(\mathbf{u}), \tag{5.2}$$

et $\forall \mathbf{u} \in \mathcal{U}, \widetilde{I}_v(\mathbf{u}) \in \widetilde{\mathcal{I}}_v^{(k)}$ (même raisonnement pour obtenir $m_{\mathcal{I}^*}$ et $\widetilde{\mathcal{I}}^*$ à partir de $\mathcal{I}^*$), $\mathcal{U}$ défini en équation (3.47). Les intensités $\widetilde{\mathcal{I}}_v^{(k)}$ et $\widetilde{\mathcal{I}}^*$ sont donc des nombres réels, gardés comme tel dans le calcul du coût SSD qui devient ainsi ZNSSD pour "Zero-mean Normalized SSD", reprenant la terminologie associée à la ZNCC (Partie 3.4.1) :

$$\begin{aligned}
\mathscr{C}_{ZNSSD}(\boldsymbol{\delta}\mathbf{p}, \widetilde{\mathcal{I}}_v, \{\mathbf{p}, \mathcal{X}\}, \widetilde{\mathcal{I}}^*) &= \tfrac{1}{2}\sum_{\mathbf{X}\in\mathcal{X}_v} \left(\widetilde{I}_v(\mathbf{u}(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p}, \mathbf{X}), t+\delta t) - \widetilde{I}(\mathbf{u}(\mathbf{p}, \mathbf{X}), t)\right)^2 \\
&= \tfrac{1}{2}||\widetilde{\mathbf{I}}_v(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p}, \mathcal{X})) - \widetilde{\mathbf{I}}^*(\mathbf{p}, \mathcal{X}))||^2 \\
&= \tfrac{1}{2}||\mathbf{C}_{\mathbf{ZNSSD}}(\boldsymbol{\delta}\mathbf{p}, \widetilde{\mathcal{I}}_v, \{\mathbf{p}, \mathcal{X}\}, \widetilde{\mathcal{I}}^*)||^2,
\end{aligned} \tag{5.3}$$

où $\mathcal{X}_v \subset \mathcal{X}$, tel que $\forall\ \mathbf{X}_v \in \mathcal{X}_v, \mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X}_v) \in [0, N_c - 1] \times [0, N_l - 1]$, c'est-à-dire que seuls les points visibles sont considérés dans les transformations affines des intensités. Si d'autres variantes du coût SSD nécessitent de prendre en compte la transformation des intensités dans l'expression des matrices jacobiennes calculées pour résoudre itérativement la minimisation du coût, ce n'est pas le cas pour l'équation 5.3. En effet, les valeurs d'intensité des pixels de l'image numérique sont généralement comprises entre 0 et 255 car codées sur 8 bits. En faisant abstraction de ce codage, les intensités pourraient être dans n'importe quel autre intervalle. Considérer les intensités $\widetilde{\mathcal{I}}_v$ et $\widetilde{\mathcal{I}}^*$ directement dans le coût (Eq. 5.3) est équivalent à considérer ces intensités comme directement mesurées par la caméra.

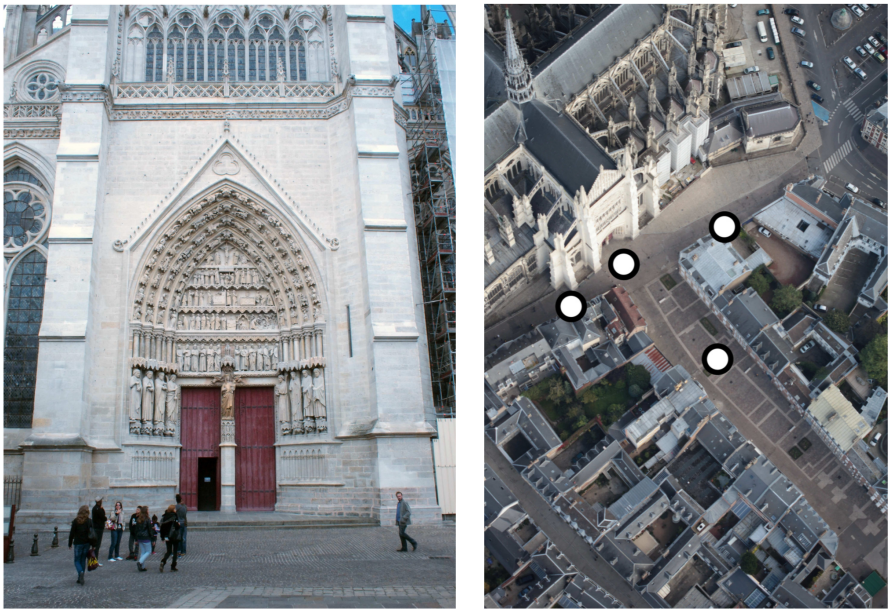
Enfin, les mêmes observations qu'en Remarque 6, sur la densité du nuage projeté dans l'image (Partie 4.2.3) sont faites pour permettre au coût d'être calculé et minimisé puisque les gradients des images ont aussi besoin d'être calculés (Eq. 3.9), pour déterminer chaque incrément de pose.

5.2 Recalage d'image perspective sur maquette 3D

5.2.1 Introduction

La numérisation 3D exhaustive d'édifices du patrimoine, est un procédé qui prend du temps, aussi bien par le volume que par la quantité de détails et la difficulté d'accès à certaines parties de ces édifices. En effet, la lasergrammétrie [Héno and Chandelier, 2014], technique reine pour ce faire, mise en oeuvre par des stations de scan laser de type Lidar (Light detection and Ranging), produit rapidement des nuages de points 3D ${}^{s_i}\mathcal{X}$ denses et précis de surfaces de l'édifice, mais locaux au repère $\mathcal{F}_{s_i}$ de la station s_i. Tout un édifice ne peut pas être numérisé en 3D instantanément, a fortiori quand une ou quelques stations Lidar sont mises en oeuvre dans ce but. Le nombre de stations Lidar est généralement faible, principalement à cause de leur coût, et pour obtenir une maquette virtuelle complète d'un édifice historique, des centaines d'acquisitions sont réalisées à des poses différentes (un extrait de quelques poses d'acquisitions du portail sud de la cathédrale d'Amiens est indiqué en figure 5.1), forcément à des moments (jusqu'à des années, dans certains cas) différents, posant des problèmes d'apparence visuelle, notamment en terme de couleur (ou d'intensité visible du spectre électromagnétique), de la maquette virtuelle $\mathcal{M}$, assemblant les nuages de points 3D ${}^{s_i}\mathcal{X}$ des $N_s \in \mathbb{N}^*$ acquisitions.

L'intensité du Lidar associée à chaque point 3D mesuré ${}^{s_i}\mathbf{X} \in {}^{s_i}\mathcal{X}$ étant liée à une longueur d'onde précise, parfois hors du spectre visible, elle ne peut, seule, apporter une couleur à associer à ${}^{s_i}\mathbf{X}$ pour donner au nuage de points 3D une apparence visuelle proche du réel. Les stations Lidar sont donc pourvues d'une ou plusieurs caméras, alignées mécaniquement ou dont les poses relatives au centre du Lidar sont pré-étalonnées. Ces caméras balayent aussi l'environnement, mais pour obtenir un ensemble haute définition d'images utilisées pour associer un triplet RVB ${}^{s_i}\mathbf{I_3} = \begin{bmatrix} I_R & I_V & I_B \end{bmatrix}^\top \in [\![0, N_I]\!]^3$ (intervalles entiers avec $N_I = 255$, généralement),



(a) (b)

FIGURE 5.1 – (a) Portail sud de la cathédrale d'Amiens pour la numérisation duquel (Indication d'échelle : la galerie quasi-horizontale est à 20 m du sol), (b) quatre poses (points blancs cerclés de noir) de la même station Lidar ont été considérées.

d'intensités de Rouge I_R, de Vert I_V et de Bleu I_B, à chaque ${}^{s_i}\mathbf{X}$, donnant ainsi une apparence visuelle de qualité photographique au nuage de points 3D. Les ${}^{s_i}\mathbf{I_3}$ engendrent une *qualité photographique* et pas une qualité *photo-réaliste* car les couleurs $\mathbf{I_3}$ dépendent de la pose de vue $\mathbf{p}_{s_i}$, des conditions lumineuses de l'environnement et des propriétés de matière de la scène au point $\mathbf{X}$, non mesurées, en plus des caractéristiques de la formation photométrique de l'image (Partie 2.3). Par conséquent, même en appliquant un changement de repère quelconque au nuage de points 3D, c'est-à-dire à toutes les paires $({}^{s_i}\mathbf{X}, {}^{s_i}\mathbf{I_3})$, par exemple ${}^{c}\mathbf{M}_{s_i} \in SE(3)$ vers un repère de caméra virtuelle pour synthétiser une image, aucune connaissance ne permet de transformer ${}^{s_i}\mathbf{I_3}$ pour donner un aspect visuel proche de la réalité, pour le point de vue de la caméra c, au pixel $\mathbf{u}$ où se projette $\mathbf{X}$. La conséquence majeure, c'est que, même si un unique point 3D $\mathbf{X}$ pouvait être exactement mesuré à partir de deux poses différentes $\mathbf{p}_{s_1}$ et $\mathbf{p}_{s_2}$ de station Lidar $({}^{c}\mathbf{M}_{s_1}{}^{s_1}\mathbf{X} = {}^{c}\mathbf{M}_{s_2}{}^{s_2}\mathbf{X})$, les couleurs ${}^{s_1}\mathbf{I_3}$ et ${}^{s_2}\mathbf{I_3}$, différentes dans le cas général, resteraient différentes après le changement de repère (Fig. 5.2(a)).

En pratique, de deux poses différentes $\mathbf{p}_{s_1}$ et $\mathbf{p}_{s_2}$, le même point 3D $\mathbf{X}$ n'est jamais mesuré deux fois, on mesure plutôt deux points $\mathbf{X}_1$ et $\mathbf{X}_2$, très proches. C'est un avantage pour densifier encore la géométrie de la maquette virtuelle, mais leurs couleurs mesurées ${}^{s_1}\mathbf{I_3}_1$ et ${}^{s_2}\mathbf{I_3}_2$ peuvent être plus différentes que dans la réalité, et c'est généralement le cas. L'avantage géométrique est donc en défaveur de l'aspect visuel de la maquette virtuelle. Des écarts faibles entre les ${}^{s_1}\mathbf{I_3}_1$ et ${}^{s_2}\mathbf{I_3}_2$

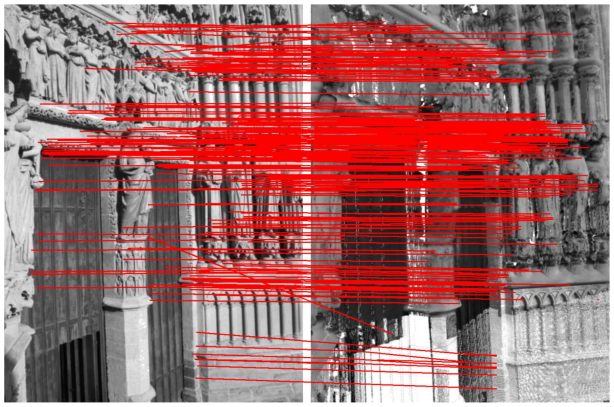
(a) (b)

FIGURE 5.2 – (a) Extraits de quatre nuages de points 3D colorés par les images de la station Lidar Leica Geosystems TPS400 acquises à chacune des quatre poses illustrées en figure 5.1(b). (b) Assemblage des quatre nuages de points en une seul maquette virtuelle.

peuvent être compensés en post-traitement par une transformation des couleurs calculée à partir des écarts de couleurs des points voisins issus des deux nuages de points 3D, dans leurs zones de recouvrement [Gui Yun Tian et al., 2002]. A titre d'indication, cette transformation permet d'améliorer suffisamment la qualité de l'apparence visuelle pour permettre une mise en correspondance d'une centaine de descripteurs ASIFT [Morel and Yu, 2009] (Fig. 5.3(b)), là où aucune correspondance correcte n'était possible avant la correction (Fig. 5.3(a)), entre une image réelle et une image synthétisée à partir de la maquette virtuelle. Cependant, et même si le contexte est différent, tout comme les variantes des approches directes pures en vision directe rapportées en partie 3.3.4.2, la transformation complète des couleurs ne peut être obtenue à partir des couleurs elles-mêmes seules, a fortiori quand elles est ont été acquises saturées (Fig. 5.2(a)), sans avoir mesuré de nombreux autres paramètres, de la position des sources lumineuses à la fonction de réponse de la caméra de la station Lidar.

Par conséquent, quand les nuages de points 3D ont déjà été acquis, remplacer l'ensemble des couleurs $\mathbf{I}_3$ de la maquette virtuelle $\mathcal{M} = \mathcal{X} \times \mathbf{I}_3$ (Fig. 5.2(b)) apparait une solution intéressante pour donner une apparence de qualité photographique (Fig. 5.1(a)) à $\mathcal{M}$, à pleine définition, au moins satisfaisante pour la visualisation par un usager (Partie 4.2). L'idée est donc de faire des acquisitions complémentaires d'images, à partir de poses de caméras judicieusement choisies pour qu'un pixel ne colore qu'un point 3D. En plaçant judicieusement la caméra, ses acquisitions de nouvelles images permettent, de plus, de s'affranchir de problèmes de résolution et

(a)



(b)

FIGURE 5.3 – (a) Image acquise (gauche) et image de synthèse (droite) de la maquette 3D aux couleurs brutes du portail sud de la cathédrale d'Amiens, obtenues à des poses similaires. (b) Illustration de la mise en correspondance de descripteurs ASIFT, après uniformisation des couleurs entre les quatre nuages de points 3D formant la maquette virtuelle.

de visibilité de scène qui interviendraient en considérant des images acquises par la station Lidar à l'une de ses poses d'acquisition.

Une fois les nouvelles images acquises, elles doivent être recalées précisément sur le nuage de points 3D pour une correspondance géométrique parfaite entre les pixels de l'image acquise et les points 3D de la maquette virtuelle, afin d'associer la bonne couleur au bon point 3D. Pour atteindre une telle précision, l'algorithme de calcul de pose direct pur (Partie 5.1) est mis en oeuvre pour recaler les images perspectives, d'une caméra dont les paramètres intrinsèques sont connus, sur la maquette virtuelle pourvue des couleurs acquises par la station Lidar.

Dans ce qui suit, malgré la qualité limitée des couleurs d'origine par rapport à l'image qui doit servir à les remplacer, la partie 5.2.2 montre que les considérer est une bonne solution, même avec le coût $\mathscr{C}_{ZNSSD}()$ (Eq. (5.3)), par rapport aux

autres critères existants pour traiter cette problématique. Ensuite, l'évaluation de la précision de recalage dans la partie 5.2.3 confirme la pertinence du coût, avant de conclure (Partie 5.2.4).

5.2.2 Etude de la fonction de coût

Cette étude expérimentale de la fonction de coût $\mathscr{C}_{ZNSSD}()$ (Eq. (5.3)), dans le cas d'une image désirée I^* acquise par une caméra réelle et d'images de synthèse $I_v(\boldsymbol{\delta p})$ obtenues par rendu de la maquette virtuelle, faite des nuages de points 3D assemblés aux couleurs brutes, se fait relativement aux critères de l'état de l'art dédiés au recalage d'image sur nuage de points 3D.

$\mathscr{C}_{ZNSSD}()$ (Eq. (5.3)) est alors confrontée à $\mathscr{C}_{ZNCC}()$ (Eq. (3.129)) et à l'information mutuelle calculée entre l'image acquise et les cartes de normales $MI_N()$ et de réflection $MI_R()$ du rendu de la maquette virtuelle [Corsini et al., 2009], méthodes de référence dans ce contexte au moment de l'étude en 2013.

La pose $\mathbf{p}$ menant à une image $I_v(\boldsymbol{\delta p})$ proche de l'image de référence $I^*()$ est déterminée manuellement. Les deux premiers degrés de liberté de $\boldsymbol{\delta p}$ (t_X et t_Y) sont utilisés pour déplacer la caméra virtuelle sur des intervalles de -1 m à 1 m. Les fonctions $\mathscr{C}_{ZNSSD}(\boldsymbol{\delta p})$, $-\mathscr{C}_{ZNCC}(\boldsymbol{\delta p})$, $-MI_N(\boldsymbol{\delta p})$ et $-MI_R(\boldsymbol{\delta p})$ sont tracées en figure 5.4 pour faciliter la comparaison et montrer l'intérêt du coût ZNSSD sur les autres. On constate les mêmes caractéristiques des fonctions $\mathscr{C}_{ZNSSD}()$ et $\mathscr{C}_{ZNCC}()$, au prix d'une complexité calculatoire supérieure pour le critère ZNCC. Ces deux dernières ont un minimum plus prononcé et sont moins bruitées que les fonctions $MI_N(\boldsymbol{\delta p})$ et $MI_R(\boldsymbol{\delta p})$, exploitant uniquement la géométrie de la maquette virtuelle.

Une étude plus vaste, sur plus de critères, incluant SSD et $MI()$ (Eq. (3.145)), et plus de degrés de liberté, faite dans la thèse de Nathan Crombez [Crombez, 2015]

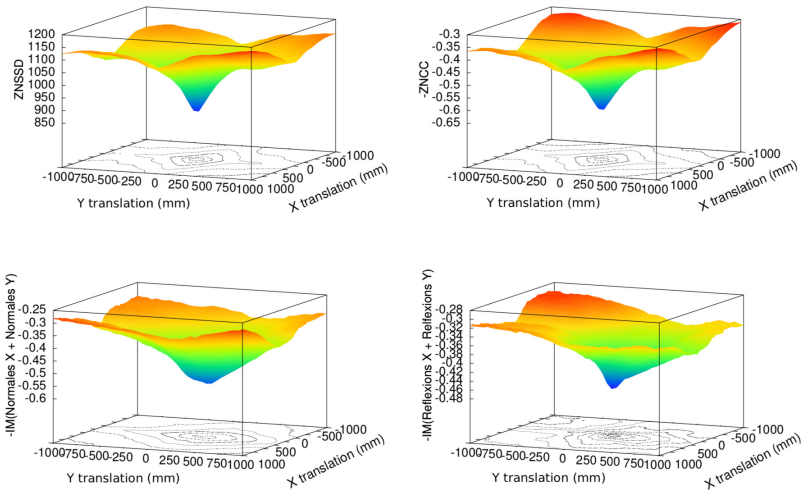


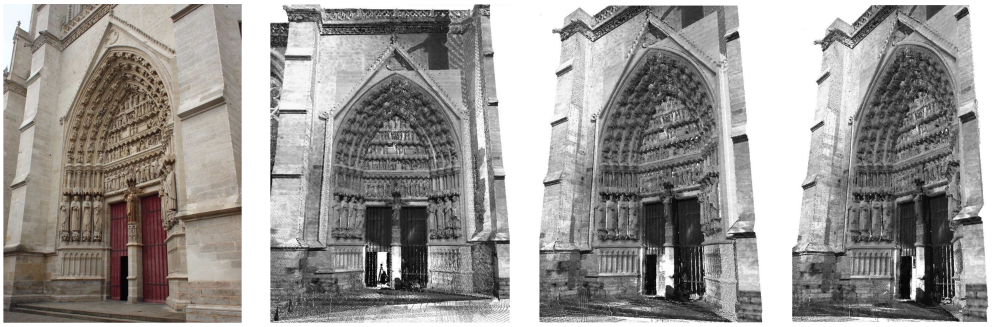
FIGURE 5.4 – Etude comparative des fonctions de coûts $\mathscr{C}_{ZNSSD}(\boldsymbol{\delta p})$, $-\mathscr{C}_{ZNCC}(\boldsymbol{\delta p})$, $-MI_N(\boldsymbol{\delta p})$ et $-MI_R(\boldsymbol{\delta p})$, dans le sens de lecture.

confirme ces quelques observations en faveur de la fonction de coût $\mathscr{C}_{ZNSSD}()$, vis-à-vis de son allure et de la complexité calculatoire, faible par rapport aux autres critères. Cela valide son choix pour le recalage d'image acquise sur la maquette virtuelle. D'autre part, le domaine de convergence apparent est d'environ 1 m de rayon, pour les deux degrés de liberté considérés. Les mêmes observations ont été faites sur quelques échantillons, ce qui ne prouve rien, mais donne une tendance favorable au coût $\mathscr{C}_{ZNSSD}()$ en permettant d'autoriser une initialisation approximative de la pose initiale de la caméra virtuelle à optimiser.

5.2.3 Résultats

Grâce au domaine de convergence du coût $\mathscr{C}_{ZNSSD}()$ considéré pour le calcul de pose, une initialisation grossière $\boldsymbol{\delta}\mathbf{p}^{(0)}$ de la pose est suffisante pour permettre de converger vers la pose précise. $\boldsymbol{\delta}\mathbf{p}^{(0)}$ est obtenue à partir de deux passes d'un calcul de pose basé points [DeMenthon and Davis, 1995] grâce aux correspondances ASIFT [Morel and Yu, 2009] entre l'image I^* (Fig. 5.5(a)) et une image de synthèse de la maquette virtuelle aux couleurs homogénéisées, rendue à partir d'une pose $\boldsymbol{\delta}\mathbf{p}_{manual}$ grossièrement définie, manuellement (Fig. 5.5(b)). La première passe produit une première pose optimale $\widehat{\boldsymbol{\delta}\mathbf{p}}_{asift1}$ basée points, qui permet de rendre une image de synthèse $I_v(\widehat{\boldsymbol{\delta}\mathbf{p}}_{asift1})$ bien plus proche de l'image réelle (Fig. 5.5(c)) qu'avec la pose $\boldsymbol{\delta}\mathbf{p}_{manual}$ (Fig. 5.5(b)). La deuxième passe calcule de nouveaux descripteurs ASIFT entre l'image I^* et $I_v(\widehat{\boldsymbol{\delta}\mathbf{p}}_{asift1})$, menant à une deuxième pose optimale $\widehat{\boldsymbol{\delta}\mathbf{p}}_{asift2}$ (Fig. 5.5(c)) basée points, mise en entrée de l'algorithme de calcul de pose direct exploitant le coût $\mathscr{C}_{ZNSSD}()$, c'est-à-dire $\boldsymbol{\delta}\mathbf{p}^{(0)} = \widehat{\boldsymbol{\delta}\mathbf{p}}_{asift2}$.

Même en deux passes, le calcul de pose basé points à partir de descripteurs



(a) (b) (c) (d)

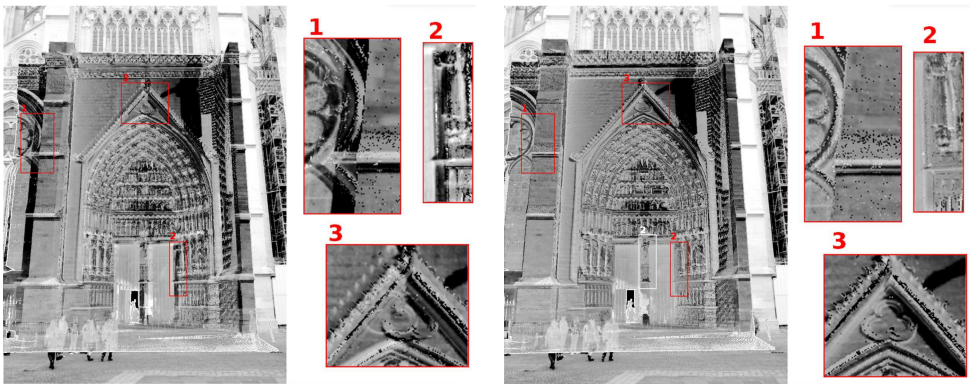
FIGURE 5.5 – (a) Image du portail sud de la cathédrale d'Amiens, acquise de trois-quart. (b) Image de synthèse de la maquette virtuelle aux couleurs originales homogénéisées du même portail, rendu de face. (c) Image de synthèse rendue à la pose optimale minimisant l'écart entre les points des descripteurs ASIFT de (a) et (b). (d) Image de synthèse rendue à la pose optimale minimisant l'écart entre les points des descripteurs ASIFT de (a) et (c).

ASIFT, quoique robustes aux transformations affines, n'est pas assez précis pour l'application de remplacement des couleurs des points 3D de la maquette virtuelle $\mathcal{M}$. En effet, des décalages persistent entre l'image I^* et l'image $I_v(\widehat{\boldsymbol{\delta \mathbf{p}}}_{asift2})$ (Fig. 5.6(a)). Néanmoins, l'ampleur de ces décalages traduit que la pose $\widehat{\boldsymbol{\delta \mathbf{p}}}_{asift2}$ a de fortes chances d'être dans le domaine de convergence du coût $\mathscr{C}_{ZNSSD}()$, ce qui se vérifie à convergence de sa minimisation en $\widehat{\boldsymbol{\delta \mathbf{p}}}$ (Fig. 5.6(b)), validant ainsi le calcul de pose direct basé ZNSSD.

La précision de la pose obtenue $\widehat{\boldsymbol{\delta \mathbf{p}}}$ permet de plaquer les couleurs des pixels d'une image acquise par une caméra sur le nuage de points 3D de la maquette $\mathcal{M}$ et d'en améliorer grandement l'apparence visuelle (Fig. 5.7(a)) par rapport aux couleurs brutes (Fig. 5.2(b)).

Pour une couverture maximale de la maquette virtuelle, comme c'était déjà le cas pour la numérisation 3D avec la station Lidar, plusieurs images doivent être acquises mais la rapidité avec laquelle trois images sont acquises par rapport à trois scans permet d'avoir une colorimétrie des images suffisamment proches pour que les algorithmes d'homogénéisation des couleurs [Gui Yun Tian et al., 2002] corrigent parfaitement les légères différences. En appliquant le calcul de pose pour les trois images séparément, leur placage sur la maquette virtuelle permet de vérifier la précision du calcul de pose car toute erreur résiduelle serait visible sur les zones de jonction entre les trois images plaquées. Ce n'est pas le cas dans les évaluations faites (Fig. 5.7(c)), ce qui indique que l'erreur résiduelle est inférieure à la résolution du nuage de point 3D.

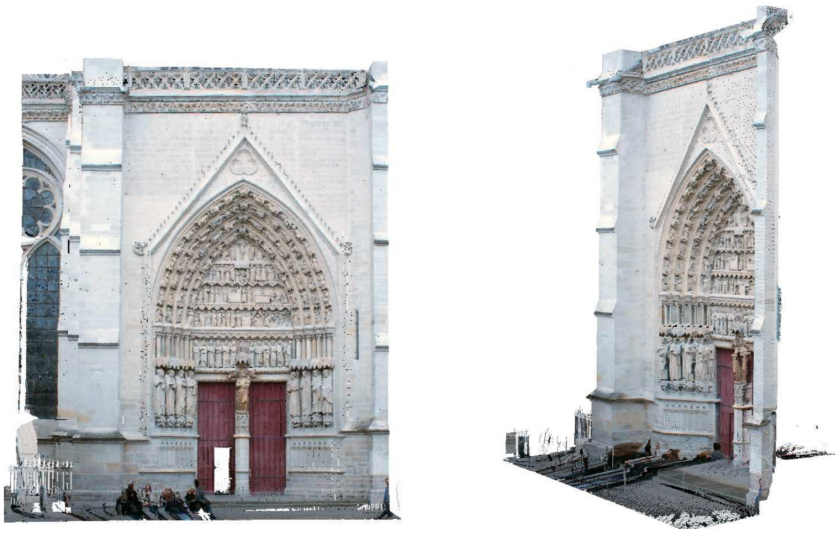
Enfin, pour quantifier l'amélioration de l'apparence visuelle, on utilise la corrélation photométrique (ZNCC) entre une nouvelle image acquise (Fig. 5.7(d)) et l'image de synthèse obtenue à la pose minimisant $\mathscr{C}_{ZNSSD}()$, avant (Fig. 5.7(b)) et après



(a) (b)

FIGURE 5.6 – (a) Erreur résiduelle qualitative du calcul de pose basé points (ASIFT). (b) Erreur résiduelle qualitative du calcul de pose direct basé ZNSSD. Pour chaque cas, les trois détails permettent de mieux comparer l'augmentation de la précision de recalage apportée par le calcul de pose basé ZNSSD.

(Fig. 5.7(c)) remplacement des couleurs. Dans cette dernière évaluation, la nouvelle image acquise n'a pas été utilisée pour remplacer les couleurs de la maquette, et, pour évaluation plus exigeante, son point de vue, en contre-plongée, est très différent de ceux, horizontaux, des images ayant servi à remplacer les couleurs. Sur l'exemple du visage de la statue de la Vierge Dorée du portail de la cathédrale d'Amiens, le remplacement des couleurs engendre un accroissement de la corrélation photométrique de 30%, validant ainsi l'amélioration nette de la qualité photographique de la maquette virtuelle, déjà clairement visible qualitativement en comparant les figures 5.7(b) et 5.7(c).



(a)

(b) (c) (d)

FIGURE 5.7 – (a) Images de synthèse à deux poses différentes de la maquette virtuelle aux couleurs remplacées par celles issues d'une photo globale du portail sud de la cathédrale d'Amiens. (b) Image de synthèse de la maquette virtuelle aux couleurs brutes. (c) Image de synthèse à la même pose que (b) de la maquette virtuelle dont les couleurs ont été remplacées par celles issues de trois images acquises à des points de vue différents pour une couverture maximale des courbures. (d) Image acquise à la même pose que (b) et (c) pour l'évaluation qualitative et quantitative de l'amélioration de la qualité de l'apparence visuelle de la maquette virtuelle.

Résultats en vidéo :
http://mis.u-picardie.fr/~g-caron/videos/2014_Crombez_recalagePersp3D.mp4.

5.2.4 Conclusion partielle

La déclinaison du coût direct pur pour le calcul de pose 3D de caméra basé maquette virtuelle d'apparence visuelle photographique permet d'atteindre une précision aussi haute que ce qui avait déjà pu être montré en asservissement visuel photométrique (Partie 3.3.2.4) ou en suivi direct d'objet plan (Partie 3.3.2.3). La considération des intensités centrées et normalisées des images dans le calcul du coût et dans sa minimisation permet d'agrandir le domaine de convergence grâce à l'invariance des intensités aux changements d'illumination affines globaux. Les résultats, obtenus sur des jeux de données difficiles où l'apparence visuelle d'origine de la maquette virtuelle est très perturbée, montrent que même en présence de variations d'illumination non-linéaires, la méthode conserve ses qualités, encore une fois grâce à la redondance d'information apportée par la considération de tous les pixels.

Enfin, la précision du recalage des images acquises sur la maquette virtuelle est telle que le remplacement des couleurs de la maquette par plusieurs images n'engendre pas d'artefact visible et accroît considérablement la qualité photographique de l'apparence visuelle de la maquette virtuelle. Pour atteindre ces précisions, il faut, bien entendu, que la caméra soit étalonnée précisément, mais la généralisation de la méthode de calcul de pose à l'étalonnage de caméra, en ajoutant les degrés de liberté du modèle de projection, est aussi envisageable et représente une perspective de ce travail.

5.3 Suivi visuel panoramique basé maquette virtuelle 3D

5.3.1 Introduction

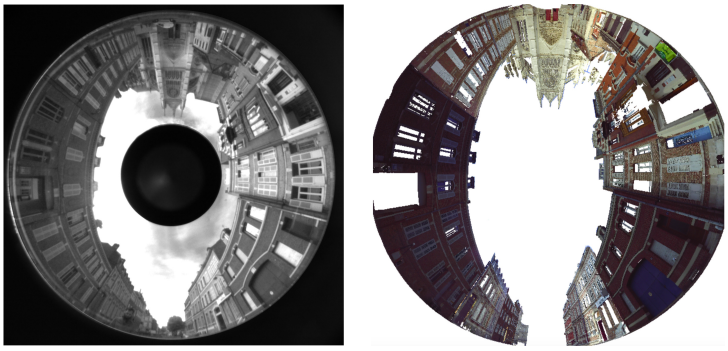
Le même coût $\mathscr{C}_{ZNSSD}()$, introduit dans ce chapitre (Partie 5.1), s'adapte aux caméras panoramiques, de la même façon que l'asservissement visuel photométrique ou le suivi de plan s'étend à ces mêmes caméras (Partie 3.3). Dans l'équation (5.3) du coût $\mathscr{C}_{ZNSSD}()$, on considère le modèle de projection central unifié (Partie 2.1.2), adapté à ces caméras pour exprimer $\mathbf{u}(\mathbf{p} \oplus \boldsymbol{\delta}\mathbf{p}, \mathbf{X})$, et une adaptation du calcul des gradients de l'image panoramique (similaire à la fin de la partie 3.3.2.3).

Une modification de la caméra virtuelle des mêmes outils logiciels utilisés dès la partie 4.1.4 permet de rendre une image de synthèse, dont la géométrie est très similaire à celle acquise par une caméra panoramique centrale (Fig. 5.8), en utilisant le modèle de projection centrale unifié et ses paramètres, identiques à ceux de la caméra réelle obtenus par étalonnage.

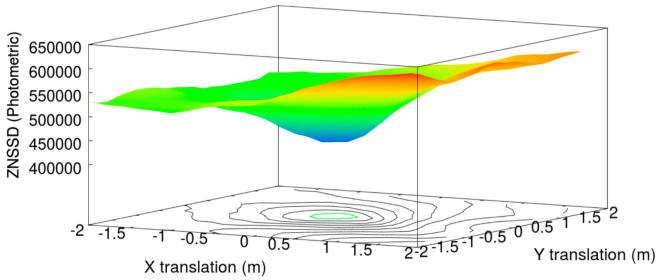
Les mêmes conclusions en faveur du coût de type ZNSSD se font en vision panoramique qu'en vision perspective (Partie 5.2.2) : domaine de convergence le plus large et le plus lisse parmi les critères considérés et complexité calculatoire la plus faible. Sans pouvoir généraliser à partir de quelques exemples, l'ampleur minimum

du domaine de convergence apparaît similaire à celui de la vision perspective mais plus étendu dans certains cas. Par exemple, la partie droite de la figure 5.8(c) montre que le domaine de convergence en translation dépasse les 2 m, dans ce cas. Cette observation est similaire aux résultats qui avaient été obtenus par le passé en asservissement visuel photométrique panoramique central [Caron et al., 2013] où des écarts de pose de plus d'1,2 m avaient pu être corrigés alors qu'en vision perspective, ce n'était pas le cas.

La figure 5.8(c) illustre un autre détail intéressant apporté par la vision panoramique centrale sur la vision perspective en ce qui concerne la forme de la fonction de coût ZNSDD autour du minimum. En effet, cette dernière est beaucoup plus proche de la parabole qu'en vision perspective où la forme de la fonction de coût ZNSSD autour du minimum est beaucoup plus pointue (Fig. 5.4). Cette observation indique que les algorithmes d'optimisation découlant de la méthode de Newton devraient mieux se comporter en vision panoramique centrale qu'en vision perspective.



(a) (b)



(c)

FIGURE 5.8 – (a) Image panoramique centrale acquise à l'aide d'un objectif catadioptrique à double-miroir RemoteReality. (b) Image de synthèse rendue à une pose similaire. (c) Tracé de la fonction de coût $\mathscr{C}_{ZNSSD}(\boldsymbol{\delta p})$ en faisant varier les deux premiers degrés de liberté de -2 m à +2 m.

5.3.2 Suivi visuel direct pur basé maquette virtuelle 3D

Les quelques observations succinctes faites en partie 5.3.1 vont dans le sens de l'intérêt de la vision panoramique en robotique, récemment prouvée mieux adaptée pour l'odométrie visuelle [Zichao Zhang et al., 2016], par exemple. C'est pourquoi, dans cette partie, on s'intéresse à appliquer le calcul de pose direct en vision panoramique centrale dans le suivi visuel basé maquette virtuelle d'environnement urbain, permettant ainsi la localisation du robot qui embarque la caméra panoramique.

Sur l'exemple d'environnement urbain de quatre rues de la ville d'Amiens visible en figure 5.9, la longueur cumulée de ces rues représente 400 m. Un scanner Lidar Faro Focus 3D a été utilisé à 13 poses distinctes dans ces rues pour produire une maquette virtuelle simplifiée d'environ 10 millions de points 3D, aux couleurs homogénéisées. Cette simplification, qui permet d'avoir une maquette de taille mémoire raisonnable, tout en ayant une densité suffisante pour synthétiser des images panoramiques denses de 350×350 pixels au milieu des rues, représente néanmoins trop de données à traiter en direct à chaque itération du calcul de pose qui se fait pour chaque image acquise. Par conséquent, la maquette est subdivisée en sous-nuages de points 3D locaux, obtenus par rendus sphériques au centre des rues avec un pas régulier (3 m dans les expérimentations faites), expérimentalement choisi. Le rendu permet de ne considérer, localement, que les points 3D visibles, à la fois relativement aux auto-occultations potentielles et relativement à la résolution variable dans l'image panoramique. Le pas régulier doit permettre de conserver une densité suffisante dans les images, même si la caméra réelle est entre deux nuages de points 3D locaux.

En résumé, par rapport au suivi visuel basé maquette virtuelle 3D de la partie 4.1, seule une étape préliminaire de pré-sélection du sous-nuage de points 3D le plus proche de la précédente pose optimale est requise, avant de lancer le calcul de pose. Grâce à cette structuration de la maquette virtuelle et à l'utilisation du coût direct pur, 4 poses par seconde peuvent ainsi être calculées.

5.3.3 Résultats

Tout d'abord, le suivi visuel direct pur basé maquette virtuelle a été appliqué avec succès à l'intérieur de la cathédrale d'Amiens, sur une trajectoire en forme de "U" de 20 m de longueur pour environ 600 images acquises par la caméra embarquée sur un robot mobile Pioneer 3AT. La maquette virtuelle consiste en un nuage de points 3D simplifié d'environ 30 millions de points 3D obtenus par la station Lidar Leica Scanstation C10 à partir de 50 poses.

Ensuite, comme évoqué dans la partie 5.3.2, une maquette virtuelle de quatre rues est considérée pour la localisation du robot mobile par suivi visuel direct pur. Sur les 350 m de trajectoire du robot (Fig. 5.10(a), vérité terrain, obtenue par SLAM avec nappe Lidar SICK LMS-200, en rouge), 21000 images sont acquises. Les poses successives estimées (Fig. 5.10(a), bleu), indiquent une grande proximité du suivi visuel par rapport à la vérité terrain et ce, malgré des conditions d'illu-

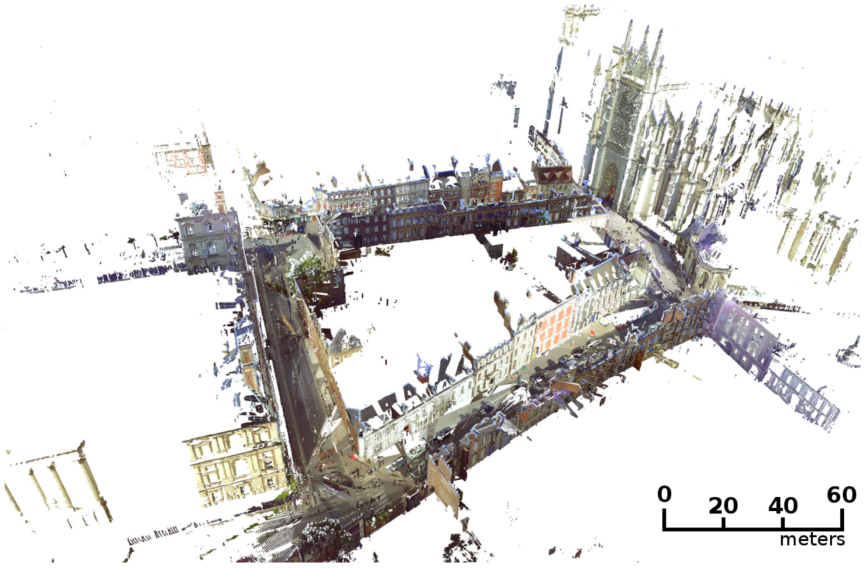


FIGURE 5.9 – Vue plongeante sur la maquette virtuelle obtenue par scan Lidar des rues Victor Hugo, Lesueur, Robert de Luzarches et Cormont, à proximité de la cathédrale d'Amiens.

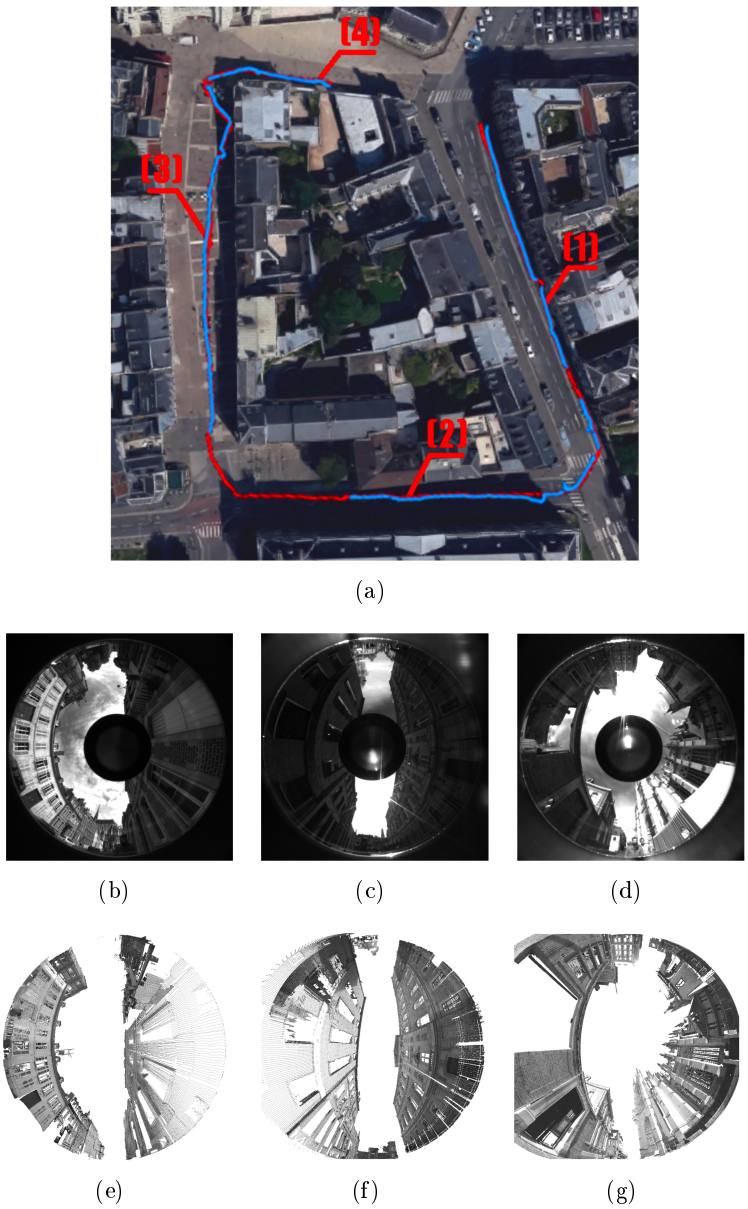
mination différentes et rendues difficiles à cause du soleil par rapport au jour d'acquisition des nuages de points 3D. En effet, les figures 5.10(b) à 5.10(d) montrent trois images acquises à trois endroits différents de la trajectoire du robot et les figures 5.10(e) à 5.10(g), les images de synthèse correspondantes aux poses optimales. Les figures 5.10(b) et 5.10(g) montrent au moins un des deux côtés de la rue trop ou insuffisamment exposé par rapport à la maquette virtuelle. Le cas de la figure 5.10(c), est plus facile à corriger par le critère ZNSSD car l'ensemble de l'image est plus sombre que dans la maquette, et pas uniquement une partie. Le suivi échoue quand les rotations du robot sont trop rapides, quand il descend ou monte un trottoir, en présence d'occultation majeure, par exemple quand un camion occulte totalement un côté de la rue en passant et que l'autre côté n'est pas exploitable car trop sombre par rapport à la maquette. En résumé, le suivi visuel direct basé maquette virtuelle est un succès sur environ 85 %, représentant 297 m cumulés, de la trajectoire dans cette expérimentation, malgré le fait que la maquette virtuelle de l'environnement ait été acquise à partir d'un capteur différent de celui utilisé pour la localisation.

Résultats en vidéo :

http://mis.u-picardie.fr/~g-caron/videos/2016_Crombez_suiviPano3D.mp4.

5.3.4 Conclusion partielle

La généralisation du calcul de pose direct pur basé maquette virtuelle 3D au modèle de projection centrale unifié a permis de l'appliquer aux caméras panoramiques centrales. La précision à convergence apportée par le coût direct pur engendre une



(a)

(b) (c) (d)

(e) (f) (g)

FIGURE 5.10 – (a) Image satellite des quatre mêmes rues qu'en figure 5.9 avec les trajectoires du robot mobile estimées par SLAM basé nappe Lidar (rouge) et par la méthode de suivi visuel direct de cette partie (bleu). (b - d) Trois images acquises aux points 1, 2 et 4 indiqués en (a) et (e - g) les images de synthèse rendues à leurs poses optimales respectives (images du point 3 : 5.8(a) et 5.8(b)).

localisation précise du robot mobile qui embarque une telle caméra, malgré le fait que la maquette virtuelle n'ait pas été obtenue à partir de la même caméra, contrairement aux approches équivalentes de type SLAM. Néanmoins, la structuration de la maquette sous forme de nuages de points 3D locaux par le biais de rendus d'images de synthèses sphériques s'apparente aux images sphériques clés augmentées issues de ces derniers travaux [Meilland et al., 2015]. Elle permet un suivi visuel basé maquette virtuelle proche du temps-réel.

Cependant, le suivi échoue dans les cas d'occultation majeure, ou de mouvement trop rapide engendrant un écart trop important entre images successives pour que le minimum global du coût $\mathscr{C}_{ZNSSD}$ soit déterminé par la méthode d'optimisation itérative de Levenberg-Marquardt. Ce dernier point fait apparaître le besoin d'un domaine de convergence plus important.

5.4 Conclusion du chapitre

Décliner le coût direct pur pour le calcul de pose basé maquette virtuelle en vision perspective et en vision panoramique a permis de le mettre à l'épreuve dans de nouveaux contextes où les intensités ou les couleurs de référence n'ont pas été acquises par la même caméra que les images courantes et confirme, une fois de plus, son intérêt. La qualité croissante de l'apparence visuelle des maquettes virtuelles a permis ce succès, là où, auparavant, des critères plus robustes aux imperfections de la maquette devaient être employés (Partie 4.1). D'autres travaux de recherche plus récents confirment aussi à nouveau son intérêt, tels que ceux menés en odométrie visuelle éparse directe [Engel et al., 2018]. Ses avantages majeurs sont la précision à convergence et la relative simplicité théorique et pratique, comparé à d'autres critères. Ces avantages, combinés à un domaine de convergence plus étendu que les autres critères, dans le cas idéal, sont cependant tempérés par la moins bonne robustesse aux variations d'illumination ou aux occultations du critère direct pur, par rapport à ceux issus de la théorie de l'information, par exemple. Mais elle offre un compromis intéressant.

Cependant, pour améliorer le taux de succès du suivi visuel basé maquette virtuelle, l'analyse des résultats en vision panoramique (Partie 5.3.3) fait apparaître, parmi les besoins, celui d'un domaine de convergence plus étendu pour pouvoir pallier les mouvements brusques, parfois imprévisibles, engendrant un mouvement trop important entre images successivement acquises.

Un domaine de convergence plus étendu représenterait aussi un intérêt pour le calcul de pose en vision perspective, même pour l'application spécifique abordée dans la partie 5.2, car cela permettrait d'éliminer l'initialisation basée points et descripteurs ASIFT, de même que l'uniformisation des couleurs des nuages de points de maquette virtuelle, afin que le calcul de pose soit réellement direct de bout en bout.

C'est de cette problématique d'élargissement du domaine de convergence que traite le chapitre 6 suivant.

Chapitre 6

Vision robotique directe basée mélange de potentiels

Sommaire

6.1 Le Mélange de Potentiels Photométriques 136
6.1.1 Problématique 136
6.1.2 Modélisation du mélange de potentiels photométriques 137
6.2 Suivi visuel de plan adaptatif dans l'espace d'échelle 138
6.2.1 Modélisation du problème 138
6.2.2 Evaluation 140
6.2.2.1 Détermination des valeurs des paramètres 140
6.2.2.2 Mouvement projectif 141
6.2.3 Conclusion partielle 143
6.3 Asservissement visuel direct basé mélange de potentiels . . . 144
6.3.1 Définition du coût et de la loi de commande 144
6.3.2 Matrice jacobienne relative à l'échelle 145
6.3.3 Matrice jacobienne relative à la pose de la caméra 145
6.3.4 Stratégie de mise à jour de l'étendue d'attraction 147
6.3.5 Résultats 148
6.3.6 Conclusion partielle 152
6.4 Gyroscope visuel sphérique direct 152
6.4.1 Vue d'ensemble 152
6.4.2 Représentation d'image et mélange de potentiels sphériques . . . 153
6.4.3 Estimation directe de rotations 155
6.4.4 Résultats 157
6.4.4.1 Quels paramètres pour un vaste domaine de convergence ? 157
6.4.4.2 Evaluation du gyroscope visuel 159
6.4.5 Conclusion partielle 161
6.5 Conclusion du chapitre 162

Ce chapitre synthétise les travaux menés pour accroître significativement le domaine de convergence des méthodes de vision robotique directe. L'idée fondamentale repose sur la transformée des images en mélanges de potentiels photométriques. Cette approche directe étendue s'apparente aux approches raisonnant dans un espace d'échelle (Partie 3.5.4) et peut être vue comme une généralisation des approches de suivi direct reposant sur une représentation pyramidale (gaussienne) des images, grâce à deux éléments clés :

— considérer une pyramide continue d'échelle, plutôt que des niveaux disjoints

— faire varier l'échelle automatiquement en fonction du coût à minimiser

Dans ce qui suit, la transformée de l'image en mélange de potentiels photométriques (MPP) est tout d'abord introduite (Partie 6.1). Dans un deuxième temps, son exploitation dans une formulation numérique pour le suivi visuel de région plane est proposée (Partie 6.2) avant d'en exprimer une formulation analytique pour l'asservissement visuel (Partie 6.3). Ensuite, la partie 6.4 présente une adaptation sphérique du MPP pour l'estimation directe d'orientation de caméra sphérique. Enfin, ce concept des mélanges de potentiels est appliqué à la saillance visuelle pour le cadrage automatique direct d'objet inconnu en réel, comme en virtuel (Partie **??**) avant de conclure le chapitre (Partie 6.5).

La transformée de l'image en mélange de potentiels photométriques et son utilisation en entrée de l'asservissement visuel *[Crombez et al., 2015a, Crombez et al., 2019]* sont des travaux initiés et réalisés dans la thèse de Nathan Crombez au laboratoire MIS de l'UPJV, que j'ai co-encadrée, sur allocation ministérielle. Ensuite, sa déclinaison au suivi de région plane *[Ahmine et al., 2019]* est l'un des travaux de thèse de Yassine Ahmine au laboratoire MIS de l'UPJV et au laboratoire LTSS de l'UATL, en Algérie, en co-tutelle et que je co-encadre. L'adaptation sphérique du MPP *[Caron and Morbidi, 2018]* est un travail que j'ai mené en parallèle de mes co-encadrements de thèse. Enfin, l'application de la transformée en mélange de potentiels à la saillance visuelle *[Habibi et al., 2015, Habibi et al., 2015]* a été réalisée dans la thèse de Zaynab Habibi au laboratoire MIS de l'UPJV, que j'ai co-encadrée, dans le cadre du projet CR Picardie "Assiduitas", qui s'inscrit dans le programme e-Cathédrale du laboratoire.

6.1 Le Mélange de Potentiels Photométriques

6.1.1 Problématique

Chaque intensité $I(\mathbf{u})$ au pixel de coordonnées $\mathbf{u} \in \mathcal{U}$ (Eq. 3.47) d'une image numérique $I()$ est le fruit de la convolution du signal image continu $\mathrm{I}(\mathbf{u}_\mathbb{R}) : \mathbb{R} \longmapsto \mathbb{R}_+$ avec l'impulsion de Dirac $d_\mathbb{R}()$ (similaire à l'équation (3.138) mais de $\mathbb{R}$ dans $\mathbb{N}$) tel que [Gonzalez and Woods, 2018] :

$$I(\mathbf{u}) = \int \mathrm{I}(\mathbf{u}_\mathbb{R}) d(||\mathbf{u} - \mathbf{u}_\mathbb{R}||)\, \mathrm{d}\mathbf{u}_\mathbb{R}. \tag{6.1}$$

Par conséquent, dans l'image $I()$, toutes les intensités des pixels sont indépendantes, ce qui conduit les coûts directs purs (Partie 3.3) à avoir un domaine de convergence étroit en dehors des cas particuliers de dégradés d'intensités. A l'autre extrême, une image très simple, où un seul pixel a une intensité différente de tous les autres, engendre le domaine de convergence le plus étroit d'un seul pixel de rayon, alors que l'on pourrait penser que ce cas est le plus simple à traiter et que l'on pourrait faire converger une estimation directe de mouvement sans difficulté.

6.1.2 Modélisation du mélange de potentiels photométriques

Pour résoudre ce problème, on peut lier les intensités des pixels de l'image numérique entre elles en combinant les intensités du signal image pour obtenir $I(\mathbf{u})$. Afin de garder des variations d'intensité dans l'image et une certaine discriminance de chaque pixel, on remplace l'impulsion de Dirac de l'équation (6.1) par une fonction $f_p : \mathbb{R}^2 \longmapsto \mathbb{R}_+$ continue, positive, décroissante et monotone :

$$I_p(\mathbf{u}) = \int \mathrm{I}(\mathbf{u}_\mathbb{R}) f_p(\mathbf{u} - \mathbf{u}_\mathbb{R}) \, \mathrm{d}\mathbf{u}_\mathbb{R}. \tag{6.2}$$

Concrètement, cela est équivalent à faire l'acquisition d'une image numérique floue, qui permet, certes, d'agrandir le domaine de convergence mais au détriment de la précision à convergence [Kallem et al., 2007]. C'est pourquoi les approches multi-échelles pyramidales (Partie 3.5.4) considèrent une image numérique acquise nette, pour assurer la précision à convergence, et la génération d'échelles plus grossières pour agrandir le domaine de convergence. Selon ce principe largement validé, on peut s'inspirer de l'équation (6.2) pour transformer l'image numérique acquise (nette), où les intensités $I(\mathbf{u})$ sont indépendantes, en une nouvelle image $I_{MPP}()$, dont l'intensité de chaque pixel dépend de toutes les intensités de $I(\mathbf{u})$:

$$I_{MPP}(\mathbf{u}_p) = \sum_{\mathbf{u} \in \mathcal{U}} I(\mathbf{u}) f_p(\mathbf{u}_p - \mathbf{u}). \tag{6.3}$$

L'expression est exactement la même que l'expression générale de la convolution discrète, que l'on retrouve, notamment, en vision directe basée noyaux photométriques [1] (Partie 3.5.3.2) et pyramide de niveaux d'échelle (Partie 3.5.4). Cette expression est simplement présentée d'une façon différente, non pas *uniquement* pour réduire le bruit et rendre dérivable l'image, mais pour qu'un pixel d'un coin du rectangle image ait un potentiel d'attraction jusqu'au coin opposé.

Ce potentiel d'attraction est modélisé par la fonction $f_p()$ (Eq. (6.3)). Un potentiel d'attraction est centré en chaque pixel $\mathbf{u}$ de l'image acquise et son étendue dépend de l'expression de $f_p()$ elle-même, au moins paramétrable par un scalaire $\lambda_p \in \mathbb{R}_+$. En considérant ce paramètre λ_p d'étendue de potentiel, on définit l'intensité $I_{PP}(\mathbf{u}_p, \mathbf{u}, \lambda_p)$, de potentiel photométrique de $I(\mathbf{u})$ en $\mathbf{u}_p$, par le contenu de la somme de l'équation (6.3) :

$$I_{PP}(\mathcal{I}, \mathbf{u}_p, \mathbf{u}, \lambda_p) = I(\mathbf{u}) f_p(\mathbf{u}_p - \mathbf{u}, \lambda_p), \tag{6.4}$$

1. Au détail près que f_p n'est pas forcément isotrope dans l'équation (6.3)

et donc :

$$I_{MPP}(\mathcal{I}, \mathbf{u}_p, \lambda_p) = \sum_{\mathbf{u} \in \mathcal{U}} I_{PP}(\mathcal{I}, \mathbf{u}_p, \mathbf{u}, \lambda_p), \tag{6.5}$$

n'est autre que l'intensité du mélange de potentiels photométriques, obtenue par la somme des intensités de potentiels photométriques, dépendant donc des intensités $\mathcal{I}$ de toute l'image (ou de toute la région considérée).

Selon la valeur de λ_p, c'est-à-dire l'étendue de $f_p()$, quand λ_p tend vers 0, l'attraction d'un pixel ne s'applique qu'en sa position donc on aura $I_{MPP}() = I()$ (similaire à la base des pyramides gaussiennes). A l'inverse, quand λ_p est grand, l'attraction d'un pixel s'appliquera aux autres pixels de l'image, mais si λ_p est trop grand, tous les pixels de l'image auront le même potentiel d'attraction sur tous les autres pixels de l'image, perdant ainsi toute capacité de discriminer un pixel d'un autre.

Toute fonction noyau (à support infini) issue des statistiques non-paramétriques [Titterington et al., 1985] pourrait théoriquement être employée mais les fonctions de potentiel $f_p()$ considérées dans la suite du chapitre sont gaussiennes car dérivables partout et communément considérées dans la représentation d'image en espace d'échelle [Lindeberg, 1994], facilitant ainsi le parallèle avec les approches qui y sont associées. Les parties 6.2 et 6.4 considèrent une gaussienne normale, pour un point de coordonnées $\mathbf{a} \in \mathbb{R}^n$:

$$f_p(\mathbf{a}, \lambda_p) = \frac{1}{\sqrt{(2\pi)^n \det(\lambda_p^2 \mathbf{I}_{n\times n})}} \exp\left(-\frac{1}{2}\left(\mathbf{a}^\top \left(\lambda_p^2 \mathbf{I}_{n\times n}\right)^{-1} \mathbf{a}\right)\right) \tag{6.6}$$

alors que la partie 6.3 considère une gaussienne sans le facteur de normalisation à gauche de l'exponentielle, essentiellement pour simplifier les expressions et réduire les temps de calcul au maximum pour permettre l'asservissement visuel.

6.2 Suivi visuel de plan adaptatif dans l'espace d'échelle

6.2.1 Modélisation du problème

Le problème de suivi visuel de région plane dans l'image adaptant automatiquement l'échelle de cette dernière se formalise en combinant la transformée de l'image en mélange de potentiels photométriques (Partie 6.1) au coût du suivi de région selon un mouvement projectif dans l'image $\mathscr{C}_{BM}()$ (Eq. (3.66)) :

$$\begin{aligned} \mathscr{C}_{A+}(\mathcal{D}_{A+}, \mathcal{I}, \mathcal{P}_{A+}, \mathcal{I}^*) &= \tfrac{1}{2} \sum_{\mathbf{u}_p^* \in \mathcal{R}^*} \left(I_{MPP}(\mathcal{I}, {}^c\mathbf{H}_{c^*}(\mathbf{h})\, \tilde{\mathbf{u}}_p^*, \lambda_p) - I_{MPP}^*(\mathcal{I}^*, \tilde{\mathbf{u}}_p^*, \lambda_p^*)\right)^2 \\ &= \tfrac{1}{2} ||\mathbf{I_{MPP}}(\mathbf{h}, \lambda_p) - \mathbf{I_{MPP}}^*||^2 \\ &= \tfrac{1}{2} ||\mathbf{C_{A+}}(\mathbf{h}, \lambda_p)||^2, \end{aligned} \tag{6.7}$$

avec $\mathcal{D}_{A+} = \{\mathbf{h}, \lambda_p\}$ et $\mathcal{P}_{A+} = \{\mathcal{R}^*, \lambda_p^*\}$.

La fonction de potentiel $f_p()$, utilisée pour calculer $I_{MPP}()$ et $I_{MPP}^*()$, s'exprime comme en équation (6.6) en dimension 2 (avec n $n = 2$). En dehors des spécificités

liées aux mélanges de potentiels photométriques, la paramétrisation de l'évolution du mouvement est légèrement différente par rapport à $\mathscr{C}_{BM}()$ car, à chaque itération de la minimisation de $\mathscr{C}_{A+}()$ (Eq. (6.7)), ses degrés de liberté sont mis à jour par approche additive.

Dans l'équation (6.7), on distingue le paramètre λ_p, d'étendue de la fonction de potentiel, du mélange de potentiels photométriques courant du désiré λ_p^*. Ce dernier est fixe et proche de zéro pour considérer un mélange de potentiels photométriques similaire aux intensités $\mathcal{I}^*$ comme référence. Mais le point le plus important est de considérer λ_p dans les degrés de liberté du problème de minimisation du coût $\mathscr{C}_{A+}()$:

$$\begin{bmatrix} \widehat{\mathbf{h}} \\ \widehat{\lambda}_p \end{bmatrix} = \underset{\mathbf{h},\lambda_p}{\arg\min}\, \mathscr{C}_{A+}\big(\{\mathbf{h},\lambda_p\},\mathcal{I},\{\mathcal{R}^*,\lambda_p^*\},\mathcal{I}^*\big)\,. \tag{6.8}$$

On résout itérativement l'équation (6.8), par la méthode de Gauss-Newton qui nous permet d'exprimer l'incrément des degrés de liberté à l'itération k par :

$$\begin{bmatrix} \dot{\mathbf{h}}^{(k)} \\ \dot{\lambda}_p^{(k)} \end{bmatrix} = -\lambda \mathbf{J}^{\dagger}_{\mathbf{h}\boldsymbol{\lambda}_\mathbf{p}{}^{(k)}} \mathbf{C}_{\mathbf{A}+}(\mathbf{h}^{(k)},\lambda_p^{(k)}), \tag{6.9}$$

avec :

$$\mathbf{J}_{\mathbf{h}\boldsymbol{\lambda}_\mathbf{p}{}^{(k)}} = \left[\left.\frac{\partial \mathbf{I_{MPP}}(\mathbf{h},\lambda_p)}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{h}^{(k)}} \;\middle|\; \left.\frac{\partial \mathbf{I_{MPP}}(\mathbf{h},\lambda_p)}{\partial \lambda_p}\right|_{\lambda_p=\lambda_p^{(k)}} \right], \tag{6.10}$$

où :

$$\left.\frac{\partial \mathbf{I_{MPP}}(\mathbf{h},\lambda_p)}{\partial \mathbf{h}}\right|_{\mathbf{h}=\mathbf{h}^{(k)}} = \left. \begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{u}_p} I_{MPP} \frac{\partial \mathbf{u}_p}{\partial \mathbf{h}}\right)\Big|_{\mathbf{u}_p=\mathbf{u}_{p_i}} \\ \vdots \end{bmatrix} \right|_{\mathbf{h}=\mathbf{h}^{(k)}}, \tag{6.11}$$

et :

$$\left.\frac{\partial \mathbf{I_{MPP}}(\mathbf{h},\lambda_p)}{\partial \lambda_p}\right|_{\lambda_p=\lambda_p^{(k)}} = \left. \begin{bmatrix} \vdots \\ \left(\frac{\partial I_{MPP}(\mathbf{u}_p,\lambda_p)}{\partial \lambda_p}\right)\Big|_{\mathbf{u}_p=\mathbf{u}_{p_i}} \\ \vdots \end{bmatrix} \right|_{\lambda_p=\lambda_p^{(k)}}. \tag{6.12}$$

En considérant que c'est directement le mélange de potentiels photométriques qui subit la transformation projective, $\partial \mathbf{u}_p/\partial \mathbf{h}$ a la même expression que l'expression générale de l'équation (3.70).

Les gradients spatiaux et d'échelle de $I_{MPP}()$ sont obtenus par une approximation aux différences finies :

$$\nabla_{\mathbf{u}_p} I_{MPP} = \begin{bmatrix} \nabla_{u_p} I_{MPP} & \nabla_{v_p} I_{MPP} \end{bmatrix} \tag{6.13}$$

avec :

$$\begin{cases} \nabla_{u_p} I_{MPP} \approx \dfrac{I_{MPP}\left(\begin{bmatrix} u_p + L_u & v_p \end{bmatrix}^\top, \lambda_p\right) - I_{MPP}\left(\begin{bmatrix} u_p - L_u & v_p \end{bmatrix}^\top, \lambda_p\right)}{2L_u} \\ \nabla_{v_p} I_{MPP} \approx \dfrac{I_{MPP}\left(\begin{bmatrix} u_p & v_p + L_v \end{bmatrix}^\top, \lambda_p\right) - I_{MPP}\left(\begin{bmatrix} u_p & v_p - L_v \end{bmatrix}^\top, \lambda_p\right)}{2L_v} \end{cases}, \tag{6.14}$$

et, enfin :

$$\begin{aligned} \frac{I_{MPP}(\mathbf{u}_p, \lambda_p)}{\lambda_p} &\approx \frac{I_{MPP}(\mathbf{u}_p, \lambda_p + L_\lambda) - I_{MPP}(\mathbf{u}_p, \lambda_p - L_{\lambda_p})}{2L_{\lambda_p}} \\ &\approx \frac{1}{2L_{\lambda_p}} \sum_{\mathbf{u} \in \mathcal{U}} I(\mathbf{u}) \left(f_p(\mathbf{u}_p - \mathbf{u}, \lambda_p + L_{\lambda_p}) - f_p(\mathbf{u}_p - \mathbf{u}, \lambda_p - L_{\lambda_p}) \right). \end{aligned} \tag{6.15}$$

L_u, L_v et L_{λ_p} paramètrent les pas de chacune des dérivées. Dans les évaluations qui suivent, $L_u = L_v = 1$ et $L_{\lambda_p} = 0,5$.

Remarque 7 (Convolution avec la fonction de potentiel) *Dans ce travail, en pratique, le support de la fonction de potentiel gaussienne $f_p()$, dans sa convolution de l'image $I(\mathbf{u})$ pour obtenir $I_{PP}(\mathbf{u}_p, \mathbf{u})$, est limité à une distance $2 * \lambda_p^{(k)}$ du centre, à chaque itération k. En effet, 95,4% de la surface sous la gaussienne étant comprise dans un rayon de $2*\lambda_p$, c'est une approximation classique qui permet, notamment, de réduire considérablement les temps de calcul, en particulier quand l'optimisation tend vers la convergence. Dans ce dernier cas, λ_p devient proche de zéro, donc presque tout le potentiel d'attraction d'un pixel est dans ce pixel lui-même. Par conséquent, à convergence, le coût $\mathscr{C}_{A+}$ (Eq. (6.7)) devient similaire au coût direct pur $\mathscr{C}_{BM}()$ (Eq. (3.66)), dont le domaine de convergence est largement suffisant pour corriger une erreur de l'ordre du pixel (cf. Partie 6.1.1).* ◇

6.2.2 Evaluation

6.2.2.1 Détermination des valeurs des paramètres

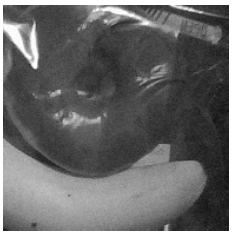
Pour l'évaluation de la méthode de suivi adaptative en échelle, l'étendue de la fonction de potentiel dans le mélange de potentiels photométriques de référence est fixé à $\lambda_p^* = 0,5$. Une valeur proche de zéro permet de conserver suffisamment de détail pour atteindre le même niveau de précision qu'avec une approche directe pure. Les valeurs de la paire $\lambda_p^{(0)}$ et λ, du calcul de l'incrément (Eq. (6.9)) des degrés de liberté, est déterminée expérimentalement en cherchant le meilleur compromis entre domaine de convergence, précision à convergence et nombre d'itérations. Ainsi, $\lambda \in [0,2; 0,3]$ et $\lambda_p^{(0)} \in [3; 6]$ mènent aux meilleures performances pour une transformation projective réduite à la translation pure dans le plan image, avec un taux de convergence supérieur à 85% sur 5000 images, contre un peu moins de 50% pour le suivi sans adaptation de l'espace d'échelle (Partie 3.3.2.1). Ces taux sont donnés pour une précision à convergence d'un dixième de pixel. Les 5000 images proviennent

du jeu de données variées MS-COCO [Lin et al., 2014] et, pour chaque, une image de référence est obtenue en appliquant une translation tirée aléatoirement d'une distribution uniforme dans l'intervalle [-10 ; 10] pixels pour chaque axe. La région, de 29 pixels de côté, à suivre est centrée en un point de l'image dont les coordonnées ont elles aussi été tirées aléatoirement.

6.2.2.2 Mouvement projectif

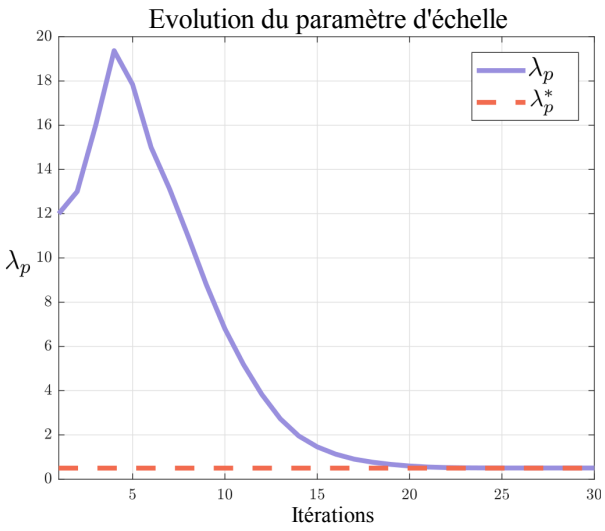
Suite à l'étude de la partie 6.2.2.1, on fixe $\lambda_p^* = 0,5$ et $\lambda = 0,3$. Pour λ_p, l'algorithme de suivi est très tolérant à son initialisation $\lambda_p^{(0)}$, comme le montre la figure 6.1(c) dans un cas où λ_p, initialisé à une valeur trop basse pour permettre de converger, augmente automatiquement au fil des itérations afin de trouver "l'entrée" du domaine de convergence avant de diminuer pour être précis à convergence. Cependant, même si ce dernier comportement est gage de robustesse à un ordre de grandeur d'amplitude de mouvement non anticipée, initialiser λ_p directement à une valeur compatible avec l'ordre de grandeur des transformations recherchées permet de réduire le nombre d'itérations. Par conséquent, dans cette partie, puisque les tirages aléatoires (indirects) des transformations sont faits dans l'intervalle [-42 ; +42] pixels, l'intervalle idéal des $\lambda_p^{(0)}$, déterminé en partie 6.2.2.1, est multiplié par 4, menant à $\lambda_p^{(0)} \in [12; 24]$ et fixer $\lambda_p^{(0)} = 12$ mène aux meilleures résultats.

Toujours sur les 5000 images du jeu de données MS-COCO, le suivi visuel adaptatif dans l'espace d'échelle atteint les meilleures performances en précision et en taux de convergence comparé au variantes additive, compositionnelle, composition-



(a)

(b)



(c)

FIGURE 6.1 – (a) Image dans laquelle trouver (b) la région de référence. (c) Evolution de l'étendue du potentiel d'attraction λ_p, d'abord positive pour "accrocher" le domaine de convergence, puis négative pour la précision à convergence.

nelle inverse et additive avec trois niveaux de pyramide gaussienne du coût $\mathscr{C}_{BM}()$ (Eq. (3.66)) pour le suivi de plan sur $\mathbb{P}^2$, de même que la formulation sur $\mathfrak{se}(3)$ exploitant la méthode d'optimisation ESM (Partie 3.2.4). Enfin, comparaison est aussi faite avec une approche géométrique utilisant des correspondances SIFT et estimant l'homographie par RANSAC [Fischler and Bolles, 1981], avec un taux de succès similaire à celui de l'ESM de 20% pour une erreur tolérée d'un pixel sur la moyenne des erreurs géométriques des quatre coins de la région suivie (Fig. (6.2)). Toutes les autres variantes du coût $\mathscr{C}_{BM}()$ mènent à des taux de succès de l'ordre de 5% pour la même tolérance sur l'erreur d'estimation, sauf l'approche à trois niveaux de pyramide qui monte à 60%, rejointe par la méthode base SIFT et RANSAC, si on tolère une erreur de trois pixels. Dans ces conditions, la méthode proposée de suivi visuel adaptatif dans l'espace d'échelle atteint 82% de succès quand une erreur d'un pixel est tolérée et dépasse les 85% quand trois pixels d'erreur sont tolérés.

Enfin, la méthode proposée est confrontée aux approches directes basées apprentissage, notamment le Conditionnal-LK [Lin et al., 2016] et SDM (Supervised descent method) [Xiong and De la Torre, 2015], ayant atteint des taux de convergence en suivi de régions correspondant à un visage supérieurs à l'approche compositionnelle inverse résolvant la minimisation de $\mathscr{C}_{BM}()$ (Eq. (3.66)) sur $\mathbb{P}^2$, jusqu'à +40% de taux de convergence quand un nombre suffisamment pertinent d'exemples, soit 1000 images, a été considéré dans l'apprentissage. En reproduisant le processus d'apprentissage du Conditionnal-LK sur 6 visages du jeu de données de Yale [Belhumeur et al., 1997], on retrouve des courbes de taux de convergence en fonction de l'amplitude de la transformation d'homographie, cette fois-ci représentée

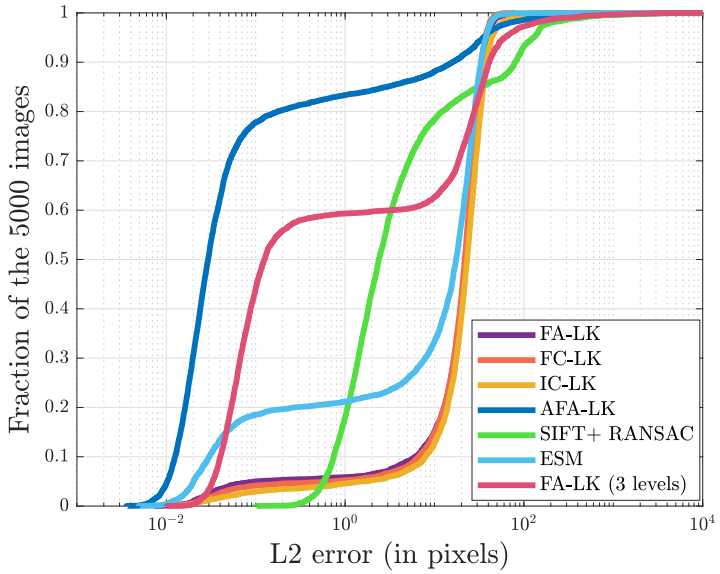
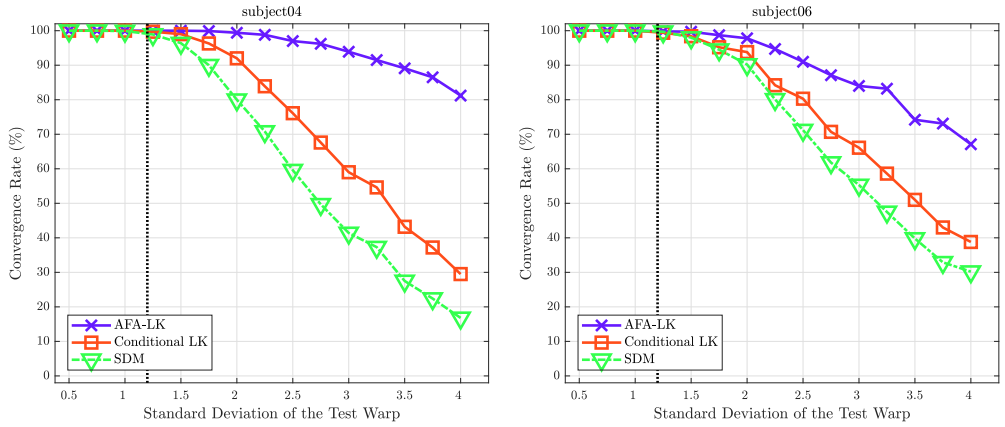


FIGURE 6.2 – Distribution cumulée de l'erreur moyenne sur les quatre coins de la région à suivre pour plusieurs algorithmes (AFA-LK : suivi visuel adaptatif dans l'espace d'échelle).



(a) (b)

FIGURE 6.3 – Extraits des taux de convergence pour deux des six sujets considérés du jeu de données de Yale, calculés sur un total de 15000 paires d'images pour chaque sujet (AFA-LK : suivi visuel adaptatif dans l'espace d'échelle).

par l'écart type σ de la distribution de laquelle 1000 tirages aléatoires de perturbations des quatre coins de la région à suivre, plus une perturbation globale sur les 4. L'apprentissage est réalisé pour $\sigma = 1,2$ pixels à mettre en regard des régions de 20×20 pixels à suivre. Ensuite, une base de test de 15000 images par visage est générée sur le même principe, 1000 par valeur de σ, pour 15 σ entre 0,5 et 4 pixels, avec un pas de 0,25.

Pour les images de la base de test qui ont été générées avec $\sigma < 1,2$, les trois méthodes donnent 100% de taux de convergence. Pour $\sigma = 1,5$, les méthodes basées apprentissages descendent sous les 100% alors que le suivi dans l'espace d'échelle se maintient. De $\sigma = 1,75$ à 4, tous les taux de convergences décroissent, mais avec une bien meilleure robustesse du suivi dans l'espace d'échelle aux transformations les plus fortes : taux de convergence de +30% (Fig. 6.3(a)) à +50% (Fig. 6.3(b)) par rapport aux approchent basées apprentissage.

6.2.3 Conclusion partielle

La généralisation des approches de suivi visuel direct basées pyramide multi-échelle au suivi visuel adaptatif dans l'espace d'échelle, en considérant non plus un nombre arbitraire de niveaux échantillonnant grossièrement l'espace d'échelle, mais un coefficient d'échelle, le paramètre d'étendue d'attraction λ_p, pouvant prendre toute valeur réelle positive, a permis de surpasser largement de nombreuses méthodes directes, géométriques et basées apprentissage de l'état-de-l'art. En particulier, en présence de mouvement important de la région à suivre dans l'image de l'ordre, indicatif puisque le mouvement est projectif, de 40 pixels dans une image de 128 pixels de côté, un accroissement du taux de convergence de 25% à 50%, par rapport aux méthodes les plus pertinentes, c'est-à-dire le suivi visuel direct basé

pyramide à trois niveaux d'échelle et le Conditionnal-LK, montre l'impact de considérer le paramètre d'étendue de l'attraction des pixels dans les degrés de liberté du problème d'optimisation. Il est clair que ce degré de liberté supplémentaire permet de naviguer entre les niveaux entiers d'échelles des pyramides gaussiennes au fil des itérations, et pas forcément toujours dans le même sens, selon le contenu de l'image et l'allure de la fonction de coût. En plus de ces avantages nets, ce nouveau suivi visuel direct introduit une formalisation élégante du passage d'un coefficient d'échelle à un autre, en un unique problème d'optimisation, plutôt qu'une séquence de résolutions successives de problèmes d'optimisation dans des niveaux d'échelle discontinus.

Enfin, tout comme pour les approches de vision robotique directe basées noyaux photométriques, l'introduction de la convolution par une fonction de potentiel gaussienne rend I_{MPP} plus lisse que l'image acquise, permettant d'envisager de la dériver analytiquement, plutôt que numériquement comme fait dans ce travail. C'est ce que propose la partie 6.3, dans le cadre de l'asservissement visuel.

6.3 Asservissement visuel direct basé mélange de potentiels

6.3.1 Définition du coût et de la loi de commande

Le coût à minimiser est similaire au coût $\mathscr{C}_{A+}()$ (Eq. (6.7)), à ceci près que l'on ne se restreint pas à une région de l'image, que les degrés de liberté géométriques sont les six degrés de liberté rigides $\boldsymbol{\delta}\mathbf{p}$ de la caméra, et que l'on ne connaît pas la géométrie de la scène. On suppose donc que cette dernière est constante, à convergence, comme dans le coût $\mathscr{C}_{CMC}()$ (Eq. (3.98)) de l'asservissement visuel photométrique :

$$\begin{aligned}
&\mathscr{C}_{C+}(\{\boldsymbol{\delta}\mathbf{p},\lambda_p\},\mathcal{I},\{Z^*,\gamma_u,\lambda_p^*\},\mathcal{I}^*)\\
&\quad= \tfrac{1}{2}\sum_{\mathbf{u}_p^*\in\mathcal{U}^*}\left(I_{MPP}(\mathcal{I},\mathbf{u}_p(\boldsymbol{\delta}\mathbf{p},\mathbf{u}_p^*,Z^*),\lambda_p)-I^*_{MPP}(\mathcal{I}^*,\mathbf{u}_p^*,\lambda_p^*)\right)^2\\
&\quad= \tfrac{1}{2}||\mathbf{I_{MPP}}(\boldsymbol{\delta}\mathbf{p},\lambda_p)-\mathbf{I_{MPP}}^*||^2\\
&\quad= \tfrac{1}{2}||\mathbf{C_{C+}}(\boldsymbol{\delta}\mathbf{p},\lambda_p)||^2.
\end{aligned} \tag{6.16}$$

En considérant la méthode de Gauss-Newton, la loi de commande permettant de minimiser le coût $\mathscr{C}_{C+}()$ (Eq.n (6.16)) est :

$$\mathbf{v_{C+}}=\begin{bmatrix}\dot{\boldsymbol{\delta}\mathbf{p}}^{(t)}\\ \dot{\lambda}_p^{(t)}\end{bmatrix}=-\lambda\mathbf{J}^{\dagger}_{\boldsymbol{\delta}\mathbf{p}\lambda_p^{(t)}}\mathbf{C_{C+}}(\boldsymbol{\delta}\mathbf{p}^{(t)},\lambda_p^{(t)}), \tag{6.17}$$

avec :

$$\begin{aligned}
\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}\lambda_p^{(t)}} &= \left[\left.\frac{\partial\mathbf{I_{MPP}}(\boldsymbol{\delta}\mathbf{p},\lambda_p)}{\partial\boldsymbol{\delta}\mathbf{p}}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(t)}} \;\middle|\; \left.\frac{\partial\mathbf{I_{MPP}}(\boldsymbol{\delta}\mathbf{p},\lambda_p)}{\partial\lambda_p}\right|_{\lambda_p=\lambda_p^{(t)}}\right]\\
&= \left[\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}_{MPP}^{(t)}} \;\middle|\; \mathbf{J}_{\lambda_p^{(t)}}\right].
\end{aligned} \tag{6.18}$$

6.3.2 Matrice jacobienne relative à l'échelle

La matrice jacobienne $\mathbf{J}_{\lambda_p^{(t)}}$ des dérivées de $\mathbf{I_{MPP}}$ par rapport à λ_p est similaire à celle de l'équation (6.12) et nécessite d'exprimer la dérivée de chaque intensité $I_{MPP}(\mathbf{u}_p, \lambda_p)$ du mélange de potentiels photométriques par rapport à λ_p. Dans la partie 6.2.1, ces dérivées sont approximées par des différences finies, nécessitant de fixer le pas de la *dérivée*. Dans cette partie, on les exprime analytiquement à partir de l'expression formelle de $I_{MPP}()$ (Eq. (6.3)) :

$$\begin{aligned}
\frac{\partial I_{MPP}(\mathbf{u}_p, \lambda_p)}{\partial \lambda_p} &= \sum_{\mathbf{u}\in\mathcal{U}} \frac{\partial \left(I(\mathbf{u}) f_p(\mathbf{u}_p - \mathbf{u}, \lambda_p)\right)}{\partial \lambda_p} \\
&= \sum_{\mathbf{u}\in\mathcal{U}} \left(I(\mathbf{u}) \frac{\partial f_p(\mathbf{u}_p - \mathbf{u}, \lambda_p)}{\partial \lambda_p} \right) \\
&= \sum_{\mathbf{u}\in\mathcal{U}} \left(I(\mathbf{u}) \frac{\partial}{\partial \lambda_p} \left[\exp\left(-\tfrac{1}{2} \left([\mathbf{u}_p - \mathbf{u}]^\top \left(\lambda_p^2 \mathbf{I}_{2\times 2}\right)^{-1} [\mathbf{u}_p - \mathbf{u}] \right) \right) \right] \right) \\
&= \sum_{\mathbf{u}\in\mathcal{U}} \left(I(\mathbf{u}) \frac{(\mathbf{u}_p - \mathbf{u})^2}{\lambda_p^3} \exp\left(-\frac{(\mathbf{u}_p - \mathbf{u})^2}{2\lambda_p^2} \right) \right),
\end{aligned} \tag{6.19}$$

en ayant repris l'expression de la fonction de potentiel $f_p()$ en dimension 2 ($n = 2$, Eq. (6.6)), privé du facteur devant l'exponentielle. Dans l'équation (6.19), tout est connu et $I(\mathbf{u})$ est utilisée tel quel, sans gradient d'image, permettant de calculer le gradient analytique et non numérique de $I_{MPP}()$ par rapport à λ_p.

6.3.3 Matrice jacobienne relative à la pose de la caméra

L'expression analytique des dérivées de chaque $I_{MPP}()$ par rapport à la pose $\boldsymbol{\delta}\mathbf{p}$ (ensuite empilées pour former la matrice $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}_{MPP}} = \mathbf{I_{MPP}}(\boldsymbol{\delta}\mathbf{p}, \lambda_p)/\partial\boldsymbol{\delta}\mathbf{p}$), doit s'affranchir du gradient de l'image $\nabla_{\mathbf{u}} I$, qui intervient théoriquement dans les calculs puisque $I_{MPP}()$ dépend de $I()$, qui dépend elle-même de $\mathbf{u}$ qui, physiquement, dépend de la pose $\boldsymbol{\delta}\mathbf{p}$ et du déplacement de la caméra. Pour ce faire, deux méthodes ont été étudiées.

La première méthode évite le calcul des gradients $\nabla_{\mathbf{u}} I$ en exploitant le théorème de Green, comme dans la formulation de la matrice jacobienne des moments photométriques (Partie 3.5.3.3), en admettant que les intensités des pixels des bords du rectangle image soient nuls.

La seconde méthode repose sur une autre hypothèse selon laquelle c'est le mélange de potentiels photométriques qui subit directement la transformation géométrique du contenu de l'image engendrée par le déplacement de la caméra dans l'espace, comme dans le suivi visuel adaptatif dans l'espace d'échelle (Partie 6.2). Cependant, contrairement à ce dernier, le calcul numérique par différence finie des gradients de $I_{MPP}()$ est remplacé par le gradient analytique exprimé à partir de la définition formelle de $I_{MPP}()$ (Eq. (6.3)). Ainsi, en suivant une méthodologie similaire à celle permettant l'expression de la matrice jacobienne directe pure par rapport à la pose (Partie 3.2.2), reposant, cette fois-ci sous l'hypothèse de conservation de

l'intensité du mélange de potentiels photométriques :

$$I_{MPP}(\mathbf{u}_p(\mathbf{p},\mathbf{X}),t) = I_{MPP}(\mathbf{u}_p(\mathbf{p}\oplus\boldsymbol{\delta}\mathbf{p},\mathbf{X}),t+\delta t), \tag{6.20}$$

on exprime la matrice jacobienne $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}_{MPP}^{(t)}}$ (Eq. (6.18)) par :

$$\begin{aligned}\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}_{MPP}^{(t)}} &= \left.\frac{\partial \mathbf{I_{MPP}}(\boldsymbol{\delta}\mathbf{p},\lambda_p)}{\partial \boldsymbol{\delta}\mathbf{p}}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(t)}} \\ &= \left.\begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{u}_p} I_{MPP}\dfrac{\partial \mathbf{u}_p}{\partial \boldsymbol{\delta}\mathbf{p}}\right)\Big|_{\mathbf{u}_p=\mathbf{u}_{p_i}} \\ \vdots \end{bmatrix}\right|_{\boldsymbol{\delta}\mathbf{p}=\boldsymbol{\delta}\mathbf{p}^{(t)}},\end{aligned} \tag{6.21}$$

où $\partial\mathbf{u}_p/\partial\boldsymbol{\delta}\mathbf{p}$ a la même expression qu'en équation (3.10) (avec un remplacement de $\mathbf{u}$ par $\mathbf{u}_p$, et en considérant le modèle de projection perspective dans cette partie) et :

$$\begin{aligned}\nabla_{\mathbf{u}_p} I_{MPP} &= \frac{\partial I_{MPP}(\mathcal{I},\mathbf{u}_p,\lambda_p)}{\partial \mathbf{u}_p} \\ &= \sum_{\mathbf{u}\in\mathcal{U}} \frac{\partial\left(I(\mathbf{u}) f_p(\mathbf{u}_p-\mathbf{u},\lambda_p)\right)}{\partial \mathbf{u}_p} \\ &= \sum_{\mathbf{u}\in\mathcal{U}} \left(I(\mathbf{u})\frac{\partial}{\partial \mathbf{u}_p}\left[\exp\left(-\tfrac{1}{2}\left(\left[\mathbf{u}_p-\mathbf{u}\right]^\top\left(\lambda_p^2\mathbf{I}_{2\times 2}\right)^{-1}\left[\mathbf{u}_p-\mathbf{u}\right]\right)\right)\right]\right) \\ &= -\tfrac{1}{\lambda_g^2}\sum_{\mathbf{u}\in\mathcal{U}}\left(I(\mathbf{u})\left[\mathbf{u}_p-\mathbf{u}\right]^\top \exp\left(-\tfrac{\|\mathbf{u}_p-\mathbf{u}\|^2}{2\lambda_p^2}\right)\right).\end{aligned} \tag{6.22}$$

L'équation ci-dessus confirme bien l'absence de gradient numérique de l'image acquise.

Quand la première méthode d'expression de $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}_{MPP}}$ à base du théorème de Green est développée en supposant explicitement une scène fronto-parallèle à la caméra *[Crombez et al., 2019]*, les deux méthodes, bien que reposant sur des hypothèses totalement différentes, et partant d'expressions différentes elles-aussi, mènent à des expressions très similaires de la matrice jacobienne $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}_{MPP}}$. Les éléments de cette dernière sont même strictement identiques [2] pour trois des six degrés de liberté (translations selon les axes $\mathbf{X}_c$ et $\mathbf{Y}_c$ et rotation autour de l'axe $\mathbf{Z}_c$). Pour les trois autres, l'expression de l'équation (6.21) représente une approximation de la méthode à base du théorème de Green en supprimant un terme.

Néanmoins, l'évaluation en simulation des asservissements visuels à 2 et à 6 degrés de liberté face à une scène plane avec les deux méthodes de calcul de la matrice jacobienne $\mathbf{J}_{\boldsymbol{\delta}\mathbf{p}_{MPP}}$ n'a pas montré de différence significative [3] de comportement *[Crombez et al., 2019]* (trajectoires similaires, nombres d'itérations similaires

2. Grâce au fait que les grilles de *pixels* de $I()$ et $I_{MPP}()$ coïncident parfaitement.

3. les asservissements visuels à 2 degrés de liberté donnent même exactement les mêmes résultats puisque les deux premières translations sont considérées.

pour converger complètement), montrant que la deuxième méthode est une très bonne approximation de la première, considérée comme la méthode rigoureuse. Cependant, le coût calculatoire beaucoup plus modéré de la deuxième justifie de la privilégier en expérimentations sur robot réel.

Dans les deux cas, à cause des approximations faites sur la scène, ou le fait de considérer que tous les pixels aient la même coordonnée de profondeur, constante malgré le déplacement de la caméra, la loi de commande de l'équation (6.17) n'emploie, en fait, qu'une approximation $\widehat{\mathbf{J}}_{\boldsymbol{\delta \mathbf{p}}}$ de la matrice jacobienne $\mathbf{J}_{\boldsymbol{\delta \mathbf{p}}}$.

6.3.4 Stratégie de mise à jour de l'étendue d'attraction

Le rôle des paramètres d'étendue d'attraction λ_p et λ_p^* est d'assurer une vaste étendue du domaine de convergence et une précision importante. Comme dans le suivi visuel direct dans l'espace d'échelle 6.2, la stratégie retenue se base sur une valeur proche de zéro pour λ_p^* afin d'assurer la précision à convergence et une valeur initiale plus importante pour $\lambda_p^{(0)}$ pour assurer l'importante étendue du domaine de convergence. De même, les valeurs pertinentes de $\lambda_p^{(0)}$ dépendent de l'amplitude maximale du mouvement dans l'image, c'est-à-dire la taille de l'image elle-même dans le cas de l'asservissement visuel. En effet, idéalement, on voudrait pouvoir faire converger l'asservissement visuel avec seulement un pixel commun entre l'image initiale et l'image désirée. Par conséquent, puisque 95,4% de la surface sous la gaussienne représentant le potentiel d'attraction se situe dans un cercle de rayon $2\lambda_p$ autour du pixel auquel cette gaussienne est centrée, fixer $\lambda_p^{(0)}$ à la moitié de la diagonale de l'image apparaît, intuitivement, une initialisation pertinente. En pratique, l'asservissement visuel se montre robuste au $\lambda_p^{(0)}$ au point de pouvoir l'initialiser à des valeurs plus faibles [4]

Pour agrandir encore le domaine de convergence de l'asservissement visuel basé mélange de potentiels photométriques, une stratégie en deux étapes est mise en place :

1. Fixer $\lambda_p^{(0)}$ à une valeur dans la partie haute de l'intervalle compris entre 2 et la demi-longueur de la diagonale de l'image et $\lambda_p^* = \lambda_p^{(0)}/2$.

2. Quand, au temps t_1, la première étape a convergé, fixer $\lambda_p^* = \lambda_p^{(t_1)}$ à une valeur proche de zéro ($\lambda_p^* = \lambda_p^{(t_1)} = 1$ dans toutes les expérimentations faites en environnement réel), recalculer I_{MPP}^* et continuer l'asservissement visuel jusqu'à atteindre la stabilité du coût $\mathscr{C}_{C+}()$ (Eq. 6.16).

Pour détecter la convergence de la première étape, puisque l'asservissement, comme le suivi visuel de la partie 6.2, fait tendre $\lambda_p^{(t)}$ vers λ_p^* au fil du déplacement de la caméra, l'écart $|\lambda_p^{(t)} - \lambda_p^*|$ entre les deux est testé selon un seuil, fixé empiriquement à $0,1$ dans toutes les expérimentations en environnement réel.

4. jusque $\lambda_p^{(0)} = 2$, pour des images de 40×40 à 80×80 pixels *[Crombez et al., 2019]*.

6.3.5 Résultats

De nombreuses expérimentations ont été faites en simulation, où les degrés de liberté cartésiens de la caméra (virtuelle) sont directement *actionnés*, et en réel, allant de deux degrés de liberté face à une scène ponctuelle (tous les pixels sont noirs sauf un) à 6 degrés de liberté face à une scène 3D *[Crombez et al., 2019]*.

Tout d'abord en simulation en commandant les 6 degrés de liberté cartésiens de la caméra observant une scène 3D (Fig. 6.4), on considère, comme dans le cas réel, que la profondeur de la scène est inconnue [5] et on fait l'approximation d'une même profondeur pour tous les pixels à la pose désirée et aux poses courantes (${}^cZ = 0,5m$). L'écart entre la pose désirée et la pose initiale, c'est-à-dire, implicitement le $\boldsymbol{\delta \mathbf{p}}^{(0)}$ (explicitement inconnu de la loi de commande), est de $(26,95\ m; -5,02\ m; -11,14\ m; 14,40^o; -27,54^o, -5,88^o)$. Cet écart est très important, particulièrement en rotation autour des deux premiers axes, rendant cette expérimentation très difficile. La figure 6.4(e) montre que la différence initiale dans l'image est aussi très importante, reflétant la nature très difficile de cette expérimentation. La figure 6.4(c) correspond à l'itération précédent le passage à la deuxième étape de la stratégie de mise à jour de l'étendue d'attraction (Partie 6.3.4). A la fin de l'asservissement visuel de ce cas très difficile pour toute approche directe, la caméra converge à la pose désirée (Fig. 6.4(g)) avec une erreur finale de $(1,6\ mm; 2,6\ mm; 4,1\ mm; 0,02^o; 0,01^o; 0,02^o)$.

Cette dernière expérimentation est un extrait des 20 expérimentations ayant pour but d'évaluer les taux de convergence de l'approche proposée par rapport aux méthodes de référence dans l'état de l'art, à savoir l'asservissement visuel photométrique [Collewet and Marchand, 2011] (Partie 3.3.2.4) et l'asservissement visuel basé moments photométriques [Bakthavatchalam et al., 2018] (Partie 3.5.3.3). Toutes ces expérimentations ont pour consigne la même image synthétisée à la pose désirée mais 20 tirages aléatoires donnent autant de poses initiales. L'asservissement visuel basé mélange de potentiels photométriques converge dans 100% des cas avec la même valeur de $\lambda_p^{(0)} = 25$ alors que le photométrique ne converge que dans 3 cas sur les 20 et les moments photométriques permettent de converger 11 fois sur les 20. "Converger", dans le cas présent, signifie s'être stabilisé avec une erreur à la pose désirée inférieure à $(5,00\ mm; 5,00\ mm; 5,00\ mm; 0,1^o; 0,1^o; 0,1^o)$.

Cette amélioration considérable du taux de succès de l'asservissement visuel est directement imputable à la transformée de l'image en un mélange de potentiels photométriques dont l'étendue d'attraction est automatiquement optimisée au fil du déplacement de la caméra. Cependant, sa mise en oeuvre se fait au prix d'un coût calculatoire relativement important puisqu'en considérant des images de 100 pixels de côté, 100 ms de traitements parallélisés sur GPU (Nvidia Cuda) sont nécessaires pour produire le vecteur de commande et de mise à jour du paramètre d'étendue d'attraction alors que l'approche photométrique, elle, ne demande que 20 ms de

5. La simulation donne une profondeur pour chaque pixel où un objet de la scène virtuelle se projette mais cette information est ignorée dans ce cas pour s'approcher de la réalité et valider la loi de commande.

traitements séquentiels sur CPU. Sur un système réel, ce temps de traitement limite la boucle de commande à 10Hz, ce qui impose de considérer un gain faible dans la loi de commande. C'est pourquoi, dans les expérimentations réelles, les images sont réduites à 80 pixels de côté, maximum.

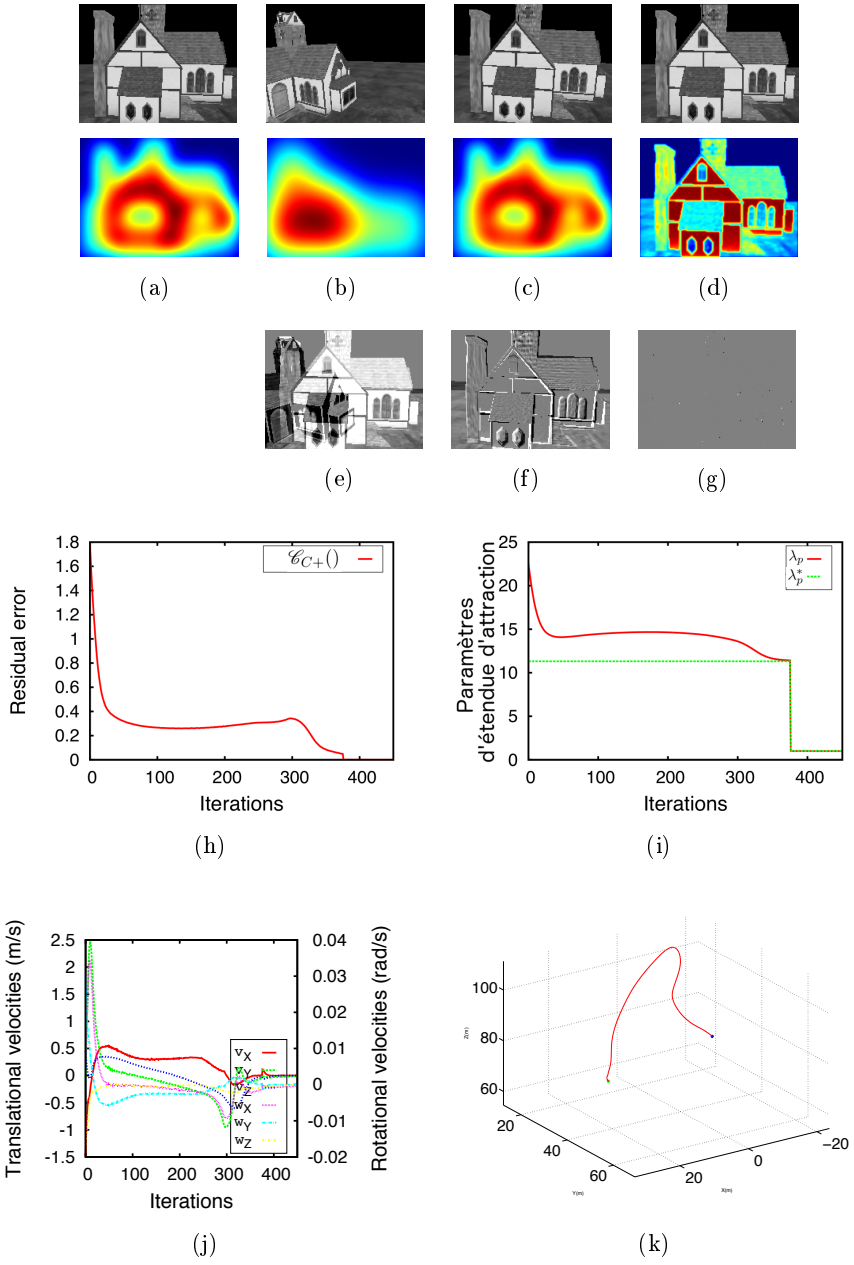
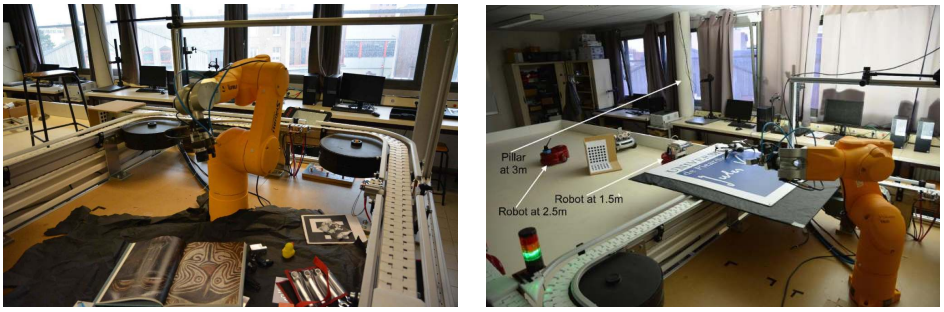


FIGURE 6.4 – Simulation d'asservissement visuel basé mélange de potentiels photométriques. Images et leurs mélanges de potentiels : (a) désirée, (b) initiale, (c) à la fin de l'étape 1, (d) finale. (e)-(g) Image de différence entre (b)-(d) et (a), (h) évolution du coût, (i) paramètres d'étendue d'attraction, (j) vitesses, (k) trajectoire.

Les expérimentations en environnement réel sont menées avec une caméra perspective montée sur un bras robotique Stäubli TX-60 dont le volume de travail utilisable est une demi-sphère de 60 cm de rayon (Fig. 6.5). De nombreuses expérimentations ont été menées dans des scènes différentes allant d'objets plats sur fond uniforme sans perturbation à des scènes 3D avec des profondeurs très différentes (entre $0,5$ m et 6 m, Fig. 6.5(b)), à éclairage variable et occultations partielles pendant l'asservissement visuel *[Crombez et al., 2019]*. L'asservissement visuel basé mélange de potentiels photométriques a convergé dans de très nombreuses situations, y compris en présence d'écarts très importants dans l'espace comme dans l'image.

A titre d'exemple, la figure 6.6 montre l'une des expérimentations où les images initiale et désirée présentent d'importantes différences tout comme les poses initiale et désirée de la caméra avec un écart de $(-0,043\ m; -0,300\ m; 0,018\ m; 20,80^o; -7,84^o; 5,97^o)$. A noter qu'avec la taille du robot et les risques d'atteindre des butées articulaires ou des singularités, non traitées par la loi de commande de l'équation (6.17), appliquée dans ces expérimentations, tout en ayant un minimum d'intersection des champs de la caméra aux poses initiale et désirée, il est difficile d'avoir un écart plus important dans l'espace. La scène contient plusieurs objets 3D disposés sur une table distante d'approximativement $0,5$ m de la caméra à la pose désirée (${}^cZ = 0,5$). Néanmoins, ces objets ne sont pas très grands menant à une variation de profondeurs de $0,02$ m environ par rapport à la table. Pendant l'asservissement, des occultations partielles ont été opérées (Figs. 6.6(h)-6.6(k)) pour observer leur impact sur la commande du robot. Ces occultations sont clairement visibles dans les courbes d'évolution du coût, des vitesses et des paramètres d'étendue d'attraction en fonction du temps (Figs. 6.6(l)-6.6(n)) aux itérations 500 et 1100. A convergence, l'image finale de différence (Fig. 6.6(g)) n'est pas parfaitement nulle à cause d'un changement d'illumination entre l'acquisition de l'image désirée et l'asservissement visuel, mais l'erreur finale de positionnement reste très faible $(1,48\ mm; 0,97\ mm; 0,54\ mm; -0,13^o; -0,21^o; 0,06^o)$, comparable aux résultats de simulation.

Résultats en vidéo : `https://youtu.be/wYYnu3PENLw`.



(a) (b)

FIGURE 6.5 – Environnement d'expérimentation (caméra sur robot Stäubli TX-60).

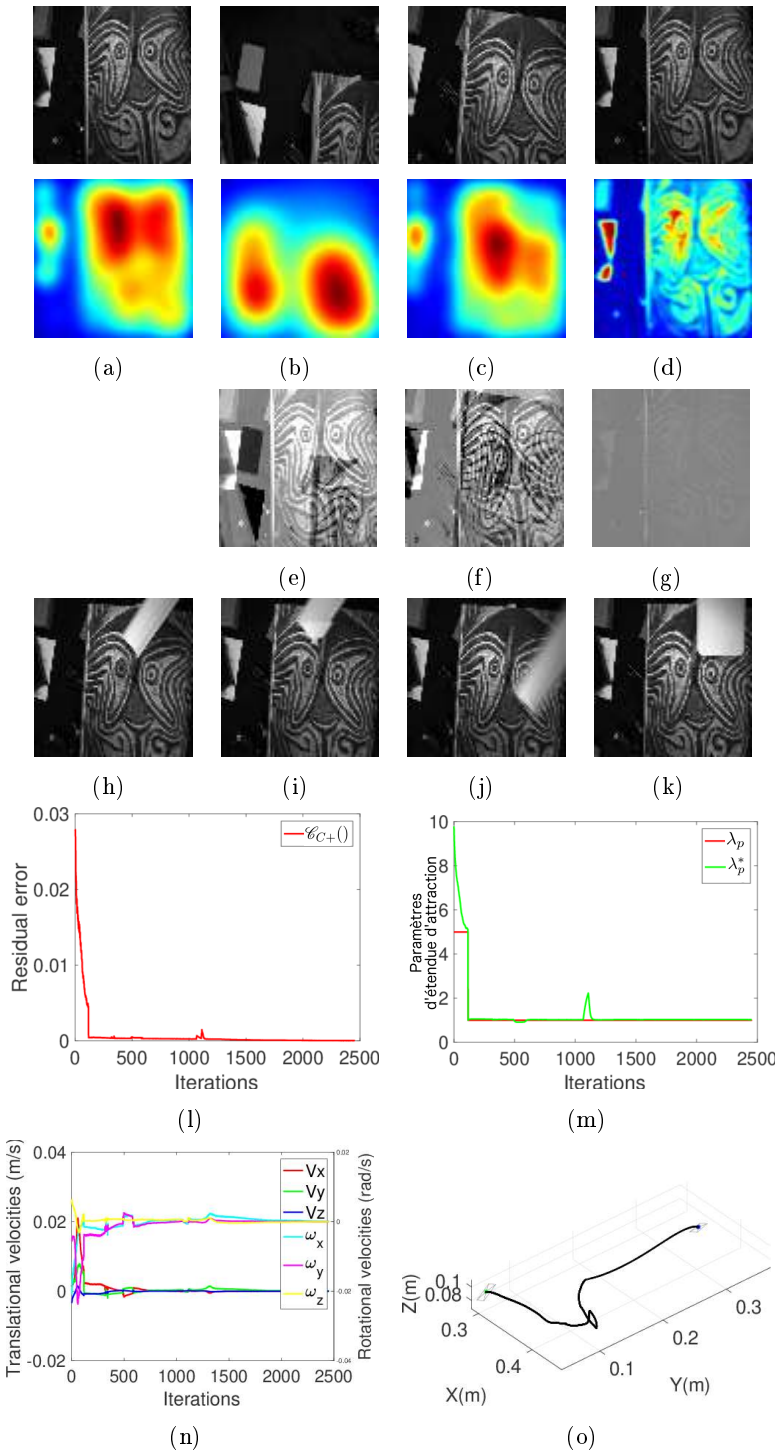


FIGURE 6.6 – Expérimentation en environnement réel. Images et leurs mélanges de potentiels : (a) désirée, (b) initiale, (c) fin de l'étape 1, (d) finale. (e)-(g) Image de différence, (h)-(k) Exemples d'occultations, (l) évolution du coût, (m) paramètres d'étendue d'attraction, (n) vitesses, (o) trajectoire.

6.3.6 Conclusion partielle

L'amélioration considérable du taux de succès, proche d'être multiplié par deux, de l'asservissement visuel direct grâce aux mélanges de potentiels photométriques est du même ordre que le gain apporté au suivi visuel direct dans l'espace d'échelle (Partie 6.2), reposant sur la même transformée des intensités des pixels des images, à quelques détails près. Cette amélioration confirme donc la pertinence de la transformée d'image en mélange de potentiels photométriques pour la vision robotique directe quand de grands déplacements dans l'image sont requis pour atteindre l'image désirée, avec précision.

Cependant, dans les cas où l'asservissement visuel basé moments photométriques converge, sa trajectoire dans l'espace est proche de la géodésique, ce qui est loin d'être le cas pour l'asservissement visuel basé mélange de potentiels photométriques. En effet, pour ce dernier, rien ne l'impose alors que la loi de commande exploitant les moments photométriques est explicitement construite dans le but de découpler les degrés de liberté (Partie 3.5.3.3). D'autre part, cette construction est aussi faite pour que la matrice jacobienne, approximée, de la loi de commande soit carrée, permettant d'envisager la stabilité globale, si l'approximation est raisonnable. L'asservissement visuel basé mélange de potentiels photométriques, quant à lui, considère autant d'intensités $I_{MPP}()$ que de pixels de l'image acquise $I()$ dans la loi de commande, bien plus que le nombre de degrés de liberté contrôlés, donc seule une stabilité asymptotique locale est attendue [Chaumette and Hutchinson, 2006]. Les simulations et expérimentations menées ont toutefois permis d'observer que cette localité est très vaste. Quoi qu'il en soit, la proximité de la trajectoire réalisée à la géodésique est une piste d'amélioration pour cet asservissement visuel.

Ce dernier problème, important quand on déplace une caméra réelle, est beaucoup plus tolérable, tant que le nombre d'itérations reste faible pour faire converger le suivi visuel de région d'image (Partie 6.2) ou pour l'estimation de mouvement de caméra, traité dans la partie 6.4 suivante.

6.4 Gyroscope visuel sphérique direct

6.4.1 Vue d'ensemble

Les images d'une caméra sphérique (cf. Partie 2.1.3) offrent le potentiel de pouvoir en observer toute rotation pure ${}^{c}\mathbf{R}_{c^*} \in SO(3)$. En effet, avec une telle caméra, l'environnement est perceptible dans toutes les directions sans exception, quelque soit ${}^{c}\mathbf{R}_{c^*}$, contrairement aux caméras à champ de vue limité, pour lesquelles une rotation ${}^{c}\mathbf{R}_{c^*}$ de trop forte ampleur engendre une absence totale de recouvrement des champs de vue de la caméra aux orientations $\mathbf{r}_c \in \mathbb{R}^3$ et $\mathbf{r}_{c^*} \in \mathbb{R}^3$, représentations vectorielles des orientations des repères caméras $\mathcal{F}_c$ et $\mathcal{F}_{c^*}$. En effet, même pour une caméra panoramique à champ de vue de 180°, deux images acquises avec un écart de rotation de 180° de la caméra autour de son axe $\mathbf{X}_c$ ne partagent aucune information directe.

Le gyroscope visuel sphérique direct est alors une méthode d'estimation de rotations entre deux images acquises par une caméra sphérique, en exploitant la transformée d'image en mélange de potentiels photométriques.

Des gyroscopes visuels partiellement sphériques ont déjà été proposés dans la littérature pour estimer tout ou partie des trois angles de rotation, mais principalement à l'aide de primitives géométriques [Mariottini et al., 2012, Bazin et al., 2012, Churchill and Vardy, 2013, Demonceaux et al., 2006], les contraignant à des environnements structurés, par exemple pourvus de droites parallèles, ou, inversement naturels et très dégagés afin que la ligne d'horizon soit visible.

Pour palier à ces problèmes et proposer des méthodes moins contraintes par l'environnement, les intensités de tous les pixels de l'image, interprétée comme un signal 2D, ont été exploitées pour estimer la pose de la caméra : les méthodes basées apparence [Labrosse, 2006, Scaramuzza and Siegwart, 2008], dans le domaine de l'image, et les méthodes basées analyse harmonique dans le domaine fréquentiel pour l'estimation d'attitude [Kyatkin and Chirikjian, 1999] et de mouvement rigide dans l'espace [Makadia et al., 2007].

Enfin, des méthodes hybrides ont récemment émergé, tel que SVO [Forster et al., 2017] (Semidirect Visual Odometry), qui bénéficient à la fois des avantages des méthodes basées primitives géométriques, généralement à domaine de convergence quasi-global quand la mise en correspondance réussit, voire global avec des méthodes d'optimisation adaptées, et des méthodes directes, à savoir la précision à convergence.

Dans cette partie, on se propose d'adapter l'expression du mélange de potentiels photométriques (Partie 6.1) à l'image sphérique pour formuler un gyroscope visuel sphérique direct, à la fois précis et à domaine de convergence très vaste, en exploitant les méthodes d'optimisation usuelles (Partie 3.2.2). En pratique, les images sphériques sont acquises par une caméra (Fig. 2.3(a)), produisant donc des images planes, plaquées sur la sphère unitaire du modèle de projection centrale unifié à deux plans images (Partie 2.1.3) pour en changer la représentation avant d'en transformer les intensités en un mélange de potentiels photométriques sphérique (Partie 6.4.2). Ensuite, cette représentation est exploitée dans un algorithme d'optimisation des trois degrés de liberté de la rotation (Partie 6.4.3) appliquant itérativement des rotations au mélange de potentiels sphériques courant pour minimiser l'écart entre le mélange courant et le mélange désiré. Enfin, l'algorithme est évalué expérimentalement en partie 6.4.4, avant de conclure ce travail (Partie 6.4.5).

La figure 6.7 donne une vue d'ensemble de l'algorithme.

6.4.2 Représentation d'image et mélange de potentiels sphériques

Les recherches en représentation d'image sphérique sont toujours actives [Adarve and Mahony, 2017] et on choisit la représentation géodésique quasi-régulière, permettant un échantillonnage de l'image sphérique presque aussi régulier que celui d'une image perspective (plane) [Li and Hai, 2010].

Ainsi, une image sphérique est représentée par un ensemble de points (pixels)

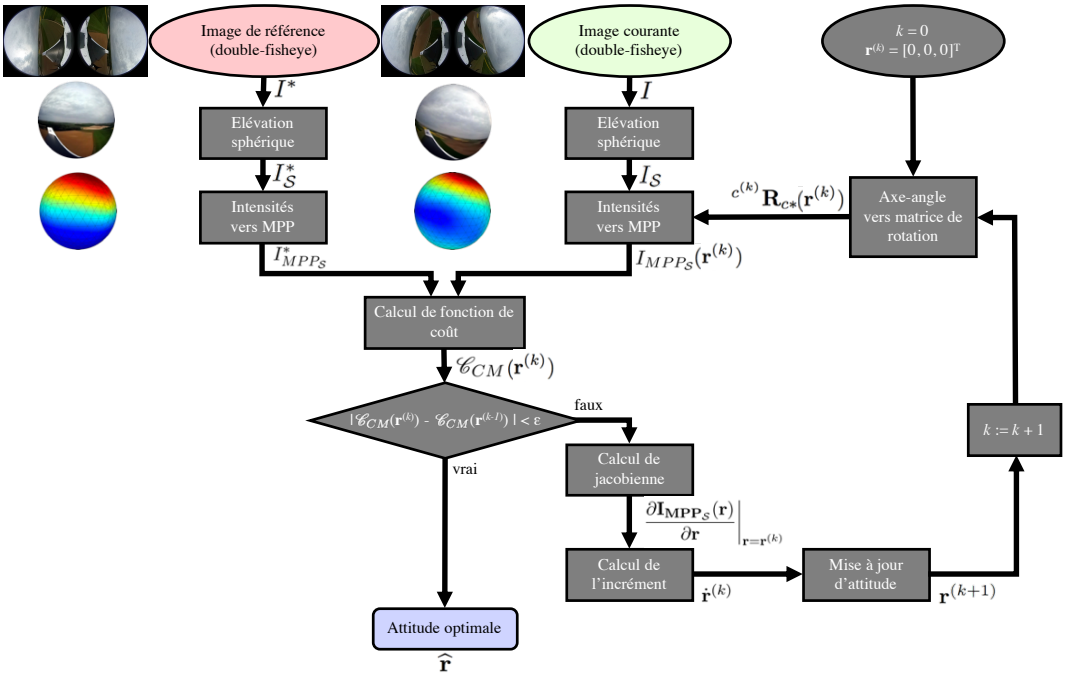


FIGURE 6.7 – Synoptique de l'algorithme d'estimation de rotations entre une image sphérique *requête* et une image sphérique de référence.

répartis uniformément, au sens géodésique, sur la surface de la sphère. L'ensemble discret $\mathcal{X}_{\mathcal{S}}$ de points sphériques ${}^c\mathbf{X}_{\mathcal{S}i} \in \mathcal{X}_{\mathcal{S}}$, $i \in \{1, 2, \ldots, |\mathcal{X}_{\mathcal{S}}|\}$ est uniformément réparti sur la surface de la sphère quand il est construit à partir d'un icosaèdre convexe régulier (un polyèdre de vingt triangles équilatéraux et douze sommets), dont les faces subissent N_{sub} subdivisions, N_{sub} étant proportionnel à la définition désirée de l'image sphérique [Li and Hai, 2010]. A noter que les sommets d'un icosaèdre convexe régulier sont régulièrement distribués sur la sphère unitaire circonscrite au polyèdre : cette propriété reste valide pour un icosaèdre dont les faces ont été divisées récursivement. Pour N_{sub} niveaux de subdivision, le polyèdre correspondant possède $|\mathcal{X}_{\mathcal{S}}| = \frac{1}{2}(20 \times 4^N_{sub}) + 2$ sommets et $20 \times 4^N_{sub}$ facettes.

Le mélange de potentiels photométriques sphériques s'écrit à partir des intensités $I_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}i}) = I(\mathbf{u}({}^c\mathbf{X}_{\mathcal{S}i}))$ [6] de chaque point ${}^c\mathbf{X}_{\mathcal{S}i}$ du polyèdre échantillonnant l'image sphérique, tel que l'ensemble $\overline{\mathcal{I}}_{\mathcal{S}}$ rassemble toutes les intensités sphériques normalisées (voir ci-dessous) :

$$I_{MPP_{\mathcal{S}}}(\overline{\mathcal{I}}_{\mathcal{S}}, {}^c\mathbf{X}_{\mathcal{S}_p}, \lambda_p) = \sum_{{}^c\mathbf{X}_{\mathcal{S}} \in \mathcal{X}_{\mathcal{S}}} I_{PP_{\mathcal{S}}}(\overline{\mathcal{I}}_{\mathcal{S}}, {}^c\mathbf{X}_{\mathcal{S}_p}, {}^c\mathbf{X}_{\mathcal{S}}, \lambda_p) \tag{6.23}$$

avec :

$$I_{PP_{\mathcal{S}}}(\overline{\mathcal{I}}_{\mathcal{S}}, {}^c\mathbf{X}_{\mathcal{S}_p}, {}^c\mathbf{X}_{\mathcal{S}}, \lambda_p) = \overline{I}_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}}) f_p(\arccos({}^c\mathbf{X}_{\mathcal{S}_p}^\top {}^c\mathbf{X}_{\mathcal{S}}), \lambda_p), \tag{6.24}$$

6. en négligeant de faire apparaître l'interpolation de $I()$ aux coordonnées réelles

et la fonction de potentiel d'attraction $f_p()$ définie comme en équation (6.6) ($n = 1$) en dimension 1 par l'introduction de la distance géodésique $\arccos({}^c\mathbf{X}_{\mathcal{S}_p}^\top {}^c\mathbf{X}_{\mathcal{S}})$ entre ${}^c\mathbf{X}_{\mathcal{S}_p}$ et ${}^c\mathbf{X}_{\mathcal{S}}$, sur la sphère unitaire $\mathcal{S}^2$. Dans l'équation (6.23), $\overline{I}_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}}) \in \mathbb{R}$ décrit l'intensité normalisée de l'image sphérique au point ${}^c\mathbf{X}_{\mathcal{S}}$ tel que $\sum_{{}^c\mathbf{X}_{\mathcal{S}} \in \mathcal{X}_{\mathcal{S}}} \overline{I}_{\mathcal{S}}({}^c\mathbf{X}_{\mathcal{S}}) = 1$. Cette dernière contrainte fait que l'équation (6.23) du mélange de potentiels photométriques est un mélange de gaussiennes (photométriques) normales, au sens strict de la définition [McLachlan and Peel, 2000, Chap. 3].

Remarque 8 (Relation entre N_{sub} et la définition de l'image numérique) *Sur un exemple de caméra polydioptrique compacte à deux objectifs panoramiques, l'image numérique obtenue à travers chaque objectif serait définie par $\frac{|\mathcal{X}_{\mathcal{S}}|}{2}$ pixels (pour un N_{sub} donné), de densité variable dans le plan image, puisque de densité quasi-régulière sur la sphère unitaire. En faisant abstraction de cette différence de densité, le rectangle image numérique serait, dans ce cas, un carré de $\sqrt{\frac{|\mathcal{X}_{\mathcal{S}}|}{2}} = \sqrt{\frac{1}{2}\left(\frac{1}{2}(20 \times 4_{sub}^N) + 2\right)}$ pixels de côté. Selon cette dernière formule, $N_{sub} = 2$, engendre deux images panoramiques de 9 pixels de côté, puis environ 18 pixels de côté pour $N_{sub} = 3$, 36 pour $N_{sub} = 4$, 72 pour $N_{sub} = 5$, etc.* ◇

6.4.3 Estimation directe de rotations

L'estimation directe de rotations cherche à calculer le vecteur $\mathbf{r} \in \mathfrak{so}(3)$, par abus de notation (cf. Partie 2.2), représentation axe-angle de la matrice ${}^c\mathbf{R}_{c^*} \in SO(3)$ de rotation inconnue entre l'image de référence $I_{\mathcal{S}}^*$ et l'image courante $I_{\mathcal{S}}$, par le biais de leurs transformées en mélanges de potentiels photométriques $I_{MPP_{\mathcal{S}}}^*$ et $I_{MPP_{\mathcal{S}}}$, respectivement. On ré-exprime donc l'intensité de mélange de potentiels photométriques sphériques (Eq. (6.23)) en fonction des points sphériques ${}^{c^*}\mathbf{X}_{\mathcal{S}} \in \mathcal{X}_{\mathcal{S}}^*$ de l'image de référence et des degrés de liberté $\mathbf{r}$:

$$I_{MPP_{\mathcal{S}}}\left(\overline{\mathcal{I}}_{\mathcal{S}}, \mathbf{r}, {}^{c^*}\mathbf{X}_{\mathcal{S}_p}, \lambda_p\right) = \sum_{{}^{c^*}\mathbf{X}_{\mathcal{S}} \in \mathcal{X}_{\mathcal{S}}^*} I_{PP_{\mathcal{S}}}(\overline{\mathcal{I}}_{\mathcal{S}}, \mathbf{r}, {}^{c^*}\mathbf{X}_{\mathcal{S}_p}, {}^{c^*}\mathbf{X}_{\mathcal{S}}, \lambda_p), \tag{6.25}$$

avec :

$$\begin{aligned} &I_{PP_{\mathcal{S}}}(\overline{\mathcal{I}}_{\mathcal{S}}, \mathbf{r}, {}^{c^*}\mathbf{X}_{\mathcal{S}_p}, {}^{c^*}\mathbf{X}_{\mathcal{S}}, \lambda_p) \\ &\quad = \overline{I}_{\mathcal{S}}({}^c\mathbf{R}_{c^*}(\mathbf{r}){}^{c^*}\mathbf{X}_{\mathcal{S}}) f_p(\arccos([{}^c\mathbf{R}_{c^*}(\mathbf{r}){}^{c^*}\mathbf{X}_{\mathcal{S}_p}]^\top [{}^c\mathbf{R}_{c^*}(\mathbf{r}){}^{c^*}\mathbf{X}_{\mathcal{S}}]), \lambda_p). \end{aligned} \tag{6.26}$$

Ensuite, on exprime le coût $\mathscr{C}_{CM}()$ de façon similaire au coût $\mathscr{C}_{A+}()$ (Eq. (6.7)) par :

$$\begin{aligned} &\mathscr{C}_{CM}(\mathbf{r}, \overline{\mathcal{I}}_{\mathcal{S}}, \{\mathcal{X}_{\mathcal{S}}^*, \lambda_p\}, \overline{\mathcal{I}}_{\mathcal{S}}^*) \\ &\quad = \tfrac{1}{2} \sum_{{}^{c^*}\mathbf{X}_{\mathcal{S}_p} \in \mathcal{X}_{\mathcal{S}}^*} \left(I_{MPP_{\mathcal{S}}}(\overline{\mathcal{I}}_{\mathcal{S}}, \mathbf{r}, {}^{c^*}\mathbf{X}_{\mathcal{S}_p}, \lambda_p) - I_{MPP_{\mathcal{S}}}^*(\overline{\mathcal{I}}_{\mathcal{S}}^*, {}^{c^*}\mathbf{X}_{\mathcal{S}_p}, \lambda_p)\right)^2 \\ &\quad = \tfrac{1}{2} ||\mathbf{I}_{\mathbf{MPP}_{\mathcal{S}}}(\mathbf{r}) - {\mathbf{I}_{\mathbf{MPP}_{\mathcal{S}}}}^*||^2 \\ &\quad = \tfrac{1}{2} ||\mathbf{C}_{\mathbf{CM}}(\mathbf{r})||^2, \end{aligned} \tag{6.27}$$

à minimiser itérativement par la méthode de Gauss-Newton, calculant les incréments $\dot{\mathbf{r}}^{(k)}$ de rotation, servant à mettre à jour $\mathbf{r}^{(k)}$ par approche compositionnelle (Partie 3.2.3), pour obtenir la rotation optimale $\widehat{\mathbf{r}}$. La différence majeure avec les autres méthodes basées mélanges de potentiels photométriques de ce chapitre, c'est qu'en l'état actuel du travail, λ_p n'est pas inclus dans les degrés de liberté du problème, essentiellement pour des raisons de coût calculatoire du mélange de potentiels sphériques par rapport au plan image en vision perspective. $I_{MPP_{\mathcal{S}}}()$ et $I^*_{MPP_{\mathcal{S}}}()$ ont donc le même λ_p, dont la partie expérimentale montera quel intervalle de valeurs il peut prendre pour atteindre les mêmes performances.

De façon similaire à tout ce qui précède, calculer $\widehat{\mathbf{r}}$ par la méthode de Gauss-Newton passe par le calcul de la matrice jacobienne du coût $\mathscr{C}_{CM}()$, exprimée par :

$$\begin{aligned}
\mathbf{J}_{\mathbf{r}}^{(k)} &= \left.\frac{\partial \mathbf{I}_{\mathbf{MPP}_{\mathcal{S}}}(\mathbf{r})}{\partial \mathbf{r}}\right|_{\mathbf{r}=\mathbf{r}^{(k)}} \\
&= \left.\begin{bmatrix} \vdots \\ \left(\nabla_{\mathbf{X}_{\mathcal{S}_p}} I_{MPP_{\mathcal{S}}} \dfrac{\partial \mathbf{X}_{\mathcal{S}_p}}{\partial \mathbf{r}}\right)\Big|_{\mathbf{X}_{\mathcal{S}_p}=\mathbf{X}_{\mathcal{S}_{p_i}}} \\ \vdots \end{bmatrix}\right|_{\mathbf{r}=\mathbf{r}^{(k)}} .
\end{aligned} \tag{6.28}$$

En supposant à nouveau que c'est directement le mélange de potentiels photométriques (sphériques) qui subit la rotation pure et en posant $D = [{}^c\mathbf{R}_{c^*}(\mathbf{r})\,{}^{c^*}\mathbf{X}_{\mathcal{S}_p}]^\top [{}^c\mathbf{R}_{c^*}(\mathbf{r})\,{}^{c^*}\mathbf{X}_{\mathcal{S}}]$, on exprime :

$$\begin{aligned}
\nabla_{\mathbf{X}_{\mathcal{S}_p}} I_{MPP_{\mathcal{S}}} &= \sum_{{}^{c^*}\mathbf{X}_{\mathcal{S}} \in \mathcal{X}^*_{\mathcal{S}}} \overline{I}_{\mathcal{S}}({}^c\mathbf{R}_{c^*}(\mathbf{r})\,{}^{c^*}\mathbf{X}_{\mathcal{S}}) \frac{\partial f_p(\arccos([{}^c\mathbf{R}_{c^*}(\mathbf{r})\,{}^{c^*}\mathbf{X}_{\mathcal{S}_p}]^\top [{}^c\mathbf{R}_{c^*}(\mathbf{r})\,{}^{c^*}\mathbf{X}_{\mathcal{S}}]),\lambda_p)}{\partial \mathbf{X}_{\mathcal{S}_p}} \\
&= \sum_{{}^{c^*}\mathbf{X}_{\mathcal{S}} \in \mathcal{X}^*_{\mathcal{S}}} \frac{\arccos(D)}{\lambda_p^2} \frac{{}^c\mathbf{R}_{c^*}(\mathbf{r})\,{}^{c^*}\mathbf{X}_{\mathcal{S}}}{\sqrt{1-D^2}} I_{PP_{\mathcal{S}}}(\overline{\mathcal{I}}_{\mathcal{S}}, \mathbf{r}, {}^{c^*}\mathbf{X}_{\mathcal{S}_p}, {}^{c^*}\mathbf{X}_{\mathcal{S}}, \lambda_p)
\end{aligned} \tag{6.29}$$

et :

$$\frac{\partial \mathbf{X}_{\mathcal{S}_p}}{\partial \mathbf{r}} = \left[\mathbf{X}_{\mathcal{S}_p}\right]_\times . \tag{6.30}$$

Tout comme la matrice jacobienne de l'asservissement visuel basé mélange de gaussiennes photométriques, exprimée sous l'approximation que c'est directement le mélange qui subit le déplacement de la caméra, aucun gradient des images sphériques $I_{\mathcal{S}}()$ ni $I^*_{\mathcal{S}}()$ n'apparait dans les équations ci-dessus. Cette caractéristique apporte encore plus d'avantages dans le cas sphérique, où, pour le calcul des gradients de l'image par approximation aux différences finies, les voisinages des pixels sphériques sont plus complexes que dans le cas perspective, menant à un accroissement significatif des temps de calcul.

6.4.4 Résultats

6.4.4.1 Quels paramètres pour un vaste domaine de convergence ?

Dans le gyroscope visuel basé mélanges de potentiels photométriques, deux paramètres sont essentiels, à savoir N_{sub}, le nombre de niveaux de subdivisions du polyèdre *maillant* l'image sphérique, concrètement la définition de l'image sphérique, et λ_p, en particulier car il n'est pas optimisé, contrairement au reste des méthodes du chapitre.

Pour déterminer la meilleure paire de valeurs pour N_{sub} et λ_p, une étude expérimentale précise et systématique a été menée en montant la caméra polydioptrique sphérique Ricoh Theta S sur l'effecteur d'un bras robotique Stäubli TX-60 (Fig. 6.8). L'étalonnage poignet-caméra est réalisé avec précision afin de pouvoir appliquer des rotations pures avec un pas fixé autour du centre optique de la caméra pour avoir des jeux d'images avec une vérité terrain très précise.

En considérant la rotation pure autour de l'axe $\mathbf{X}_c$ de la caméra sphérique, de -180^o à $+180^o$, avec un pas de $2,5^o$, on peut évaluer un échantillonnage de la fonction de coût $\mathscr{C}_{CM}()$ (Eq. (6.27)) à partir d'images acquises, pour différentes valeurs de N_{sub} et de λ_p. D'autre part, en analysant les variations de la fonction *discrète* de coût ainsi obtenue, on peut déduire les bornes du domaine de convergence $\mathscr{C}_{CM}()$

On observe tout d'abord que pour N_{sub} et λ_p faibles, par exemple $N_{sub} = 2$ et $\lambda_p = 0,01$, l'allure de la fonction de coût est la même qu'avec le coût direct

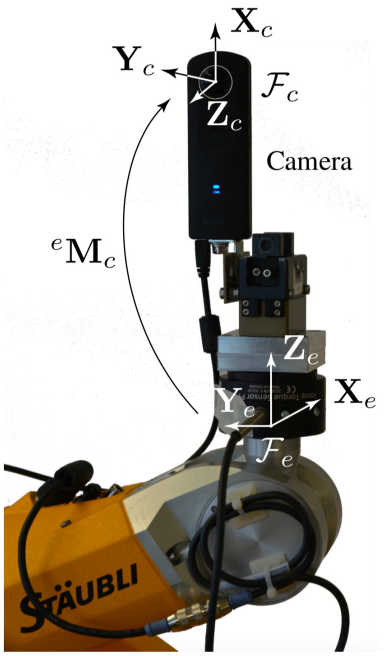
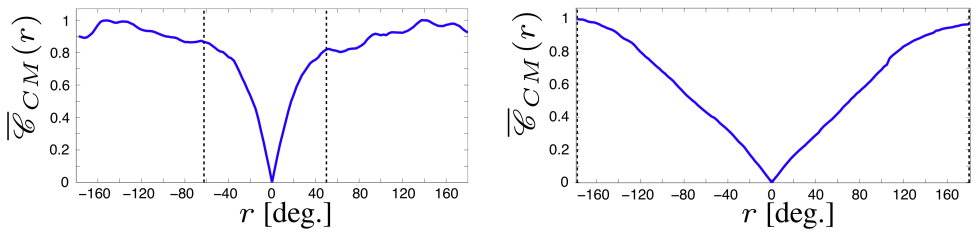


FIGURE 6.8 – Caméra Ricoh Theta montée sur l'effecteur du robot Stäubli TX-60. ${}^e\mathbf{M}_c$: transformation rigide entre le repère caméra $\mathcal{F}_c$ et celui de l'effecteur, $\mathcal{F}_e$.

pur de type SSD (Fig. 6.9(a)). D'autre part, on confirme à nouveau qu'augmenter λ_g agrandit considérablement le domaine de convergence, jusqu'à atteindre un domaine de convergence global, et même avec une pente quasi-linéaire (Fig. 6.9(b)) quand le coût $\mathscr{C}_{CM}()$ intègre un estimateur robuste[7], reposant sur une loi de Huber [Stewart, 1999]. En systématisant l'évaluation de l'étendue du domaine de convergence du coût $\mathscr{C}_{CM}()$ pour $N_{sub} = \{1; 2; \ldots; 6\}$ et $\lambda_p = \{0,075; 0,1; \ldots; 1\}$, on détermine que le domaine de convergence est global uniquement en utilisant l'estimateur robuste pour $\lambda_p \in [0,2; 0,4]$ et $N_{sub} > 3$. Quand $N_{sub} \leq 3$, l'estimateur robuste perturbe les fonctions de coût au point de réduire le domaine de convergence. Sans estimateur robuste, le domaine de convergence reste très vaste puisque supérieur à 300°, pour $N_{sub} > 2$ quand $\lambda_p > 0,25$, là où le coût direct pur ne dépasse pas les 125°. Par conséquent, pour l'algorithme du gyroscope visuel, il est aisé d'évaluer le coût pour deux rotations, par exemple $\mathbf{r}^{(0)} = \mathbf{0}$ et $\mathbf{r}^{(0)\prime} = [180^o\ 0\ 0]$, candidates à l'initialisation de l'optimisation, pour sélectionner une initialisation à 90° près, suffisamment bonne pour converger avec ou sans estimateur robuste. En effet, négliger l'estimateur robuste reste intéressant pour limiter les temps de calculs puisque celui-ci requiert au moins $N_{sub} = 4$ alors que sans estimateur robuste $N_{sub} = 3$ suffit et conduit à un domaine de convergence plus vaste qu'avec estimateur robuste. Les temps de calcul croissant de façon exponentielle en N_{sub} (directement lié au nombre de sommets du polyèdre de l'image sphérique, Partie 6.4.2), ce n'est pas négligeable. Concrètement, considérer $N_{sub} = 4$ demande, en moyenne, $3,92s$ de temps de calcul par image alors que $N_{sub} = 3$ ne demande que $0,38s$ par image, beaucoup plus proche du temps réel, surtout que l'étendue du domaine de convergence tolère des mouvements importants entre images.

Dans tous les cas, même si l'étendue du domaine de convergence change en fonction des paramètres N_{sub} et λ_p, le minimum du coût, lui, ne change pas de position, ce qui permet de justifier de ne pas placer λ_p dans les degrés de liberté du problème, contrairement au suivi de plan ou à l'asservissement visuels basés

7. Les détails sont laissés de côté mais la modélisation est la même qu'en partie 3.3.4.1.



(a) (b)

FIGURE 6.9 – Fonctions de coût normalisées (pour faciliter la comparaison qualitative) $\overline{\mathscr{C}}_{CM}(r)$ (Eq. (6.27)) pour un seul degré de liberté de rotation r (tirets noirs verticaux : limites du domaine de convergence) : (a) $N = 2$, $\lambda_g = 0,01$, sans M-estimateur ; (b) $N = 5$, $\lambda_g = 0,3$ avec M-estimateur.

mélanges de potentiels photométriques. Certes, cette observation n'est faite que pour un degré de liberté de rotation pure. On peut cependant penser que dans la mesure où la rotation pure d'une image sphérique ne change pas son contenu, pour une définition suffisante, on pourrait faire la même observation pour les 3 angles de rotation, même si le montage expérimental est plus complexe à mettre en oeuvre. Par contre, dès qu'une translation de la caméra apparaît, il est clair que l'on se retrouve dans le même cas que le suivi ou l'asservissement visuels où l'image subit une transformation projective, engendrant parfois un changement de la position du minimum quand λ_p est grand, justifiant à nouveau de l'optimiser.

6.4.4.2 Evaluation du gyroscope visuel

Après les observations de la partie précédente (Partie 6.4.4.1) permettant de fixer $\lambda_p = 0,325$ l'algorithme d'estimation de rotation est appliqué entre l'image acquise à $\mathbf{r} = \mathbf{0}_3$ et toutes les autres, sans estimation incrémentale. Pour $N_{sub} = 2$, le temps de calcul est, certes, dérisoire, mais les estimations sont si mauvaises qu'elles n'ont pas de sens. Par contre, dès $N_{sub} = 3$, l'erreur moyenne d'estimation est inférieure à 4^o et elle descend sous les 3^o pour $N_{sub} = 4$ et est légèrement supérieure à 1^o pour $N_{sub} = 5$.

Ces dernières erreurs moyennes augmentent légèrement quand la caméra subit une rotation autour des trois axes pour atteindre $7,5^o$ pour $N_{sub} = 3$, $4,2^o$ pour $N_{sub} = 4$ et $3,7^o$ pour $N_{sub} = 5$ à partir de 94 images.

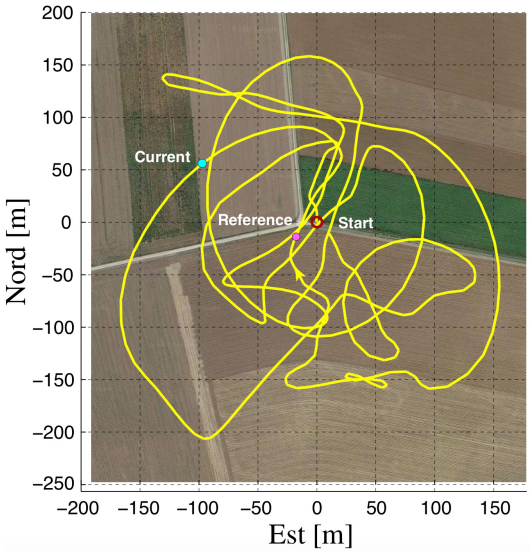
On pourrait penser qu'il en va de même en présence de translations, puisque le coût $\mathscr{C}_{CM}()$ ne les prend pas en compte dans sa formulation. Néanmoins, l'erreur commise par le gyroscope visuel, réduit à un estimateur de cap, n'augmente pas pour $N_{sub} = \{3, 4\}$, même pour des écarts en position de 40 cm (distance maximum pour le robot dans sa configuration pour cette expérimentation) dans un environnement de $10,05\ m \times 7,03\ m \times 2,70\ m$. Pour $N_{sub} = 5$, l'erreur d'estimation n'augmente pas jusqu'à une distance de 24 cm, puis elle augmente significativement de $1,1^o$ à $3,5^o$.

Pour terminer les expérimentations, la caméra Ricoh Theta a été fixée à un drone Parrot Disco (Fig. 6.10(b)) et a enregistré un fichier vidéo au format MP4 à 30 images par secondes. Le Disco est un drone d'envergure d'1,15 m, à propulsion mono-rotor, alimenté par batterie, pour une masse de 750 g.

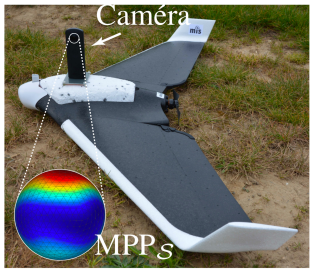
La position surélevée de la caméra à l'avant du drone permet de percevoir l'environnement en minimisant les occultations dues au drone et sans trop perturber son centre de gravité. Le Disco a été piloté au-dessus d'un champ proche de la ville d'Amiens pour un vol de 6'41" durant lequel il a parcouru 4,2 km, atteint l'altitude de 106 m et une vitesse au sol maximale de 85,9 km/h (Fig. 6.10(a)).

Contrairement au robot industriel, il n'y a pas de vérité terrain fiable pour cette expérimentation. En effet, les données de la centrale inertielle embarquée dans la Ricoh Theta se sont révélées trop imprécises pour une étude quantitative. Pour faire une évaluation quantitative de la méthode, l'image de référence a été choisie à 4'41" de la vidéo, instant où le drone est à la fois horizontal et central dans la zone par-

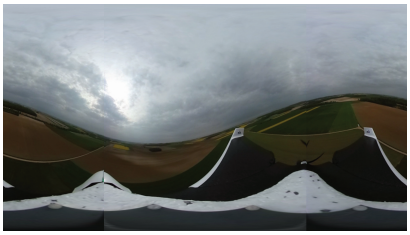
courue durant le vol. La Figure 6.10(c) montre une représentation équirectangulaire de cette image de référence à pleine définition pour en faciliter la visualisation. La Figure 6.10(d) montre une image courante (à 5'11" de la vidéo) et la Figure 6.10(e) la même image mais dont la rotation estimée, pour $N_{sub} = 3$, par rapport à l'image



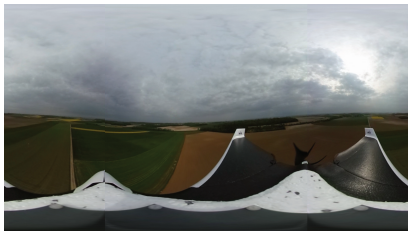
(a)



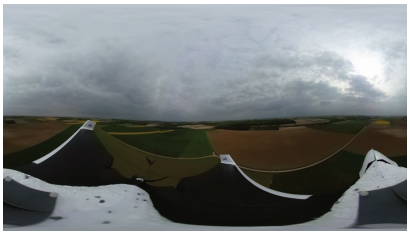
(b)



(d)



(c)



(e)

FIGURE 6.10 – (a) Trajectoire GPS du drone Disco (repère géodésique local) ; (b) Ricoh Theta embarquée sur le drone ; Représentations équirectangulaires des images (c) de référence (cf. le point rose en (a)), (d) courante sans correction (cf. le point bleu en (a)) et (e) courante après correction en utilisant l'orientation estimée par le gyroscope visuel.

de référence, a été compensée. Cette compensation est satisfaisante puisque les Figures 6.10(e) et 6.10(c) ont un aspect très similaire de ligne d'horizon et de position du soleil et ce, malgré la distance entre l'image courante et celle de référence de 108 m, pour une altitude respective de 58,3 m et 50,1 m. Un niveau de compensation comparable est observé dans 88% des images de la séquence de 6'41".

Les échecs (12% des images) sont attribuables à des minimums locaux dans la fonction de coût $\mathscr{C}_{CM}()$ (Eq. (6.27)) en présence de grandes translations.

Il est à noter que ces résultats ont été obtenus pour un drone évoluant en environnement naturel non structuré avec une initialisation et un paramétrage faits exactement comme pour les expérimentations avec le robot industriel où un environnement structuré intérieur est considéré.

Résultats en vidéo : `https://youtu.be/5WmEzkxVsMU`.

6.4.5 Conclusion partielle

L'adaptation du mélange de potentiels photométriques à la représentation sphérique complète d'image et sa considération dans un algorithme direct d'estimation de rotation entre deux mélanges de potentiels photométriques sphériques montrent à nouveau à quel point ils engendrent un vaste domaine de convergence. De plus, cela est montré dans un cadre encore différent du suivi visuel de région et de l'asservissement visuel, pour un type de caméra au champ de vue sans limite. L'absence de limite du champ de vue offre la possibilité de percevoir toujours tout l'environnement, même pour des rotations très importantes. Pour un degré de liberté de rotation, tel que le cap, le mélange de potentiels photométriques peut engendrer un domaine de convergence global. Avec les deux degrés de liberté de rotation supplémentaires, le domaine est au moins très vaste, comme le montrent les multiples expérimentations sur robot industriel et drone.

Depuis ces travaux, les moments photométriques sphériques ont été formulés pour l'estimation de rotation entre images, sans avoir besoin d'itérer [Hadj-Abdelkader et al., 2018]. Cette dernière approche apparaît donc directe à la fois en terme de lien entre les intensités des pixels des images acquises et les degrés de liberté du problème d'estimation de rotation, et en terme de méthode de résolution de problème. N'ayant pas été considérée dans le cas d'une caméra sphérique complète, il reste à montrer que le domaine de convergence peut être global.

Les réseaux de neurones convolutionnels ont aussi été adaptés aux images équirectangulaires représentant les images sphériques pour estimer directement l'orientation de la caméra à partir du flot optique. Il faut donc calculer ce flot optique en amont, à partir des intensités, ce qui limite l'approche d'estimation d'orientation à des rotations de faible amplitude [Kim et al., 2019] et l'éloigne des approches les plus directes sur lesquelles ce mémoire se concentre.

Pour poursuivre les travaux sur le gyroscope visuel sphérique basé mélanges de potentiels photométriques, optimiser λ_p comme dans le suivi et l'asservissement visuels est une évidence pour améliorer la précision à convergence. Cela n'a pas été

fait car considérer λ_p constant a permis de réduire énormément les temps de calcul à chaque itération pour s'approcher du temps réel. Cependant, en sélectionnant le niveau de subdivision N_{sub} du polyèdre support de l'image sphérique selon λ_p, moins de pixels sphériques seraient considérés pour les valeurs les plus importantes de λ_p, s'approchant ainsi d'une vrai pyramide d'échelle continue, intégrant le changement de définition de l'image I_{MPP_S} et pas uniquement son "lissage" des méthodes précédemment proposées. Dans le cas des images sphériques, cette notion de pyramide d'échelle *pleine* deviendrait une boule d'échelle.

6.5 Conclusion du chapitre

Revisiter la représentation multi-échelle pyramidale d'image sous le prisme du mélange de potentiels photométriques permet d'accroître considérablement le domaine et les taux de convergence du suivi projectif visuel direct de région, de l'asservissement visuel direct et de l'estimation de mouvement de caméra en considérant l'intégralité de chaque image qu'elle acquiert. Le domaine de convergence va jusqu'à se montrer *global* en rotation pure selon un axe de caméra sphérique. Les taux de convergence sont *doublés* en suivi et en asservissement visuel par rapport aux méthodes les plus performantes de l'état de l'art.

Clairement, cette amélioration significative des performances des algorithmes mentionnés ci-dessus est due au fait de ne plus considérer un nombre arbitraire, entier, de niveaux échantillonnant, grossièrement, l'espace d'échelle mais un coefficient, réel, d'échelle adaptatif. Ce coefficient, ajouté aux degrés de liberté du suivi projectif et de l'asservissement visuels, paramètre l'étendue d'attraction de chaque pixel. Il varie au fil des itérations de la boucle d'optimisation afin d'atteindre une échelle fine à partir d'une échelle grossière de façon non-linéaire, quasi-continue, et pas forcément monotone, selon le contenu des images.

C'est une formalisation élégante de l'évolution du coefficient d'échelle au sein d'un unique problème d'optimisation de degrés de liberté géométriques de mouvement. Elle évite la séquence habituelle de résolutions de problèmes d'optimisation dans des niveaux d'échelle successifs discontinus.

En première perspective, il reste à finaliser la généralisation des approches multi-échelles pyramidales en adaptant aussi la résolution des images, de façon continue au fil des itérations, simultanément au coefficient d'échelle. Le maillage multi-densité de l'image sphérique, évoqué en partie 6.4.5, est une piste en direction de cette perspective de travail. Ce concept rapidement évoqué, est adaptable à la vision perspective, ce qui apportera de nouveaux éléments de réflexion pour améliorer l'approche afin de faire tendre les trajectoires des caméras vers le chemin le plus court.

Une seconde perspective s'entrevoit en appliquant la modélisation à base de mélanges de potentiels visuels directs à d'autres informations denses issues des images, comme les nombreuses extensions des approches directes pures l'ont fait dans l'état de l'art (Partie 3).

Des travaux préliminaires ont été menés en ce sens, en exploitant la saillance visuelle [Itti et al., 1998, Achanta et al., 2009] de chaque pixel des images pour la commande de déplacement de caméra par un robot *[Habibi et al., 2015]* ou en environnement virtuel *[Habibi et al., 2015]*. Ces travaux, menés dans le cadre de la thèse de Zaynab Habibi [8] permettent l'exploration d'environnement et le cadrage, au sens cinématographique, d'objet d'intérêt de façon implicite, c'est-à-dire à partir de données brutes, sans sémantique, à la manière des travaux d'exploration de maquette virtuelle rapportés en partie 4.2.

Enfin, appliquer la transformée d'image de champs de descripteurs (Partie 3.5.2) en mélanges de potentiels *invariants* laisse entrevoir la possibilité d'atteindre une grande robustesse aux perturbations visuelles au sein d'une méthode directe étendue de suivi, d'estimation de pose et d'asservissements visuels à vaste domaine de convergence.

8. Pour rappel, la thèse de Zaynab Habibi, que j'ai co-encadrée, a été faite au laboratoire MIS de l'UPJV, dans le cadre du projet CR Picardie "Assiduitas", qui s'inscrit dans le programme e-Cathédrale du laboratoire.

Chapitre 7

Bilan et perspectives

Sommaire

7.1 Bilan 165
7.2 Projet de recherche 167
7.2.1 Vue d'ensemble 167
7.2.2 A court terme 167
7.2.3 A moyen terme 168
7.2.4 A long terme 169
7.2.5 Positionnement du projet scientifique 170
7.2.5.1 Contribution du projet à l'équipe PR 170
7.2.5.2 Situation du projet à l'international 171

7.1 Bilan

Après un rappel d'éléments de modélisation de la formation des images, ce mémoire s'est attaché à ré-écrire une part non négligeable de l'état de l'art en vision robotique directe, pure et étendue, sous un formalisme unique, qui a été conservé tout au long des chapitres synthétisant les contributions pour plus facilement faire apparaître leurs spécificités.

En prenant un peu de recul sur l'état de l'art, on observe que la vision robotique directe s'est construite sur la base de travaux préliminaires proposés au début des années 1980 et que des recherches de moins en moins ponctuelles au fil du temps ont apporté des contributions significatives, à la fois en terme de nombre de degrés de liberté directement optimisés ou contrôlés à partir des intensités seules des images, et en terme de critères visuels directs pour la robustesse à de nombreuses perturbations. Tous ces travaux se sont faits en repoussant le besoin apparent d'extraire et mettre en correspondance des primitives géométriques.

Le coeur des contributions scientifiques rassemblées dans ce mémoire s'est attaché à identifier et repousser les limites de l'information visuelle directe pour le suivi visuel et le calcul de pose basés maquette virtuelle 3D texturée, le suivi projectif de région d'image, l'estimation d'orientation de caméra et l'asservissement visuel.

Le chapitre 4, en étendant l'asservissement visuel basé information mutuelle au calcul de pose et au suivi visuels de maquette virtuelle 3D urbaine pour la localisation de véhicule a permis de confirmer la robustesse du critère d'information mutuelle aux changements d'illumination, occultations partielles et à la nature différente des

images acquises et des images de synthèse. Néanmoins, le réalisme des maquettes virtuelles 3D à base de maillage simplifié, sur lequel des photos sont plaquées pour le texturer, reste un frein au succès de la méthode à grande échelle.

Le chapitre 5, en proposant l'alternative d'exploiter des nuages de points 3D colorés issus de scans Lidar denses en guise de maquette virtuelle, a permis de ré-écrire le problème de suivi et de calcul de pose visuels à partir des intensités des pixels, directement. La caméra perspective et, d'autre part, la caméra panoramique centrale ont été considérées dans la formulation du problème. Le domaine de convergence un peu plus vaste qu'avec l'information mutuelle et la précision à convergence dans les conditions nominales pour l'information visuelle directe, ont permis d'améliorer l'apparence visuelle de la maquette en remplaçant ses couleurs par celles d'images de très haute qualité. Ces caractéristiques, sensiblement meilleures en vision panoramique, en terme de domaine de convergence, que précédemment, ont permis à un robot mobile de se localiser avec précision à partir des images que la caméra panoramique qu'il embarque acquiert. Dans des conditions expérimentales un peu plus difficiles que dans le chapitre 4, les échecs ponctuels de cette approche son imputables au domaine de convergence trop étroit malgré tout.

Ainsi, le chapitre 6, rassemblant les plus fortes contributions de ce mémoire, s'est attaché à revisiter élégamment la représentation multi-échelle pyramidale d'image dans le cadre d'une nouvelle information visuelle directe proposée : la transformée des intensités d'une image en mélange de potentiels photométriques. Cette dernière permet d'accroître considérablement le domaine et les taux de convergence en suivi projectif visuel direct de région, en asservissement visuel direct de bras robotique et en estimation de mouvement de caméra sphérique montée sur bras robotique aussi ou sur drone. Dans des cas particuliers, le domaine de convergence se montre même global et, dans tous les cas, les taux de convergence sont accrus de plus de 50% par rapport à l'état de l'art.

D'autre part, rapporté en quelques mots (Partie 6.5), le mélange de potentiels visuels peut être exploité autrement qu'avec les intensités brutes des images, par exemple avec la réponse dense d'un quantifieur de saillance appliqué à l'image. Maximiser la saillance visuelle dans l'image d'une caméra montée sur un bras robotique, permet de cadrer convenablement un objet ou une scène, voire l'explorer, suivant le même mode que la maximisation d'entropie en environnement virtuel du chapitre 4.

Enfin, d'autres travaux ont été menés (voir la Partie 1.1.6 pour leur description succincte et les références associées) en vision catadioptrique multimiroirs pour la reconstruction 3D, en vision plénoptique pour le suivi d'objet plan, en vision couleur et profondeur pour le suivi d'objet articulé ou encore en étalonnage de station Lidar en vue d'en exploiter la *réflectance*, même s'ils sont en dehors du paradigme de la vision robotique directe.

Avec ou sans ces travaux supplémentaires, les contributions inscrites dans ce mémoire pourraient suivre différentes trames de lecture comme les types de caméra, les types d'actionnement réels ou virtuels, les types de robot, les types d'informations de références, de la 2D à la 3D dense colorée, les thèses co-encadrées, etc. Cela

illustre, certes, la variété des activités et des travaux entrepris et réalisés mais c'est bien la trame de la recherche d'une meilleure information visuelle directe qui articule le mieux et le plus clairement mes principales contributions.

7.2 Projet de recherche

7.2.1 Vue d'ensemble

Mon projet de recherche s'inscrit sous le thème général de la vision robotique directe. Il s'appuie sur le bilan des travaux déjà réalisés (Partie 7.1).

Ce projet de recherche vise, à court terme, à continuer d'étudier la primitive directe du mélange de potentiels photométriques, notamment pour une définition adaptative d'image, d'autres types de visions, d'autres types de robots et plus de degrés de liberté. Un autre volet de la poursuite des recherches sur cette primitive consiste à étudier d'autres fonctions caractérisant l'attraction de chaque pixel.

A plus long terme, mon projet scientifique vise à proposer une nouvelle primitive visuelle directe qui ait à la fois un domaine de convergence très étendu, voire global, et qui soit robuste aux perturbations les plus difficiles à traiter comme la variation d'apparence d'une scène en fonction des saisons et des conditions météorologiques.

Enfin, j'envisage d'étendre la structuration de mes travaux de recherches à un axe supplémentaire aux trois introduits en partie 1.1.6 :

Axe 4 Conception de caméra : ajouter ou modifier des composants optiques des caméras et en créer de nouveaux.

La suite du projet de recherche s'organise en trois phases à court (Partie 7.2.2), moyen (Partie 7.2.3) et long (Partie 7.2.4) termes, c'est-à-dire des travaux à propos desquels la réflexion a déjà commencé aux travaux les plus prospectifs. Enfin, un positionnement de ce projet de recherche est proposé en regard des thématiques de recherche de l'équipe Perception Robotique (Partie 7.2.5.1), dont je suis membre au laboratoire MIS de l'UPJV, ainsi qu'une situation de ce projet à l'international (Partie 7.2.5.2).

7.2.2 A court terme

Dans un premier temps, l'objectif est d'étendre l'asservissement visuel basé mélange de potentiels photométriques à la vision omnidirectionnelle catadioptrique pour la commande de robot mobile [**Axe 3**]. Ces recherches serviront de base à la considération de la vision sphérique pour la commande de bras manipulateur et de robot mobile, puis de drone dans le futur. Ces commandes seront modélisées pour découpler les degrés de liberté de translation et de rotation afin de traiter, en partie, le problème de la non-holonomie des robots mobiles de type unicyle et assimilés, mais aussi pour autoriser des mouvements du bras permettant de s'éloigner des butées articulaires, tout en continuant à minimiser l'écart entre les mélanges de

potentiels photométriques courants et désirés. Ces travaux sont l'objet de la thèse de Seif Eddine Guerbas, que je co-encadre [**Doc.6**].

En parallèle, j'étudie l'asservissement visuel direct de drone [**Axe 3**]. Les travaux de thèse d'Eder Rodriguez, que je co-encadre [**Doc.4**], doivent montrer que malgré la dynamique d'un drone quadrirotors, l'asservissement visuel photométrique est possible pour stabiliser le drone en position et en cap. La limite de la primitive photométrique étant son domaine de convergence, il convient d'activer l'asservissement visuel photométrique dès que l'image désirée est acquise par la caméra embarquée afin que, malgré les perturbations en vol, le drone ne s'en éloigne pas trop, sous peine de sortir du domaine de convergence. L'élargissement du domaine de convergence apporté par la primitive du mélange de potentiels photométriques permettra l'alléger cette contrainte. Cependant, les temps de calcul devant être très courts pour la stabilisation du drone, l'ajout de la densité de l'échantillonnage du mélange comme degré de liberté supplémentaire à l'asservissement visuel évoqué en perspectives de la partie 6.5 est une piste sérieuse pour atteindre le temps-réel. De plus, pour la navigation autonome du drone, une nouvelle planification de chemin est en cours d'étude, considérant la planification d'un chemin visuel, constitué d'une mémoire d'images. Pour ce faire, on introduit une maquette virtuelle 2D ou 3D de la scène dans laquelle le drone va devoir naviguer. A partir de cette maquette, des vues de synthèse sont rendues pour définir, implicitement, des poses de passage au drone. Cette définition se fait sous les contraintes du modèle de mouvement du drone et du domaine de convergence de la primitive visuelle. On parle donc de planification de chemin visuel basée modèle (maquette 2D ou 3D).

7.2.3 A moyen terme

Dans un deuxième temps, sous le volet "vision non-conventionnelle" de mes activités de recherche, j'ai le projet d'exploiter les caractéristiques de la caméra plénoptique compacte pour poursuivre les efforts des recherches de l'état de l'art afin de rendre les suivis et asservissements visuels directs robustes aux matériaux très réfléchissants ou qui transmettent la lumière [**Axe 1**-**Axe 3**]. L'expérience récemment acquise pour le suivi d'objet plan, basé points d'intérêt *[Crombez et al., 2018]*, avec ce type de caméra servira de base à une reformulation du problème, directement à partir des intensités de cette dernière, qui échantillonne la fonction plénoptique par ses micro-lentilles. Un même point 3D étant perçu dans plusieurs pixels d'une image de caméra plénoptique avec des points de vue légèrement différents, l'idée serait de non pas considérer les intensités brutes (coût direct de type SSD, Partie 3.2.1), ni l'éclairement des photo-sites de la matrice photosensible de la caméra (Partie 3.3.4.2), obtenu grâce à un étalonnage photométrique de caméra, mais remonter à la luminance même du point 3D et à certaines propriétés de matière. Cela permettrait de s'affranchir des modifications de la lumière à chaque étape de son chemin de la scène au pixel de l'image. Pour atteindre cet objectif, cette recherche s'appuiera sur la littérature abondante d'estimation de la distribution de réflectance bidirectionnelle par caméra computationnelle (exemple récent : [Ngo et al., 2019]).

Ensuite, afin d'apporter plus de robustesse aux commandes ou estimateurs visuels directs, notamment aux conditions météorologiques, en particulier l'éclairage solaire, ses occultations partielles par les nuages ou des bâtiments, et les ombres qu'ils engendrent, je me propose de commencer à traiter le problème à la base, c'est-à-dire l'acquisition de l'image. En effet, plus le champ de vision est grand, plus les écarts d'illumination de parties de la scène observée peuvent être importants. La dynamique des matrices photosensibles étant encore limitée, des zones de l'image acquise sont sur-exposées et d'autres sous-exposées. Cela se vérifie en vision omnidirectionnelle (ex : Fig. 5.10(b) où les façades d'un côté de rue sont sur-exposées et celles de l'autre côté sont sous-exposées) et en vision sphérique, réduisant leur potentiel d'utilisation à l'extérieur. Par conséquent, mon projet de recherche s'étend à la conception de caméra (**Axe 4**), en combinant des filtres à densité neutre à une caméra catadioptrique de stéréovision omnidirectionnelle. Dès lors, le capteur de stéréovision n'est plus utilisé comme tel, mais comme un ensemble d'acquisitions simultanées d'images à différentes expositions. La combinaison de ces différentes images reste un problème à traiter à cause de leurs différents points de vue, contrairement à l'imagerie "classique" à large bande de dynamique, qui a cependant l'inconvénient de se cantonner aux scènes et caméras statiques. Néanmoins, considérer ce nouveau capteur de vision dans un asservissement visuel permettrait de le valider non seulement sans que les points de vue légèrement différents soient un problème, mais qu'ils deviennent un avantage [**Axe 3**].

7.2.4 A long terme

A plus long terme, je projette de lancer le nouveau champ de recherche de la vision omnidirectionnelle *adaptative* (**Axe 4**). Dans les travaux existants, la conception de capteur visuel catadioptrique se fait, généralement, en fonction d'une application visée. Cependant, dès lors que les conditions changent, ces capteurs ne sont plus optimaux. Pouvoir travailler avec des miroirs déformables permettrait de reconfigurer à volonté la caméra en déformant le miroir et ainsi obtenir une distribution de résolution adaptée au contenu de la scène. Ceci serait intéressant, par exemple pour suivre deux personnes diamétralement opposées par rapport à la caméra, tout en maximisant la surface de l'image où les personnes se trouvent pour plus de précision. Le principe est donc de déformer le miroir en fonction de l'image, ce qui pourrait être exprimé sous le formalisme de l'asservissement visuel. Ce genre d'idées dépend des technologies des miroirs déformables (miroirs mécaniques, ferrofluides magnétiques ou photosensibles ou encore tissus miroirs), dont les plus intéressantes ne semblent pas encore proposer des déformations assez importantes pour minimiser la surface inutile de l'image. Cependant ces miroirs déformables ne sont pas créés pour la vision artificielle en robotique mais, en général, pour l'observation astronomique où une faible déformation mais à une fréquence très importante est nécessaire. Il pourrait donc être intéressant de collaborer avec des équipes de spécialistes dans les miroirs déformables afin d'aboutir à un type de miroir convenable pour la vision artificielle en robotique. Une autre approche, pour commencer à étudier cette pro-

blématique, pourrait exploiter une grille de miroirs plans orientables afin de pouvoir réaliser une courbure plane par morceaux. C'est le contexte initial de la thèse de Julien Ducrocq [**Doc.7**] qui débute, et que je co-encadre.

Ensuite, de manière générale à l'ensemble des axes par lesquels je propose de structurer mes travaux de recherche, je projette de travailler à la définition de nouvelles primitives visuelles directes robustes et à vaste domaine de convergence. En effet, les primitives visuelles directes de l'état de l'art ont montré qu'il était soit possible d'avoir une grande robustesse, y compris à la modalité, mais au détriment du volume du domaine de convergence, soit possible d'avoir un domaine de convergence étendu, pour une robustesse moindre. Les résultats obtenus dans les travaux de recherche que j'ai mené abondent aussi dans ce sens. Si des estimateurs robustes peuvent être mis en place par dessus la primitive visuelle, leur impact reste limité, vis-à-vis des changements de luminosité et de la modalité. C'est pourquoi j'oriente mes recherches, à long terme, autour de nouvelles primitives visuelles directes qui soient à la fois robustes et dont le domaine de convergence d'un estimateur ou d'une loi de commande l'exploitant, soit très vaste, voire global. Pour ce faire, trois pistes sont envisagées. La première consiste à appliquer la transformée d'image de champs de descripteurs en mélanges de potentiels *invariants* (Partie 6.5), à mettre en entrée de l'optimisation des degrés de liberté du suivi, de l'estimation de pose ou de l'asservissement visuel. La deuxième consiste à étudier d'autres fonctions d'attraction dans les mélanges de potentiels photométriques que la fonction gaussienne. La troisième s'oriente vers l'étude d'autres méthodes de comparaisons de deux mélanges de potentiels photométriques, à l'instar de l'information mutuelle pour la primitive photométrique. Mes travaux gagneront probablement à combiner ces trois pistes.

Enfin, l'introduction de nouveaux capteurs de vision omnidirectionnelle adaptative permettra aussi de mettre à l'épreuve ces nouvelles primitives visuelles directes, qu'il conviendra probablement de revisiter pour les rendre utilisables pour les images acquises par ces futurs capteurs de vision.

7.2.5 Positionnement du projet scientifique

7.2.5.1 Contribution du projet à l'équipe PR

Les asservissements visuels et leurs variantes sont au coeur de mon projet scientifique. Au sein de l'équipe Perception Robotique (PR) du laboratoire MIS dans laquelle je travaille, je suis l'élément le plus actif sur ce thème, que j'ai largement contribué à introduire à mon arrivée dans l'équipe et développé depuis. En particulier, les recherches que je mène et que je compte mener à propos des primitives visuelles directes pour l'asservissement visuel sont assez complémentaires des travaux fondateurs de l'équipe Perception Robotique, plutôt orientés vers les primitives visuelles géométriques pour la localisation et la navigation de robot mobile basée vision omnidirectionnelle, avec une proximité moindre entre la mesure visuelle et la commande du robot que dans mes contributions. Ces dernières élargissent aussi le spectre robotique de l'équipe en considérant le bras manipulateur en plus du robot

mobile, mais aussi le drone, dont la commande de haut niveau sera faite par retour visuel, et le fauteuil roulant instrumenté pour les contributions naissantes.

Enfin, plutôt sur un volet méthodologique, je vais aussi développer la librairie logicielle de l'équipe Perception Robotique [**Log.2**], que j'ai récemment créé, en rassemblant et uniformisant les développements associés à mes contributions scientifiques, mais aussi y associer celles des autres membres de l'équipe. Cela permettra de pérenniser les travaux, mais aussi favoriser la reproductibilité de la recherche, aussi bien en interne que pour d'autres chercheurs, avec un partage public envisagé à terme. J'envisage aussi la question de la reproductibilité de la recherche, en systématisant le partage de jeux de données pertinents, et correctement formatés pour être réutilisables, pour chaque contribution scientifique, validant ainsi les quelques expériences récentes [**Log.3**, **Log.4**, **Log.5**, **Log.6**] que j'ai initiées.

7.2.5.2 Situation du projet à l'international

Mon projet scientifique orienté vision robotique directe s'inscrit dans une communauté scientifique de plus en plus active sur le sujet ces dernières années. En effet, si on comptait 2 articles nouveaux sur le sujet par tranche de 5 ans avant le début des années 2000, on est progressivement passé à 5 puis une dizaine d'articles pertinents par an, désormais. Le groupe de vision par ordinateur de l'Université Technique de Munich est leader en localisation et cartographie 3D directes depuis une petite dizaine d'années. L'équipe Rainbow (ex-Lagadic) d'Inria Rennes-Bretagne Atlantique est, elle, leader en asservissement visuel direct depuis plus de 10 ans. La plupart des contributions à l'asservissement visuel direct rapportées dans ce mémoire ont d'ailleurs été faites en collaboration avec l'équipe Rainbow. Ce dernier point indique une certaine visibilité des travaux dans lesquels je suis impliqué, qu'il convient de renforcer en se rapprochant aussi des autres équipes majeures travaillant activement sur le sujet qui, en élargissant le périmètre de ce mémoire et de l'état de l'art qu'il rapporte aux multicaméras et caméras actives, font apparaître d'autres acteurs tel que l'université d'Oxford, l'ETH Zurich, l'EPF Lausanne, Intel Labs ou encore Inria Sophia, sans être exhaustif.

Le volet "vision non-conventionnelle" du projet scientifique proposé et, surtout, le nouvel **Axe 4** de structuration de mes travaux à propos de conception de caméra pour apporter des solutions aux limites actuelles des approches directes pures et étendues, dès la phase d'acquisition d'image, apparaît donc plus original. La collaboration engagée avec l'Optical Media Interface lab du NAIST depuis quelques années, et qui s'amplifie progressivement, est l'un des moteurs qui permettront d'obtenir des résultats au sein de ce volet en combinant ses compétences en optique pour la vision par ordinateur à celles que j'apporte, appuyé de l'équipe Perception Robotique du laboratoire MIS, à l'interface de la vision et de la robotique.

Bibliographie

[Achanta et al., 2009] Achanta, R., Hemami, S., Estrada, F., and Susstrunk, S. (2009). Frequency-tuned salient region detection. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1597–1604. 163

[Adarve and Mahony, 2017] Adarve, J. and Mahony, R. (2017). Spherepix : a Data Structure for Spherical Image Processing. *IEEE Robot. Autonom. Lett.*, 2(2) :483–490. 153

[Ahmine et al., 2019] Ahmine, Y., Caron, G., Mouaddib, E., and Chouireb, F. (2019). Adaptive lucas-kanade tracking. *Image and Vision Computing*, 88 :1 – 8. 136

[Alismail et al., 2016] Alismail, H., Browning, B., and Lucey, S. (2016). Robust tracking in low light and sudden illumination changes. In *2016 Fourth International Conference on 3D Vision (3DV)*, pages 389–398. 91, 98

[Alj and Caron, 2015] Alj, Y. and Caron, G. (2015). Featureless omnidirectional vision-based control of non-holonomic mobile robot. In *2015 12th International Conference on Ubiquitous Robots and Ambient Intelligence (URAI)*, pages 95–100. 70, 97

[Andujar et al., 2004] Andujar, C., Vazquez, P., and Fairen, M. (2004). Way-Finder : guided tours through complex walkthrough models. *Computer Graphics Forum*. 111

[Antonakos et al., 2015] Antonakos, E., i Medina, J. A., Tzimiropoulos, G., and Zafeiriou, S. P. (2015). Feature-based lucas–kanade and active appearance models. *IEEE Transactions on Image Processing*, 24(9) :2617–2632. 91, 98

[Baker and Matthews, 2001] Baker, S. and Matthews, I. (2001). Equivalence and efficiency of image alignment algorithms. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition, CVPR'01*, pages 1090 – 1097. 52, 97

[Baker and Matthews, 2004] Baker, S. and Matthews, I. (2004). Lucas-kanade 20 years on : A unifying framework. *International Journal of Computer Vision*, 56(3) :221–255. 52, 53, 54, 97

[Baker and Nayar, 1999] Baker, S. and Nayar, S. K. (1999). A theory of single-viewpoint catadioptric image formation. *Int. Journal on Computer Vision*, 35(2) :175–196. 38

[Bakthavatchalam et al., 2013] Bakthavatchalam, M., Chaumette, F., and Marchand, E. (2013). Photometric moments : New promising candidates for visual servoing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'13*, pages 5521–5526, Karlsruhe, Germany. 98

[Bakthavatchalam et al., 2018] Bakthavatchalam, M., Tahri, O., and Chaumette, F. (2018). A direct dense visual servoing approach using photometric moments. *IEEE Transactions on Robotics*, 34(5) :1226–1239. 94, 98, 148

[Barreto and Araujo, 2001] Barreto, J. P. and Araujo, H. (2001). Issues on the geometry of central catadioptric imaging. In *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, Hawaii, USA. 38, 39

[Bateux et al., 2018] Bateux, Q., Marchand, E., Leitner, J., Chaumette, F., and Corke, P. (2018). Training deep neural networks for visual servoing. In *IEEE Int. Conf. on Robotics and Automation, ICRA'18*, pages 3307–3314, Brisbane, Australia. 83, 84, 97

[Bazin et al., 2012] Bazin, J.-C., Demonceaux, C., Vasseur, P., and Kweon, I. (2012). Rotation estimation and vanishing point extraction by omnidirectional vision in urban environment. *Int. J. Robot. Res.*, 31(1) :63–81. 153

[Belhumeur et al., 1997] Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. fisherfaces : recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7) :711–720. 142

[Benhimane and Malis, 2004] Benhimane, S. and Malis, E. (2004). Real-time image-based tracking of planes using efficient second-order minimization. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, volume 1, pages 943–948 vol.1. 56, 61, 62, 63, 97

[Benhimane and Malis, 2007] Benhimane, S. and Malis, E. (2007). Homography-based 2d visual tracking and servoing. *Int. Journal of Robotics Research*, 26(7) :661–676. 62, 97

[Bergen et al., 1992] Bergen, J. R., Anandan, P., Hanna, K. J., and Hingorani, R. (1992). Hierarchical model-based motion estimation. In Sandini, G., editor, *Computer Vision — ECCV'92*, pages 237–252, Berlin, Heidelberg. Springer Berlin Heidelberg. 95

[Bergmann et al., 2018] Bergmann, P., Wang, R., and Cremers, D. (2018). Online photometric calibration of auto exposure video for realtime visual odometry and slam. *IEEE Robotics and Automation Letters (RA-L)*, 3 :627–634. 42, 71, 77, 97

[Bonaventura et al., 2018] Bonaventura, X., Feixas, M., Sbert, M., Chuang, L., and Wallraven, C. (2018). A survey of viewpoint selection methods for polygonal models. *Entropy*, 20(5) :370. 117

[Bonaventure et al., 2011] Bonaventure, X., Feixas, M., and Sbert, M. (2011). Viewpoint information. In *Proceedings of the 21st Graphicon Int. Conf. on Computer Graphics and Vision*, pages 16–19, Moscou, Russie. 111

[Bouguet, 2000] Bouguet, J.-Y. (2000). Pyramidal implementation of the lucas kanade feature tracker. *Intel Corporation, Microprocessor Research Labs*. 96

[Burt, 1988] Burt, P. (1988). Smart sensing within a pyramid vision machine. *Proceedings of the IEEE*, 76 :1006–1015. 90

[Burt and Adelson, 1983] Burt, P. and Adelson, E. (1983). The laplacian pyramid as a compact image code. *IEEE Transactions on Communications*, 31(4) :532–540. 95

[Byravan and Fox, 2017] Byravan, A. and Fox, D. (2017). Se3-nets : Learning rigid body motion using deep neural networks. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 173–180. 83

[Caron et al., 2014] Caron, G., Dame, A., and Marchand, E. (2014). Direct model based visual tracking and pose estimation using mutual information. *Image and Vision Computing*, 32(1) :54–63. 88, 98, 99

[Caron et al., 2011] Caron, G., Marchand, E., and Mouaddib, E. (2011). Tracking planes in omnidirectional stereovision. In *IEEE Int. Conf. on Robotics and Automation, ICRA'11*, pages 6306–6311, Shanghai, China. 67, 97

[Caron et al., 2013] Caron, G., Marchand, E., and Mouaddib, E. M. (2013). Photometric visual servoing for omnidirectional cameras. *Autonomous Robots*, 35(2-3) :177–193. 69, 97, 130

[Caron and Morbidi, 2018] Caron, G. and Morbidi, F. (2018). Spherical Visual Gyroscope for Autonomous Robots using the Mixture of Photometric Potentials. In *IEEE International Conference on Robotics and Automation*, Brisbane, Australia. 41, 98, 136

[Caruso et al., 2015] Caruso, D., Engel, J., and Cremers, D. (2015). Large-scale direct slam for omnidirectional cameras. In *International Conference on Intelligent Robots and Systems (IROS)*. 70, 97

[Chaumette, 2004] Chaumette, F. (2004). Image moments : a general and useful set of features for visual servoing. *IEEE Transactions on Robotics*, 20(4) :713–723. 94

[Chaumette and Hutchinson, 2006] Chaumette, F. and Hutchinson, S. (2006). Visual servo control, Part I : Basic approaches. *IEEE Robotics and Automation Magazine*, 13(4) :82–90. 49, 95, 152

[Chaumette and Hutchinson, 2007] Chaumette, F. and Hutchinson, S. (2007). Visual servo control, Part II : Advanced approaches. *IEEE Robotics and Automation Magazine*, 14(1) :109–118. 55

[Churchill and Vardy, 2013] Churchill, D. and Vardy, A. (2013). An Orientation Invariant Visual Homing Algorithm. *J. Intell. Robot. Syst.*, 71(1) :3–29. 153

[Collewet and Marchand, 2009] Collewet, C. and Marchand, E. (2009). Photometry-based visual servoing using light reflexion models. In *IEEE Int. Conf. on Robotics and Automation, ICRA'09*, pages 701–706, Kobe, Japan, Japan. 73, 74

[Collewet and Marchand, 2011] Collewet, C. and Marchand, E. (2011). Photometric visual servoing. *IEEE Trans. on Robotics*, 27(4) :828–834. 73, 74, 83, 97, 104, 148

[Collewet et al., 2008] Collewet, C., Marchand, E., and Chaumette, F. (2008). Visual servoing set free from image processing. In *IEEE Int. Conf. on Robotics and Automation*, pages 81–86, Pasadena, California. 68, 69, 97

[Comaniciu and Meer, 2002] Comaniciu, D. and Meer, P. (2002). Mean shift : a robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5) :603–619. 92

[Comaniciu et al., 2003] Comaniciu, D., Ramesh, V., and Meer, P. (2003). Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5) :564–577. 92, 93

[Comport et al., 2006] Comport, A., Marchand, E., Pressigout, M., and Chaumette, F. (2006). Real-time markerless tracking for augmented reality : the virtual visual servoing framework. *IEEE Trans. on Visualization and Computer Graphics*, 12(4) :615–628. 100

[Comport et al., 2010] Comport, A. I., Malis, E., and Rives, P. (2010). Real-time Quadrifocal Visual Odometry. *International Journal of Robotics Research*, 29(2) :245–266. 57

[Corsini et al., 2009] Corsini, M., Dellepiane, M., Ponchio, F., and Scopigno, R. (2009). Image-to-geometry registration : a mutual information method exploiting illumination-related geometric properties. *Computer Graphics Forum*, 28 :1755–1764. 125

[Courty and Marchand, 2001] Courty, N. and Marchand, E. (2001). Computer animation : a new application for image-based visual servoing. In *International Conference on Robotics and Automation (ICRA)*, volume 1, pages 223–228. 111

[Crivellaro and Lepetit, 2014] Crivellaro, A. and Lepetit, V. (2014). Robust 3d tracking with descriptor fields. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 3414–3421. 91, 98

[Crombez, 2015] Crombez, N. (2015). *Contributions aux asservissements visuels denses : nouvelle modélisation des images adaptée aux environnements virtuels et réels*. PhD thesis, Université de Picardie Jules Verne. 125

[Crombez et al., 2018] Crombez, N., Caron, G., Funatomi, T., and Mukaigawa, Y. (2018). Reliable planar object pose estimation in light fields from best subaperture camera pairs. *IEEE Robotics and Automation Letters*, 3(4) :3561–3568. 168

[Crombez et al., 2014] Crombez, N., Caron, G., and Mouaddib, E. M. (2014). Colorisation de nuages de points 3D par recalage dense d'images numériques. *Traitement du Signal*, 31(1-2) :81–106. 72, 97, 119

[Crombez et al., 2015a] Crombez, N., Caron, G., and Mouaddib, E. M. (2015a). Photometric gaussian mixtures based visual servoing. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 5486–5491. 98, 136

[Crombez et al., 2015b] Crombez, N., Caron, G., and Mouaddib, E. M. (2015b). Using dense point clouds as environment model for visual localization of mobile robot. In *IEEE Int. Conf. on Ubiquitous Robots and Ambiant Intelligence, URAI'15*, pages 40–45, Goyang, South Korea. 72, 97, 119

[Crombez et al., 2019] Crombez, N., Mouaddib, E., Caron, G., and Chaumette, F. (2019). Visual servoing with photometric gaussian mixtures as dense features. *IEEE Transactions on Robotics*, 35(1) :49–63. 136, 146, 147, 148, 150

[Dalal and Triggs, 2005] Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893. 91

[Dame and Marchand, 2010] Dame, A. and Marchand, E. (2010). Accurate real-time tracking using mutual information. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pages 47–56. 116

[Dame and Marchand, 2011] Dame, A. and Marchand, E. (2011). Mutual information-based visual servoing. *IEEE Trans. on Robotics*, 27(5) :958–969. 87, 88, 98, 101, 102

[Dame and Marchand, 2012] Dame, A. and Marchand, E. (2012). Second-order optimization of mutual information for real-time image registration. *IEEE Transactions on Image Processing*, 21(9) :4190–4203. 88, 98

[Deguchi, 2000] Deguchi, K. (2000). A direct interpretation of dynamic images with camera and object motions for vision guided robot control. *Int. Journal on Computer Vision*, 37(1) :7–20. 96, 98

[Delabarre et al., 2012] Delabarre, B., Caron, G., and Marchand, E. (2012). Omnidirectional visual servoing using the normalized mutual information. In *10th IFAC Symposium on Robot Control, Syroco 2012*, Dubrovnik, Croatia. 88, 89, 98

[Delabarre and Marchand, 2012] Delabarre, B. and Marchand, E. (2012). Visual servoing using the sum of conditional variance. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'12*, pages 1689–1694, Vilamoura, Portugal. 87, 90, 98

[Delabarre and Marchand, 2013] Delabarre, B. and Marchand, E. (2013). Camera localization using mutual information-based multiplane tracking. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1620–1625. 88, 98

[DeMenthon and Davis, 1995] DeMenthon, D. F. and Davis, L. S. (1995). Model-based object pose in 25 lines of code. *Int. Journal of Computer Vision*, 15 :123–141. 126

[Demonceaux and Vasseur, 2009] Demonceaux, C. and Vasseur, P. (2009). Omnidirectional image processing using geodesic metric. In *Int. Conf. on Image Processing*, pages 221–224, Cairo, Egypt. 67

[Demonceaux et al., 2006] Demonceaux, C., Vasseur, P., and Pegard, C. (2006). Robust attitude estimation with catadioptric vision. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 3448–3453. 153

[Dowson and Bowden, 2006] Dowson, N. and Bowden, R. (2006). A unifying framework for mutual information methods for use in non-linear optimisation. In *European Conf. Computer Vision*, pages 365–378, Graz, Austria. 103

[Duflot et al., 2018] Duflot, L.-A., Reisenhofer, R., Tamadazte, B., Andreff, N., and Krupa, A. (2018). Wavelet and Shearlet-based Image Representations for Visual Servoing. *The International Journal of Robotics Research*. 96, 98

[Engel et al., 2018] Engel, J., Koltun, V., and Cremers, D. (2018). Direct sparse odometry. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 77, 97, 134

[Engel et al., 2014] Engel, J., Schöps, T., and Cremers, D. (2014). LSD-SLAM : Large-scale direct monocular SLAM. In *European Conference on Computer Vision (ECCV)*. 70, 71, 97

[Engel et al., 2013] Engel, J., Sturm, J., and Cremers, D. (2013). Semi-dense visual odometry for a monocular camera. In *2013 IEEE International Conference on Computer Vision*, pages 1449–1456. 70, 97

[Engel et al., 2016] Engel, J., Usenko, V., and Cremers, D. (2016). A photometrically calibrated benchmark for monocular visual odometry. Technical Report arXiv :1607.02555v2, arxiv.org. 77

[Fischler and Bolles, 1981] Fischler, M. and Bolles, R. (1981). Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6) :381–395. 142

[Forster et al., 2017] Forster, C., Zhang, Z., Gassner, M., Werlberger, M., and Scaramuzza, D. (2017). Svo : Semidirect visual odometry for monocular and multicamera systems. *IEEE Transactions on Robotics*, 33(2) :249–265. 153

[Fraissinet-Tachet et al., 2016] Fraissinet-Tachet, M., Schmitt, M., Wen, Z., and Kuijper, A. (2016). Multi-camera piecewise planar object tracking with mutual information. *Journal of Mathematical Imaging and Vision*, 56(3) :591–602. 88, 98

[Freitag et al., 2018] Freitag, S., Weyers, B., and Kuhlen, T. W. (2018). Interactive exploration assistance for immersive virtual environments based on object visibility and viewpoint quality. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 355–362. 111

[Georgel et al., 2008] Georgel, P., Benhimane, S., and Navab, N. (2008). A unified approach combining photometric and geometric information for pose estimation. In *British Machine Vision Conf., BMVC*. 100

[Geyer and Daniilidis, 2000] Geyer, C. and Daniilidis, K. (2000). A unifying theory for central panoramic systems and practical applications. In *European Conf. on Computer Vision*, Dublin, Ireland. 38

[Giancola et al., 2018] Giancola, S., Valenti, M., and Sala, R. (2018). *A Survey on 3D Cameras : Metrological Comparison of Time-of-Flight, Structured-Light and Active Stereoscopy Technologies*. SpringerBriefs in Computer Science. Springer International Publishing. 57

[Goldmannan, 2010] Goldmannan, D. B. (2010). Vignette and exposure calibration and compensation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32 :2276–2288. 43

[Golub and Van Loan, 1983] Golub, G. H. and Van Loan, C. F. (1983). *Matrix Computation*. Johns Hopkins University Press. 62

[Gonzalez and Woods, 2018] Gonzalez, R. and Woods, R. E. (2018). *Digital Image Processing, 4th Edition*. Pearson. 49, 136

[Grossberg and Nayar, 2003] Grossberg, M. D. and Nayar, S. K. (2003). Determining the camera response from images : what is knowable ? *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(11) :1455–1467. 43

[Gui Yun Tian et al., 2002] Gui Yun Tian, Gledhill, D., Taylor, D., and Clarke, D. (2002). Colour correction for panoramic imaging. In *Proceedings Sixth International Conference on Information Visualisation*, pages 483–488. 123, 127

[Habibi et al., 2015] Habibi, Z., Caron, G., and Mouaddib, E. M. (2015). 3d model automatic exploration : Smooth and intelligent virtual camera control. In *Computer Vision - ACCV 2014 Workshops, e-Heritage*, pages 612–626. 99, 115

[Habibi et al., 2015] Habibi, Z., Caron, G., and Mouaddib, E. M. (2015). Assistive visual framing in 3d dense points cloud. In *2015 Digital Heritage*, volume 2, pages 109–112. 136, 163

[Habibi et al., 2015] Habibi, Z., Mouaddib, E. M., and Caron, G. (2015). Good feature for framing : Saliency-based gaussian mixture. In *2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 3682–3687. 136, 163

[Hadj-Abdelkader, 2006] Hadj-Abdelkader, H. (2006). *Asservissement visuel en vision omnidirectionnelle*. PhD thesis, Université Blaise Pascal - Clermont II. 49

[Hadj-Abdelkader et al., 2018] Hadj-Abdelkader, H., Tahri, O., and Benseddik, H. (2018). Closed form solution for rotation estimation using photometric spherical moments. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 627–634. 98, 161

[Hager and Belhumeur, 1998] Hager, G. and Belhumeur, P. (1998). Efficient region tracking with parametric models of geometry and illumination. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 20(10) :1025–1039. 54, 60, 71, 76, 97

[Hager et al., 2004] Hager, G. D., Dewan, M., and Stewart, C. V. (2004). Multiple kernel tracking with ssd. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004.* 92, 93, 94

[Hartley and Zisserman, 2004] Hartley, R. I. and Zisserman, A. (2004). *Multiple View Geometry in Computer Vision*. Cambridge University Press, second edition. 66

[Héno and Chandelier, 2014] Héno, R. and Chandelier, L. (2014). *3D Modeling of Buildings : Outstanding Sites*. FOCUS Series. Wiley. 57, 121

[Horn and Schunck, 1981] Horn, B. K. P. and Schunck, B. G. (1981). Determining optical flow. *Artif. Intell.*, 17(1-3) :185–203. 47

[Huber, 1981] Huber, P.-J. (1981). *Robust statistics*. Wiley, New York, USA. 71

[Irani and Anandan, 1998] Irani, M. and Anandan, P. (1998). Robust multi-sensor image alignment. In *Sixth International Conference on Computer Vision*, pages 959–966. 90, 91, 98

[Ito et al., 2011] Ito, E., Okatani, T., and Deguchi, K. (2011). Accurate and robust planar tracking based on a model of image sampling and reconstruction process. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pages 1–8. 80, 97

[Itti et al., 1998] Itti, L., Koch, C., and Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(11) :1254–1259. 163

[Jeon et al., 2015] Jeon, H., Park, J., Choe, G., Park, J., Bok, Y., Tai, Y., and Kweon, I. S. (2015). Accurate depth map estimation from a lenslet light field camera. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1547–1555. 57

[Jiang, 2006] Jiang, B. (2006). Calibration-free line-based tracking for video augmentation. In *Int. Conf. on Computer Graphics & Virtual Reality, CGVR*, pages 104–110, Las Vegas, USA. 100

[Kallem et al., 2007] Kallem, V., Dewan, M., Swensen, J., Hager, G., and Cowan, N. (2007). Kernel-based visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and System*, pages 1975–1980, San Diego, USA. 94, 98, 137

[Kehl et al., 2017] Kehl, W., Manhardt, F., Tombari, F., Ilic, S., and Navab, N. (2017). SSD-6D : making rgb-based 3d detection and 6d pose estimation great again. In *IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017*, pages 1530–1538. 83

[Kendall et al., 2015] Kendall, A., Grimes, M., and Cipolla, R. (2015). Posenet : A convolutional network for real-time 6-dof camera relocalization. In *2015 IEEE International Conference on Computer Vision, ICCV 2015, Santiago, Chile, December 7-13, 2015*, pages 2938–2946. 83

[Kerl et al., 2013] Kerl, C., Sturm, J., and Cremers, D. (2013). Dense visual slam for rgb-d cameras. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2100–2106. 57

[Kim et al., 2019] Kim, D., Pathak, S., Moro, A., Komatsu, R., Yamashita, A., and Asama, H. (2019). E-cnn : Accurate spherical camera rotation estimation via uniformization of distorted optical flow fields. In *ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 2232–2236. 161

[Kim and Pollefeys, 2008] Kim, S. J. and Pollefeys, M. (2008). Robust radiometric calibration and vignetting correction. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(4) :562–576. 43

[Kyatkin and Chirikjian, 1999] Kyatkin, A. and Chirikjian, G. (1999). Pattern Matching as a Correlation on the Discrete Motion Group. *Comput. Vis. Image Und.*, 74(1) :22–35. 153

[Labrosse, 2006] Labrosse, F. (2006). The Visual Compass : Performance and Limitations of an Appearance-Based Method. *J. Field Robot.*, 23(10) :913–941. 153

[Lapresté and Mezouar, 2004] Lapresté, J. and Mezouar, Y. (2004). A Hessian approach to visual servoing. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS*, pages 998–1003, Sendai, Japan. 104

[Lepetit and Fua, 2005] Lepetit, V. and Fua, P. (2005). Monocular model-based 3d tracking of rigid objects : A survey. *Foundations and Trends in Computer Graphics and Vision*, pages 1–89. 100

[Li, 2006] Li, S. (2006). Full-View Spherical Image Camera. In *Proc. IEEE Int. Conf. Pattern Recogn.*, volume 4, pages 386–390. 40

[Li and Hai, 2010] Li, S. and Hai, Y. (2010). A Full-View Spherical Image Format. In *Proc. IEEE Int. Conf. Pattern Recogn.*, pages 2337–2340. 153, 154

[Lin et al., 2016] Lin, C.-H., Zhu, R., and Lucey, S. (2016). The conditional lucas & kanade algorithm. In Leibe, B., Matas, J., Sebe, N., and Welling, M., editors, *Computer Vision – ECCV 2016*, pages 793–808, Cham. Springer International Publishing. 142

[Lin et al., 2014] Lin, T.-Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., and Zitnick, C. L. (2014). Microsoft coco : Common objects in context. In Fleet, D., Pajdla, T., Schiele, B., and Tuytelaars, T., editors, *Computer Vision – ECCV 2014*, pages 740–755, Cham. Springer International Publishing. 141

[Lindeberg, 1994] Lindeberg, T. (1994). *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA. 138

[Lindner et al., 2007] Lindner, M., Kolb, A., and Hartmann, K. (2007). Data-fusion of pmd-based distance-information and high-resolution rgb-images. In *2007 International Symposium on Signals, Circuits and Systems*, volume 1, pages 1–4. 57

[Lindner, 2016] Lindner, W. (2016). *Digital photogrammetry : A practical course, fourth edition*. Remote sensing/photogrammetry. Springer-Verlag Berlin Heidelberg. 57

[Lowe, 1999] Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1150–1157. 91

[Lucas and Kanade, 1981] Lucas, B. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Int. Joint Conf. on Artificial Intelligence, IJCAI'81*, pages 674–679. 51, 58, 59, 97

[Ma et al., 2004] Ma, Y., Soatto, S., Košecká, J., and Sastry, S. (2004). *An invitation to 3D vision*. Springer. 40, 42, 52

[Makadia and Daniilidis, 2003] Makadia, A. and Daniilidis, K. (2003). Direct 3d-rotation estimation from spherical images via a generalized shift theorem. In *2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings.*, volume 2, pages II–217. 96, 98

[Makadia et al., 2007] Makadia, A., Geyer, C., and Daniilidis, K. (2007). Correspondence-free structure from motion. *International Journal of Computer Vision*, 75(3) :311–327. 96, 98, 153

[Malis, 2004] Malis, E. (2004). Improving vision-based control using efficient second-order minimization techniques. In *IEEE Int. Conf. on Robotics and Automation*, pages 1843–1848, New Orleans, USA. 55

[Marchand, 2019] Marchand, E. (2019). Subspace-based visual servoing. *IEEE Robotics and Automation Letters*, 4(3) :2699–2706. 96, 98

[Marchand and Chaumette, 2002] Marchand, E. and Chaumette, F. (2002). Virtual visual servoing : A framework for real-time augmented reality. *Computer Graphics Forum*, 21(3) :289–298. 102

[Mariottini et al., 2012] Mariottini, G., Scheggi, S., Morbidi, F., and Prattichizzo, D. (2012). An accurate and robust visual-compass algorithm for robot-mounted omnidirectional cameras. *Robot. Autonom. Syst.*, 60(9) :1179–1190. 153

[Matthies et al., 1988] Matthies, L., Szeliski, R., and Kanade, T. (1988). Incremental estimation of dense depth maps from image sequences. In *Proceedings CVPR '88 : The Computer Society Conference on Computer Vision and Pattern Recognition*, pages 366–374. 70, 97

[McLachlan and Peel, 2000] McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. Wiley. 155

[Mei et al., 2006] Mei, C., Benhimane, S., Malis, E., and Rives, P. (2006). Homography-based tracking for central catadioptric cameras. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 669–674. 62, 63, 64, 65, 97

[Mei et al., 2008] Mei, C., Benhimane, S., Malis, E., and Rives, P. (2008). Efficient homography-based tracking and 3-d reconstruction for single-viewpoint sensors. *IEEE Transactions on Robotics*, 24(6) :1352–1364. 66, 67, 72, 97

[Mei and Reid, 2008] Mei, C. and Reid, I. (2008). Modeling and generating complex motion blur for real-time tracking. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. 78, 79, 97

[Meilland et al., 2015] Meilland, M., Comport, A., and Rives, P. (2015). Dense omnidirectional rgb-d mapping of large scale outdoor environments for real-time localisation and autonomous navigation. *Journal of Field Robotics, "Special Issue on Ground Robots Operating in dynamic, unstructured and large-scale outdoor environments"*, 32(4) :474–503. 134

[Ming Li et al., 2004] Ming Li, Xin Chen, Xin Li, Bin Ma, and Vitanyi, P. M. B. (2004). The similarity metric. *IEEE Transactions on Information Theory*, 50(12) :3250–3264. 89

[Morbidi and Caron, 2017] Morbidi, F. and Caron, G. (2017). Phase correlation for dense visual compass from omnidirectional camera-robot images. *IEEE Robotics and Automation Letters*, 2(2) :688–695. 96, 98

[Morel and Yu, 2009] Morel, J.-M. and Yu, G. (2009). Asift : A new framework for fully affine invariant image comparison. *SIAM J. Img. Sci.*, 2 :438–469. 108, 123, 126

[Nayar et al., 1996] Nayar, S., Nene, S., and Murase, H. (1996). Subspace methods for robot vision. *IEEE Trans. on Robotics and Automation*, 12(5) :750–758. 96, 98

[Newcombe et al., 2011] Newcombe, R. A., Lovegrove, S. J., and Davison, A. J. (2011). Dtam : Dense tracking and mapping in real-time. In *Proceedings of the 2011 International Conference on Computer Vision*, ICCV '11, pages 2320–2327, Washington, DC, USA. IEEE Computer Society. 70, 97

[Ngo et al., 2019] Ngo, T.-T., Nagahara, H., Nishino, K., Taniguchi, R.-i., and Yagi, Y. (2019). Reflectance and shape estimation with a light field camera under natural illumination. *International Journal of Computer Vision*. 168

[Ojala et al., 1996] Ojala, T., Pietikäinen, M., and Harwood, D. (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern Recognition*, 29(1) :51 – 59. 91

[Ok et al., 2016] Ok, K., Greene, W. N., and Roy, N. (2016). Simultaneous tracking and rendering : Real-time monocular localization for mavs. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4522–4529. 110

[Park et al., 2017] Park, S., Schöps, T., and Pollefeys, M. (2017). Illumination change robustness in direct visual slam. In *2017 IEEE International Conference on Robotics and Automation (ICRA)*, pages 4523–4530. 91, 98

[Park et al., 2012] Park, Y., Lepetit, V., and Woo, W. (2012). Handling motion-blur in 3d tracking and rendering for augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 18(9) :1449–1459. 79, 97

[Pascoe et al., 2015] Pascoe, G., Maddern, W., and Newman, P. (2015). Direct visual localisation and calibration for road vehicles in changing city environments. In *IEEE International Conference on Computer Vision : Workshop on Computer Vision for Road Scene Understanding and Autonomous Driving*, Santiago, Chile. 89, 98

[Pascoe et al., 2017] Pascoe, G., Maddern, W., Tanner, M., Pinies, P., and Newman, P. (2017). NID-SLAM : Robust monocular SLAM using normalised information distance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI. 89, 98

[Peretroukhin and Kelly, 2018] Peretroukhin, V. and Kelly, J. (2018). Dpc-net : Deep pose correction for visual localization. *IEEE Robotics and Automation Letters*, 3(3) :2424–2431. 83

[Petit et al., 2012] Petit, A., Marchand, E., and Kanani, K. (2012). Tracking complex targets for space rendezvous and debris removal applications. In *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, IROS'12*, pages 4483–4488, Vilamoura, Portugal. 106

[Phong, 1975] Phong, B. T. (1975). Illumination for computer generated pictures. *Commun. ACM*, 18(6) :311–317. 73

[Pickering et al., 2009] Pickering, M. R., Muhit, A. A., Scarvell, J. M., and Smith, P. N. (2009). A new multi-modal similarity measure for fast gradient-based 2d-3d image registration. In *2009 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 5821–5824. 85

[Polonsky et al., 2005] Polonsky, O., Patané, G., Biasotti, S., Gotsman, C., and Spagnuolo, M. (2005). What's in an image ? *The Visual Computer*, 21(8) :840–847. 111

[Pressigout and Marchand, 2007] Pressigout, M. and Marchand, E. (2007). Real-time hybrid tracking using edge and texture information. *Int. Journal of Robotics Research, IJRR*, 26(7) :689–713. 100

[Rad and Lepetit, 2017] Rad, M. and Lepetit, V. (2017). Bb8 : A scalable, accurate, robust to partial occlusion method for predicting the 3d poses of challenging objects without using depth. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 3848–3856. 83

[Richa et al., 2011] Richa, R., Sznitman, R., Taylor, R., and Hager, G. (2011). Visual tracking using the sum of conditional variance. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2953–2958. 86, 89, 98

[Rosten and Drummond, 2005] Rosten, E. and Drummond, T. (2005). Fusing points and lines for high performance tracking. In *IEEE Int. Conf. on Computer Vision*, volume 2, pages 1508–1511. 100

[Royer et al., 2005] Royer, E., Lhuillier, M., M., D., and Chateau, T. (2005). Localization in urban environments : Monocular vision compared to a differential gps sensor. In *IEEE Conf. on Computer Vision and Pattern Recognition, CVPR*, volume 2, pages 114–121, Washington, DC, USA. 100

[Salazar-Garibay et al., 2009] Salazar-Garibay, A., Malis, E., and Mei, C. (2009). Visual tracking of planes with an uncalibrated central catadioptric camera. In *IEEE/RSJ Int. Conf. on Intelligent RObots and Systems*, Saint-Louis, USA. 65, 97

[Saputra et al., 2018] Saputra, M. R. U., Markham, A., and Trigoni, N. (2018). Visual slam and structure from motion in dynamic environments : A survey. *ACM Comput. Surv.*, 51(2) :37 :1–37 :36. 57

[Scandaroli et al., 2012] Scandaroli, G. G., Meilland, M., and Richa, R. (2012). Improving ncc-based direct visual tracking. In *Computer Vision – ECCV 2012*, pages 442–455, Berlin, Heidelberg. Springer Berlin Heidelberg. 81, 82, 97

[Scaramuzza and Siegwart, 2008] Scaramuzza, D. and Siegwart, R. (2008). Appearance-Guided Monocular Omnidirectional Visual Odometry for Outdoor Ground Vehicles. *IEEE Trans. Robot.*, 24(5) :1015–1026. 153

[Shannon, 1948] Shannon, C. E. (1948). A mathematical theory of communication. *Bell Syst. Tech. J.*, 27 :379–423. 85, 89

[Shi and Tomasi, 1994] Shi, J. and Tomasi (1994). Good features to track. In *1994 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 593–600. 60, 97

[Shum and Szeliski, 2000] Shum, H.-Y. and Szeliski, R. (2000). Systems and experiment paper : Construction of panoramic image mosaics with global and local alignment. *International Journal of Computer Vision*, 36(2) :101–130. 51

[Siciliano et al., 2008] Siciliano, B., Sciavicco, L., Villani, L., and Oriolo, G. (2008). *Robotics : Modelling, Planning and Control*. Springer Publishing Company, Incorporated, 1st edition. 54

[Silveira, 2014] Silveira, G. (2014). On intensity-based 3d visual servoing. *Robotics and Autonomous Systems*, 62(11) :1636 – 1645. Special Issue on Visual Control of Mobile Robots. 69

[Silveira and Malis, 2007] Silveira, G. and Malis, E. (2007). Real-time visual tracking under arbitrary illumination changes. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–6. 74, 75, 97

[Steinbrücker et al., 2011] Steinbrücker, F., Sturm, J., and Cremers, D. (2011). Real-time visual odometry from dense rgb-d images. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 719–722. 57

[Stewart and Newman, 2012] Stewart, A. and Newman, P. (2012). Laps - localisation using appearance of prior structure : 6-dof monocular camera localisation using prior pointclouds. In *Proceedings - IEEE International Conference on Robotics and Automation*, pages 2625–2632. 90

[Stewart, 1999] Stewart, C. (1999). Robust parameter estimation in computer vision. *SIAM Rev.*, 41 :513–537. 72, 82, 158

[Studholme et al., 1999] Studholme, C., Hill, D., and Hawkes, D. (1999). An overlap invariant entropy measure of 3d medical image alignment. *Pattern Recognition*, 32(1) :71 – 86. 89

[Swaminathan and Nayar, 1999] Swaminathan, R. and Nayar, S. K. (1999). Polycameras : camera clusters for wide angle imaging. Technical Report CUCS-013-99, Columbia university, Computer science. 40

[Szeliski, 2011] Szeliski, R. (2011). *Computer Vision - Algorithms and Applications*. Texts in Computer Science. Springer. 57

[Sünderhauf et al., 2018] Sünderhauf, N., Brock, O., Scheirer, W., Hadsell, R., Fox, D., Leitner, J., Upcroft, B., Abbeel, P., Burgard, W., Milford, M., and Corke, P. (2018). The limits and potentials of deep learning for robotics. *The International Journal of Robotics Research*, 37(4-5) :405–420. 84

[Tekin et al., 2018] Tekin, B., Sinha, S. N., and Fua, P. (2018). Real-time seamless single shot 6d object pose prediction. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. 83, 97

[Titterington et al., 1985] Titterington, D., Smith, A., and Makov, U. (1985). *Statistical Analysis of Finite Mixture Distributions*. John Wiley & Sons. 138

[Vázquez et al., 2001] Vázquez, P.-P., Feixas, M., Sbert, M., and Heidrich, W. (2001). Viewpoint selection using viewpoint entropy. In *Proceedings of the Vision Modeling and Visualization Conference 2001*, VMV '01, pages 273–280. Aka GmbH. 111

[Xiang et al., 2018] Xiang, Y., Schmidt, T., Narayanan, V., and Fox, D. (2018). Posecnn : A convolutional neural network for 6d object pose estimation in cluttered scenes. In *Robotics : Science and Systems XIV, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA, June 26-30, 2018*. 83

[Xiong and De la Torre, 2015] Xiong, X. and De la Torre, F. (2015). Global supervised descent method. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2664–2673. 142

[Ying and Hu, 2004] Ying, X. and Hu, Z. (2004). Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model. In *European Conf. on Computer Vision*, pages 442–455, Prague, Czech Republic. 38

[Zhang et al., 2019] Zhang, F., Leitner, J., Ge, Z., Milford, M., and Corke, P. (2019). Adversarial discriminative sim-to-real transfer of visuo-motor policies. *The International Journal of Robotics Research*, 38(10-11) :1229–1245. 84

[Zhong et al., 2018] Zhong, L., Lu, M., and Zhang, L. (2018). A direct 3d object tracking method based on dynamic textured model rendering and extended dense feature fields. *IEEE Transactions on Circuits and Systems for Video Technology*, pages 1–1. 91, 98

[Zichao Zhang et al., 2016] Zichao Zhang, Rebecq, H., Forster, C., and Scaramuzza, D. (2016). Benefit of large field-of-view cameras for visual odometry. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 801–808, Stockholm, Suède. 131

Résumé : Une caméra collecte la luminance de la scène qu'elle observe, sous la forme d'une grille d'intensités. Utiliser les pixels de cette grille en entrée de lois de commande et d'algorithmes de suivi d'objet évite les habituels traitements intermédiaires de détection, mise en correspondance de primitives et d'estimation de paramètres. Cette dernière se transforme ainsi en commande directe de robot et en suivi et calcul de pose directs à partir des réelles mesures de la caméra. Ces problématiques de commande et de suivi directs forment le cadre de la vision robotique directe.

Après un résumé des activités que j'ai entreprises à ce sujet, ce document propose une étude ciblée, mais détaillée, des travaux de recherche en vision robotique directe, réécrits sous un formalisme unique. Ensuite, les principales contributions faites avec les doctorants que j'ai co-encadrés, sont détaillées. Principalement, elles étendent des critères directs, photométriques et basés entropie photométrique, à de nouveaux contextes ou types de caméras. Elles définissent aussi théoriquement de nouveaux critères, comme le mélange de potentiels photométriques, repoussant significativement les limites des domaines de convergence des approches directes. Les caméras perspectives, panoramiques et sphériques sont considérées dans des contextes de commande, navigation et localisation de robot et en environnement virtuel.

Enfin, le bilan des contributions est discuté, en regard de l'état de l'art et des travaux en cours au sein de la communauté scientifique, avant de proposer un projet scientifique pour traiter les problèmes restants.

Mots clés : Vision par ordinateur, vision panoramique, vision sphérique, robotique, robotique mobile, drone, information visuelle, primitive directe, entropie, information mutuelle, potentiels photométriques, gaussiennes photométriques, espace d'échelle, suivi visuel, calcul de pose, asservissement visuel, maquette virtuelle, nuages de points 3D.

Direct robotic vision

Abstract : A camera collects the radiance of the observed scene, as a grid of intensities. Using the pixels of that grid as input of robot control laws and object trackers avoids the usual intermediate processes of features detection, matching and parameters estimation. The latter is, thus, rethought as direct robot control and direct tracking and pose estimation from camera actual measurements. These direct control and tracking issues are gathered as direct robotic vision.

After summing up the activities that I have undertaken in that frame, this document proposes a focused, but detailed, survey of works related to direct robotic vision, rewritten under a unique formalism. Then, the main contributions made with co-supervised PhD students, are detailed. Mainly, they extend direct criteria, such as the photometric and photometric entropy-based ones, to new contexts or camera types. They also theoretically define new criteria, as mixtures of photometric potentials, drastically overcoming the convergence domain limits of direct approaches. Perspective, panoramic and spherical cameras are considered in contexts of robot control, navigation and localization as well as virtual environments.

Finally, the outcome of the contributions is discussed, facing the state-of-the-art and ongoing works in the research community, before proposing a scientific project about the remaining open problems.

Keywords : Computer vision, panoramic vision, spherical vision, robotics, mobile robotics, drone, visual information, direct feature, entropy, mutual information, photometric potentials, photometric Gaussians, scale-space, visual tracking, pose computation, visual servoing, virtual model, 3D point clouds.