



Estimation de la rotation 3D : Approches denses et génériques basées images sphériques

Houssem-Eddine Benseddik

► To cite this version:

Houssem-Eddine Benseddik. Estimation de la rotation 3D : Approches denses et génériques basées images sphériques. Traitement du signal et de l'image [eess.SP]. Université Paris-Saclay; Université d'Evry-Val-d'Essonne; Université Aboubakr Belkaïd - Tlemcen, 2017. Français. NNT : 2017SACLE045 . tel-01759180

HAL Id: tel-01759180

<https://hal.science/tel-01759180>

Submitted on 5 Apr 2018

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

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



Estimation de la rotation 3D : Approches denses et génériques basées images sphériques

Thèse de doctorat de l'Université Paris-Saclay
préparée à l'université d'Evry Val d'Essonne

École doctorale n°580 Sciences et technologies
de l'information et de la communication (STIC)
Spécialité de doctorat: Robotique

Thèse présentée et soutenue le 21/12/2017, par

Houssem-Eddine Benseddik

Composition du Jury :

Abdelhafid Bessaid Professeur, Université de Tlemcen (– LGBM)	Président
Youcef Mezouar Professeur, Institut Pascal, (– SIGMA)	Rapporteur
Oualid Djekoune, Maître de recherche HDR, Centre de développement des technologies avancées CDTA (– DPR)	Rapporteur
Omar Tahri Maître de conférences HDR, INSA Centre Val de Loire, Bourges (– PRISME)	Examineur
Michèle Gouiffès Maître de conférences, Université Paris Sud (–LIMSI)	Examineur
Samia Bouchafa--Bruneau Professeur, Université d'Evry (– IBISC)	Directeur de thèse
Brahim Cherki Professeur, Université de Tlemcen (– LAT)	Directeur de thèse
Hicham Hadj-Abdelkader Maître de conférences, Université d'Evry (– IBISC)	Co-Directeur de thèse

الجمهورية الجزائرية الديمقراطية الشعبية

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة أبي بكر بلقايد – تلمسان

Université Aboubakr Belkaïd – Tlemcen –

Faculté de TECHNOLOGIE



THESE

Présentée pour l'obtention du **grade** de **DOCTEUR EN SCIENCES**

En : Automatique

Par : Housseem-Eddine Benseddik

Sujet

***Estimation de la rotation 3D : Approches denses et génériques
basées images sphériques***

Soutenue publiquement, le 21 / 12 / 2017, devant le jury composé de :

M. Abdelhafid Bessaid	Professeur	Univ. Tlemcen	Président
Mme. Samia Bouchafa- Bruneau	Professeur	Univ. d'Evry Val d'Essonne, France	Directeur de thèse
M. Brahim Cherki	Professeur	Univ. Tlemcen	Directeur de thèse
M. Hicham Hadj- Abdelkader	MCA	Univ. d'Evry Val d'Essonne, France	Co-Directeur de thèse
M. Youcef Mezouar	Professeur	Institut Pascal, SIGMA, France	Examineur
M. Oualid Djekoune	MRA	CDTA, Alger	Examineur
M. Omar Tahri	MCA	INSA Centre Val de Loire, Bourges, France	Examineur
Mme. Michèle Gouiffès	MCA	Univ. Paris Sud, France	Examineur

REMERCIEMENTS

Les travaux présentés dans ce mémoire ont été menés aux seines de Laboratoire Informatique, Biologie Intégrative et Systèmes Complexes (IBISC) de l'Université d'Évry Val d'Essonne - Évry, France, et le Laboratoire d'Automatique de Tlemcen (LAT) de l'Université Abou Bekr Belkaid - Tlemcen, Algérie, dans le cadre d'une thèse en cotutelle internationale. Que ceux qui ont contribué de près ou de loin à l'aboutissement et à la bonne réalisation de ces travaux, trouvent ici mes remerciements les plus sincères.

Je remercie l'ensemble des membres du jury pour avoir accepté de participer à mon jury de thèse et pour l'intérêt qu'ils ont manifesté à l'égard de ce travail de thèse. Je remercie mes rapporteurs Monsieur Youcef Mezouar, professeur des universités à l'Université Blaise Pascal - Clermont 2 de Clermont-Ferrand et Monsieur Oualid Djekoune, maître de recherche au centre de développement des technologies avancées CDTA à Alger, pour leurs remarques judicieuses et constructives qui m'ont permis d'améliorer considérablement la qualité de ce manuscrit. Je remercie également Monsieur Abdelhafid Bessaid, professeur à l'Université Abou Bekr Belkaid de Tlemcen, pour avoir présidé le jury. Je tiens à remercier Madame Michelle Michèle Gouiffès, maître de conférences à l'université de Paris Sud et Monsieur Omar Tahri, maître de conférences à l'Institut National des Sciences Appliquées, d'avoir examinés minutieusement et rigoureusement ce travail de thèse, tout en y prodiguant leurs remarques, critiques et propositions.

J'exprime ma profonde gratitude à mes directeurs de thèse Madame Samia Bouchafa-Bruneau, professeur des universités à l'Université d'Évry Val d'Essonne et Monsieur Brahim Cherki, professeur à l'Université Abou Bekr Belkaid de Tlemcen, grâce à leur implications efficaces, leur conseils instructifs et leur rigueurs scientifiques, tout en laissant un important espace à l'initiative personnelle, c'est grâce à vous que nous avons pu réaliser ces travaux et de mener à terme cette thèse. J'ai été extrêmement ravie de travailler avec vous et je souhaite à travers ces lignes que vous receviez toutes les marques de ma reconnaissance.

J'ai un remerciement tout particulier à mon co-directeur de thèse Hicham Hadj-Abdelkader, maître de conférences à l'université d'Évry Val d'Essonne. D'abord, car cette thèse n'aurait pas eu lieu sans lui, ensuite de m'avoir apporté tout son soutien, son aide et son expérience durant toutes les années de cette thèse. Il a su créer une ambiance chaleureuse, confiante et surtout de joué un rôle plus important que celui d'un encadrant. Je tiens également à souligner ses qualités humaines qui m'ont beaucoup inspiré.

Je souhaite aussi adresser un grand merci à Amine Hadj-Abdelkader pour sa disponibilité et sa gentillesse quand j'en avais besoin, notamment pour s'occuper des diverses

formalités administratives au niveau de l'université de Tlemcen, à chaque occasion où je n'y étais pas présent.

Une bonne partie de ce travail a été entamée au CDTA, à qui je suis très reconnaissant, mes remerciements vont également à tous mes collègues du CDTA, notamment les membres de la division "Productique et Robotique" qui ont contribué à l'élaboration d'une ambiance chaleureuse agréable au travail.

Je ne citerai pas, mais ils se reconnaîtront, tous mes amis, qui m'ont soutenu et encouragé, même de loin. Je leur en suis reconnaissant.

Je remercie vivement le gouvernement Algérien de son aide financière, qui m'a été offert dans le cadre d'une bourse du Programme National Exceptionnel PNE.

Pour finir, je ne saurais exprimer à quel point je remercie ma famille pour son soutien affectif et ses encouragements depuis toujours. Plein de petits mercis à mon frère et ma petite sœur. Un immense merci à vous, mes très chers parents, je vous suis éternellement reconnaissant et redevable. Un merci tout spécial à mon épouse, qui a su comprendre l'importance de mon travail pour moi, de m'avoir toujours soutenue et pour la patience dont elle a fait preuve, notamment lors de la phase de rédaction de ce manuscrit.

DÉDICACE

TABLE DES MATIÈRES

LISTE DES FIGURES	xi
LISTE DES TABLEAUX	xvi
INTRODUCTION	1
1 Imagerie omnidirectionnelle et estimation de mouvement - état de l'art	5
1.1 Imagerie omnidirectionnelle	6
1.1.1 Les cameras polydioptriques	6
1.1.2 Les cameras fisheye	7
1.1.3 Les cameras catadioptriques	8
1.1.4 Modèle de projection unifié	11
1.1.5 Étalonnage du capteur	13
1.2 La géométrie multi-vues	14
1.2.1 Homographie entre deux vues	15
1.2.2 La géométrie épipolaire	17
1.2.3 La matrice essentielle	18
1.2.4 Estimation du mouvement par la matrice essentielle	19
1.2.5 L'estimation robuste du mouvement	21
1.2.6 La matrice fondamentale	22
1.3 L'estimation de l'égo-mouvement	22
1.3.1 Les capteurs de l'odométrie visuelle	24
1.3.2 Les méthodes de l'odométrie visuelle	28
1.4 Conclusion	35

2 Estimation de la rotation 3D basée sur la représentation en harmoniques sphériques SPHARM	37
2.1 Introduction	38
2.2 Génération du maillage surfacique 3D	42
2.2.1 Échantillonnage de l'image sphérique	42
2.2.2 Génération de nuages de points 3D	42
2.2.3 Construction du maillage surfacique 3D	43
2.3 Réparation du maillage	46
2.3.1 Topologie des surfaces	47
2.3.2 Algorithme de réparation du maillage	51
2.4 Le descripteur SPHARM de formes 3D	52
2.4.1 Paramétrisation sphérique du maillage 3D	52
2.4.2 Expansion SPHARM	55
2.5 Estimation de la rotation 3D	58
2.5.1 Estimation initiale de la rotation	59
2.5.2 Raffinement par l'ICP	62
2.6 Résultats expérimentaux	63
2.6.1 Expérimentation sur des images synthétiques	64
2.6.2 Expérimentation sur des images réelles	70
2.6.3 Efficacité et coût en temps de calcul	75
2.7 Conclusion	76
3 Estimation de la rotation 3D basée moments sphériques	77
3.1 Introduction	78
3.2 Les moments géométriques	79

3.3	Matrice d'interaction	83
3.3.1	Matrice d'interaction d'un point	84
3.4	Matrices d'interaction associées aux moments 2D (cas continu)	85
3.5	Matrices d'interaction associées aux moments 2D (cas discret)	87
3.6	Matrice d'interaction des moments sphériques	88
3.6.1	Expression des moments en projection sphérique	89
3.6.2	Matrice d'interaction d'un point en projection sphérique	89
3.6.3	Projection d'un élément de surface sur la sphère unitaire	90
3.6.4	Matrice d'interaction des moments sphériques	91
3.7	Estimation de la rotation basée moments photométriques sphériques . . .	93
3.7.1	Première méthode (M1)	97
3.7.2	Deuxième méthode (M2) :	98
3.7.3	Troisième méthode (M3) :	98
3.8	Expérimentations	99
3.8.1	Données synthétiques	99
3.8.2	Données réelles	104
3.9	Conclusion	114
4	Résultats expérimentaux	115
4.1	Environnement intérieur	115
4.2	Environnement extérieur	119
4.3	Discussion	122
	CONCLUSION	125
A	Annexes	128

A.1 La transformée de Fourier sphérique	128
A.2 Le calcul de la variation temporelle des moments	131
A.3 Étude comparative de performances des détecteurs/descripteurs SIFT et SURF	134

LISTE DES FIGURES

1.1	(a) Le capteur LadyBug (b) Image acquise avec un système multi-caméras .	7
1.2	(a) lentille fisheye (b) Image obtenue à l'aide d'un objectif fisheye	7
1.3	(a) Le capteur catadioptrique (b) Image acquise avec un système catadioptrique	8
1.4	Les modèles de projection des cameras catadioptriques (a) centrales (b) non-centrales	9
1.5	Géométrie de formation de l'image acquise par caméra catadioptrique. (a) Miroir conique (b) Miroir sphérique (c) Miroir hyperbolique (d) Miroir parabolique. . . .	12
1.6	Double projection sur une sphère unitaire.	14
1.7	Passage entre deux vues exprimé par l'homographie	16
1.8	Géométrie épipolaire pour le cas planaire.	17
1.9	Géométrie épipolaire pour le cas sphérique.	18
1.10	Une illustration de la contrainte épipolaire.	21
1.11	Algorithme d'un système d'OV basé primitives	30
2.1	Schéma général de l'approche proposée pour estimer la rotation optimale entre deux images sphériques.	41
2.2	Représentation géométrique de l'intensité normalisée du pixel de l'image sphérique.	43
2.3	Exemple d'un maillage triangulaire et sa représentation (arête, faces et sommets).	44
2.4	(a) Image sphérique (b) Nuage de points 3D (c) Maillage surfacique 3D.	45

2.5	Génération de l'objet en fonction du B	45
2.6	Une tasse est homéomorphe à un tore car il existe une déformation continue entre eux.	48
2.7	Un exemple sur les composantes de bord et les trous sur un maillage (a) le trou dans le tore est délimité par deux anneaux représentant les bords (en rouge). (b) Le trou est rempli en obtenant un tore avec une topologie en anneau.	49
2.8	Un exemple sur l'orientabilité d'un maillage. Orientation anti-horaire : (C,D,A), (A,D,B), (C,B,D), orientation horaire : (C,A,D), (D,A,B), (B,C,D) et Non orienté : (C,D,A), (D,A,B), (C,B,D).	49
2.9	Objets de genres 0, 1 et 3.	50
2.10	L'algorithme de réparation du maillage : (a) Maillages 3D avec trous (b) Voxellisation du maillage (b) Maillage 3D après réparation.	52
2.11	Exemple de paramétrisation entre une surface et un domaine.	53
2.12	Paramétrisation sphérique de l'objet 3D.	55
2.13	Représentation en trois dimensions des harmoniques sphériques allant jusqu'à l'ordre 3, où les portions bleus correspondent aux régions, où la fonction est positive et les portions rouges aux régions où la fonction est négative.	56
2.14	(a) et (b) représentent l'image sphérique et l'objet 3D. (c), (d), (e) et (f) la reconstruction SPHARM pour $L_{max}=1, 5, 10$ et 15 , respectivement.	58
2.15	La rotation d'une image sphérique avec l'exemple de l'intensité lumineuse normalisée d'un pixel.	59
2.16	(a) L'image sphérique originale. (b) Image après une rotation de $\mathbf{R}(45^\circ; 60^\circ; 45^\circ)$	64
2.17	Estimation des rotations (a) autour de l'axe (Z). (b) autour de l'axe (Y).	65
2.18	Exemple de raffinement de pose sur l'estimation de la rotation. (a) : deux objets séparés par une rotation, (b) l'alignement après l'estimation initiale. (c) l'alignement après le raffinement de l'estimation.	66
2.19	Performance en terme d'erreur angulaire absolue pour l'estimation de la rotation autour les axes Z et Y avec différents niveaux de bruits gaussiens ajoutés : (a) $\sigma^2 = 1$, (b) $\sigma^2 = 5$, (c) $\sigma^2 = 10$ and (d) $\sigma^2 = 20$	68

2.20 Performance en terme d'erreur angulaire absolue pour l'estimation de la rotation autour les axes Z et Y avec différents niveaux de changements de luminosité : (a) +40%, (b) +20%, (c) -20% and (d) -40%.	68
2.21 Performance en terme d'erreur angulaire absolue pour l'estimation de la rotation autour les axes Z et Y avec différents niveaux de Facteur de Qualité de la compression JPEG : (a) QF = 10, (b) QF = 30, (c) QF = 60 and (d) QF = 90. .	68
2.22 Kumotek VS-C42U-300-TK montée sur un trépied gradué.	71
2.23 Exemple de détection du plan pour le calcul de l'homographie entre les deux images.	71
2.24 Erreur angulaire absolue de l'estimation de la rotation autour des axes : (a) : Z et (b) : Y.	72
2.25 Première colonne : les images avec occultation (a) L'image de référence non-occultée, (b) Une personne, (c) Deux personnes et (d) Trois personnes dans la scène. Deuxième rangée : le niveaux d'occultation introduit.	74
3.1 Orientation d'un objet et définition des paramètres usuels d'une ellipse. . .	82
3.2 Modèle de projection d'un point.	84
3.3 (a) Illustration des variations temporelles d'un contour d'image fermé $C(t)$ (b) détails	85
3.4 Élément de surface en projection sphérique.	90
3.5 Exemple d'un filtre gaussien sphérique.	98
3.6 La mise en correspondance d'un point d'intérêt pour la méthode M3. . . .	99
3.7 Voisinages sphériques des points d'intérêts utilisés pour la méthode M3. . .	99
3.8 Estimation des rotations α autour de l'axe (Z), β autour de l'axe (Y) et γ autour de l'axe (X).	100
3.9 Performance en terme de l'erreur angulaire absolue de l'estimation de la rotation autour les axes Z, Y et X avec différents niveaux du bruit gaussien ajouté : (a) $\sigma^2 = 5$, (b) $\sigma^2 = 15$, (c) $\sigma^2 = 25$ and (d) $\sigma^2 = 50$. Première, deuxième et troisième lignes correspond aux méthodes M1, M2 et M3 respectivement	102

3.10 Performance en terme de l'erreur angulaire absolue de l'estimation de la rotation autour les axes Z, Y et X avec différents niveaux du changement de luminosité : (a) +40%, (b) +20%, (c) -20% and (d) -40%. Première, deuxième et troisième ligne correspond aux méthodes M1, M2 et M3 respectivement .	103
3.11 Performance en terme de l'erreur angulaire absolue de l'estimation de la rotation autour les axes Z, Y et X avec différents niveaux du Facteur de Qualité de la compression JPEG : (a) QF = 10, (b) QF = 30, (c) QF = 60 and (d) QF = 90. Première, deuxième et troisième lignes correspond aux méthodes M1, M2 et M3 respectivement	105
3.12 Manipulateur <i>Stäubli</i> TX60.	106
3.13 Le robot Pioneer 3-AT équipé d'une camera catadioptrique	106
3.14 Une image catadioptrique du scénario 2 qu'est obtenue en milieu intérieur.	107
3.15 Erreur angulaire absolue de l'estimation de la rotation représentée par les angles d'Euler autour des axes : Z, Y et X et comparaison avec la méthode "5 primitives + RANSAC"	108
3.16 Une image catadioptrique du scénario 2 qu'est obtenue en milieu extérieur.	108
3.17 Illustration graphique de la structure de l'ensemble de données (les 25 points de collections)généré avec le manipulateur Staubli.	109
3.18 Stabilisation de la rotation sur des images catadioptrique par les méthodes proposées. (a) Orientation de référence. (b, f, j) Les images initiales acquises après les rotations suivantes : 17.8°, 185.3° et 266.3°, respectivement. Chaque ligne correspond à l'alignement de rotation par rapport à l'orientation de (a) par les méthodes M1, M2 et M3. Les flèches colorées illustrent une direction initiale (vert) ou alignée M1 (cyan), M2 (bleu) et M3 (magenta). Pour une visualisation plus claire, une seule direction est tracée, mais l'alignement rotatif consiste non seulement en une rotation 1D, mais plutôt en une transformation de rotation 3D complète.	113
4.1 Robot d'expérimentation en environnement intérieur : Turtlebot 2, équipé de la camera catadioptrique	116
4.2 Le plan de masse de l'environnement intérieur dont la trajectoire parcourue par le robot a été superposée. La sélection de 8 poses parmi les images de la séquence.	117

4.3	Résultats d'estimation de déplacement du robot obtenus par les approches suivantes : (a) SPHARM, (b) M1, (c) M2 et (d) M3.	118
4.4	(a) L'image initiale de scénario 3 (b) L'image 137 de scénario 3 (c) La trajectoire de forme 8	120
4.5	L'estimation des angles d'Euler pour la trajectoire de 8 par les approches suivantes : SPHARM (marron), M1 (bleu), M2 (vert) , et M3 (rose), par rapport à la vérité de terrain (noir).(a) roll, (b) pitch et (c) yaw	121
4.6	La nomenclature des angles d'Euler : Roll, Pitch et Yaw, indiquant les angles de rotation autour des trois axes XYZ.	121
A.1	Detection et description des points d'interets par (a) SIFT+RANSAC (b) SURF+RANSAC.	134
A.2	Évaluation des performances de SIFT et SURF par rapport au changement de l'angle de vue et l'échelle. (a) Nombre de correspondance (b) Répétabilité.	135
A.3	Les courbes de Recall en fonction de la precision	136

LISTE DES TABLEAUX

1.1	Configurations réalisables de capteur à PVU	11
2.1	Résumé de la comparaison entre la méthode proposée, DNE, NPM, NCC + DNE et NCC + NPM, pour l'estimation des angles d'Euler. Une valeur de ∞ signifie que la minimisation non linéaire n'a pas convergé.	69
2.2	Résultats de l'estimation en présence d'occultations. La première colonne correspond aux rotations mesurées à partir de la graduation du trépied. La deuxième colonne correspond aux rotations obtenues à partir de la relation d'homographie (vérité terrain). Les colonnes 3-5 sont les résultats de rotation de chaque image avec clutter (4 %, 9 % et 12 %), respectivement.	73
2.3	AAE de l'estimation de la rotation estimation avec des translations entre les images réelles.	74
3.1	Résultats de l'estimation de la rotation suivant les trois axes par l'approche dense en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.	111
3.2	Résultats de l'estimation de la rotation suivant les trois axes par l'approche gaussien-dense en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.	111
3.3	Résultats de l'estimation de la rotation suivant les trois axes par l'approche dense-local en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.	112
4.1	Résultats de l'estimation de la rotation suivant les trois axes par l'approche dense en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.	122

A.1 Performances des détecteurs	137
A.2 Performances des descripteurs	137

INTRODUCTION

Introduction

Donner la vue à une machine est une tâche difficile, alors que voir est une activité tellement naturelle pour les êtres humains et les animaux que l'ampleur de la difficulté est souvent minorée. Voir est une tâche innée pour nous, mais c'est aussi un processus inconscient que nous ne sommes pas capables de l'expliquer et encore moins de le reproduire artificiellement. L'invention de la caméra et de l'ordinateur ont permis de franchir un premier pas vers la vision artificielle, mais le plus dur reste à faire. La caméra est capable d'acquérir une image, un peu à la manière de ce que fait l'œil. Mais pour interpréter l'image, l'ordinateur et les algorithmes mis au point par l'homme sont loin de pouvoir égaler ce dont est capable le cerveau. Pourtant, la recherche en vision par ordinateur est active depuis plus d'une cinquantaine d'années et elle s'appuie sur des connaissances plus anciennes. Les ambitions premières ont dû être revues à la baisse, et plutôt que de vouloir reproduire la totalité du système visuel humain, les recherches se concentrent sur des tâches plus particulières et plus simples comme reconnaître un objet particulier, ou retrouver la structure tridimensionnelle d'une scène. Ces travaux ont d'ores et déjà donné lieu à un certain nombre d'applications dans divers domaines : surveillance, métrologie, modélisation pour la synthèse d'images, robotique, . . . etc.

En effet, la théorie de la vision artificielle n'a réellement trouvée ses fondements que récemment avec en parallèle la compréhension de la nature même de l'image en terme de signal et la formalisation de la géométrie sous-jacente à l'acquisition des images. L'enjeu durant les années à venir va être de montrer que ces modèles mathématiques et photométriques sont effectivement à même d'appréhender la complexité d'images réelles. Clairement, autant sur un plan théorique que sur le plan des applications, la vision par ordinateur se situe au carrefour de nombreuses sciences de l'ingénieur, telles que les mathéma-

tiques fondamentales et appliquées, l'intelligence artificielle, le traitement de signal, l'automatique et l'informatique. Avec les avancées technologiques, ainsi que l'émergence de l'électronique et de l'informatique, la capacité de traitement ouvre de larges perspectives en terme d'applications tant dans le domaine de l'analyse de scènes et de la modélisation d'objets que dans le domaine de l'utilisation de l'information visuelle dans des boucles de commande. Elle va permettre également de confronter les avancées théoriques et les efforts de formalisation qui ont eu lieu dans les dix dernières années à la dure réalité de la complexité du monde réel. Toujours pour les mêmes raisons liées au progrès de l'informatique et du fait de la "maturité" de la théorie de la vision, on assiste aujourd'hui à une explosion des domaines applicatifs, comme par exemple, l'aide à la conduite ou la conduite automatique de véhicule, la construction de modèle tridimensionnel comme dans le cas d'images géologiques ou médicales, la détection et la caractérisation du mouvement d'objets ou de véhicules, l'indexation d'images ou de séquences vidéos dans des bases de données.

La localisation d'un robot est une des tâches aux quelles la vision par ordinateur peut apporter une réponse. Se localiser est une de ces choses que nous faisons naturellement en utilisant notre vue. Cela ne nous aide pas beaucoup pour construire un algorithme utilisable par un ordinateur, mais cela nous montre tout de même qu'il existe une solution performante au problème de la localisation par vision. Se localiser est une tâche importante qu'un robot doit être capable de réaliser pour accomplir un certain nombre de missions qu'il devra réaliser de manière autonome. C'est même incontournable pour un robot mobile qui doit atteindre son but et revenir à son point de départ une fois le travail effectué. Le premier champ d'application des robots est la réalisation de tâches dans des lieux où l'homme ne peut pas aller. Il peut s'agir de zones difficilement accessibles (exploration planétaire ou sous-marine), dangereuses (zones radioactives, robots militaires), ou à des échelles différentes (robot à l'intérieur d'une canalisation, voire robot chirurgical à l'intérieur du corps humain).

A l'heure actuelle, la camera à grand champ de vue est le capteur visuel qui connaît le plus d'intérêt, de la part de la communauté roboticienne, du fait de sa propriété de pouvoir saisir une perception à très large champ de vue de l'environnement, en une seule image. Le but central de ce type de camera est de répondre aux deux principales limitations de la vision perspective classique que sont la limitation du champ de vue, le manque de redondance et ainsi que certains types d'applications, exigent un champ de vision plus large que celui des caméras standards. La vision omnidirectionnelle peut être obtenue en plaçant un miroir de révolution, par exemple un miroir sphérique, face à l'objectif de la caméra ou en utilisant un objectif à large champ de vue de type fisheye. Le champ de vue dépend du type d'équipement utilisé et il est possible d'atteindre un champ de vue supérieur à 180° et engendrant une perception à 360° autour de l'axe optique de la caméra. En

effet, les images apportent beaucoup d'informations et permettent d'estimer la position et l'orientation 3D et de réaliser une localisation de robot.

Cadre applicatif et but de ce travail

Le problème de l'estimation de la pose a pris l'attention des chercheurs dans le domaine de la vision par ordinateur. Les applications concernées sont plusieurs. Nous citons ici par exemple la robotique, la réalité virtuelle, la reconnaissance et la classification des objets, Dans le cadre de la robotique mobile par exemple, les caméras sont utilisées comme un moyen pour corriger les dérives de la localisation basée capteurs proprioceptifs. Les méthodes de la reconstruction de déplacement basée vision peuvent être séparées en deux catégories : les méthodes basées apparence (template based) et les méthodes basées primitives (feature based). La première catégorie concerne la reconstruction de petit déplacement où les variations dans l'image sont petites. L'utilisation de toutes les informations dans le signal image rend le calcul de déplacements efficace. Cependant, ces techniques sont valables pour des petits déplacements. La seconde catégorie traite les grands déplacements entre les prises de vues en exploitant les matrices essentielles (ou d'homographie si la scène observée est plane) entre les correspondances de deux vues. Ces méthodes nécessitent des phases d'extraction et de mise en correspondance de primitives visuelles. Cependant, l'estimation de déplacement est biaisée par les erreurs de détection entre les correspondances.

Dans le cadre de cette thèse, nous nous intéresserons aux approches denses d'estimation de la rotation 3D valables pour les grands déplacements. Nous souhaitons utiliser toutes les informations dans le signal image pour accomplir l'estimation. Afin d'obtenir des algorithmes génériques et valides pour des caméras conventionnelles et omnidirectionnelles, nous considérerons les modèles de projection sphérique.

Des approches directes utilisant toute l'information contenue dans l'image, faisant appel aux représentations en harmoniques sphériques et les moments sphériques des fonctions sphériques, extraits à partir des deux images sphériques consécutives, ont été utilisées lors de l'estimation du mouvement relatif qui sépare les deux images, sans détection et sans mise en correspondance de primitives.

Nos travaux ont donc permis de développer plusieurs approches d'estimation de la rotation 3D d'images sphériques, fondées sur une extraction globale de l'information depuis les images obtenues par un système mono-caméra. Afin de bien cerner les différentes solutions proposées pour remédier à la problématique traitée dans le cadre de ce travail de thèse, le manuscrit est structuré en quatre chapitres et une conclusion générale.

Le premier chapitre présente les modèles géométriques existants pour les caméras perspectives et omnidirectionnelles ainsi que la géométrie epipolaire et multi-vues. Les approches majeures développées dans la littérature en réponse aux problèmes de calcul de pose, sont détaillées à travers un état de l'art.

Le deuxième chapitre décrit en détail notre première méthode d'estimation de la rotation entre deux images sphériques. Tout d'abord, ce chapitre introduira la nouvelle représentation proposée, en générant un objet 3D représentant l'intensité lumineuse de l'image. Les notions liées aux maillages 3D ainsi que la topologie des surfaces ont été introduites. Ensuite, nous présenterons la technique SPHARM et son intérêt qui réside dans le fait qu'elle pourra obtenir une description globale de la forme géométrique 3D. Par la suite, nous expliquerons comment estimer la rotation à travers les coefficients harmoniques obtenus de la représentation SPHARM. Enfin, nous présenterons les résultats de la méthode, et en particulier, une analyse de performance illustrée sur l'estimation de la rotation depuis des séquences d'images synthétiques et réelles.

Le troisième chapitre présentera une solution analytique de l'estimation de la rotation d'images sphériques. La méthode se base sur la matrice d'interaction calculée à partir des moments photométriques sphériques. Puis, nous montrerons comment la méthode peut être généralisée à la proposition de trois différentes méthodes, dont le concept s'articule sur la solution proposée. En effet, la première méthode consiste à calculer les moments photométriques à partir de l'image sphérique globale. Dans la deuxième méthode qui représente un raffinement de l'estimation fournie par la première méthode. En filtrant l'image par une fonction de pondération, quant à elle une gaussienne, permettant d'attribuer plus de poids à la zone contenant plus d'information sur la rotation par rapport aux autres. La troisième méthode, consiste à appliquer la solution proposée sur le voisinage d'un couple de points d'intérêts, extraits et mis en correspondance, depuis les images originales. Enfin, une analyse des capacités et performances des 3 méthodes sera détaillée.

Au quatrième chapitre, nous décrirons les résultats obtenus avec les approches développées dans cette thèse. Les expérimentations seront effectuées dans des environnements d'intérieur et d'extérieur. Une analyse critique des résultats sera présentée afin d'établir les forces et les faiblesses des algorithmes proposés. Nous terminerons ce document par une conclusion générale ainsi que des perspectives pour des futurs travaux.

Chapitre 1

Imagerie omnidirectionnelle et estimation de mouvement - état de l'art

Sommaire

1.1 Imagerie omnidirectionnelle	6
1.1.1 Les cameras polydioptriques	6
1.1.2 Les cameras fisheye	7
1.1.3 Les cameras catadioptriques	8
1.1.4 Modèle de projection unifié	11
1.1.5 Étalonnage du capteur	13
1.2 La géométrie multi-vues	14
1.2.1 Homographie entre deux vues	15
1.2.2 La géométrie épipolaire	17
1.2.3 La matrice essentielle	18
1.2.4 Estimation du mouvement par la matrice essentielle	19
1.2.5 L'estimation robuste du mouvement	21
1.2.6 La matrice fondamentale	22
1.3 L'estimation de l'égo-mouvement	22
1.3.1 Les capteurs de l'odométrie visuelle	24
1.3.2 Les méthodes de l'odométrie visuelle	28
1.4 Conclusion	35

1.1 Imagerie omnidirectionnelle

Les images conventionnelles, acquises par des cameras standards, possèdent un champ de vue relativement limité, en général de l'ordre de 40° à 60° . En raison de leur champ de vue restreint, un bon nombre d'applications est affectée par cette contrainte de champ de vue. Alternativement, les images omnidirectionnelles permettent de pallier à cet inconvénient en fournissant un champ de vue panoramique allant jusqu'à $180^\circ \times 360^\circ$. En effet, les capteurs omnidirectionnels ont l'avantage d'avoir beaucoup plus d'informations sur la scène que les capteurs classiques.

Cette particularité a permis à ce type de capteurs d'être largement utilisés dans différents domaines, comme la vidéo surveillance [2], la reconstruction 3D [3], la navigation autonome en robotique [4] ou la capture d'événements sportifs [5] ...

L'imagerie omnidirectionnelle est un domaine en effervescence dont les premiers travaux remontent en 1994 avec le dispositif COPIS proposé par Yasushi Yagi[6]. En associant une caméra perspective et un miroir conique, une vue panoramique de la scène a été obtenue. Depuis, plusieurs systèmes ont été proposés afin d'acquérir des images omnidirectionnelles. Pour en obtenir, on distingue trois classes de systèmes possibles dans la littérature, soit en utilisant :

- Des images multiples (reçues par des caméras conventionnelles)
- Des lentilles spéciales (fisheye)
- Des miroirs convexes (catadioptrique).

1.1.1 Les cameras polydioptriques

Les caméras polydioptriques sont quant à elles constituées d'un ensemble de caméras permettant de couvrir un champ de vue plus large. Les transformations inter-caméras sont généralement connues à l'aide d'un calibrage préalable du capteur permettant d'associer les images provenant de toutes les caméras afin de former l'image omnidirectionnelle. L'avantage avec ces systèmes de vision est qu'ils permettent l'acquisition d'images panoramiques de très haute résolution et contrairement aux systèmes faisant intervenir des caméras rotatives, ils sont tout à fait exploitables sur des robots mobiles. On peut toutefois nuancer la praticabilité de ces capteurs qui nécessitent une synchronisation suffisamment précise entre les caméras, une phase de calibrage plus contraignante ainsi qu'un coût élevé.

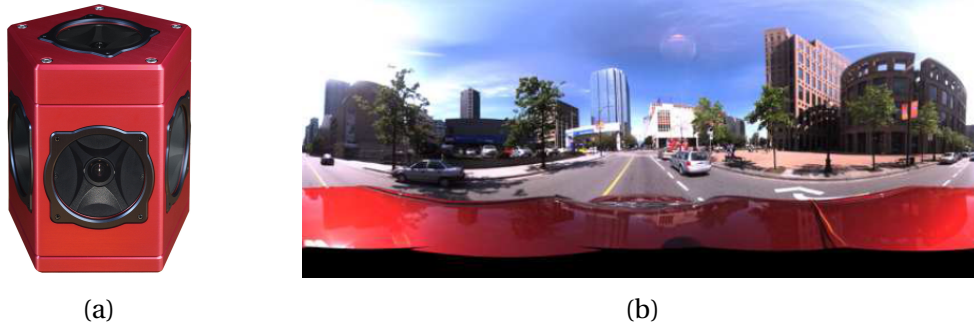


FIGURE 1.1 – (a) Le capteur LadyBug (b) Image acquise avec un système multi-caméras

Les systèmes de vision polydioptriques peuvent être constitués de caméras perspectives, mais aussi d'un ensemble de caméras omnidirectionnelles, on parle alors de système poly-omnidirectionnelle [7]. De nombreux systèmes commerciaux tel que la Ladybug (voir figure 1.1), le Panono ou encore la Bublcam démocratisent l'utilisation de ces caméras non seulement pour les entreprises ("Google street view" reposait sur l'utilisation de la ladybug jusqu'en 2008) mais aussi pour les particuliers. De nombreux travaux portent actuellement sur la navigation de véhicules autonomes équipés de caméras polydioptriques [8, 9].

1.1.2 Les cameras fisheye

On appelle caméra fisheye toute caméra équipée d'un objectif de type fisheye (ou œil de poisson en français, voir figure 1.2-(a)). Ce type objectif a une focale très courte, qui permet par conséquent d'obtenir un champ de vue très large pouvant atteindre 220° . L'acquisition de ce type d'image hémisphérique se fait au détriment d'une forte distorsion radiale déformant les lignes droites qui se trouvent on bordure de l'image (phénomène clairement visible sur la figure 1.2-(b)).



FIGURE 1.2 – (a) lentille fisheye (b) Image obtenue à l'aide d'un objectif fisheye

De même que les caméras catadioptriques que nous allons décrire dans la section sui-

vante, la résolution de l'image est plus forte en son centre qu'à sa périphérie. Par ailleurs, l'usage de ce genre de capteur permet d'exploiter la zone aveugle, la partie noire centrale qui caractérise les images catadioptriques (1.2-(b))). Néanmoins, les miroirs catadioptriques possèdent un bien large champ de vue, avec une perception de 360° autour de l'axe optique contre 180° en diagonale pour un fisheye.

Les capteurs fisheye sont composés d'une lentille qui ne fait pas partie des dispositifs à point de vue unique [10]. Par conséquent, les rayons lumineux ne convergent pas en un point unique mais en une zone appelée diacaustique. Il existe cependant différents modèles permettant de modéliser les distorsions radiales et tangentielles de ces lentilles. Dans [11], il a été montré qu'il est possible d'utiliser le modèle unifié qu'on décrit plus loin, pour ce type de capteur. Cette modélisation est possible si l'erreur de calibration est en adéquation avec l'application désirée. A base de cette proposition, Mei dans [12] a présenté une méthode de calibration basée sur la sphère unitaire. Cette méthode de calibration est destinée aux capteurs à point de vue unique et aux fisheyes.

1.1.3 Les cameras catadioptriques

L'idée consiste à utiliser un miroir pour récolter une vue large atteignant les 360° (figure 1.3), associé à une caméra (classique) qui capture la scène projetée sur le miroir. Le dispositif caméra + miroir est appelé capteur catadioptrique (figure 1.3-(a)). La catoptrique est la science des surfaces réfléchissantes (miroirs) alors que la dioptrique est la science des éléments réfractant (lentilles) [13]. Le concept de ces capteurs est apparu pour la première fois dans la présentation de René Descartes en 1637 en Discours de la Méthode [14].

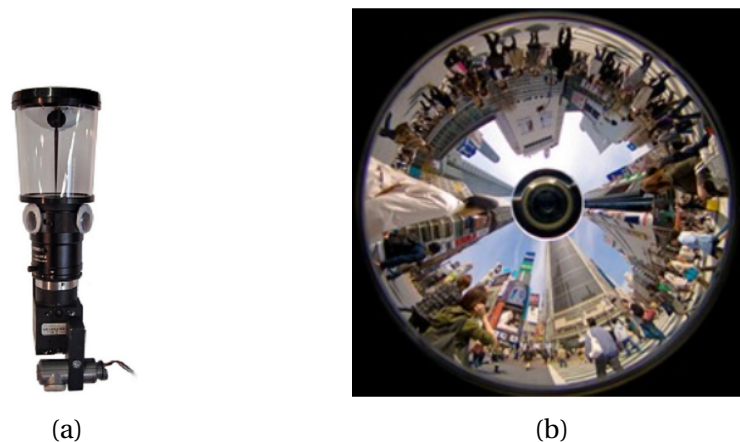


FIGURE 1.3 – (a) Le capteur catadioptrique (b) Image acquise avec un système catadioptrique

Les capteurs catadioptriques présentent l'avantage de fournir une vue large d'une manière relativement simple, efficace et plus rapide [15]. Lors de l'utilisation dans la robo-

tique mobile, un capteur catadioptrique permet de caractériser une position dans une scène indépendamment de la direction du robot, il enlève aussi l'ambiguïté présente (translation ou rotation?) dans le cas des caméras conventionnelles pour l'estimation du mouvement [16, 15, 17].

Cependant, l'image résultante comporte des distorsions qui varient selon la proximité au centre de l'image. Dans ce qui suit, nous verrons différents types de capteurs catadioptriques qui, selon Baker et Nayar [18], sont scindés en deux catégories : systèmes catadioptriques centraux et autres non centraux.

1.1.3.1 Modèle de caméras centrales

Les capteurs catadioptriques centraux sont des systèmes qui vérifient la contrainte du Point de Vue Unique (PVU) de sorte que les rayons réfléchis convergent vers le même point (figure 1.4-(a)). Un des principaux avantages des modèles de caméras centrales est de pouvoir utiliser un certain nombre de méthodes applicables aux caméras conventionnelles. Aussi, les systèmes satisfaisants cette contrainte présentent d'autres avantages par rapport aux non centraux, parmi eux : la possibilité d'éliminer les distorsions, ainsi que la simplicité relative à la modélisation et du calibrage de cameras. Ainsi, il est possible, après adaptation, d'utiliser une géométrie épipolaire permettant une recherche efficace des epipoles que ce soit pour les cameras monoculaires ou stéréoscopiques, et l'estimation d'homographies par exemple. Contrairement aux caméras conventionnelles, les points utilisés ne sont pas situés dans le plan de l'image mais doivent être portés par une autre surface, telle que celle du miroir ou d'une sphère virtuelle. Il est ainsi possible par exemple de construire une mosaïque d'images issues de caméras catadioptriques pour générer des images omnidirectionnelles à partir d'un nombre réduit d'images [19].

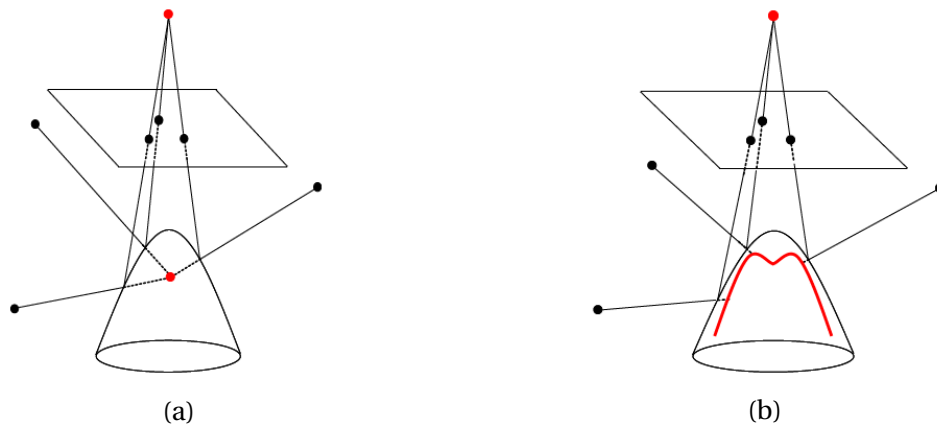


FIGURE 1.4 – Les modèles de projection des caméras catadioptriques (a) centrales (b) non-centrales

Baker et Nayar [20, 18, 21] ont présenté une classe complète des capteurs catadioptriques ayant un point de vue unique. Ces capteurs disposant des miroirs plans, paraboloides, hyperboloïdes ou ellipsoïdes qui admettent un point de vue unique. Les miroirs sphériques et coniques admettent aussi un point de vue unique, mais le montage miroirs + caméras n'en a pas [22].

Parmi ces différents couples miroir/caméra, seules les deux configurations impliquant un miroir de forme convexe présentent une véritable utilité dans le cadre de la vision omnidirectionnelle. D'un point de vue pratique, la configuration utilisant une caméra perspective est souvent préférée pour des raisons liées au coût de construction et à la taille moins encombrante du capteur. D'autres configurations de caméra catadioptrique à PVU sont théoriquement valides mais ne peuvent pas être mise en œuvre physiquement, ces différentes configurations sont exposées dans [10].

Chaque forme des miroirs convexes, présentées sur la figure 1.5 : conique, sphérique, hyperbolique ou parabolique, a des spécificités propres à cette forme concernant la satisfaction du PVU, l'importance des aberrations et la complexité de réalisation. Dans ce qui suit, nous allons décrire les différents miroirs existants ainsi que les diverses configurations possibles.

Les miroirs coniques ont un point de vue unique au sommet du cône. Cela signifie que dans le but d'accomplir la contrainte PVU, le sténopé de la caméra doit être placé à l'extrémité du cône. Le miroir conique est donc une solution dégénérée pour un système catadioptrique central.

Les miroirs sphériques respectent la contrainte PVU seulement dans le cas où une caméra de modèle sténopé est placée au centre de la sphère. La caméra dans ce cas n'acquerra que sa propre image. Ce genre de miroirs n'est pas utile pour les systèmes qui exigent le respect de la contrainte du point de vue unique. Néanmoins, les miroirs de forme sphérique représentent un choix attrayant, en raison de leurs caractéristiques telles que le faible coût de fabrication, les spécificités de la densité de résolution d'image qui est en fonction de l'angle d'élévation, ainsi que leurs tailles.

Les miroirs hyperboliques, la contrainte de PVU est satisfaite lorsque le sténopé et le point de vue du miroir sont placés sur l'un des deux foyers de l'hyperboloïde. Par conséquent, la conception de ce type de capteurs est délicate. La caméra perspective doit être placée soigneusement par rapport au miroir afin que son centre optique soit confondu avec l'un des foyers et que son axe optique passe par l'autre foyer.

Les miroirs paraboliques sont des miroirs dont la forme est une portion de paraboloides de révolution. Une parabole est l'ensemble de tous les points dans le plan situés

à égale distance de la directrice L , et du foyer F . Le f paramètre est la distance entre la directrice et le foyer, donnée par $f = 2a$, où a est la distance du sommet à la directrice ou le foyer. La surface de révolution obtenue par rotation d'une parabole autour de son axe de symétrie est appelé un paraboloïde. La forme parabolique est une solution de la contrainte PVU dans le cas particulier d'une projection orthographique. Par conséquent, un miroir parabolique doit être utilisé en conjonction avec une caméra orthographique.

Dans la figure 1.5 nous avons présenté la géométrie de formation de l'image catadioptrique, pour chaque type de miroir énuméré précédemment.

Le tableau 1.1 résume les propriétés des différents miroirs de révolution utilisés pour concevoir une caméra catadioptrique à PVU.

TABLEAU 1.1 – Configurations réalisables de capteur à PVU

Caméra	Miroir
Perspective	Convexe hyperbolique
Perspective	Plan
Perspective	Concave ellipsoïdale
Orthographique	Convexe parabolique

1.1.3.2 Modèle de caméras non-centrales

Dans le cas des capteurs catadioptriques non centraux, les rayons avant réflexion ne sont pas contraints de passer par un point unique (figure 1.4-(b)). Pour ce type de caméra non centrale, le miroir sphérique est souvent utilisé parce qu'il est facile à fabriquer. Il arrive que le miroir conique soit aussi utilisé dans cas d'une caméra non centrale. Dans [23], il a été proposé un système catadioptrique à point de vue non unique avec un miroir sphérique.

Tout de même, les capteurs catadioptriques non centraux sont utilisés dans plusieurs applications de robotique et de vidéo-surveillance. Dans [24] et [25], les auteurs utilisent un miroir sphérique pour la localisation du robot.

1.1.4 Modèle de projection unifié

Lors du choix du modèle représentant des capteurs centraux, un compromis doit être fait pour choisir un modèle qui reflète bien la vérité physique, mais qu'il ne soit pas trop complexe pour être utilisable et que ses paramètres puissent être estimables.

Comme nous l'avons évoqué auparavant, les caméras à point de vue unique forment

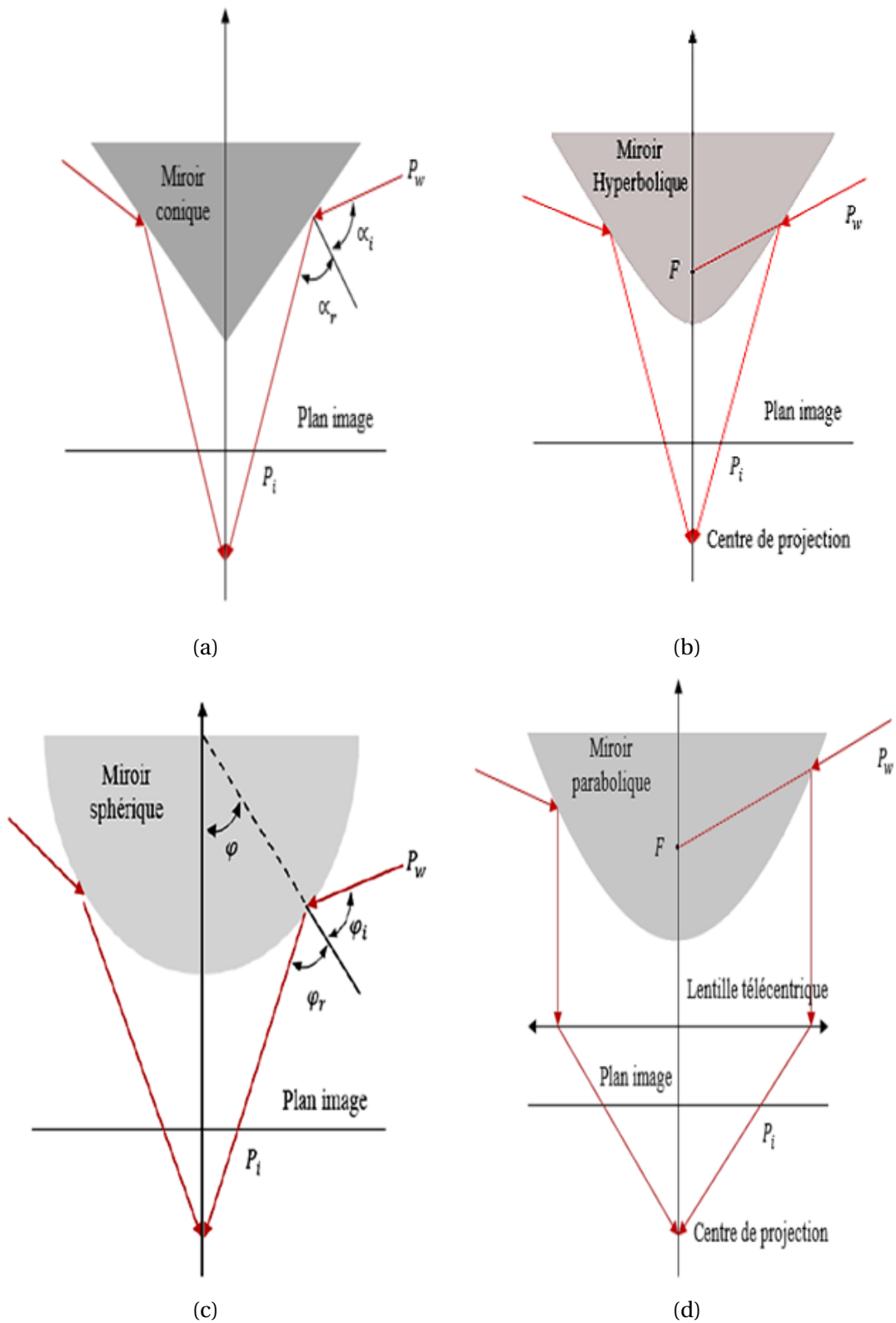


FIGURE 1.5 – Géométrie de formation de l'image acquise par caméra catadioptrique.
(a) Miroir conique (b) Miroir sphérique (c) Miroir hyperbolique (d) Miroir parabolique.

un large spectre de capteurs de vision, en comprenant les caméras perspectives, les caméras catadioptriques centrales et les caméras équipées de lentille fisheye.

Toutefois, un problème se pose pour ce type de cameras est celui de la distorsion introduite par le miroir de certaines d'entre elles. Cette déformation est beaucoup plus importante par rapport aux images issues des cameras perspectives classiques. Par conséquent, les outils classiques conçus pour les images perspectives, comme le calcul de la géométrie épipolaire [26], la détection de contour [27] ou encore la mise en correspondance de points caractéristiques [28], ne sont pas adaptés aux images omnidirectionnelles, car ne tenant pas compte de la géométrie de miroirs.

Plusieurs travaux ont été proposés dont l'idée était de trouver un espace de substitution à celui du plan. Geyer et Daniilidis [29], ont montré que la projection qu'effectue une caméra para-catadioptrique est équivalente à une double projections. Tout d'abord, une projection sur la surface d'une sphère de rayon 1 depuis son centre. Elle est suivie d'une projection stéréographique (projection perspective depuis le pôle nord) vers un plan. Ils ont montré la même équivalence pour d'autres capteurs catadioptriques à centre de projection unique avec comme seule différence que la seconde projection s'effectue depuis un autre point que le pôle nord, situé sur le segment défini par le pôle nord et le centre de la sphère.

La projection est visible sur l'exemple de la figure 1.6. Soit S la sphère de rayon 1 centrée en O . Le point 3D $P = (X, Y, Z)^T$ est tout d'abord projeté à la surface de S via O en un point P' , lui même projeté vers le plan image en $p = (x, y)$ par une projection perspective de centre F situé au pôle supérieur de S . p est bien la projection catadioptrique de P .

1.1.5 Étalonnage du capteur

Dans le cadre général de la vision par ordinateur, le calibrage doit toujours être réalisé. Ce calibrage est décomposé en deux parties : la détermination des paramètres internes du système (focale, coordonnées du point principal...), appelés paramètres intrinsèques, et la détermination des paramètres de transformation 3D/2D, appelés paramètres extrinsèques. Dans le cas de la vision catadioptrique, ce calibrage s'enrichit des éléments relatifs au miroir, et au montage (position relative du miroir par rapport à la caméra).

Le rôle du calibrage d'un capteur catadioptrique, est d'estimer la valeur des paramètres du modèle de capteur afin de modéliser le processus de formation des images captées. En effet, cette modélisation se contente d'estimer les paramètres permettant de passer d'un point dans un repère métrique spatial aux coordonnées dans l'image, et vice-versa (d'un pixel de l'image à une direction dans l'espace 3D) [30, 31].

Pour étalonner une caméra, on utilise notamment une image comportant une mire. Les coins de la mire sont détectés et fonctionnent en tant que points de contrôle. Connais-

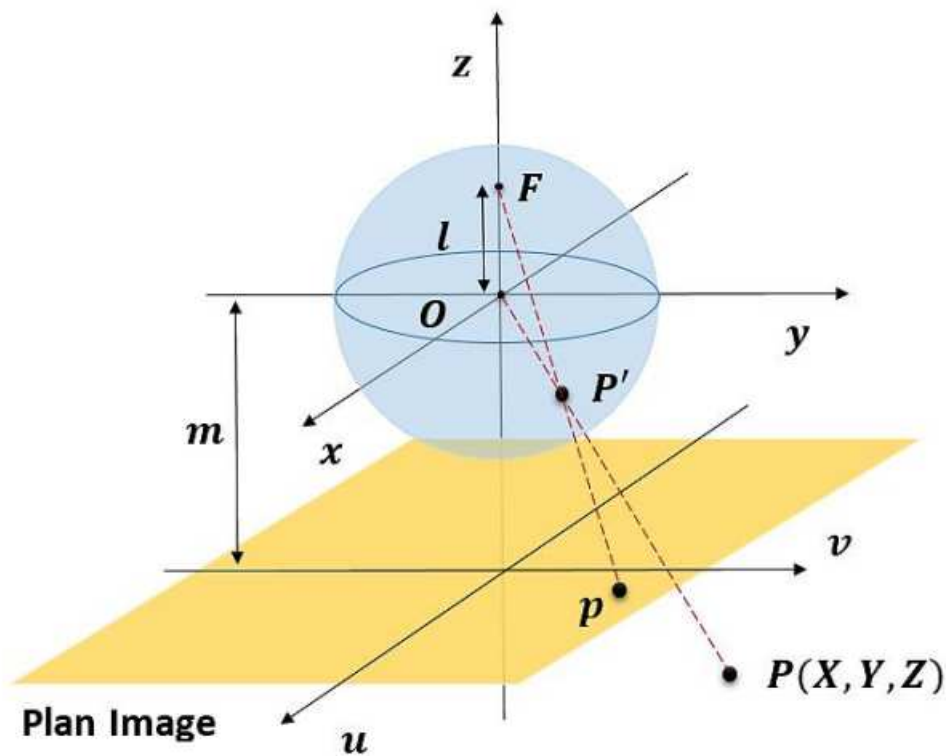


FIGURE 1.6 – Double projection sur une sphère unitaire.

sant a priori les éléments de cette mire (modèle 3D de la mire), on peut obtenir des informations 3D et des erreurs associées. Ensuite, en prenant une série d'images dans différentes positions, les erreurs (accidentelles) sont minimisées et les paramètres intrinsèques sont ainsi obtenus.

Étant donné que les paramètres intrinsèques du capteur fiables et précis sont connus, il existe des méthodes pour mesurer la position et l'orientation relative entre deux vues, et qui sont les paramètres extrinsèques du système.

1.2 La géométrie multi-vues

Le but de cette section est de fournir des définitions et des notions relatives à la géométrie multi-vues, c'est-à-dire les relations qui existent entre la projection des points de la scène dans plusieurs vues. Ces notions sont fondamentales pour la suite de ce manuscrit. Tout d'abord, nous aborderons, en premier lieu, le cas faisant intervenir la géométrie de deux vues, connue par la géométrie épipolaire. Puis en deuxième lieu, nous entamerons d'une façon à être la plus générique possible, les formalismes qui se suivent pour tout type de caméra satisfaisant la contrainte PVU (perspectives et catadioptrique) ou assimilables (panoramique, fisheye, ...).

Un modèle géométrique de caméra modélise le processus de formation de l'image au niveau géométrique et permet d'exprimer les relations existantes entre les points 3D d'une scène et leurs projections sur les images obtenues par la caméra. Le modèle est nécessaire lorsque l'on souhaite connaître la position pixellique dans l'image d'un point de l'espace. La projection d'un point 3D $P(X, Y, Z)$ sur les deux images I_1 et I_2 respectivement en p_1 et p_2 peut s'exprimer à l'aide des matrices de projection associées à chacune des caméras. Si le repère lié à la scène correspond au repère de la première caméra, on obtient les relations suivantes :

$$\begin{aligned} p_1 &\sim \mathbf{M}_1 \mathbf{P}^\top \implies p_1 \sim \mathbf{K}_1 [\mathbf{I} | \mathbf{0}] \mathbf{P}^\top \\ p_2 &\sim \mathbf{M}_2 \mathbf{P}^\top \implies p_2 \sim \mathbf{K}_1 [\mathbf{R} | \mathbf{t}] \mathbf{P}^\top \end{aligned} \quad (1.1)$$

Un modèle peut être commun à un ensemble de caméras, l'adaptation à chacune d'elles se faisant en fonction du lien entre la matrice de projection $\mathbf{M}$ et les grandeurs physiques dont il dépend. Ceux-ci sont communément décomposés en deux types de paramètres :

Les paramètres intrinsèques sont des caractéristiques internes propres à la caméra, qui ne dépendent pas du point de vue. Dans le cas du modèle pinhole, la matrice $\mathbf{K}$ des paramètres intrinsèques, exprime la relation entre l'image et les points exprimés dans un repère lié à la caméra. Ce repère est positionné sur le centre de projection et ses axes x et y sont alignés respectivement avec les axes des lignes et des colonnes u et v du plan image. L'expression de $\mathbf{K}$ peut être plus ou moins complexe, en fonction du nombre de paramètres pris en compte (pixels carrés ou non, orthogonalité des lignes et colonnes, etc ...).

Les paramètres extrinsèques, qu'en à eux, définissent la position et l'orientation de la caméra, par une transformation rigide entre le repère monde, c'est à dire le repère global dans lequel sont exprimées les coordonnées des points 3D, et le repère caméra. Cette transformation Euclidienne comprenant six degrés de liberté de la matrice $\mathbf{M}$. Ceux-ci représentent respectivement la matrice de rotation $\mathbf{R}$ et la translation $\mathbf{t}$ modélisant la pose de la caméra.

1.2.1 Homographie entre deux vues

Étant qu'il existe une relation directe entre un plan et sa projection dans l'image, il existe une homographie liant la projection d'un plan sur deux prises de vue. Le but de l'estimation de l'homographie projective 2D entre les deux plans est d'en extraire la trans-

formation qui lie les deux images contenant les deux plans. Cependant, dans la plupart des cas, il n'est pas possible d'extraire directement le déplacement de la caméra lié à cette homographie projective. Le but est de déterminer une matrice 3×3 non singulière euclidienne relative au plan de référence π , qui caractérise le déplacement de la caméra entre les deux prises de vue. L'obtention des paramètres qui caractérisent le déplacement de la caméra, la rotation et la translation, à partir de la matrice homographique, a fait l'objet de plusieurs travaux. En premier lieu, Tsai dans les années 1980, a publié un travail [32] présentant l'estimation des paramètres de mouvement 3D en décomposant en valeurs singulières la matrice d'homographie. Zhang et Hanson ont proposés dans [33], une méthode basée à la fois sur la décomposition en valeurs singulières et sur la décomposition en valeurs propres de la matrice d'homographie.

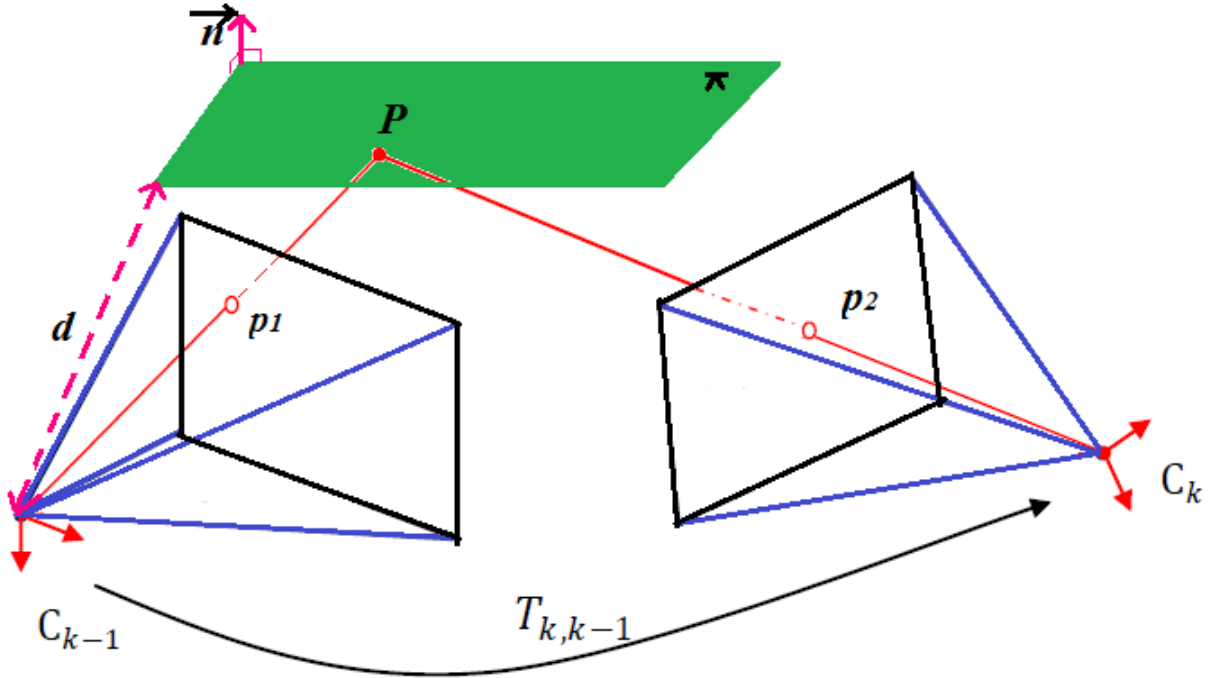


FIGURE 1.7 – Passage entre deux vues exprimé par l'homographie .

Nous illustrons maintenant l'expression d'une homographie planaire. Tout d'abord, rappelons qu'un plan π est défini par un point P et un vecteur normal $\vec{n}$. Notons d , la distance entre ce plan et le centre de projection C_1 (Figure 1.7). Le point P est projeté sur les deux plans image respectivement aux points p_1 et p_2 . Il existe une homographie $\mathbf{G}_{12}$ respectant :

$$p_2 \sim \mathbf{G}_{12} p_1 \quad (1.2)$$

avec cette fois une matrice d'homographie dépendante des paramètres intrinsèques de la caméra $\mathbf{K}$:

$$\mathbf{G}_{12} \sim \mathbf{K} \mathbf{H}_{12} \mathbf{K}^{-1} \quad (1.3)$$

Avec

$$\mathbf{H}_{12} = \mathbf{R} - t \frac{\mathbf{n}^\top}{d} \quad (1.4)$$

1.2.2 La géométrie épipolaire

La géométrie épipolaire correspond au modèle mathématique liant deux images d'une même scène capturée sous deux points de vue différents. La géométrie épipolaire se base sur l'intersection des plans images avec le plan épipolaire π' formé par les centres focaux des caméras et le point 3D P projeté sur les deux images aux points de correspondance p_1 et p_2 (voir figure 1.8). En absence d'a priori sur la scène la position 3D du point P est inconnu entraînant une ambiguïté sur la position du point de correspondance de p_1 dans l'autre image. On peut toutefois affirmer que celui-ci sera nécessairement localisé sur la ligne épipolaire l_2 résultante, $p_1 \mapsto l_2$ (phénomène illustré sur la figure 1.8). Tous les plans épipolaires possibles passent par les épipoles e_1 et e_2 formés par l'intersection de la baseline et des plans images. Cette contrainte est particulièrement utile lorsqu'on doit par exemple trouver des points de correspondance entre images [34]. Mais nous verrons également qu'elle est essentielle dans le calcul de pose des caméras ou pour la rectification d'image [35].

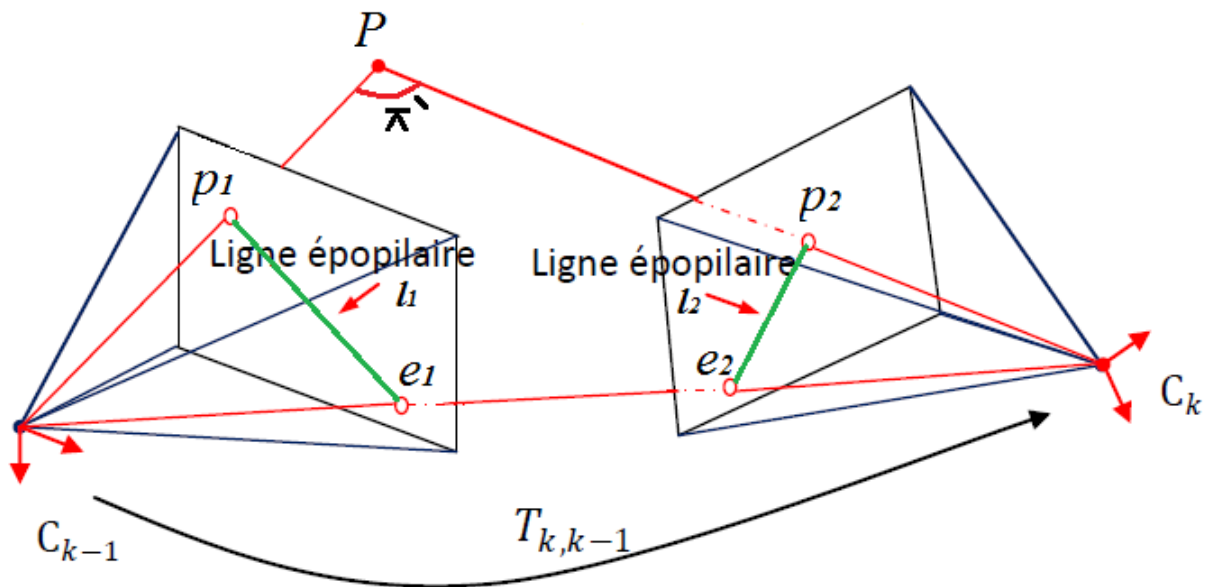


FIGURE 1.8 – Géométrie épipolaire pour le cas planaire.

La contrainte épipolaire est une propriété de la projection centrale jouant un rôle primordial dans la reconstruction euclidienne d'un déplacement d'une caméra entre deux prises de vue. Cette contrainte géométrique peut être appliquée aux capteurs catadioptriques centraux [36]. Svoboda et Padjla sont les premiers à avoir étudié la géométrie épipolaire pour les capteurs catadioptriques centraux à miroirs hyperbolique [37] et parabolique [38]. Une formulation générale de cette géométrie pour tout capteur à projection centrale est présentée dans [39]. En particulier, il a été noté que les capteurs à point central unique calibrés admettent une géométrie épipolaire similaire à celles des caméras conventionnelles [37]. Ce résultat est important car les algorithmes d'estimation des géométries projective et euclidienne conçus pour les caméras conventionnelles, peuvent être assez directement utilisés pour des capteurs catadioptriques centraux. En effet, la matrice d'homographie réalise une transformation linéaire de $\mathbb{S}^2$ dans $\mathbb{S}^2$ des points images relatifs à la projection catadioptrique des points de la scène appartenant au plan de référence, sur la sphère unitaire. Lorsqu'une caméra perspective est considérée, on obtient une relation homographique identique mais réalisant une transformation de $\mathbb{P}^2$ dans $\mathbb{P}^2$ des points images relatifs à la projection perspective des points de la scène appartenant au plan de référence.

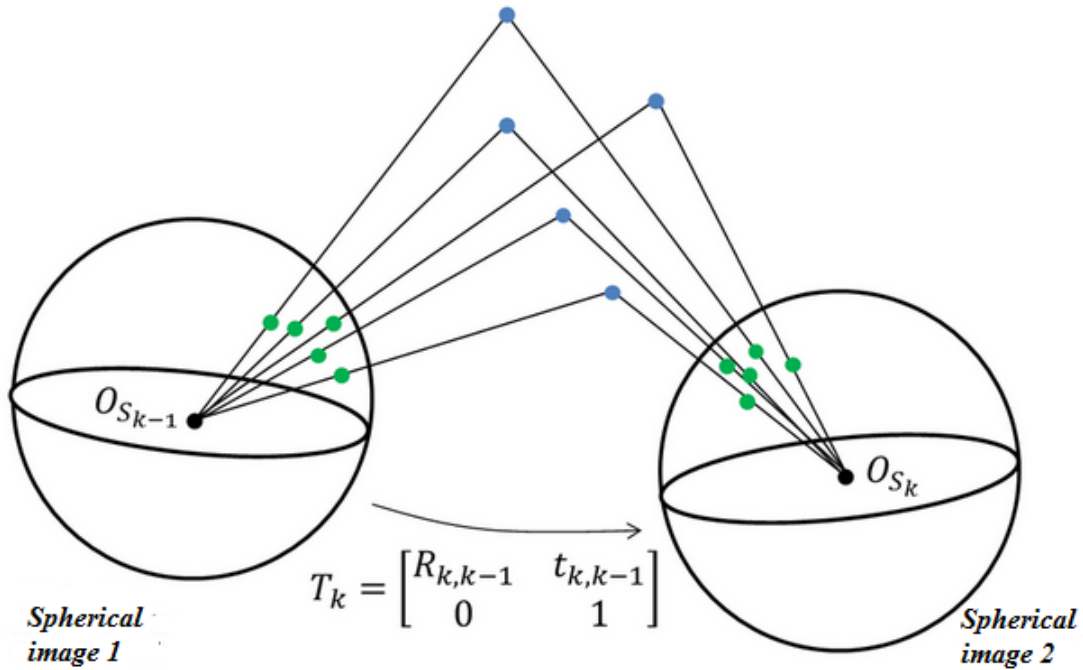


FIGURE 1.9 – Géométrie épipolaire pour le cas sphérique.

1.2.3 La matrice essentielle

La contrainte épipolaire peut être formulée de manière algébrique à l'aide de la matrice essentielle $\mathbf{E}$, qui relie les points de correspondance d'images extraites à partir des

deux vues (voir figure 1.9). Considérons maintenant deux positions du capteur décrites par les deux positions du centre de la caméra C_1 et C_2 . Ces deux positions sont liées par une rotation $\mathbf{R}$ et une translation $\mathbf{t}$. Étant donné un point P de la scène 3D, se projette sur le plan d'image pour les deux positions de la caméra en des points de coordonnées normalisées p_1 et p_2 . Définissons le plan épipolaire par le point de la scène P et les centres de la caméra. Clairement, les projections de P sur les images dans les deux positions appartiennent également au plan épipolaire. Étant donné le point d'image p_1 dans le système de coordonnées de la caméra C_1 , sa position p_2 dans le système de coordonnées C_2 est donnée par :

$$p_2 = \mathbf{R}p_1 + \mathbf{t} \quad (1.5)$$

Étant donnée que p_2 et $p_2^\top [\mathbf{t}]_x$ sont perpendiculaires, la multiplication de 1.5 par $p_2^\top [\mathbf{t}]_x$ donne :

$$p_2^\top [\mathbf{t}]_x \mathbf{R}p_1 = p_2^\top \mathbf{E}p_1 = 0 \quad (1.6)$$

où la matrice essentielle 3×3 $\mathbf{E} \sim [\mathbf{t}]_x \mathbf{R}$ et $[\mathbf{t}]_x$ est la matrice antisymétrique du vecteur de translation $\mathbf{t}$. De nombreux algorithmes permettent d'estimer la matrice essentielle à partir d'un ensemble de correspondances entre les images ont été proposés. Nister dans [40] a démontré qu'à partir de cinq correspondances, le calcul de la matrice essentielle peut être établi.

1.2.4 Estimation du mouvement par la matrice essentielle

L'estimation du mouvement est une étape primordiale, le mouvement de la caméra est calculé entre l'image courante et précédente. Par concaténation de tous les mouvements, on peut rétablir la trajectoire complète de la caméra. Dans cette partie, nous rappelons comment estimer la transformation homogène $\mathbf{T}_k$, entre deux images I_{k-1} et I_k . La relation géométrique entre les images d'une caméra calibrée est décrite par la matrice essentielle $\mathbf{E}$, qui contient les paramètres du mouvement de la caméra à un facteur d'échelle inconnu pour la translation :

$$\mathbf{E}_k \propto \mathbf{t}_{xk} \mathbf{R}_k \quad (1.7)$$

$$t_k = \begin{bmatrix} t_x & t_y & t_z \end{bmatrix}^\top \text{ et } \mathbf{t}_{\mathbf{x}k} = \begin{bmatrix} \mathbf{0} & -t_z & t_y \\ t_z & \mathbf{0} & -t_x \\ -t_y & t_x & \mathbf{0} \end{bmatrix}$$

Le symbole $\propto$ est utilisé pour indiquer que l'équivalence est à un scalaire multiplicatif. La matrice essentielle peut être calculée à partir de correspondance de primitives, et la rotation et la translation peuvent être directement extraites de $\mathbf{E}$. La principale propriété de l'estimation de mouvement basée 2D/2D est la contrainte épipolaire, qui détermine la droite sur laquelle le point d'intérêt $p'^\top$ correspondant p se trouve dans l'autre image (figure 1.10).

Cette contrainte peut être formulée par :

$$p'^\top \mathbf{E} p = 0$$

La matrice essentielle peut être calculée à partir des correspondances des primitives en utilisant la contrainte épipolaire. Le cas minimal est d'utiliser cinq correspondances pour trouver une solution [41]. L'algorithme de 5-points de Nister est devenu un standard pour l'estimation de mouvement 2D/2D en présence de valeurs aberrantes. Une solution simple et directe pour n points non coplanaires $n \geq 8$ est donnée par l'algorithme de Longuet-Higgins 8-points [42].

À l'aide de l'équation (1.7), on trouve directement un moyen pour calculer $\mathbf{t}$. Notamment, $\mathbf{t}_k$ est effectivement un vecteur du noyau de $\mathbf{E}$. Cette solution nous donne $\mathbf{t}_k$ seulement à un facteur d'échelle près. Tout ce que l'on peut obtenir, est donc la direction du déplacement. Le calcul de la matrice de rotation $\mathbf{R}$ est un peu plus compliqué que celui de $\mathbf{t}$, même si, ayant calculé le vecteur $\mathbf{t}$, l'équation (1.7) permettrait d'établir un système d'équations linéaires. Dans la suite, nous présentons donc une méthode pour extraire $\mathbf{R}$ et $\mathbf{t}$, basée sur la décomposition en valeurs singulières de la matrice essentielle. Soit cette dernière donnée par :

$$\mathbf{E} = \mathbf{U} \mathbf{S} \mathbf{V}^\top$$

Où $\text{diag}(\mathbf{S}) = \text{diag}\{s, s, 0\}$, ce qui signifie que les valeurs singulières première et deuxième sont égales et la troisième est égale à zéro.

Pour obtenir une matrice $\mathbf{E}$ qui satisfait les contraintes, la solution doit être projetée sur l'espace de matrices essentielles valides. La matrice essentielle de projection est :

$$\mathbf{E} = \mathbf{U} \text{diag}\{1, 1, 0\} \mathbf{V}^\top$$

À partir de l'estimation de la matrice essentielle $\mathbf{E}$, on peut extraire les parties de rotation et de translation. Le vecteur t est alors donné par la troisième colonne de $\mathbf{V}$ (il est facile de

vérifier que $\mathbf{E} \cdot \mathbf{t}$ donne alors le vecteur nul). Quant à la matrice de rotation, il existe deux solutions :

$$\mathbf{R} = \mathbf{U} \begin{bmatrix} \mathbf{0} & \pm \mathbf{1} & \mathbf{0} \\ \pm \mathbf{1} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{1} \end{bmatrix} \mathbf{V}^T$$

Si $\det \mathbf{R} = -1$, multiplier $\mathbf{R}$ par -1 .

Finalement, l'algorithme de huit points fonctionne à la fois avec des caméras calibrées (perspective ou omnidirectionnel) et non calibrées (uniquement perspective), alors que l'algorithme de cinq points suppose que la caméra (point de vue ou omnidirectionnel) est calibrée [43].

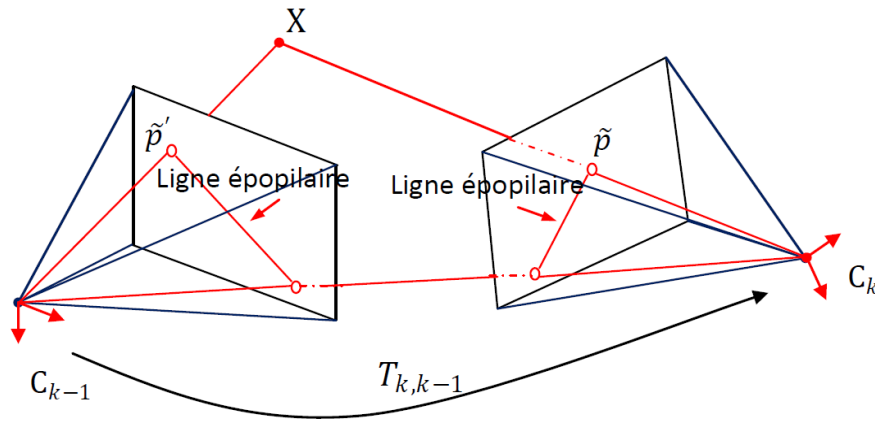


FIGURE 1.10 – Une illustration de la contrainte épipolaire.

1.2.5 L'estimation robuste du mouvement

La présence de fausse mise en correspondances entre deux nuages de points d'intérêts issus de deux images, peut ralentir la convergence de l'algorithme d'estimation de $\mathbf{R}$ et $\mathbf{t}$ elle peut même causer la convergence de l'algorithme vers une mauvaise solution. Parmi les approches classiques qui existent dans la littérature pour rejeter les outliers, l'algorithme RANSAC (RANdom SAMpling Consensus) [44] est le plus populaire. Il s'agit d'une méthode itérative permettant d'estimer les paramètres d'un modèle mathématique à partir d'un ensemble de données observées qui contient des valeurs aberrantes [44]. Dans le cas de l'odométrie visuelle, RANSAC est utilisé pour déterminer les correspondances les plus correctes qui permettent d'estimer la transformation entre les deux images. Autrement dit, RANSAC permet d'éliminer les outliers (les fausses appariements). L'algorithme consiste à tirer aléatoirement parmi l'ensemble des correspondances 5 candidats qui permettent d'estimer la matrice essentielle [41] et de comptabiliser aux autre élé-

ments qui vérifient à la tolérance fixée au préalable après chacune des estimations aléatoires. La solution est celle qui regroupe le plus grand nombre d'éléments sous le même modèle. L'estimation de la matrice essentielle est alors affinée en tenant en compte seulement de l'ensemble les inliers (les vraies correspondances).

1.2.6 La matrice fondamentale

De l'équation 1.6, les points d'image p_1 et p_2 peuvent être liés aux positions de pixel u_1 et u_2 respectivement, par l'intermédiaire de la matrice de calibration de la caméra $\mathbf{K}$:

$$p_1 \sim \mathbf{K}^{-1} u_1 \quad (1.8)$$

Cela signifie que la contrainte épipolaire (équation 1.6), peut être réécrite en termes des coordonnées en pixel :

$$u_2^\top \mathbf{F} u_1 = 0 \quad (1.9)$$

telle que $\mathbf{F} \sim (\mathbf{K}^{-1\top} \mathbf{E} \mathbf{K}^{-1})$ est une matrice 3×3 qui est la matrice fondamentale. On peut l'estimer linéairement par huit points de correspondance ou plus. Une attention considérable a été accordée au problème de l'estimation précise de la matrice fondamentale à partir des images bruitées et en présence des fausses appariements dans [45].

1.3 L'estimation de l'égo-mouvement

L'estimation de l'égo-mouvement, qui est également connu par l'odométrie visuelle (VO), est le procédé permettant de déterminer le mouvement d'une plateforme mobile (par exemple, véhicule, robot, etc.) en utilisant les images obtenues à partir d'une ou plusieurs caméras embarquées sur la plateforme. L'analogie avec l'odométrie des roues provient de la nature de calcul incrémental du mouvement relatif. Développé depuis les années 1980, ce procédé est employé dans de nombreux domaines dont la navigation robotique autonome, la réalité augmentée et la stabilisation des vidéos numériques. L'avantage de l'odométrie visuelle par rapport à l'odométrie des roues, ce n'est pas seulement que l'odométrie des roues peut être affectée par le mauvais contact des roues sur un terrain accidenté, mais les entrées visuelles permettent une variété de tâches très utiles. Par exemple, le robot peut générer une reconstruction 3D, détecter des objets d'intérêt, classi-

fication des terrains, etc. Toutes ces tâches peuvent être réalisées avec une consommation peu coûteuse et faible, ce qui rend l'OV idéale pour les plates-formes robotisées.

Par ailleurs, il y a différents facteurs à considérer afin que l'OV fonctionne dans les meilleurs conditions. Les deux facteurs majeurs exigés sont que la scène devrait être suffisamment éclairée et qu'il devrait y avoir assez de texture dans la scène. En outre, la présence de chevauchement des scènes captées par les images consécutives, va permettre de calculer le mouvement à partir des similitudes existantes entre ces images.

Le problème d'estimation de la pose relative d'une caméra et la structure de la scène, à partir d'un ensemble d'images (calibré ou non calibré) est connu dans la communauté de vision par ordinateur par SfM (Reconstruction 3D à partir d'images prises sous différents points de vue). Ses origines remontent à des travaux dans [42] et [46]. L'OV est un cas particulier de la SfM [43]. L'algorithme SfM est plus général et aborde le problème de la reconstruction 3D. Il détermine à la fois la structure et la pose de caméra à partir d'un ensemble d'images ordonnées ou non ordonnées. La structure finale et la pose de la caméra sont généralement raffinées avec une optimisation hors ligne (par exemple, l'ajustement de faisceaux [47, 48]), dont le temps de calcul augmente avec le nombre d'images. Tandis que, l'OV s'intéresse seulement à l'estimation du mouvement 3D de la caméra pour chaque image acquise de manière séquentielle et en temps réel. L'ajustement de faisceau peut être utilisé pour affiner l'estimation locale de la trajectoire. Le problème de l'estimation de l'égo-mouvement d'un véhicule, à partir de l'entrée visuelle seulement, a été introduit dans les années 1980 par Moravec [49]. Il est intéressant de constater que la plupart des premières recherches sur l'OV [50] ont été appliquées sur les Rovers planétaires qui ont été motivés par le programme d'exploration de Mars de la NASA.

L'OV peut être classée en deux catégories selon la configuration utilisée ou bien la méthode appliquée pour estimer le mouvement relatif. Dans la première catégorie, on trouve les configurations possibles existantes conformément au type de la camera portée par la plateforme mobile : mono ou stéréo ainsi que les caméras RGB-D. La deuxième catégorie, concerne le type de méthodes appliquées pour en extraire le mouvement. On peut distinguer deux types d'approches dans cette catégorie : les méthodes basées primitives qui sont extraites et suivies dans les deux images successives, et les méthodes globales (ou directes) basées sur l'apparence de toute l'image où les intensités de pixels sont utilisées pour minimiser les erreurs photométriques entre les images

1.3.1 Les capteurs de l'odométrie visuelle

1.3.1.1 Capteur Stéréoscopique

Les systèmes stéréoscopiques reposent sur l'utilisation de deux caméras séparées de quelques dizaines de centimètres. Le banc stéréo permet généralement une reconstruction dense de la scène, qui peut être exploitée via des mesures 3D. De nombreux travaux de recherche effectués dans l'OV ont été réalisées en utilisant des caméras stéréoscopiques s'appuyant sur le travail de Moravec [49]. Matthies et Shafer [51, 52], ont utilisé un système binoculaire et la procédure de Moravec pour détecter et suivre les coins. Olson et Matthies [50, 53] ont montré que l'utilisation d'une seule caméra pour estimer l'égo-mouvement provoque des erreurs d'estimation qui augmente en fonction de la distance parcourue. Une approche différente de l'estimation de mouvement et l'élimination des valeurs aberrantes pour un robot mobile tout-terrain a été proposées par Milella et Siegwart [54]. Ils ont utilisé l'approche Shi-Tomasi [55] pour la détection de coin. Puis, ils ont conservé ces points avec une grande confiance dans la carte de la disparité. L'estimation de mouvement a ensuite été résolu en utilisant d'abord un procédé des moindres carrés, puis l'algorithme itératif de plus proche (ICP) [56]. Ce dernier est un algorithme populaire pour les enregistrements 3D à partir de balayages laser. Pour des raisons de robustesse une étape d'élimination des valeurs aberrantes (fausse association) a été intégrée dans l'ICP.

Les travaux mentionnés auparavant présentent un aspect commun : la triangulation de points 3D pour chaque paire d'image stéréoscopique, et le mouvement relatif est obtenu par un algorithme d'alignement de point 3D/3D. Une approche d'implémentation complètement différente a été proposée en 2004 par Nister et al [40], qui présentent la première validation de l'OV pour une longue distance en temps réel, avec un rejet robuste des valeurs aberrantes d'appariement. Le mouvement relatif calculé est vu comme un problème 3D/2D. Un système d'estimation de mouvement différent a été ensuite introduit par Comport et al. [57]. Au lieu d'utiliser l'alignement de points 3D/3D ou les techniques d'estimation de la pose de caméra 3D/2D, le tenseur quadri-focal a été utilisé, pour calculer le mouvement à partir de la transformation des correspondances 2D/2D d'images sans appliquer l'étape de triangulation. L'avantage d'utiliser directement les points 2D bruts réside dans un calcul plus précis de mouvement.

1.3.1.2 Capteur Monoculaire

les systèmes monoculaires requièrent une seule caméra pour la localisation, sont utilisés généralement pour des raisons de simplicité de mise en œuvre, ou par contraintes

applicatives. La différence par rapport à un système stéréoscopique est que dans l'OV monoculaire, le mouvement relatif devrait être calculer à partir des images acquises par la caméra elle-même. Comme l'échelle absolue est inconnue, la distance entre les deux premières poses de la caméra est généralement fixée à une constante. Lorsque une nouvelle image arrive, l'échelle relative et la pose de la caméra sont déterminés par les deux premières images en utilisant soit la connaissance de la structure 3D ou le tenseur tri-focal [48]. Des résultats intéressants avec une seule caméra sur de longues distances (jusqu'à plusieurs kilomètres) ont été obtenus dans la dernière décennie en utilisant à la fois des caméras perspectives et omnidirectionnelles [58, 59].

Allant des méthodes basées points d'intérêts (primitives) [40] et les méthodes fondées sur l'apparence [60] ont été les deux proposées afin d'estimer le mouvement via les cameras monoculaires. D'autres techniques hybrides utilisent à la fois des primitives et de l'apparence comme dans [4]. Cependant, plusieurs travaux d'OV ont été spécialement conçus pour les véhicules où le mouvement est contraint qu'il soit sur un plan. L'avantage est de diminuer le temps de calcul et d'améliorer la précision de l'estimation du mouvement [61, 62]. Zhang a fourni dans [63], le point initial de toute procédure de la structure de mouvement basée sur la matrice essentielle. Cette méthode identifiée certaines contraintes utilisées lors de la reconstruction et se base sur le raffinement des paramètres de mouvement extraits de la matrice essentielle. Svoboda et al, ont proposé une méthode également basée sur la matrice essentielle pour extraire des informations de mouvement à partir de deux images sphériques [64]. L'accent est mis cependant sur l'outil d'extraction du mouvement efficace d'une caméra de modèle sphérique, qui permet la distinction entre les pures translations et rotations.

Dans [65], une méthode pour calculer la pose de caméra à partir d'une séquence d'images sphériques a été présentée. L'approche proposée est quasi similaire à celle qui utilise la matrice essentielle pour trouver la géométrie épipolaire des paires images initiales. Toutefois, la différence principale réside dans l'utilisation de l'estimation grossière de la trajectoire de la caméra en tant qu'une entrée supplémentaire au système pour calculer les positions de caméra par la technique de la carte de corrélation. Les techniques basées points d'intérêt et des descripteurs locaux, telles que SIFT, ont été adaptées aux images omnidirectionnelles en raison de leur performance sur les images planaires [66, 4]. Récemment, plusieurs travaux ont été réalisés pour développer des algorithmes spécifiquement conçus pour traiter les images sphériques dans le but d'estimer le déplacement de la caméra [67, 68].

Bloesch et al. présentent dans [69] un algorithme d'odométrie visuelle-inertielle (RO-VIO) qui utilise directement les erreurs d'intensité des pixels des patches d'image. Pour cela, les primitives à patch de plusieurs niveaux, sont détectées, suivies et couplées au

filtre de Kalman étendu EKF. L'estimation des points de repères 3D est robocentrique c'est-à-dire leurs emplacements sont toujours estimés directement dans le repère du robot, qui évolue instantanément, au lieu d'une représentation dans le repère de référence fixe. En outre, la position d'un point de repère est paramétrée par un vecteur de roulement et une distance. Par conséquent, le processus d'initialisation pourrait être omis. L'approche présentée dans ce travail a permis d'obtenir des performances de suivi précises tout en présentant un niveau de robustesse très élevé.

Makadia et al, ont présenté dans [70, 71] une approche pour estimer la rotation entre deux images sphériques. Elle utilise les propriétés du groupe spécial orthogonal des rotations $SO(3)$, en se basant sur le contenu de l'image. L'une des images est mise en rotation au cours de la recherche jusqu'à ce qu'il corresponde à l'autre dans l'espace des coefficients harmoniques. Le critère de correspondance utilisé pour estimer les rotations est basé corrélation. L'optimisation sur le groupe $SO(3)$ est également discutée dans [72]. Une autre alternative des méthodes d'estimation du mouvement sans correspondance entre les primitives, proposée par Makadia [73, 74], repose sur la transformée de Radon où toute l'image est utilisée pour estimer le mouvement relatif entre deux prises de vue, à travers l'analyse des harmoniques sphériques.

1.3.1.3 Capteur RGB-D

Le lancement des caméras RGB-D a offert une alternative puissante pour la perception 3D à faible coût. En effet, le capteur RGB-D a suscité un intérêt particulier dans la communauté de la vision par ordinateur, et de la robotique notamment pour le SLAM visuel [75]. Son introduction dans la résolution du problème a donné lieu à plusieurs travaux que nous pouvons en citer quelques applications dans ce qui suit :

Huang et al. ont proposé dans [76] l'un des systèmes initiaux d'odométrie et de cartographie basées caméra RGB-D, appelé l'odométrie rapide à partir de la vision (FOVIS). Le principe de la technique est basé sur le calcul de la pose initiale à partir de données visuelles, puis le raffinement de l'estimation est obtenu en utilisant des informations 3D. La détection de points d'intérêt dans les images et le suivi de ces points lors du déplacement de la caméra, est assuré par l'algorithme FAST. La solution FOVIS a été proposée afin d'améliorer l'odométrie visuelle, dans le but de remplacer l'alignement des images consécutives par un alignement par rapport à une image clef suffisamment proche de l'image courante. En effet, l'utilisation de cette technique permet de réduire considérablement la dérive de l'odométrie visuelle surtout quand la caméra effectue des mouvements lents.

Parmi les premières méthodes d'odométrie visuelle dense minimisant l'erreur des in-

tensités des pixels, nous citons les travaux de Kerl et al. [77]. Avec l'apparition des caméras 3D de type RGB-D, ils ont proposé une méthode qui repose directement sur la minimisation des erreurs photométriques entre deux images consécutives. Cette méthode est différente aux méthodes basées primitives, car ne nécessite pas de détection et de la mise en correspondance des primitives. En effet, elle s'appuie sur l'hypothèse de la consistance photométrique et utilise plusieurs variantes des méthodes de moindres carrés pondérées basées sur M-estimateurs pour réduire l'influence des points aberrants dans la procédure de l'estimation. Une étude comparative de cette approche et de méthodes alternatives est présentée dans [78].

Zhang et al. [79, 80, 81, 82] ont publié divers travaux sur l'OV en utilisant la profondeur soit de la caméra RGB-D ou du scanner laser. Les auteurs proposent de détecter les coins de Harris uniformément sur l'image, puis le suivi de ces primitives est effectué à l'aide du flot optique d'une image à l'autre. En se basant sur la transformation d'un corps rigide, chaque coin avec sa profondeur si elle est connue contribue à deux équations, alors que dans le cas où elle est inconnue contribue à une équation. En reportant l'ensemble des équations, la résolution du système est effectuée en utilisant la méthode Levenberg-Marquardt (LM) qui permet d'en extraire la translation et la rotation. La carte de profondeur est acquise par la caméra RGB-D ou le scanner laser, mais si la profondeur de la primitive est inconnue, on recherche le patch planaire local contenant la primitive elle-même et trois d'autres primitives, dont leurs profondeurs sont connues. En outre, si la primitive est suivie à longue distance, la profondeur est obtenue par triangulation. Le mouvement image par image est raffiné par ajustement du faisceau.

Zheng et al. dans [83] et [84], se focalisent sur l'application d'un vol autonome dans environnement dégradé à travers un capteur RGB-D. Dans [83], ils comparent et évaluent plusieurs méthodes existantes dans l'état de l'art, d'odométrie visuelle en temps réel, à faire face aux différents scénarios difficiles. Les méthodes étudiées incluent l'OV basée primitives et l'OV dense. Les scénarios d'évaluation contiennent moins de primitives ou même pour certains cas sans primitives. En analysant chaque méthode, ils décrivent les avantages et les inconvénients de chaque catégorie en ce qui concerne leur robustesse et leur précision. Dans [84], ils proposent encore une méthode rapide d'odométrie visuelle fonctionnant dans un environnement dégradé en fonction de l'équation de contrainte de taux de portée et la métrique d'erreur photométrique. Un filtre à particule est déployé pour réduire la dérive à l'aide d'une carte 3D donnée.

1.3.2 Les méthodes de l'odométrie visuelle

L'odométrie visuelle peut être considérée comme l'estimation du mouvement effectué par une caméra mobile par rapport à un repère initial. Ce repère peut être par exemple la première image d'une séquence, ou alors un objet connu dans la scène, suivi dans les autres images.

En effet, pour deux images obtenues à deux instants successifs par la caméra I_n et I_{n+1} , l'odométrie visuelle permet d'estimer le déplacement réalisé par la caméra entre ces deux instants. Ce processus d'estimation doit être établi à chaque acquisition d'une nouvelle image, et par conséquent la trajectoire effectuée par la caméra durant le déplacement est obtenue.

L'OV est un sujet qui a été considérablement traité dans la littérature. Ce sujet étant devenu de plus en plus actif ces dernières années, de nombreuses méthodes d'OV ont été proposées. Cependant, toutes ces méthodes peuvent être classées dans deux grandes catégories [85] : La première catégorie appelée éparses regroupe toutes les méthodes qui utilisent un sous ensemble de primitives extraites de l'image pour estimer le mouvement de la caméra. La deuxième catégorie appelée directe ou dense, regroupe toutes les méthodes qui utilisent l'information intégrale existante dans l'image en estimant le mouvement.

Dans la suite, nous fournissons le principe de fonctionnement de chaque catégorie des méthodes d'estimation de mouvement.

1.3.2.1 Méthodes éparses

Les méthodes éparses (feature-based) se basent essentiellement sur l'extraction de points d'intérêts et leurs mises en correspondances précises. Ce type d'approches cherchent à extraire des primitives visuelles locales dans les images. L'objectif de cette extraction des primitives est de localiser les points les plus importants de la scène qui présentent une forte variabilité dans le signal visuel. En effet, il s'agit de points riches en termes d'informations sélectionnées selon un critère précis, on peut citer comme exemple les coins de Harris [86].

Pour la description des primitives, les points d'intérêt détectés sont décrits localement en se basant sur leurs voisinages par des algorithmes de description locale. Par ailleurs, on a souvent besoin de descripteurs robustes, qui sont à la fois capables d'identifier un point parmi plusieurs milliers d'autres, et assez robustes pour conserver ce pouvoir de description malgré des changements d'apparence.

Puis, l'étape d'appariement ensuite doit être entamée, c'est à dire mettre en correspondance un à un les points d'intérêts de chaque image. En général, la similarité est définie selon le coût obtenu en calculant une des distances, euclidienne, de Manhattan, ou de Hamming, entre les descripteurs qui décrivent le voisinage des points d'intérêt. Pour rendre cette étape d'appariement plus robuste, qui est souvent critique pour ce type de méthodes, il est possible d'utiliser des descripteurs locaux robustes aux changements d'échelle et aux rotations, tels que SIFT [87] et SURF [88].

Le mouvement relatif entre les images est en général obtenu par une méthode robuste de type RANSAC [89]. RANSAC est une méthode itérative permettant l'estimation robuste des paramètres d'un modèle contre les mauvais appariements dans les données d'entrée. En effet, les méthodes robustes peuvent être utilisées pour diminuer l'effet des données aberrantes à partir d'une estimation initiale. En utilisant de telles méthodes, même si l'ensemble de données contient de nombreuses valeurs aberrantes, l'algorithme converge avec une grande probabilité vers l'estimation calculée uniquement à partir des bons appariements. Dans le contexte de l'estimation de la pose, ce genre d'approches sont les plus fréquemment utilisées pour raffiner l'estimation initiale de la transformation rigide entre les primitives. Cela revient d'estimer la transformation géométrique séparant deux images.

1.3.2.1.1 Le processus d'OV Considérant une série d'images prises par un système de vision rigidement fixé sur le robot mobile, dans des instants de temps k . L'ensemble d'images prises à l'instant k est noté $I_{0:n} = \{I_0, \dots, I_n\}$.

Deux positions successives de caméra à des instants $k-1$ et k sont reliées par la transformation du corps rigide $T_k \in \mathbb{R}^{4 \times 4}$ de la forme suivante :

$$T_k = \begin{bmatrix} \mathbf{R}_k & \mathbf{t}_k \\ \mathbf{0} & \mathbf{1} \end{bmatrix}$$

Où $\mathbf{R}_k \in \text{SO}(3)$ est la matrice de rotation, et $\mathbf{t}_k \in \mathbb{R}^{3 \times 1}$ est le vecteur de translation. L'ensemble $\mathbf{T} = \{\mathbf{T}_1, \dots, \mathbf{T}_n\}$ contient tous les déplacements subséquents.

Finalement, l'ensemble des poses de la caméra $\mathbf{C}_{0:n} = \{\mathbf{C}_0, \dots, \mathbf{C}_n\}$ contient les transformations de la caméra par rapport aux coordonnées initiales du repère à l'instant $k=0$. La pose courante $\mathbf{C}_n$ peut-être calculer en concaténant les transformations $\mathbf{T}_k$ telle-que $k = (1, \dots, n)$, et par conséquent, $\mathbf{C}_n = \mathbf{C}_{n-1} \mathbf{T}_n$, avec $\mathbf{C}_0$ étant la pose de la caméra à l'instant $k=0$. La tâche principale de l'OV est de calculer les transformations relatives $\mathbf{T}_k$ à partir des images I_k et I_{k-1} et puis les concaténer, pour récupérer la trajectoire complète $\mathbf{C}_{0:n}$ de la caméra. Cela signifie que l'OV récupère la trajectoire progressivement, d'une

pose à autre.

Comme mentionné auparavant, il existe deux approches principales pour calculer le déplacement relatif T_k . Les méthodes basées apparence (ou globales), qui utilisent l'intensité de tous les pixels comme information dans les deux images d'entrées. Et les méthodes basées primitives (points d'intérêts), qui utilisent uniquement les points d'intérêts extraits et suivis à travers les images captées.

L'organigramme de l'algorithme d'OV est résumé dans la figure 1.11. Pour chaque nouvelle image I_k , les deux premières étapes consistent à détecter et apparier les points d'intérêts avec celles de l'image précédente I_{k-1} . Notons qu'il existe une différence entre la mise en correspondance et le suivi des points d'intérêts. La première consiste à détecter les points d'intérêts indépendamment de toutes les images, puis les apparier en se basant sur quelques similarités métriques. La seconde consiste à suivre des points d'intérêts (primitives) dans des images successives en utilisant une technique de recherche locale. La troisième étape consiste à calculer le mouvement relatif T_k , entre les images I_{k-1} et I_k . La pose de la caméra C_k est alors calculée par concaténation de T_k avec la pose précédente. Enfin, une technique itérative d'optimisation doit être établie (RANSAC) pour obtenir une estimation plus précise de la trajectoire locale.

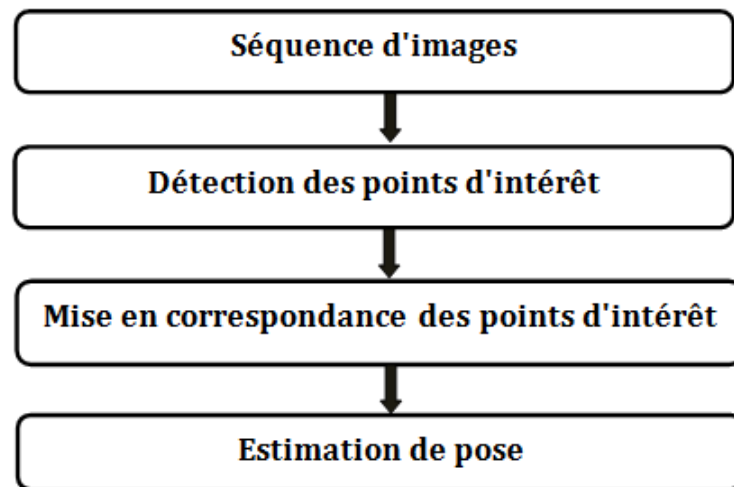


FIGURE 1.11 – Algorithme d'un système d'OV basé primitives

1.3.2.1.2 Évaluation des détecteurs et descripteurs de points d'intérêt La performances et l'efficacité des algorithmes sont évaluées via deux aspects : le pouvoir ainsi la précision de la détection des points caractéristiques et de l'exactitude de l'appariement. Toutefois, ces évaluations comparent les détecteurs et les descripteurs en présence de la vérité terrain entre les images dites requêtes et l'image de référence. Afin d'évaluer les performances, on étudie indépendamment plusieurs critères tels que : l'invariance aux transformations géométriques (la rotation, changement d'échelle et point de vue). Ces invariances sont assurées soit par le détecteur, soit par le descripteur, ou bien par les deux.

De nombreuses évaluations des détecteurs et descripteurs locaux ont été menées pour plusieurs applications différentes, notamment la reconnaissance d'objets, le tracking, etc. Selon les applications pour lesquelles la détection et la description des points d'intérêt sont sollicitées, certaines invariances sont importantes. Vue l'importance des points d'intérêts pour l'odométrie visuelle, il est donc indispensable que les détecteurs et les descripteurs de points d'intérêt sélectionnés, soient invariants aux transformations suivantes : la rotation, l'échelle et le changement de point de vue. Pour ces raisons, nous avons mené dans [90] une analyse et comparaison des performances des deux détecteurs et descripteurs les plus réputées dans le domaine de la vision par ordinateur, et qui sont SIFT et SURE. Il s'agit d'une étude comparative des propriétés et des performances de ces détecteurs de points d'intérêt utilisés pour l'odométrie visuelle. Le travail est décrit dans l'annexe A.3.

Dans la littérature, de nombreux détecteurs et descripteurs très récents ont été proposés, notamment les descripteurs binaires tels que FREAK [91], BRISK [92], etc. Nous n'avons pas pris en compte ces derniers lors de cette évaluation, car ce travail a été effectué au cours de la même période de la proposition de ce nouveau type de détecteurs et descripteurs binaires. Cependant, il y a d'autres travaux qui ont apparus très récemment où ils ont mené l'étude des performances de ce type de descripteurs [93].

1.3.2.2 Méthodes directes

Les méthodes directes, n'ont pas de phase de sélection de points d'intérêts ou de primitives visuelles. Le mouvement de la caméra est directement obtenu en minimisant l'erreur d'intensités commune aux deux images à l'aide d'une transformation paramétrique. Dans ce cas l'estimation du mouvement et la mise correspondance des pixels s'effectuent simultanément lors de l'optimisation.

La modélisation paramétrique du mouvement global entre deux images nécessite, le plus souvent, une minimisation d'une fonction de coût définie sur toute l'image. Elle est composée au minimum d'un terme de données dense. Les sections qui suivent présentent les principales techniques de détermination du mouvement global dans une séquence d'images.

1.3.2.2.1 Flot optique Considérons une séquences d'images $I(x(t), y(t), t)$. Le mouvement 2D apparent dans l'image, plus communément appelé **flot optique**, est la vitesse de déplacement $\vec{v}(x(t), y(t), t)$ des pixels dans la séquence d'images. Le flot optique est une variable cachée et n'est pas accessible que par la variations d'intensités en fonction

du temps. C'est pourquoi il est nécessaire de faire une hypothèse sur la variation de cette intensité quand t , le temps, varie. Cette hypothèse, dite contrainte du flot optique, permettant de calculer le flot optique, elle stipule que l'intensité d'un pixel dans une image ne varie pas au fil du temps. En effet un changement de luminosité dans la scène se traduit par un mouvement dans le calcul du flot optique.

Ces contraintes rendent l'hypothèse forte, en considérant que les variations temporelles de luminance sont uniquement dues au mouvement. Cette hypothèse possède des limitations évidentes et ne tient pas compte d'un certain nombre de phénomènes tels que les variations d'illumination, les occultations ou le bruit, et plus généralement de tout ce qui peut produire des variations assez considérables d'illumination de la scène. Afin de prendre en compte ces phénomènes, l'hypothèse de conservation de la luminance peut être enrichie de manière à considérer les variations de luminance et de contraste. Cette hypothèse est connue sous le nom d'Équation de Contrainte du Mouvement Apparent (ECMA), sa formulation mathématique est :

$$\vec{\nabla}I((x, y), t) \cdot \vec{v}((x, y), t) + \frac{\partial I((x, y), t)}{\partial t} = 0 \quad (1.10)$$

Où $\vec{\nabla}I$ est le gradient spatial de l'image.

De nombreuses études ont été menées pour résoudre ou bien contourner la contrainte de flot optique. Ces méthodes regroupées en deux familles :

1.3.2.2.2 Méthodes fréquentielles Ces méthodes étudient l'ECMA dans le domaine fréquentiel. Considérons la séquence d'images $I(x, y, t)$ ayant la transformée de Fourier $\hat{I}(f_x, f_y, f_t)$. Les séries de Fourier sont parfaitement conçues à la représentation de fonctions continues échantillonnées. Cette décomposition fournit une représentation compacte, et un signal étant en général décrit par relativement peu de coefficients de Fourier. La décomposition en séries de Fourier opère une analyse multi-résolution du signal. La description est alors globale, et permettrait éventuellement une description du mouvement à deux échelles spatiales et temporels. Cette méthode a pour conséquence de lisser le flot simultanément dans l'espace fréquentiel et temporel. Ce qui peut parfois engendrer des problèmes en estimant le mouvement pour des séquences d'images rapides et saccadées.

1.3.2.2.3 Méthodes de Block-matching Ce type de technique est utilisé pour le suivi d'une surface planaire Lucas et Kanade [60], ou alors le suivi d'un ensemble de surfaces planaires Mei et al. [94], Shi et Tomasi [95], Caron et al [96].

Cette méthode est basée corrélation, cherchant le bloc le plus ressemblant dans l'image de référence au bloc courant. En effet, l'image courante est divisée en blocs non chevauchés de tailles identiques, qui sont vus comme des blocs indépendants, où les pixels composant un bloc ont la même allure de mouvement. Afin de déterminer le meilleur des blocs similaires, l'algorithme compare la différence entre le bloc source et les autres blocs à l'aide d'un critère de qualité tel que la SSD (*Sum of Squared Differences*) et le bloc ciblé qui minimise cette règle est retenu.

1.3.2.2.4 Méthode par transformée de Hough Au domaine de la vision par ordinateur, cette transformée permet la détection des droites, des cercles ou des ellipses en exploitant la primitive de type segments de contour. Dans ce qui va suivre, nous présentons comment la transformée de Hough classique peut être généralisée pour l'estimation de mouvement.

Pour détecter une forme par cette transformée, l'espace doit être changé. En passant de l'espace cartésien (c'est à dire l'équation de la forme avec les coordonnées x et y) à l'espace de Hough c'est à dire aux coordonnées polaires (l'angle par rapport à l'axe x θ et la distance par rapport à l'origine ρ). En effet, pour qu'une primitive détectée sur l'image, soit considérée comme un point parcouru par une droite, elle doit vérifier le vote suivant :

$$G(\rho, \theta) = \sum_x \sum_y g(x, y) \delta(\rho - x \cos(\theta) - y \sin(\theta))$$

Après que l'espace de vote soit rempli, le maximum global de ce dernier, est considéré pour localiser la ligne la plus dominante dans l'image, tandis que le reste des maxima locaux, correspondent éventuellement à d'autres lignes avec des poids plus faibles.

Trois conditions doivent être prises en compte, afin de généraliser cette formulation pour l'estimation de mouvement entre deux images sphériques [97] :

- 1) L'espace des mouvements de caméra doit être finement échantillonné. On peut associer les mouvements caméra au couple $(\mathbf{R}, \mathbf{t})$, où $\mathbf{R}$ est une rotation 3D ($\mathbf{R} \in \text{SO}(3)$) et la direction $\mathbf{t}$ est un vecteur unitaire ($\mathbf{t} \in \mathbb{S}^2$), pour le cas des images sphériques. Il existe nombreuses méthodes d'échantillonnage de $\text{SO}(3)$ et $\mathbb{S}^2$.
- 2) Les «données» qui font le vote doivent être identifiés. Pour la détection de ligne, les données étaient l'ensemble des primitives du contour. Pour l'estimation de mouvement, elles peuvent être les primitives visuelles.
- 3) La contrainte doit être identifiée. Pour la détection de ligne, la contrainte nécessite

un point contour se trouve au long de la ligne. Pour l'estimation de mouvement, la seule contrainte qui convient au lien entre la position des primitives sur l'image et aux mouvements de la caméra c'est la contrainte épipolaire, qui est définie pour un couple de primitives (p, q) par : $(\mathbf{R}p \times q)^\top \mathbf{t} = 0$.

Alors, dans ce cas là, la transformée de Hough classique peut être généralisée pour l'estimation de mouvement :

$$G(\mathbf{R}, \mathbf{t}) = \sum_{p, q} g(p, q) \delta((\mathbf{R}p \times q)^\top \mathbf{t})$$

Pour effectuer une accumulation robuste, il faudrait un moyen de filtrage qui collecte toutes les paires (p, q) de la fonction de similitude g qui satisfont la géométrie épipolaire.

La définition la plus simple est construite à partir de la contrainte épipolaire :

$$\Delta_{(\mathbf{R}, \mathbf{t})} = \delta((\mathbf{R}p \times q)^\top \mathbf{t})$$

Ici $\delta(x)$ est le train d'impulsion de Dirac :

$$\delta(x) = \begin{cases} 1 & \text{si } x=0 \\ 0 & \text{ailleurs.} \end{cases}$$

L'accumulation robuste doit être écrite comme un filtrage pour les paires de primitives définies sur les deux images sphériques :

$$G(\mathbf{R}, \mathbf{t}) = \int_{p \in S^2} \int_{q \in S^2} g(p, q) \Delta_{(\mathbf{R}, \mathbf{t})}(p, q) dp dq$$

Effectivement, $G(\mathbf{R}, \mathbf{t})$ est une fonction de probabilité globale où tous les mouvements relatifs appartenant à cette grille. Le vrai mouvement de la caméra est prévu pour coïncider avec le pic global dans cette grille.

Par ailleurs, la représentation d'image sur la sphère unitaire, a été utilisée dans une série des travaux de Makadia [71, 74, 72], dans le but d'estimer le mouvement entre les images sphériques sans mise en correspondances de primitives. Dans [74], la transformée de Radon (équivalente à la transformée de Hough), a été utilisée pour estimer le mouvement de la camera. Contrairement à l'étape d'appariement des points d'intérêts entre deux images, ici, les paires primitives sont pondérées pour indiquer la probabilité d'une

vraie correspondance. Le vote doit être effectué sur un espace de coefficient de cinq dimensions, après avoir appliqué la transformée de Fourier sphérique. La méthode a donné des résultats satisfaisants et a permis de détecter et d'estimer les mouvements indépendamment des objets mobiles.

Cependant, l'inconvénient des approches basées sur le vote, est la taille de l'espace de variables à estimer (par exemple, un minimum de 5 dimensions pour la récupération de la matrice essentielle). Le degré de l'accumulateur de vote augmente en augmentant la dimension de l'espace des solutions. Par conséquent, le remplissage de l'accumulateur, exige un grand nombre des données.

1.4 Conclusion

Dans ce chapitre, nous avons présenté les différents aspects des systèmes de vision omnidirectionnelle. Tout d'abord, nous avons introduit les différents types des capteurs de vision permettant d'acquérir des images à grand champ de vue. Les modèles géométriques des systèmes catadioptriques ont été présentés, en mettant l'accent sur le modèle de projection unifié.

La géométrie projective permettant d'établir la relation géométrique entre deux images d'une scène, à savoir la géométrie épipolaire, a ensuite été abordée en décrivant un certain nombre de contraintes qui y sont liées.

Ensuite, un aperçu des différentes classes de méthodes concernant l'analyse de l'égomouvement a été présenté. Les méthodes éparées sont basées sur la mise en correspondance de primitives, qui consistent premièrement à extraire un ensemble de primitives de l'image, puis leur mise en correspondance temporelle. Finalement, les paramètres du mouvement sont souvent calculés en utilisant la géométrie de deux vues, en résolvant un système d'équations dont les inconnues sont les paramètres du mouvement. Les primitives obtenues et leurs correspondances servent à résoudre ce système d'équations. Les méthodes denses (ou basées apparences) reposent essentiellement sur les changements temporels des niveaux de gris entre les images. Ce calcul est en général mené sur la totalité de l'image.

Chaque famille d'approches présente des avantages et des inconvénients. Si on retient les méthodes éparées, les phases d'extraction et la mise en correspondance des primitives, sont primordiales. Cependant, elles ont l'avantage de pouvoir opérer sur des mouvements importants. En outre, les méthodes denses permettent d'en appliquer directement sur les

images, par contre, ce type d'approches s'opère sur un mouvement lisse avec de faibles amplitudes.

En effet, vue la quantité de données que représente une image, notamment omnidirectionnelle, il serait préférable de disposer des méthodes qui permettent d'en extraire précisément le mouvement établi, qu'il s'agisse des petits ou larges déplacements entre les images, tout en étant plus dense. La construction de ces méthodes est véritablement le sujet des prochains chapitres.

Chapitre 2

Estimation de la rotation 3D basée sur la représentation en harmoniques sphériques SPHARM

Sommaire

2.1 Introduction	38
2.2 Génération du maillage surfacique 3D	42
2.2.1 Échantillonnage de l'image sphérique	42
2.2.2 Génération de nuages de points 3D	42
2.2.3 Construction du maillage surfacique 3D	43
2.3 Réparation du maillage	46
2.3.1 Topologie des surfaces	47
2.3.2 Algorithme de réparation du maillage	51
2.4 Le descripteur SPHARM de formes 3D	52
2.4.1 Paramétrisation sphérique du maillage 3D	52
2.4.2 Expansion SPHARM	55
2.5 Estimation de la rotation 3D	58
2.5.1 Estimation initiale de la rotation	59
2.5.2 Raffinement par l'ICP	62
2.6 Résultats expérimentaux	63
2.6.1 Expérimentation sur des images synthétiques	64
2.6.2 Expérimentation sur des images réelles	70
2.6.3 Efficacité et coût en temps de calcul	75

2.1 Introduction

Dans le domaine de la vision par ordinateur, le problème de l'estimation de la rotation d'un capteur visuel à partir de deux prises de vue, est un domaine largement couvert depuis quelques années. De nombreuses méthodes proposées dans la littérature, permettent d'estimer la rotation des caméras. Ces méthodes sont suffisamment précises et efficaces lorsqu'il s'agit de capteur conventionnel. Cependant, dans le cas des caméras dites non-conventionnelles (catadioptrique, fisheye, ...), les approches classiques sont inadaptées à la géométrie des images, pour estimer la rotation de manière efficace et robuste. Dans ce contexte, nous proposons une approche directe et générique permettant d'estimer le déplacement en rotation de caméras conventionnelles et non-conventionnelles, en exploitant la représentation sphérique des images.

Les méthodes directes cherchent à déterminer le mouvement de la caméra en examinant le changement d'apparence d'images, c-à-d sans passer par la phase de mise en correspondance de points d'intérêts ou de primitives visuelles. Dans ces méthodes, le mouvement entre deux images est directement obtenu, en minimisant les erreurs d'intensité, par l'estimation d'une transformation paramétrique.

Les premiers résultats proposés dans ce contexte sont ceux de [98], où la corrélation de phase (Phase-Only Correlation) dans le domaine de Fourier, permet de déterminer le déplacement translationnel entre deux images planaires. Kostelec et al. [99] ont développé des méthodes efficaces pour le calcul de la transformée de Wigner-D et son inverse, qui permet d'obtenir le mouvement de rotation dans le domaine spectral, entre les fonctions sphériques et également les fonctions qui sont définies sur le groupe de rotations $SO(3)$. Dans [100], Makadia et al. ont proposé une approche permettant d'estimer la rotation 3D entre les images sphériques en recherchant le maximum dans la grille de corrélation de l'espace des angles d'Euler, tout en utilisant l'analyse harmonique. Une autre approche [101] a été proposée pour l'estimation de la rotation 3D de l'image sphérique. En effet, cette approche est basée sur le calcul de la transformée de Radon dans l'espace de Fourier sphérique, et l'appariement de primitives de type "droites". Afin d'accélérer les calculs et exploiter la géométrie de l'image sphérique, l'algorithme a été défini dans le cadre de l'algèbre géométrique conforme (en anglais *conformal geometric algebra*), qui permet d'unifier les matrices, les vecteurs, les transformations, les nombres complexes dans une théorie unique, utilisant le produit géométrique avec ses produits internes et triangulaires, pour générer un concept mathématique classique connu par l'algèbre géo-

métrique conforme.

La puissance de la transformée de Fourier dans l'espace sphérique, garantissant une reconstruction relativement fidèle des objets tridimensionnels à partir d'un nombre limité d'échantillons, a fait déployer intensivement cette transformée dans plusieurs domaines. On peut citer comme exemples : la représentation des champs de température et de pression en météorologie, les déformations organiques en imagerie médicale, les données omnidirectionnelles issues des antennes équi-directives, les positionnements par satellites (GPS) et les objets tridimensionnels en vision par ordinateur. En effet, les applications de la transformée dans les domaines cités, ont concerné l'estimation de pose [102, 103], l'analyse d'images médicales [104, 105] ainsi que l'alignement et l'enregistrement d'images et d'objets 3D [106].

La représentation SPHARM (*SP*herical *HAR*Monics *transform*) des objets 3D proposée dans [107], a fait l'objet de nombreuses publications, étendant ainsi son application à divers domaines tels que la reconnaissance de formes et la vision par ordinateur. Shen et al, ont proposé dans [108, 109] un descripteur de formes 3D à surface fermée (de topologie sphérique). Ce descripteur s'inscrit plus généralement dans la famille des descripteurs géométriques globaux des objets 3D et se base essentiellement sur les coefficients harmoniques sphériques pour décrire la surface de l'objet 3D. Plusieurs travaux ont été menés par la suite, pour calculer précisément la rotation entre les données 3D, en se basant sur l'analyse harmonique sphérique. Osteen et al. [110] ont introduit une méthode d'estimation de l'ego-mouvement entre deux images RGB-D, basée sur les coefficients harmoniques sphériques obtenus à partir de l'Image de Gauss Étendue (Extended Gaussian Image) définie sur la sphère. Une solution directe et robuste d'estimation de la rotation entre les objets 3D basée SPHARM, a par ailleurs été proposée par Althloothi et al. [111], sans détection et mise en correspondance de primitives.

Récemment, la représentation de l'intensité d'une image par un maillage triangulaire a connue un intérêt considérable, en raison d'un large éventail d'applications potentielles dans les domaines du traitement d'images, de la vision par ordinateur et de l'infographie. Le but d'une telle représentation est de permettre une visualisation explicite du contenu de l'image en utilisant un maillage approximatif [112, 113]. L'avantage de ce maillage approximatif ou représentatif, est qu'il possède habituellement beaucoup moins de pixels que l'image d'origine, tout en gardant une représentation compacte de l'image. Il permet aussi, de les paramétrer dans un autre domaine et de les analyser avec des outils plus performants que ceux élaborés pour les images conventionnelles. Ainsi, la représentation géométrique d'images permet donc, de se placer dans le cadre de l'analyse géométrique, et d'exploiter la géométrie euclidienne. Dans [112], Garcia et al. ont montré comment transformer les pixels d'une image en points 3D. Leurs coordonnées X, Y représentent di-

rectement les coordonnées 2D de la grille image, tandis que la coordonnée Z représente l'intensité lumineuse (niveau de gris normalisé).

Une autre approche a été proposée dans [113], et repose sur la transformation des images 2D en espace 3D en fonction de la transformée en distance. Les harmoniques sphériques sont obtenus pour les données 3D et utilisés comme descripteurs pour les images 2D sous-jacentes.

Dans [113], Sajjanhar et al ont proposé une approche de recherche d'images par le contenu à partir de ses caractéristiques globales. Cette approche se base sur la représentation de l'image 2D par un objet 3D. En utilisant la carte de distance, les intensités des pixels ont été transformées sous forme de distances 3D. Puis, la décomposition en harmoniques sphériques a été calculée pour les objets 3D, permettant ainsi de déterminer une description globale d'objet 3D, ce qui représente également la description de l'image 2D, du fait que la forme de l'objet a été générée à partir de l'intensité lumineuse de l'image. En effet, ce concept de représentation de l'intensité lumineuse d'images par des objets 3D, semble très prometteur, car il permettrait d'adapter de nombreuses méthodes destinées aux objets 3D pour des applications qui concernent les images. Malheureusement, ce type de représentation, semble peu exploré dans la littérature.

Ainsi, l'estimation de mouvement peut également être réalisé de manière dense, c'est-à-dire que tous les pixels de l'image sont alors mis à contribution. Les méthodes éparées, en revanche, sont évidemment généralisables, mais leur utilisation nécessitent souvent le suivi de primitives. En effet, les méthodes éparées se basent sur une optimisation locale, qui ne prend pas en compte la cohérence globale du mouvement des pixels de l'image. Ces approches peuvent être mises en défaut lors du suivi de zones uniformes, dans lesquelles les textures sont moins présentes au niveau d'images. Toutefois, l'utilisation de tout le signal image, améliore la précision et la robustesse de l'estimation du mouvement. En outre, les approches denses permettent d'éviter les problèmes liés aux phases d'extraction et d'appariement des primitives. Or, les méthodes basées primitives ont souvent le désavantage d'être facilement perturbées par les faux appariements, ce qui risque de converger vers une solution erronée lors de l'estimation du déplacement.

Dans ce chapitre, nous proposons un nouvel algorithme, permettant de représenter l'intensité lumineuse des pixels de l'image sphérique, par un maillage surfacique 3D de topologie sphérique. Par la suite, nous suggérons d'utiliser le descripteur de formes 3D basé harmoniques sphériques, pour l'estimation de la rotation entre les images sphériques. Un aperçu général de l'approche proposée est illustré par la figure 2.1. Les phases principales de l'algorithme proposé, qu'on les verra en détail dans la suite de ce chapitre, sont alors :

- Génération de maillages 3D à partir des images sphériques ;
- Réparation de la topologie de maillages obtenus afin de garantir une topologie sphérique ;
- Paramétrisation sphérique du maillage 3D ;
- Calcul de coefficients harmoniques sphériques SPHARM ;
- Décomposition SVD de la matrice de covariance croisée entre les coefficients harmoniques correspondants à chaque maillage ;
- Extraction de la matrice de rotation qui lie les deux images sphériques.

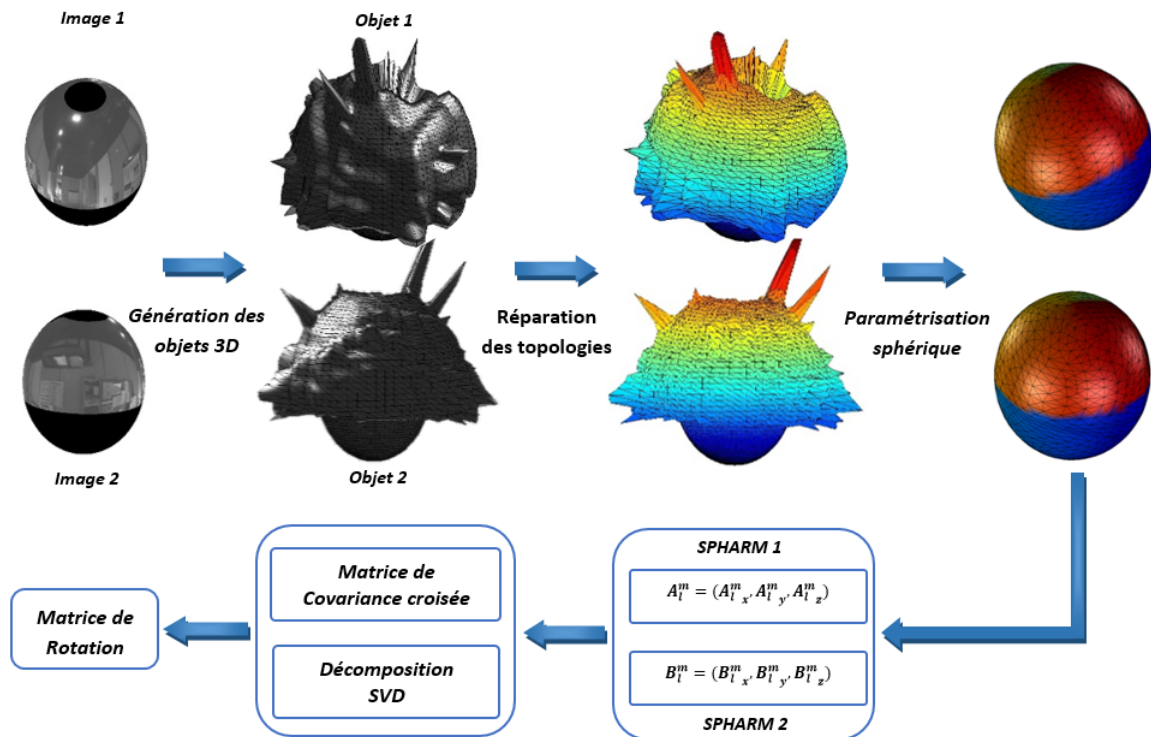


FIGURE 2.1 – Schéma général de l'approche proposée pour estimer la rotation optimale entre deux images sphériques.

La structure de ce chapitre se présente ainsi : dans un premier temps, nous développons la nouvelle représentation d'images sphériques, en générant des maillages 3D à partir de l'intensité lumineuse de chaque point de l'image sphérique. Par la suite, nous détaillons la description SPHARM utilisée pour estimer la rotation entre les images sphériques. Cette étude s'achèvera par l'application de l'approche proposée pour l'estimation de la rotation, sur une série d'expérimentations menée sur des données synthétiques et réelles.

2.2 Génération du maillage surfacique 3D

Dans cette partie, nous allons décrire notre algorithme qui estime la rotation 3D entre deux images représentées sur une sphère unitaire. Pour cela, une nouvelle représentation de l'image sphérique par un maillage surfacique 3D de topologie de genre-0 est introduite. Contrairement aux approches précédentes, cette représentation est dense, c'est-à-dire que tout le signal image peut être utilisé pour générer l'objet 3D de topologie sphérique. Pour permettre l'utilisation de techniques robustes de description globale de l'objet généré, nécessaire par la suite à l'approche dense d'estimation de la rotation, l'augmentation de l'image avec l'information d'intensité lumineuse, associée à chaque point de l'image de la sphère est indispensable. Cela amène à définir, ce qui sera appelé par la suite, le maillage surfacique 3D généré. Nous détaillerons, dans ce qui suit, les principales étapes de notre approche, afin de générer à partir de l'image sphérique un maillage surfacique 3D de topologie sphérique.

2.2.1 Échantillonnage de l'image sphérique

Nous commençons par donner brièvement les notations qui seront utilisées dans cette partie. Une image sphérique sera considérée comme une fonction sur la sphère unitaire, de paramètres $\theta \in [0, \pi[$ qui est l'angle de colatitude et $\varphi \in [0, 2\pi[$ qui est l'angle de longitude. D'une manière générale, nous définissons $\eta(\theta, \varphi) \in \mathbb{S}^2$ pour représenter le vecteur unitaire associée à un point sur la sphère.

La première étape de notre algorithme consiste à échantillonner l'image sphérique selon une grille sphérique équi-angulaire, où θ_k et φ_k sont donnés en fonction de la largeur de bande B par :

$$\theta_k = \frac{(2k+1)\pi}{4B} \quad \text{et} \quad \varphi_k = \frac{k\pi}{B} \quad (2.1)$$

avec $k = 0 \cdots 2B - 1$

2.2.2 Génération de nuages de points 3D

Dans l'étape suivante et après l'échantillonnage de l'image sphérique, pour chaque couple $\eta(\theta, \varphi)$ sur la grille de la sphère unitaire, l'information contenue en ce point sera alors utilisée pour en représenter explicitement l'intensité lumineuse. Le niveau de gris normalisé I_N correspondant au point sera, alors, attribué sous forme d'un segment qui

s'aligne avec le rayon en ce point, dont l'amplitude du segment est I_N . Dans la suite de ce travail, nous nous intéressons seulement au point de l'extrémité supérieure du segment.

Le résultat, en appliquant cette représentation sur l'ensemble de points sphériques obtenus après l'échantillonnage de l'image, est un nuage de points 3D. La figure 2.2 illustre un exemple de point 3D résultant. En effet, un point de ce nuage peut être représenté par les coordonnées sphériques (I_R, θ, φ) où I_R désigne le vecteur normal à la surface de la sphère unitaire au point (θ, φ) , dont la norme est la somme de l'intensité normalisée I_N avec le rayon de la sphère unitaire :

$$I_R = 1 + I_N \quad (2.2)$$

Le nuage de points peut être représenté en coordonnées cartésiennes (x, y, z) :

$$\begin{cases} x = I_R \cos \varphi \sin \theta \\ y = I_R \sin \varphi \sin \theta \\ z = I_R \cos \theta \end{cases} \quad (2.3)$$

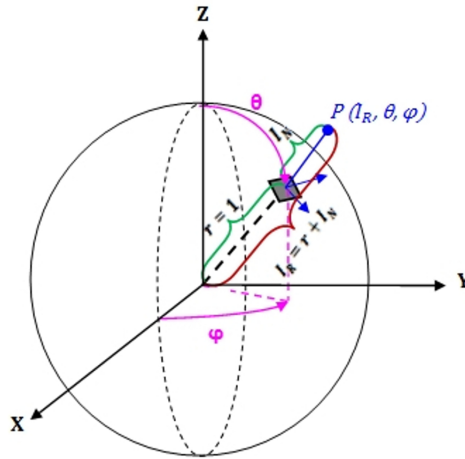


FIGURE 2.2 – Représentation géométrique de l'intensité normalisée du pixel de l'image sphérique.

2.2.3 Construction du maillage surfacique 3D

La vaste majorité des modèles géométriques utilisés dans de nombreuses applications (simulation, calcul numérique, ingénierie, biomédical, ...) emploient une représentation surfacique, ces modèles sont donc définis uniquement par leur surface. Les maillages de polygones sont probablement la représentation surfacique la plus populaire. Une telle approximation est caractérisée par une information géométrique qui définit la position

des sommets dans l'espace euclidien et une information topologique qui décrit la relation d'adjacence entre les sommets traduisant la manière dont ils sont connectés entre eux.

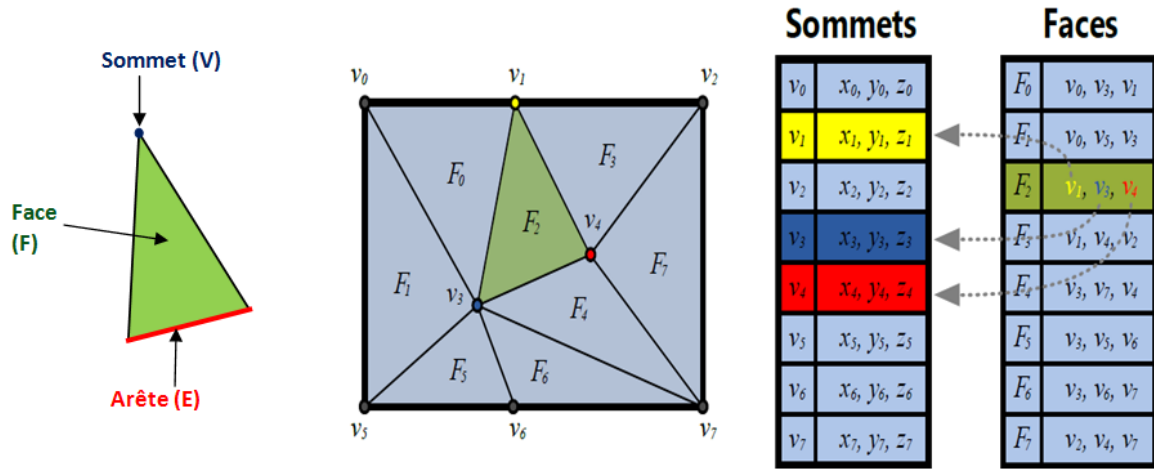


FIGURE 2.3 – Exemple d'un maillage triangulaire et sa représentation (arête, faces et sommets).

Les maillages surfaciques triangulaires sont les plus utilisés pour la représentation d'objets dans l'espace tridimensionnel. Ils sont en effet, en passe de devenir la représentation standard pour la modélisation d'objets géométriques grâce à leur simplicité et leur efficacité. Les sommets (Vertices) sont représentés par des points de l'espace. Les arêtes (Edges) sont des connexions entre deux sommets, et sont situés sur la frontière d'une face. Les faces (Faces) sont des triangles du plan ou de l'espace, dont les sommets et les arêtes appartiennent au maillage. Une structure de données représentant un maillage triangulaire doit stocker plusieurs types d'éléments : des sommets, des arêtes et des faces, comme il est présenté sur la figure 2.3.

Nous nous intéressons ici aux maillages réguliers qui peuvent être considérés comme des maillages triangulaires 3D dont les sommets deviennent réguliers en termes de connectivité (c'est-à-dire reliés à six sommets voisins). Ces maillages permettent d'obtenir des objets 3D avec une représentation simplifiée. Ils présentent aussi un coût de construction faible, afin d'obtenir un maillage uniforme qui peut être représenté par une fonction spatiale $f(x, y, z)$.

La figure 2.4 permet d'appréhender la génération du maillage 3D à partir de l'image en niveau de gris. La figure 2.4-(a) correspond à l'image sphérique d'une scène. La figure 2.4-(b) présente quant à elle les points sphériques associés à cette image, où les niveaux de gris normalisés correspondants sont considérés comme des points 3D afin de générer le nuage de points. On note dans la figure 2.4-(c) après la construction du maillage, que les nombreux "pics" correspondent à de fortes variations locales du signal en niveaux de gris. Notons que, en considérant que la seule source de variation de ce signal est le mouvement de rotation de la caméra, il est possible de constater que la forme de ces

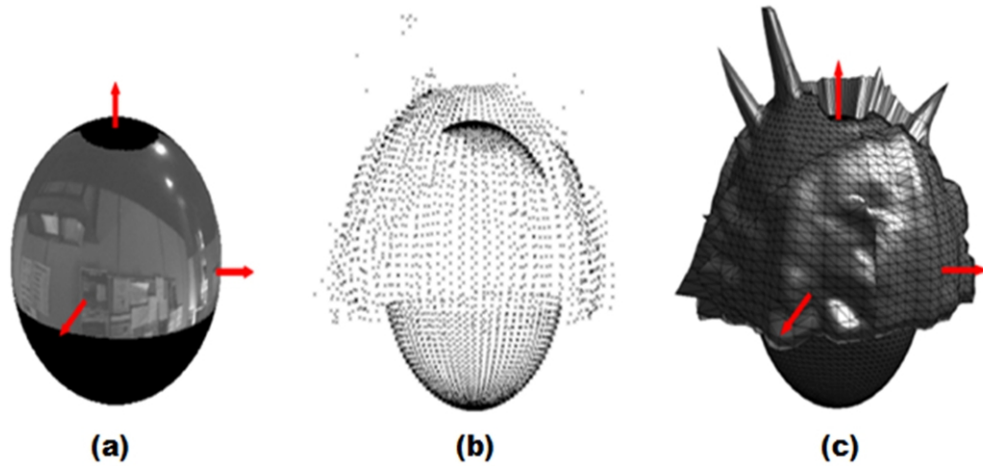


FIGURE 2.4 – (a) Image sphérique (b) Nuage de points 3D (c) Maillage surfacique 3D.

zones restent identiques. On peut donc espérer retrouver dans une nouvelle image une même distribution locale de la fonction d'intensité. L'utilisation de cette représentation de l'image permet de diminuer très fortement la quantité de données à traiter.

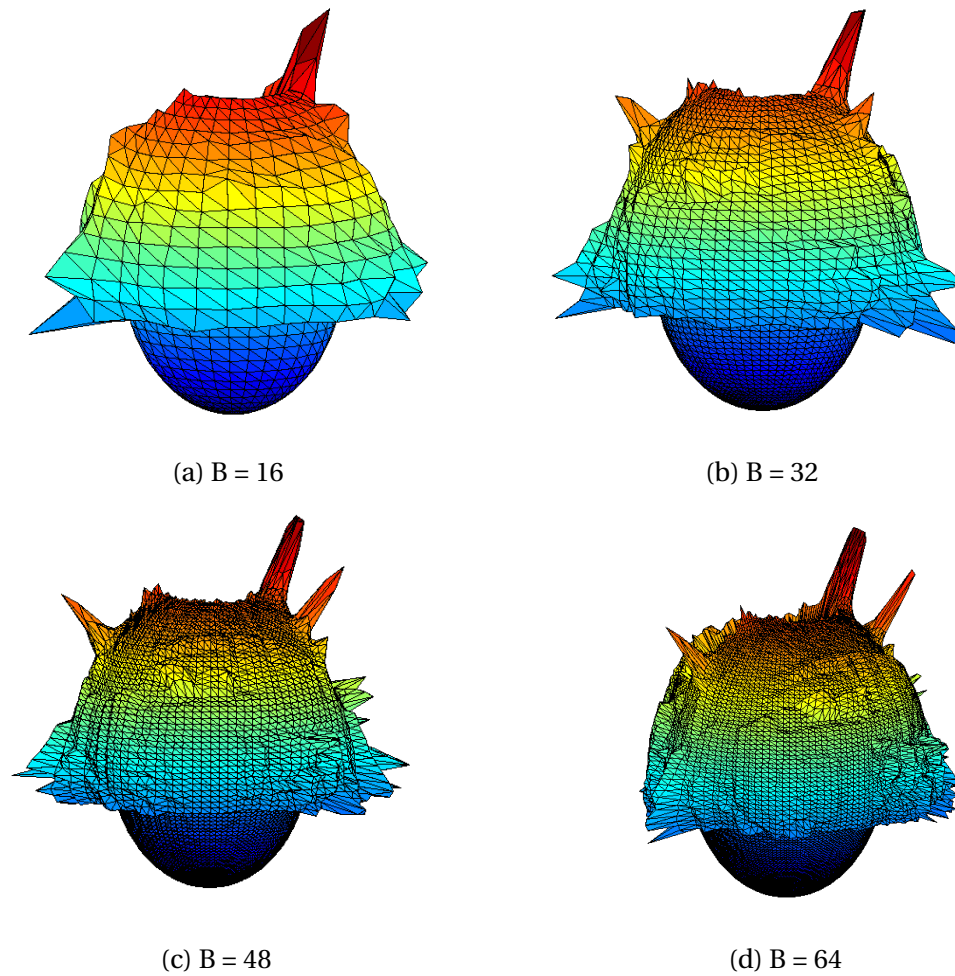


FIGURE 2.5 – Génération de l'objet en fonction du B

Le passage à une représentation géométrique du contenu de l'image, se fait en réa-

lisant une discrétisation des coordonnées spatiales du signal lumineux de l'image sphérique par un échantillonnage et une quantification. Cette discrétisation détermine la précision de la représentation des détails de l'image. Plus la largeur de bande B est élevée (plus la résolution est élevée), mieux les détails seront représentés.

Les figures 2.5-(a), (b), (c) et (d), illustrent l'influence de l'échantillonnage de l'image sur la qualité d'objet généré. La première est considérée comme étant de largeur de bande faible ($B = 16$). Pour les figures suivantes, la largeur de bande B est multipliée à chaque fois par 2. Au fur et à mesure, des détails de plus en plus fins apparaissent. Bien que plus significatif, un maillage fin est très coûteux en temps et fastidieux pour des surfaces de géométrie et/ou de topologie complexe. Or il est souvent préférable de réaliser un maillage moins dense avec une représentation globale de la géométrie 3D de l'objet, plutôt qu'un maillage précis où les détails ne sont pas nécessaire. De cette manière, la largeur de bande a été choisie empiriquement pour la suite du travail. Par conséquent, la complexité de calcul est significativement réduite, sans que la qualité ne soit altérée.

Dans le cadre de ce travail, nous cherchons un algorithme de complétion de maillages applicable au cas des images sphériques. Par ailleurs, il doit être générique pour que nous puissions intégrer nos contraintes spécifiques pour obtenir un maillage 3D de topologie sphérique.

2.3 Réparation du maillage

Cette partie présente la réparation topologique d'un maillage surfacique 3D. Nous allons tout d'abord définir les notions élémentaires utilisées dans la suite de ce chapitre. Puis, nous introduisons les différentes phases du processus de réparation du maillage.

En effet, le calcul des coefficients harmoniques sphériques d'un objet 3D nécessite la coopération de nombreux algorithmes. Un algorithme de paramétrisation sphérique permet de trouver une bijection, qu'est appelée aussi la fonction de mapping, entre la surface 3D et une sphère unitaire. Cette paramétrisation ne peut être établi que si la topologie de maillage est sphérique (c-à-d un maillage surfacique fermé, sans bords, sans trous et les liaisons sont régulières entre les sommets). Tandis que, ces contraintes ne sont pas toujours vérifiées lors de la construction d'un maillage, la réparation de la topologie du maillage est une étape cruciale, elle permet d'imposer une topologie adéquate au maillage, qu'est sphérique dans notre cas.

Nous énumérerons donc dans un premier temps quelques notions élémentaires sur la topologie des surfaces.

2.3.1 Topologie des surfaces

La représentation d'un maillage contient deux types d'information :

- Des informations géométriques : les coordonnées des sommets.
- Des informations topologiques concernant la connectivité du maillage c'est-à-dire la description des éléments du maillage, à partir de laquelle peuvent être déduites toutes les relations de voisinage entre les sommets.

2.3.1.1 Topologie

Ce qu'on appelle aujourd'hui "topologie générale" (en grec *topos* : place et *logos* : étude) est la branche des mathématiques qui étudie les propriétés d'un espace qui sont préservées lors d'une déformation continue indépendamment des grandeurs métriques [114]. Bien que la topologie soit issue de la géométrie, elle ne s'intéresse pas aux propriétés métriques, telles que la position de chaque sommet dans l'espace 3D. La topologie tente d'élaborer des théories sur les notions de proximité, frontière, localité, continuité, etc ... et leurs liens mutuels avec l'espace étudié, tandis qu'un espace topologique (dans la littérature parfois nommé aussi topologie) est une structure mathématique permettant la définition des propriétés d'assemblage entre les différents éléments de l'espace, telles que la connectivité, la continuité et l'orientation. Une topologie se compose de plusieurs sous-domaines : en ce qui concerne les objets discrétisés sont les sommets, les arêtes et les faces. La topologie combinatoire permet de déterminer certaines propriétés et théorèmes concernant la manière dont ces éléments sont connectés entre eux.

Plus intuitivement, la topologie se distingue de la géométrie euclidienne par la conception de l'équivalence entre deux objets. En géométrie euclidienne, deux objets sont équivalents si l'on peut transformer l'un vers l'autre à l'aide d'isométries (rotations, translations, réflexions, etc...) alors qu'en topologie, deux objets sont homéomorphes (c'est-à-dire topologiquement équivalents) s'ils ont simplement les mêmes propriétés topologiques (le nombre de bords et trous). En effet, ils doivent avoir le même nombre de morceaux, de trous, d'intersections etc...

En topologie, il est permis de déformer continûment les objets par étirement, plissement et compression, mais sans les rompre, ni séparer ce qui est uni, ni coller ce qui est séparé. Ce passage d'un objet à l'autre avec ces transformations se nomme homéomorphisme. Lorsqu'il est possible de passer d'un objet à un autre en appliquant ce type de transformations bi-continues, on dit que les deux objets sont homéomorphes, ou topolo-

giquement équivalents. Comme exemple, en terme de topologie, la tasse de la figure 2.6 est homéomorphe à un tore.

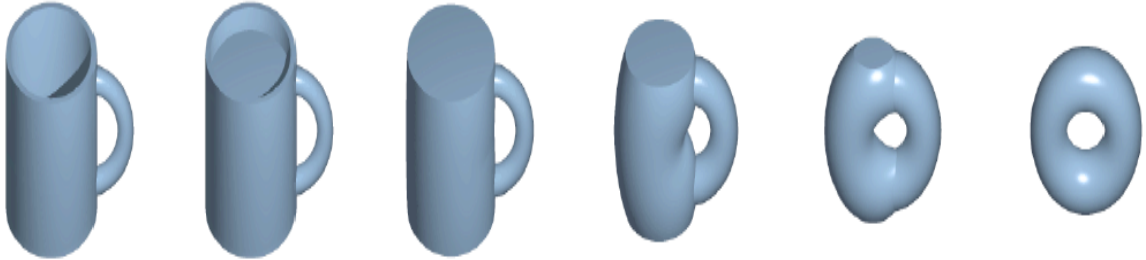


FIGURE 2.6 – Une tasse est homéomorphe à un tore car il existe une déformation continue entre eux.

2.3.1.2 Surface

Le terme "*surface*" possède plusieurs définitions dont celle utilisée en mathématiques, c-à-d un ensemble de points sur lesquels il est localement possible de se repérer à l'aide de deux coordonnées réelles, comme par exemple dans le plan (avec l'abscisse et l'ordonnée) ou sur une sphère (avec la latitude et la longitude). Dans l'espace euclidien tridimensionnel, on dit qu'une surface est un objet qui peut être décrit localement par un plan. En fait, les surfaces interviennent naturellement dans divers domaines et ses définitions diffèrent relativement, selon le domaine d'utilisation. Nous nous intéressons dans cette partie, uniquement à la notion de la surface topologique.

Une surface topologique est un espace topologique localement homéomorphe au plan ou au demi-plan, dont tout point possède un voisinage homéomorphe à $\mathbb{R}^2$. Il est aussi possible de définir la surface comme sous-variété de l'espace à trois dimensions (ou plus). Généralement, on dit qu'une partie de $\mathbb{R}^n$ est une variété de dimension d si elle ressemble au voisinage de chacun de ses points à un sous-espace affine de dimension d (d -variété). Par exemple, le tore ou la sphère sont des variétés de dimension 2 (2-variété) dans $\mathbb{R}^3$. Par le même raisonnement, un cercle est une variété de dimension 1 (1-variété) dans $\mathbb{R}^2$, car toute courbe observée suffisamment près, prendre l'apparence d'une droite.

Comme nous ne nous intéressons pas à la forme géométrique d'une surface, seules trois informations topologiques suffisent pour classer la surface, suivant son type homéomorphisme : le nombre de composantes de bord (b) qu'elle possède, l'orientabilité, et sa caractéristique d'Euler.

- Une composante de bord est un ensemble de points connectés faisant partie du



FIGURE 2.7 – Un exemple sur les composantes de bord et les trous sur un maillage (a) le trou dans le tore est délimité par deux anneaux représentant les bords (en rouge). (b) Le trou est rempli en obtenant un tore avec une topologie en anneau.

bord. On dit qu'une surface qui possède au moins un point sur le bord, est ouverte. Elle est fermée si elle ne possède aucun point de bord. La figure 2.7 présente un exemple sur la composante de bord.

- Une surface est dite orientable si le concept de rotation horaire (ou anti-horaire) est maintenu de manière cohérente sur toute la surface. Pour le cas d'un maillage triangulaire, l'orientabilité globale peut être assurée en ordonnant dans la même direction tous les sommets à l'intérieur des faces (horaire ou anti-horaire). La figure 2.8 illustre un exemple d'un maillage orienté anti-horaire, horaire et non orienté. Il sera impossible pour une surface non orientable de faire une distinction claire entre l'intérieur et l'extérieur de celle-ci.

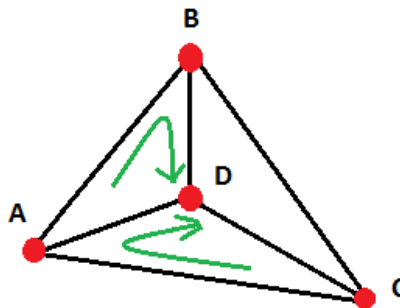


FIGURE 2.8 – Un exemple sur l'orientabilité d'un maillage. Orientation anti-horaire : (C,D,A), (A,D,B), (C,B,D), orientation horaire : (C,A,D), (D,A,B), (B,C,D) et Non orienté : (C,D,A), (D,A,B), (C,B,D).

- Dans le prochain paragraphe, nous présentons le genre d'une surface, qui permet de définir les caractéristiques d'Euler.

2.3.1.3 Genre

En topologie, deux surfaces n'ayant pas le même genre ne sont pas homéomorphes. La manière probablement la plus intuitive pour introduire la notion mathématique de

genre g , est de considérer qu'il s'agit du nombre de trous que contient une surface, en précisant toutefois qu'il faut que celle-ci soit connexe, orientable et sans bord. Comme illustre effectivement la figure 2.9, on peut constater que la sphère est de genre 0 ($g = 0$), le tore est de genre 1 ($g = 1$) et la surface d'un bretzel est de genre 3 ($g = 3$). Toutefois, une surface de genre n peut être vue comme une sphère sur laquelle sont greffés n tores. Le nombre de tores correspond au genre de la surface.

Dans le reste du manuscrit, le terme surface désignera en général une telle surface orientable, dont le type d'homéomorphisme est donné par (g, b) où g est le genre et b est le nombre de composantes de bord.

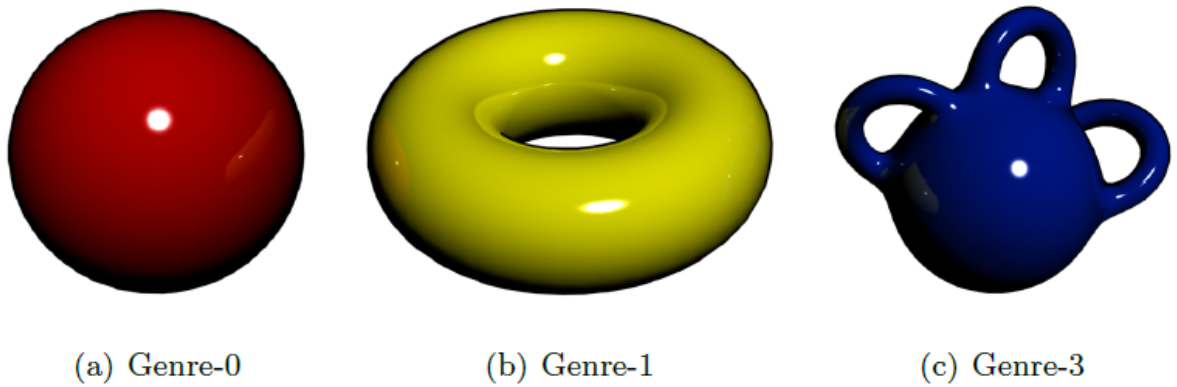


FIGURE 2.9 – Objets de genres 0, 1 et 3.

Nous allons maintenant montrer que les notions de surface triangulée (maillage surfacique) et de surface topologique sont liées par *la caractéristique d'Euler*. Soit S une surface triangulée, c'est-à-dire obtenue en recollant des triangles le long de la surface. On associe à cette triangulation le nombre :

$$\chi = V - E + F \quad (2.4)$$

où V , E , F sont respectivement le nombre de sommets (Vertices), d'arêtes (Edges) et de Faces de la triangulation (voir figure 2.3). Le nombre χ est un invariant topologique : sur une surface donnée S , toutes les triangulations donne le même nombre. Ce nombre χ est appelé *caractéristique d'Euler* ou bien *la variété* de la surface S . Pour un maillage surfacique triangulaire, il est possible de déterminer la variété à partir de genre g de cette surface, dont la relation est définie par :

$$\chi = 2 - 2g \quad (2.5)$$

Prenant maintenant comme exemple le cas d'un maillage de topologie sphérique, où le genre est zéro : $g = 0$, la variété χ est égal à 2, qu'elle est notée le plus souvent surface à 2-variété. L'équation 2.4 peut être écrite sous cette forme :

$$V - E + F = 2 \quad (2.6)$$

De façon plus générale, on appellera n -variété tout espace topologique qui avoisine localement à l'espace euclidien $\mathbb{R}^n$. Par exemple, une sphère est une 2-variété car à chaque point appartenant à la sphère, son voisinage local avoisine un plan.

2.3.2 Algorithme de réparation du maillage

Afin de calculer les coefficients harmoniques sphériques d'un maillage surfacique 3D donné, la surface de ce maillage doit être de topologie sphérique. Comme noté dans la section précédente, une telle surface est dite qu'elle a une topologie sphérique, si elle est fermée, connectée, sans bord, orientée et sans trous. Cependant, ces contraintes topologiques peuvent ne pas être satisfaites par le maillage obtenu à la section 2.2. Les singularités géométriques (comme exemple, deux faces dont les intérieurs se coupent ou un arête qui se partage entre plus de deux faces) ou topologiques (présence des trous ou des petits tunnels entre les sommets), peuvent être présentes lors de la triangulation initiale. Nous proposons dans cette section, de transformer un maillage triangulaire qui peut comporter des défauts géométriques et topologiques en une 2-variété, dont la topologie doit être réparée pour qu'elle soit sphérique.

Pour établir la réparation d'un maillage, nous avons choisi d'utiliser la représentation voxelique intermédiaire, afin d'éliminer toutes les singularités topologiques ou géométriques, ainsi pour garantir que le maillage surfacique de sortie sera une 2-variété de genre-0 [115]. En premier lieu, nous transformons le maillage surfacique 3D initial en une représentation voxelique basée sur la voxelisation des sommets de ce maillage. Cette étape produit une structure quadrilatère uniforme intermédiaire qui garantit le remplissage de tous les trous sur la surface initiale, voir l'exemple de la Figure 2.10-(b). En deuxième lieu, nous retransformerons la représentation voxelique de l'objet 3D en un maillage surfacique triangulaire 3D. Cette étape nous garantit la réparation des défauts géométriques du maillage initial. Par conséquent, le maillage obtenu finalement est un maillage surfacique triangulaire de topologie sphérique (les trous sont remplis comme illustré à la Figure 2.10-(c)). Ces étapes ont été assurées en utilisant la bibliothèque open-source iso2mesh [116].

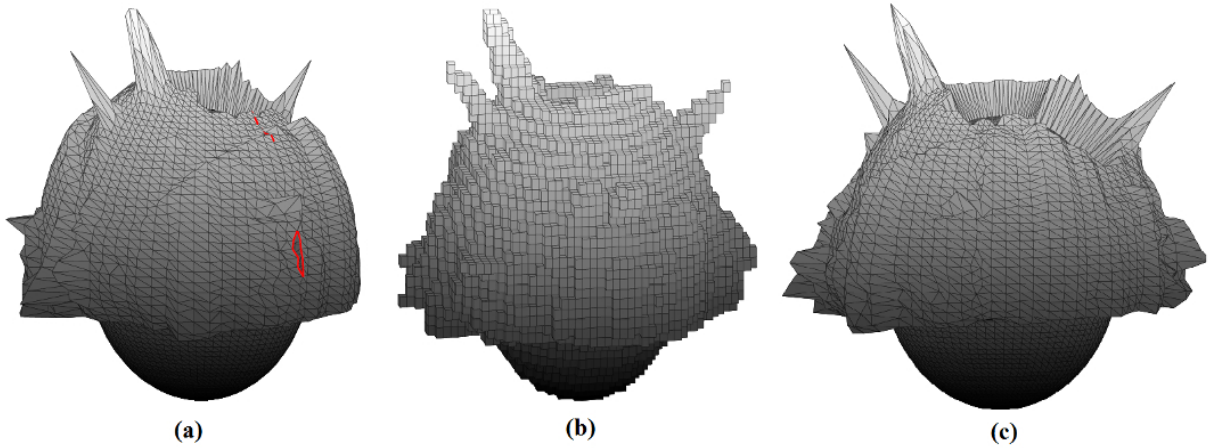


FIGURE 2.10 – L’algorithme de réparation du maillage : (a) Maillages 3D avec trous (b) Voxelisation du maillage (c) Maillage 3D après réparation.

2.4 Le descripteur SPHARM de formes 3D

La description en harmoniques sphériques (SPHARM) est l’une des méthodes les plus efficaces, fournissant une information riche, précise, utilisable pour la description globale des formes 3D. C’est ce type de descripteur qui va être utilisé dans l’algorithme proposé, pour estimer la rotation 3D. Le descripteur SPHARM décompose les formes 3D de topologie équivalente à la sphère en un ensemble de coefficients harmoniques, calculés par la transformée de Fourier sphérique. En appliquant la paramétrisation sphérique, les fonctions sphériques sont obtenues. Par la suite et de façon analogue à une décomposition de Fourier, permettant le passage de l’espace cartésien à l’espace fréquentiel, les fonctions sphériques sont décomposées en une somme des coefficients harmoniques [102]. Étant donnée que la rotation ne change pas la norme des coefficients harmoniques, la signature de chaque fonction sphérique est définie par l’ensemble de ces coefficients. Ainsi calculés pour chacune des trois sphères $x(\theta, \varphi)$, $y(\theta, \varphi)$, $z(\theta, \varphi)$, les différentes signatures sont combinées afin de définir le descripteur SPHARM du maillage surfacique 3D. Le calcul de la description se décompose en deux étapes principales : la paramétrisation sphérique du maillage 3D et l’expansion SPHARM.

2.4.1 Paramétrisation sphérique du maillage 3D

L’objectif principal d’une paramétrisation, est de réaliser une mise en correspondance bijective entre deux surfaces qui sont topologiquement similaires. En vision par ordinateur, on s’intéresse habituellement à faire la correspondance entre la surface d’un modèle tridimensionnel et une autre surface de même topologie (variété) qui s’appelle domaine de paramétrisation. Traditionnellement, les modèles sont représentés par des maillages

de triangles dans un espace 3D, et le domaine est généralement planaire 2D. Lorsqu'une des surfaces est un maillage de polygones, l'opération est appelée paramétrisation de maillage (voir la figure 2.11). Le but central de cette correspondance est d'utiliser la surface paramétrée pour effectuer diverses tâches et opérations en vision par ordinateur.

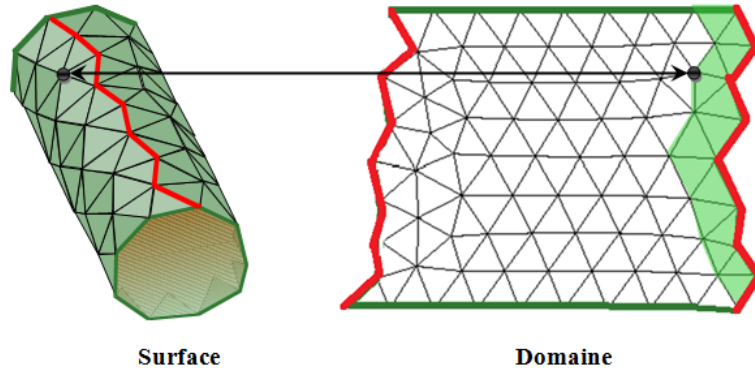


FIGURE 2.11 – Exemple de paramétrisation entre une surface et un domaine.

Ainsi, la paramétrisation sphérique peut être illustrée comme étant la projection de la connectivité d'un maillage surfacique 3D sur une sphère unitaire. L'opération s'appelle la paramétrisation sphérique du maillage. Formalisons-la par une correspondance $F : D \leftrightarrow M$ entre le domaine de paramétrisation qu'est la sphère unitaire $D \subset \mathbb{S}^2$ et la surface $M \subset \mathbb{R}^3$ est le modèle tridimensionnel (maillage surfacique 3D). L'une des propriétés importantes d'une paramétrisation est sa bijectivité, c'est-à-dire qu'un point sur la surface de M n'est associé qu'à un seul point du domaine de paramétrisation D , qui est dans ce cas la sphère, et vice-versa. Une façon simple de vérifier la bijectivité d'une paramétrisation est de s'assurer qu'elle ne contient aucun croisement d'arêtes ou aucune face renversée dans D . C'est dans ce contexte que la fonction sphérique est créée en projetant la surface du maillage 3D (sommets) sur la sphère unitaire, les sommets d'un maillage étant affectés aux paramètres sphériques (θ, φ) .

Dans la littérature, la paramétrisation sphérique est souvent considérée comme la plus appropriée aux surfaces qui ne contiennent pas de trous c-à-d de genre-0 (topologie sphérique) [117]. Plusieurs algorithmes ont été proposés dans le domaine de l'informatique graphique pour optimiser la paramétrisation. Certains algorithmes, comme dans [118, 119], partent d'un maillage triangulaire, en assignant des paramètres initiaux pour chaque sommet, puis en modifiant ces paramètres afin d'optimiser la zone triangulaire et les angles par un lissage itératif de la projection. Cependant, ce type de méthodes sont peu fiables lorsque le modèle de la forme est très complexe. D'autres travaux, tels que celui proposé par Brechbuhler et al. [107] partent d'un maillage volumique basé voxel. Shen et Makedon, dans [104], l'ont adapté au maillage surfacique triangulaire et ont montré leur efficacité pour des formes complexes. L'avantage de cet algorithme présenté en ce travail, il se base sur une optimisation non linéaire sous contraintes pour minimiser les

distorsions de la surface et de la topologie du maillage lors de l'étape de projection. La projection initiale de l'objet 3D sur l'espace de paramétrisation est construite en utilisant l'équation de Laplace discrète pour résoudre le problème de Dirichlet correspondant. La paramétrisation initiale est optimisée sous deux contraintes [120] :

- Préservation des zones : chaque région dans l'espace objet doit correspondre à une région de la zone proportionnelle dans l'espace de paramétrisation, en s'appuyant sur la méthode de Newton Raphson.
- Distorsion minimale : chaque triangle doit correspondre à un triangle sphérique dans l'espace de paramétrisation, en optimisant une fonction objective par l'utilisation de la méthode du gradient conjugué.

Nous avons sélectionné cet algorithme pour paramétrer notre maillage surfacique. La figure 2.12 montre une illustration du paramétrage sphérique. Le lecteur intéressé par une description complète de la paramétrisation et l'algorithme d'optimisation peut se référer aux références [104, 105].

La paramétrisation sphérique est cruciale pour obtenir une représentation SPHARM précise. Le résultat de cette paramétrisation est une projection bijective entre chaque sommet $v = (x, y, z)^T$ sur la surface du maillage 3D et le couple des coordonnées sphériques (θ, φ) . La fonction de paramétrisation peut être définie par :

$$\begin{aligned} v : S &\rightarrow M \\ (\theta, \varphi) &\mapsto (x, y, z) \\ v(\theta, \varphi) &= (x(\theta, \varphi), y(\theta, \varphi), z(\theta, \varphi)) \end{aligned}$$

En construisant la paramétrisation sphérique d'un maillage de genre zéro, la surface du maillage est transformée en trois signaux sphériques [104, 108]. Chaque signal correspond à une des trois coordonnées cartésiennes x , y et z , qui représente les coordonnées de S dans l'espace objet. La paramétrisation peut être représentée comme suit :

$$S(\theta, \varphi) = f \begin{pmatrix} x(\theta, \varphi) \\ y(\theta, \varphi) \\ z(\theta, \varphi) \end{pmatrix} \quad (2.7)$$

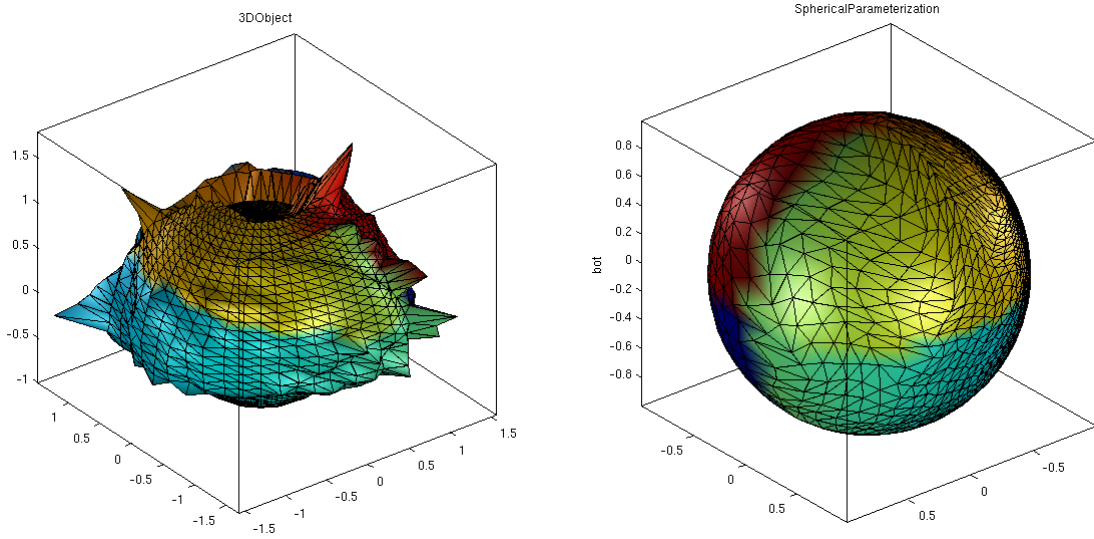


FIGURE 2.12 – Paramétrisation sphérique de l'objet 3D.

2.4.2 Expansion SPHARM

Une fois la paramétrisation sphérique du maillage 3D établie, il est possible par la suite de calculer l'expansion en harmoniques sphériques pour décomposer la surface de l'objet 3D en une série de coefficients et de base des harmoniques sphériques pour chacune des coordonnées x , y et z . En effet, la transformée de Fourier sphérique permet de calculer le spectre fréquentiel des signaux représentés sur la sphère, en utilisant des décompositions dans de bases d'harmoniques sphériques. Un aperçu sur la transformée de Fourier sphérique est donné dans l'annexe (A.1). Les harmoniques sphériques Y_l^m définissent une base de fonctions orthonormale sur la sphère. Originellement introduite pour résoudre l'équation de Laplace, elles sont définies par :

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos\theta) e^{\pm im\varphi}$$

Avec $l \in \mathbb{N}$, $|m| \leq l$ et P_l^m est le polynôme de Legendre associé (voir annexe (A))

Les fonctions harmoniques sont ordonnées par bande l contenant chacune $2l+1$ fonctions harmoniques de fréquences similaires (voir figure 2.13) notées de $m = -l$ à $m = +l$. Ainsi, pour un ordre L_{max} donné, une fonction sphérique est représentée par $(L_{max} + 1)^2$ coefficients.

Ayant la position cartésienne de chaque sommet et ses paramètres sphériques correspondants, la forme 3D de l'objet peut être décrite et reconstruite en calculant les coefficients harmoniques sphériques à l'ordre L_{max} par l'équation 2.8 :

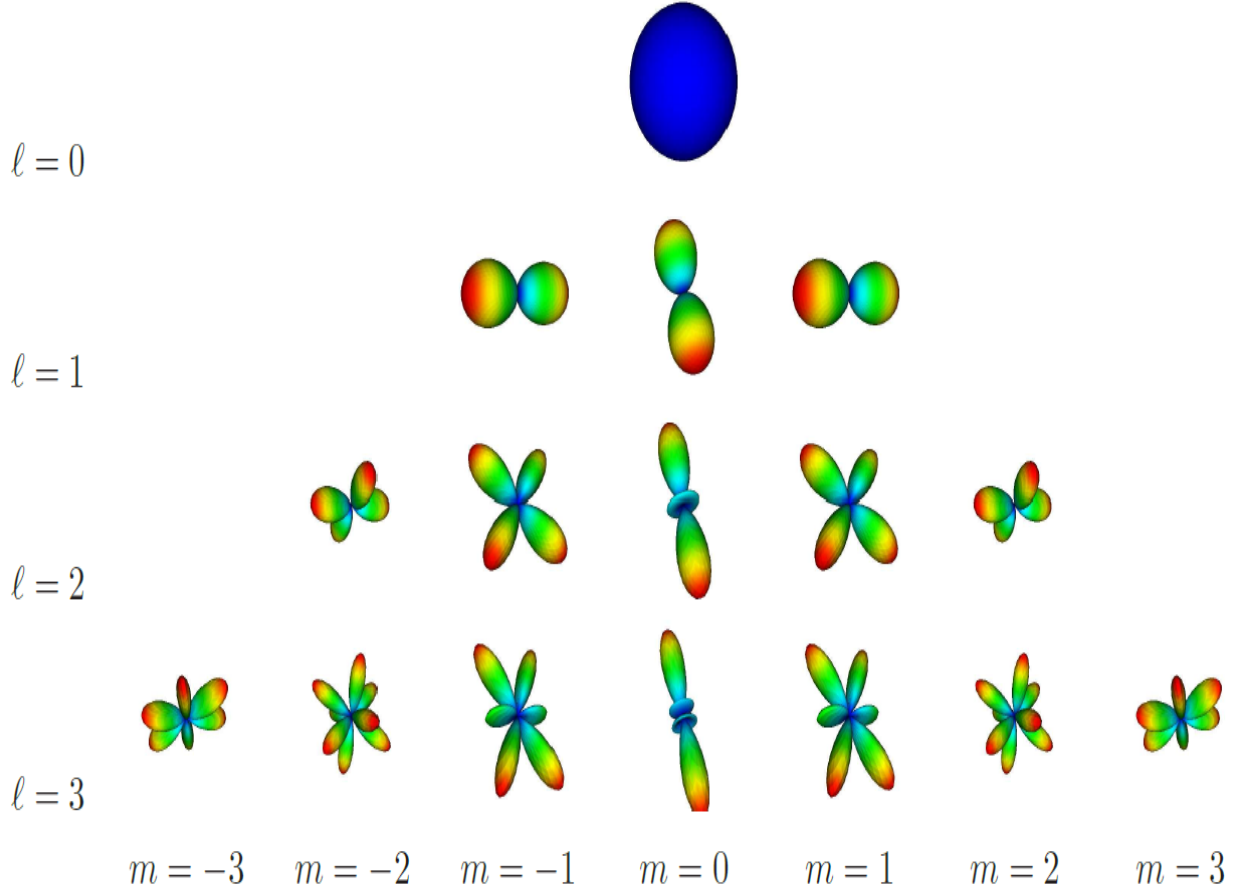


FIGURE 2.13 – Représentation en trois dimensions des harmoniques sphériques allant jusqu'à l'ordre 3, où les portions bleues correspondent aux régions, où la fonction est positive et les portions rouges aux régions où la fonction est négative.

$$S(\theta, \varphi) = \sum_{l \in \mathbb{N}} \sum_{|m| \leq l} c_l^m Y_l^m \quad (2.8)$$

Les coefficients d'harmoniques sphériques $c_l^m = (c_{l_x}^m, c_{l_y}^m, c_{l_z}^m)^\top$ sont liés à la décomposition de $S(\theta, \varphi) = (x(\theta, \varphi), y(\theta, \varphi), z(\theta, \varphi))^\top$, qui est la position de la surface originale (i.e. les sommets du maillage surfacique 3D). Les coefficients SPHARM sont obtenus en résolvant le problème par la méthode des moindres carrés, dont les étapes sont données comme suit :

- Création d'une matrice B qui a comme éléments les bases harmoniques Y_l^m de taille $(N_{ver} \times N_Y)$, où N_{ver} est le nombre de sommets du maillage 3D et $N_Y = l^2 + l + m$

$$B = \begin{pmatrix} Y_{0(1)}^0 & Y_{1(1)}^{-1} & Y_{1(1)}^0 & Y_{1(1)}^1 & \cdots & Y_{l(1)}^m \\ Y_{0(2)}^0 & Y_{1(2)}^{-1} & Y_{1(2)}^0 & Y_{1(2)}^1 & \cdots & Y_{l(2)}^m \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ Y_{0(N_{ver})}^0 & Y_{1(N_{ver})}^{-1} & Y_{1(N_{ver})}^0 & Y_{1(N_{ver})}^1 & \cdots & Y_{l(N_{ver})}^m \end{pmatrix}$$

- Création d'une matrice S de taille $(N_{ver} \times 3)$ qui contient $S(\theta, \varphi)$

$$S = \begin{pmatrix} x_1(\theta, \varphi) & y_1(\theta, \varphi) & z_1(\theta, \varphi) \\ x_2(\theta, \varphi) & y_2(\theta, \varphi) & z_2(\theta, \varphi) \\ \vdots & \vdots & \vdots \\ x_{N_{ver}}(\theta, \varphi) & y_{N_{ver}}(\theta, \varphi) & z_{N_{ver}}(\theta, \varphi) \end{pmatrix}$$

- Puis le calcul de la matrice C de taille (N_Y) par la multiplication de la pseudo-inverse de B par S :

$$C = \begin{pmatrix} cx_0^0 & cy_0^0 & cz_0^0 \\ cx_1^{-1} & cy_1^{-1} & cz_1^{-1} \\ \vdots & \vdots & \vdots \\ cx_l^m & cy_l^m & cz_l^m \end{pmatrix} = B^+ S$$

Où B^+ est la pseudo-inverse de la matrice B, tel que $B^+ = (B^T B)^{-1} B^T$.

En effet, les coefficients de basses fréquences correspondent aux traits grossiers de l'objet 3D alors que les hautes fréquences correspondent aux détails plus fins et au bruit. Les composantes correspondantes aux basses fréquences suffisent pour représenter la forme globale de l'objet. Ainsi, un ensemble de ces coefficients forment effectivement une description globale du maillage surfacique 3D.

Dans la figure 2.14, nous montrons la représentation du maillage d'objet 3D avec différents degrés de SPHARM. Les figures 2.14-(a), (b) sont l'image sphérique et l'objet 3D original généré, respectivement. Les figures 2.14 (b), (c), (d) et (e) sont des reconstructions avec $L = 0$, $L = 5$, $L = 10$ et $L = 15$, respectivement.

Comme le montre ces figures, nous pouvons voir que la reconstruction est améliorée en augmentant le degré de SPHARM. Une reconstruction presque parfaite peut être obtenue avec $L = 15$. Cela signifie que la forme de l'objet 3D peut être représentée par seulement 768 coefficients ($3 \times (15 + 1)^2 = 768$), tandis que, nous avons besoin d'utiliser un vecteur de dimension 12675 (pour 4225 sommets) pour représenter le maillage surfa-

cique de cet objet 3D.

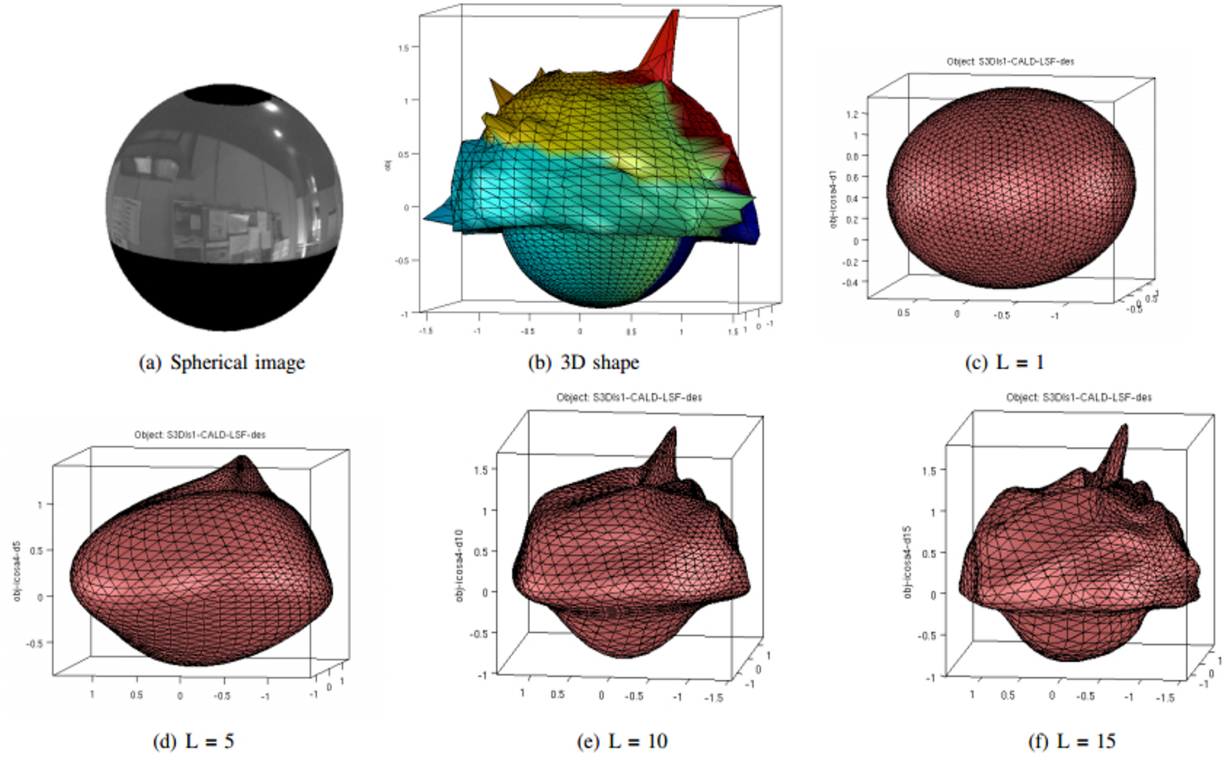


FIGURE 2.14 – (a) et (b) représentent l'image sphérique et l'objet 3D. (c), (d), (e) et (f) la reconstruction SPHARM pour $L_{max}=1, 5, 10$ et 15 , respectivement.

La description en harmoniques sphériques (SPHARM) sera utilisée dans la section suivante, pour estimer la rotation entre deux formes 3D.

2.5 Estimation de la rotation 3D

Dans cette section, nous détaillons l'approche proposée permettant d'estimer la rotation qui lie deux images sphériques. Cette approche est basée sur les coefficients SPHARM obtenus à partir des objets 3D y correspondants. Tandis que l'estimation initiale est obtenue dans le domaine paramétrique spectral, une étape de raffinement a été ajoutée afin de trouver la rotation précise entre les objets 3D.

Dans ce cadre, nous allons expliquer dans la première partie l'estimation initiale obtenue par une solution directe à partir des coefficients harmoniques et de la décomposition SVD. Ensuite, nous allons présenter l'algorithme ICP que l'on intègre dans notre méthode pour le raffinement de l'estimation initiale.

2.5.1 Estimation initiale de la rotation

L'objectif de cette section est de présenter la solution qui permet de retrouver dans le domaine spectral la rotation entre deux images sphériques. Partant de la matrice de rotation $\mathbf{R} \in \text{SO}(3)$, nous définissons la rotation d'une image sphérique $\eta(\zeta(\theta, \varphi))$ avec $\zeta(\theta, \varphi) \in S^2$ par l'opérateur $\Lambda_{\mathbf{R}}$ agissant sur l'image $\eta(\theta, \varphi)$ comme suit :

$$\Lambda_{\mathbf{R}} \eta(\zeta(\theta, \varphi)) = \eta(\mathbf{R}^T \zeta(\theta, \varphi)) \quad (2.9)$$

Le problème abordé dans ce qui suit, est celui de l'estimation par une approche dense la rotation 3D entre les images sphériques, en utilisant la représentation 3D intermédiaire proposée dans la section précédente. L'interrogation essentielle pour la résolution d'un tel problème concerne l'adaptation et l'extension des méthodes robustes proposées pour les objets 3D pour retrouver la rotation entre les images. La figure 2.15 fait apparaître un exemple montrant l'équivalence entre la rotation d'un pixel et un point 3D généré à partir de l'intensité normalisée de ce pixel, comme indiqué dans l'équation 2.3.

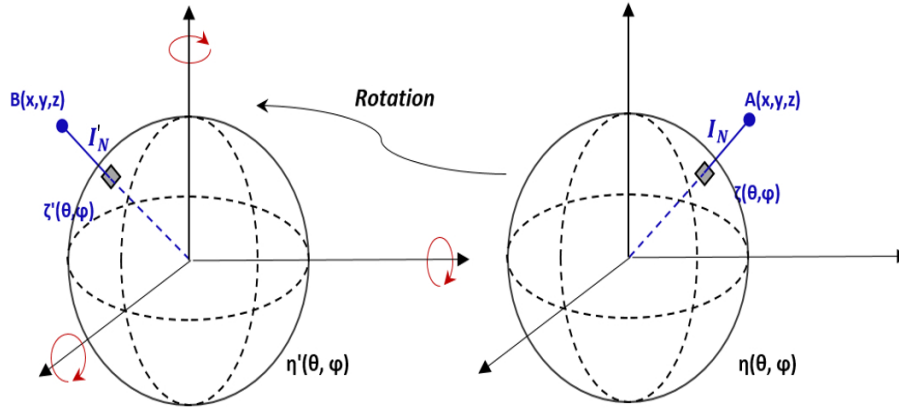


FIGURE 2.15 – La rotation d'une image sphérique avec l'exemple de l'intensité lumineuse normalisée d'un pixel.

La relation entre deux points sphériques $\zeta(\theta, \varphi)$ et $\zeta'(\theta, \varphi)$ liés par la rotation $\mathbf{R}$, est donnée par :

$$\zeta' = \mathbf{R} \zeta \quad (2.10)$$

Supposant la contrainte de conservation de l'intensité lumineuse entre les images, qui est décrite par la relation : $\varepsilon = I_N(\zeta(\theta, \varphi)) = I'_N(\zeta'(\theta, \varphi)) = \text{constante}$, et représente l'intensité lumineuse normalisée des deux images sphériques $\eta(\theta, \varphi)$ et $\eta'(\theta, \varphi)$ liées par la

rotation $\mathbf{R}$.

En multipliant les deux cotés de l'équation 2.10 par la constante ϵ nous obtenons :

$$\begin{aligned}\epsilon \zeta' &= \epsilon \mathbf{R} \zeta \\ &= \mathbf{R} \epsilon \zeta\end{aligned}\tag{2.11}$$

En se basant sur les équations 2.1 et 2.11, on peut conclure que l'estimation de la rotation entre deux objets 3D générés à partir des intensités des images revient à estimer la rotation entre les images sphériques. En effet, l'algorithme d'estimation de la rotation 3D proposé dans cette section est un schéma générique qui peut être appliqué aussi bien aux images sphériques qu'aux objets 3D. Ainsi, la description SPHARM des formes 3D est une représentation concise qui permet de trouver la rotation 3D, en utilisant les coefficients d'harmoniques sphériques qui offrent une caractérisation globale, compacte et robuste attachée à la forme 3D. Par conséquent, la précision de cette estimation de rotation dépend fortement de la bande d'harmoniques sphériques choisies pour représenter l'objet 3D dans le domaine spectral.

Soient A et B les deux objets 3D ($A, B \in \mathbb{R}^{N_\nu \times 3}$, avec N_ν le nombre de sommets du maillage surfacique 3D) obtenus à partir de deux images sphériques $\eta(\theta, \varphi)$ et $\eta'(\theta, \varphi)$ respectivement. La matrice de rotation qui minimise la L_2 -distance entre A et B , peut être obtenue en résolvant l'équation suivante [121] :

$$\min_{\mathbf{R}} \|A - \mathbf{R}B\|_F\tag{2.12}$$

Avec $\|\cdot\|_F$ représente la norme de Frobenius.

Sachant que $\|D\|_F^2 = \text{tr}(D^\top D)$ et que $\mathbf{R}$ est orthogonale, on peut donc écrire :

$$\begin{aligned}\|A - \mathbf{R}B\|_F^2 &= \text{tr}(A^\top A) + \text{tr}(B^\top \mathbf{R}^\top \mathbf{R} B) - 2\text{tr}(\mathbf{R}^\top B^\top A) \\ &= \text{tr}(A^\top A) + \text{tr}(B^\top B) - 2\text{tr}(\mathbf{R}^\top B^\top A)\end{aligned}\tag{2.13}$$

L'équation 2.12 est minimisée lorsque le dernier terme de l'équation 2.13 est maximisé, ce qui revient à maximiser : $\text{tr}(\mathbf{R}^\top \mathbf{M})$, où $\mathbf{M}$ est la matrice de covariance croisée :

$$\mathbf{M} = B^\top A\tag{2.14}$$

La matrice $\mathbf{R}$ qui maximise $\text{tr}(\mathbf{R}^\top \mathbf{M})$ peut être calculée en appliquant la décomposition SVD de $\mathbf{M} = \mathbf{B}^\top \mathbf{A}$:

$$\text{tr}(\mathbf{R}^\top \mathbf{M}) = \text{tr}(\mathbf{R}^\top \mathbf{U}_\mathbf{M} \Sigma_\mathbf{M} \mathbf{V}_\mathbf{M}^\top)$$

où $\mathbf{U}_\mathbf{M}$ et $\mathbf{V}_\mathbf{M}$ sont des matrices qui contiennent les deux vecteurs singuliers gauche et droite de $\mathbf{M}$, respectivement, tandis que $\Sigma_\mathbf{M}$ est une matrice diagonale des valeurs singulières de $\mathbf{M}$ dans l'ordre décroissant. Ainsi, la rotation est obtenue en définissant :

$$\mathbf{R} = \mathbf{U}_\mathbf{M} \mathbf{V}_\mathbf{M}^\top \quad (2.15)$$

Dans notre approche, la matrice $\mathbf{M}$ est construite à partir de la covariance croisée entre les deux ensembles des coefficients d'harmoniques sphériques

$$\mathbf{A}_l^m = (A_{l_x}^m, A_{l_y}^m, A_{l_z}^m)^\top \text{ et } \mathbf{B}_l^m = (B_{l_x}^m, B_{l_y}^m, B_{l_z}^m)^\top, \text{ avec } 0 \leq l \leq L_{max} \text{ et } -l \leq m \leq l.$$

Afin d'estimer la matrice de rotation dans le domaine spectral, la matrice de covariance croisée $\mathbf{M}$ entre les deux ensembles de coefficients SPHARM est définie par :

$$\mathbf{M} = \sum_{xyz} \sum_{k=1}^N [(\mathbf{A}_k - \mu_\mathbf{A})(\mathbf{B}_k - \mu_\mathbf{B})^\top] \quad (2.16)$$

Où $N = (L_{max} + 1)^2$ et $\mu_\mathbf{A}$ et $\mu_\mathbf{B}$ sont les valeurs moyennes des coefficients SPHARM, définies par :

$$\mu_\mathbf{A} = \frac{1}{N} \sum_{k=1}^N \mathbf{A}_k \quad \text{and} \quad \mu_\mathbf{B} = \frac{1}{N} \sum_{k=1}^N \mathbf{B}_k$$

La matrice de rotation qui aligne le mieux les deux objets 3D dans le domaine spectral (c-à-d entre l'objet de référence et celui obtenu par un mouvement de rotation) peut être retrouvée.

Afin d'obtenir une estimation précise entre les modèles des surfaces 3D, nous proposons d'améliorer l'estimation comme décrite précédemment par l'algorithme ICP. L'utilisation de cet algorithme permettant d'assurer la rotation 3D optimale dans les cas où l'estimation initiale est grossière.

2.5.2 Raffinement par l'ICP

L'algorithme ICP (Iterative Closest Point) a été introduit par Besl [122], dans le but d'effectuer un recalage rigide en 3D. L'algorithme a montré son efficacité pour recaler tout modèle d'objet défini par un ensemble des primitives 3D (points, droites, courbes et surface maillées). Le principe de l'algorithme ICP se base essentiellement sur la minimisation itérative de la distance euclidienne entre le modèle de référence et l'objet de test. La solution atteinte est obtenue lorsque l'erreur résiduelle de distance entre les deux modèles est inférieure à un certain seuil fixé au préalable, afin d'en garantir la convergence vers une solution optimale.

Le concept de l'algorithme ICP peut être résumé en trois étapes :

1. Calculer l'appariement entre les deux modèles qui minimise l'erreur métrique basée sur la distance euclidienne entre les deux objets 3D.
2. Calculer la transformation géométrique qui sépare les deux modèles après l'appariement.
3. Itérer jusqu'à ce que l'erreur soit inférieure à un seuil donné a priori qui spécifie la précision.

Les principaux avantages de l'algorithme ICP sont : (1) La simplicité de la recherche itérative locale de la transformation géométrique optimale qui aligne deux modèles de données. (2) La convergence rapide, en termes de nombre d'itérations, quand une bonne estimation initiale de transformation géométrique est fournie.

Par ailleurs, l'ICP a deux limites majeures : (1) Il nécessite une initialisation raisonnablement précise (minimum global) pour atteindre la solution optimale, sans laquelle l'algorithme peut facilement converger vers un minimum local. (2) Il est coûteux en terme de temps de calcul, avec son processus itératif, pour aligner deux objets représentés par un nombre important de primitives.

Avec l'approche proposée et décrite dans la section précédente, les deux problèmes de l'algorithme ICP peuvent être surmontés. La solution fournie par la représentation SPHARM permettant de s'approcher de la rotation optimale. Cette solution peut être introduite comme une bonne initialisation pour le raffinement de l'estimation. Cependant, la solution trouvée uniquement par l'ICP, c-à-d sans appliquer l'estimation initiale, peut être loin de la solution optimale et conduit à une estimation erronée. Ainsi, le problème de coût de calcul peut être réduit grâce à la représentation SPHARM, qui est capable de

représenter un objet 3D en utilisant seulement les coefficients harmoniques sphériques au lieu d'utiliser de milliers de sommets du maillage surfacique de l'objet 3D original.

Par conséquent, l'étape de raffinement par l'ICP a été aisément combinée avec l'approche précédemment décrite, pour tirer profit de leurs avantages complémentaires. Une fois l'estimation de la matrice de rotation dans l'espace de paramétrisation est déterminée, qui se base sur les coefficients SPHARM et la décomposition SVD, elle doit être considérée, pour initialiser l'alignement final. Par la suite, l'algorithme ICP est appliqué dans l'espace objet afin de raffiner l'alignement et atteindre une estimation plus précise. En effet, la recherche locale de la technique ICP est entièrement sollicitée à minimiser la distance entre les deux objets 3D et définir l'alignement le plus fin entre eux. Une fois que l'ICP converge vers un minimum local, on peut obtenir une estimation optimale et plus précise de la rotation $\mathbf{R}$.

Il est important de noter que cette méthode a permis de résoudre le problème de l'estimation des rotations de grande amplitude par une approche dense. Quant à la précision, elle est assurée par une approche itérative robuste de type ICP.

Le schéma général de l'approche proposée pour l'estimation de la rotation 3D entre les images sphériques est résumé dans Algorithme 1.

Algorithm 1:

Data: Deux images sphériques $\eta(\theta, \varphi)$ et $\eta'(\theta, \varphi)$ liées par la rotation 3D $\mathbf{R}$.

Result: La matrice de rotation optimale $\mathbf{R}$

- 1 Construire deux objets 3D de topologie sphérique, basés sur l'intensité des images :
 $A(x, y, z) \leftarrow \eta(\theta, \varphi)$ et $B(x, y, z) \leftarrow \eta'(\theta, \varphi)$
- 2 Appliquer la paramétrisation sphérique : $A(x(\theta, \varphi), y(\theta, \varphi), z(\theta, \varphi)) \leftarrow A(x, y, z)$ et
 $B(x(\theta, \varphi), y(\theta, \varphi), z(\theta, \varphi)) \leftarrow B(x, y, z)$
- 3 Calculer les coefficients harmoniques sphériques SPHARM ($A_{l_x}^m, A_{l_y}^m, A_{l_z}^m$) et
 $(B_{l_x}^m, B_{l_y}^m, B_{l_z}^m)$
- 4 Estimer la rotation initiale en utilisant les coefficients SPHARM et la décomposition SVD.
- 5 Raffiner l'estimation en utilisant l'algorithme ICP

2.6 Résultats expérimentaux

Les performances de la nouvelle approche basée sur une représentation 3D de l'intensité d'images et la description SPHARM ont été validées et évaluées sur deux séquences d'images différentes : une séquence synthétique et une séquence réelle. Dans un premier temps, on présente les résultats obtenus quant à l'étude de l'efficacité et la robustesse

de notre approche vis-à-vis des trois types de perturbations différentes appliquées sur l'image : le bruit, le changement de luminosité et la compression des images. En outre, la précision de notre algorithme a été évaluée en la comparant avec quatre autres algorithmes compétitifs proposés dans la littérature pour estimer la rotation 3D d'images sphériques. Par ailleurs, la validation à travers des séquences d'images réelles a portée sur l'évaluation de l'algorithme lorsque la contrainte de conservation de l'intensité lumineuse est violée. Le cout en termes de temps de calcul de l'algorithme proposé, a aussi été étudié à la fin de cette section.

2.6.1 Expérimentation sur des images synthétiques

Les données synthétiques ont été générées en appliquant différentes rotations à une image omnidirectionnelle réelle de résolution de 1027×768 , acquise par une caméra catadioptrique centrale. Étant donnée que ce type de caméra est conforme au modèle de projection central [123], l'image sphérique correspondante à l'image omnidirectionnelle peut être générée. Diverses transformations sont appliquées sur l'image sphérique obtenue : géométrique, tel que la rotation, et photométrique, comme le changement d'intensité lumineuse de pixel (par bruit, changement de luminosité, compression, ...), ou bien les deux à la fois. La figure 2.16 montre l'exemple d'une image sphérique avant et après transformation par une rotation 3D.

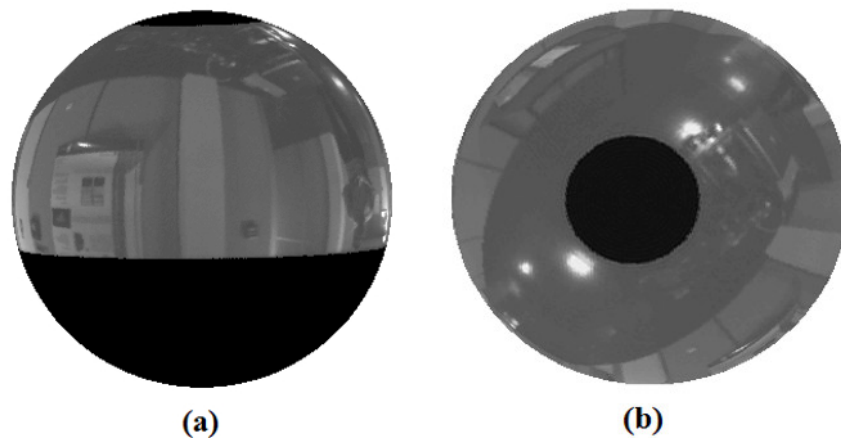


FIGURE 2.16 – (a) L'image sphérique originale. (b) Image après une rotation de $\mathbf{R}(45^\circ; 60^\circ; 45^\circ)$

2.6.1.1 Précision de l'estimation de la rotation

Afin d'évaluer l'efficacité de notre approche, la présence des données de vérité terrain est indispensable. Ces données comprennent une séquence d'images dont les transformations entre ces images sont parfaitement connues. Ainsi, la validité et la précision de

l'estimation de la rotation entre les images sphériques peuvent être vérifiées. Pour cela, nous avons pris le soin d'appliquer des rotations significatives autour de l'axe optique entre les images successives, en incrémentant à chaque fois l'angle de rotation par un pas de 10° dans l'intervalle de 0° à 90° . La figure 2.17 présente les résultats de l'estimation de la rotation 3D en utilisant la méthode proposée. Les rotations estimées par notre algorithme sont proches des données vérité terrain. De même pour la rotation autour de l'axe (Y), où des valeurs négligeables ont été trouvées.

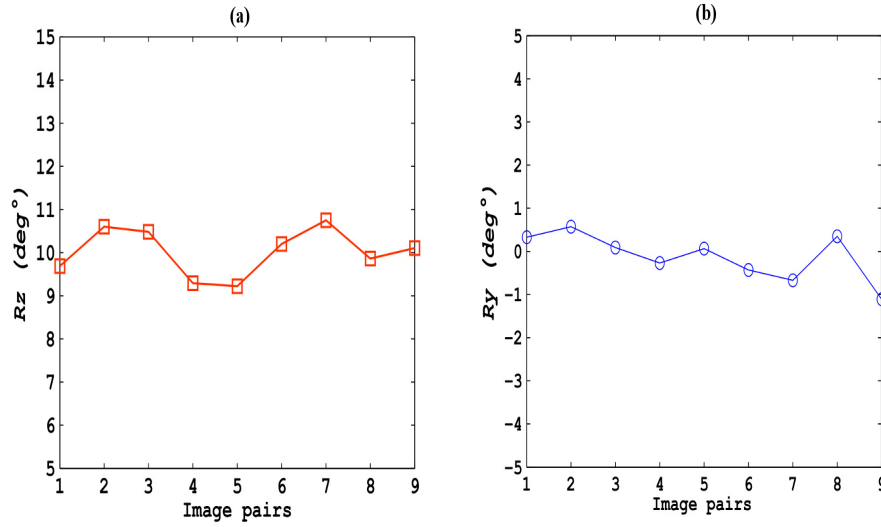


FIGURE 2.17 – Estimation des rotations (a) autour de l'axe (Z). (b) autour de l'axe (Y).

Afin de bien étudier la robustesse de l'algorithme proposé, nous avons également décidé de tester la performance de notre approche vis-a-vis de différentes perturbations appliquées simultanément sur l'image synthétique.

Les perturbations d'image sont plus connues dans la communauté de traitement d'images par les attaques d'images. En effet, les attaques les plus courantes, à savoir l'ajout du bruit, la variation de luminosité et la compression tel que JPEG, ont été testées à travers trois bases de données synthétiques. Pour cela, une base de données générique a été construite dans un premier temps, par la rotation de l'image omnidirectionnelle réelle de 45° à 180° avec un pas de 45° autour de l'axe Z. Par la suite, trois autres bases de données ont été générées à partir de la base générique, en appliquant à chaque fois une des perturbations indiquées précédemment. Ces perturbations nous permettent d'analyser la robustesse de notre algorithme. Le degré maximal L_{max} de SPHARM retenu pour ces tests est fixé à 12 et la bande passante d'échantillonnage de l'image sphérique $B = 32$.

L'estimation initiale est très importante lors du raffinement et de l'alignement des objets 3D. Dans le cas de l'algorithme ICP, la distance initiale entre les deux objets est l'un des facteurs clés qui détermineront la convergence soit vers un minimum local ou vers un minimum global. Cependant, le but de l'estimation initiale est d'avoir une meilleure

estimation initiale et de garantir la convergence vers un minimum global.

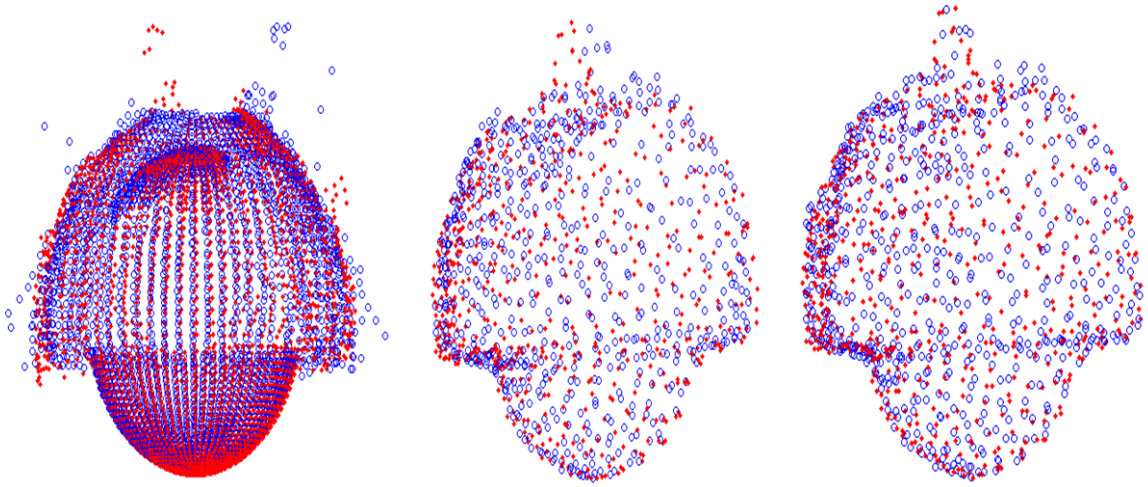


FIGURE 2.18 – Exemple de raffinement de pose sur l'estimation de la rotation. (a) : deux objets séparés par une rotation, (b) l'alignement après l'estimation initiale. (c) l'alignement après le raffinement de l'estimation.

La figure 2.18 illustre un exemple d'étapes d'estimation de la rotation en appliquant notre algorithme sur les deux objets représentés en 2.18-(a). L'alignement des deux objets 3D (le premier en bleu et le deuxième en rouge), en utilisant la solution obtenue lors de l'estimation initiale, est représenté sur la figure 2.18-(b). On peut voir que la position initiale des deux objets 3D est davantage rapprochée. On peut également voir sur la figure 2.18-(c) que, suite au raffinement de l'estimation, l'algorithme converge maintenant vers un minimum global qui est beaucoup plus près du minimum recherché : les deux objets sont alors parfaitement alignés. Par conséquent, on peut conclure qu'une meilleure estimation initiale a permis à l'algorithme de raffinement de converger vers une meilleure solution.

2.6.1.2 Robustesse vis-à-vis du bruit

Afin d'évaluer la robustesse de notre algorithme par rapport au bruit, nous rajoutons à chaque fois à la base de donnée générique d'images avec rotations, un bruit gaussien de différentes variances : $\sigma^2 = 1, 5, 10$ et 20 . Ensuite, nous estimons la rotation 3D par l'algorithme proposé par rapport à l'image de référence (sans bruit). Par conséquent, les erreurs angulaires absolues sont calculées pour chaque image affectée par les différentes variances du bruit Gaussien additif. Les résultats obtenus sont présentés en figure 2.19. Dans cette figure, les barres bleues représentent l'amplitude de l'erreur d'estimation de la rotation autour de l'axe Z, tandis que les barres rouges correspondent à l'axe Y.

Nous pouvons observer que pour des variances de bruit allant jusqu'à 10, notre algorithme est capable d'estimer la rotation avec une erreur absolue moyenne inférieure à 3° ,

ce qui reste une précision raisonnable. Cependant, pour un bruit de variance très importante, le changement introduit au niveau de l'image devient significatif et l'hypothèse de conservation de l'intensité entre les images sphériques n'est plus valable.

2.6.1.3 Robustesse au changement de la luminosité des images

L'expérimentation est effectuée sur la base de données génériques perturbée par le changement de luminosité globale. La luminosité a été réduite de 40% et 20%, puis augmentée de 20% et 40%. La figure 2.20, illustre les résultats des erreurs angulaires absolues (AAE) entre les angles de rotation réels (vérité terrain) et estimés. Comme le montrent ces résultats, la valeur la plus élevée de l'erreur d'estimation est d'environ $2,5^\circ$. Nous pouvons conclure que la précision de l'estimation de la rotation, en utilisant l'approche proposée, n'est pas altérée par la présence du changement de luminosité des images.

2.6.1.4 Robustesses vis-à-vis de la compression d'images

Dans ce dernier test, nous voulons évaluer l'influence du changement de la résolution entre les images sur la précision de l'estimation de la rotation. Ainsi une compression JPEG a été sélectionnée pour dégrader la qualité des images. Comme nous l'avons fait pour les évaluations précédentes, nous introduisons la compression JPEG avec quatre facteurs de qualité différents (QFs = 10, 30, 60 et 90) à l'ensemble des images de la base de données générique. Comme le montre la figure 2.21, on remarque que l'erreur d'estimation augmente progressivement lorsque le facteur de qualité de compression JPEG augmente. Cela est prévisible car la qualité de l'image se dégrade avec l'augmentation du QF. Néanmoins, notre méthode donne généralement des résultats d'estimation appropriés (l'erreur d'estimation est inférieure à $2,5^\circ$) malgré un QF élevé d'environ 90%. Ces résultats sont prévisibles, puisque les objets 3D utilisés pour estimer la rotation, ont été déjà générés à partir des images sphériques échantillonnées à basse résolution.

2.6.1.5 Comparaison avec d'autres algorithmes

Dans le but d'analyser l'efficacité de notre approche, nous l'avons comparée avec d'autres algorithmes similaires. Néanmoins, on recense peu de travaux dans la littérature sur l'estimation de la rotation entre les images sphériques en général et à travers les approches denses en particulier. Seuls les travaux présentés par Makadia et al dans [124], traitent du problème de l'estimation dense de la rotation.

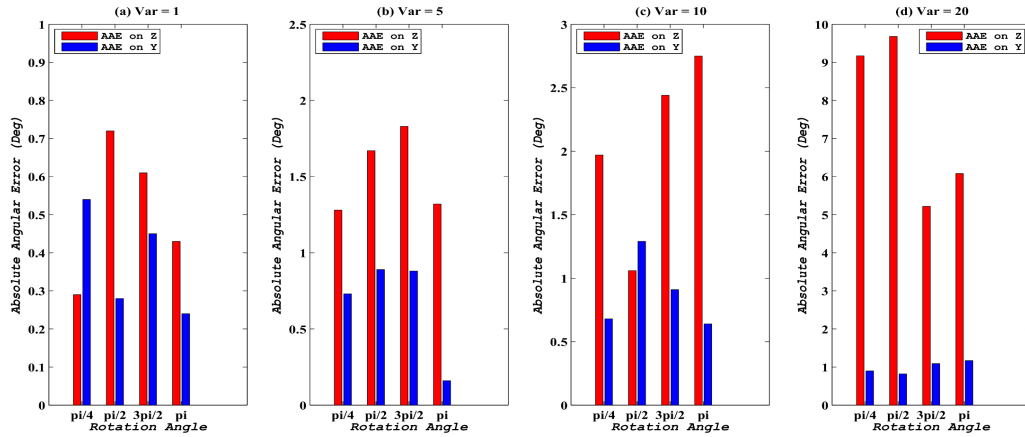


FIGURE 2.19 – Performance en terme d'erreur angulaire absolue pour l'estimation de la rotation autour les axes Z et Y avec différents niveaux de bruits gaussiens ajoutés : (a) $\sigma^2 = 1$, (b) $\sigma^2 = 5$, (c) $\sigma^2 = 10$ and (d) $\sigma^2 = 20$.

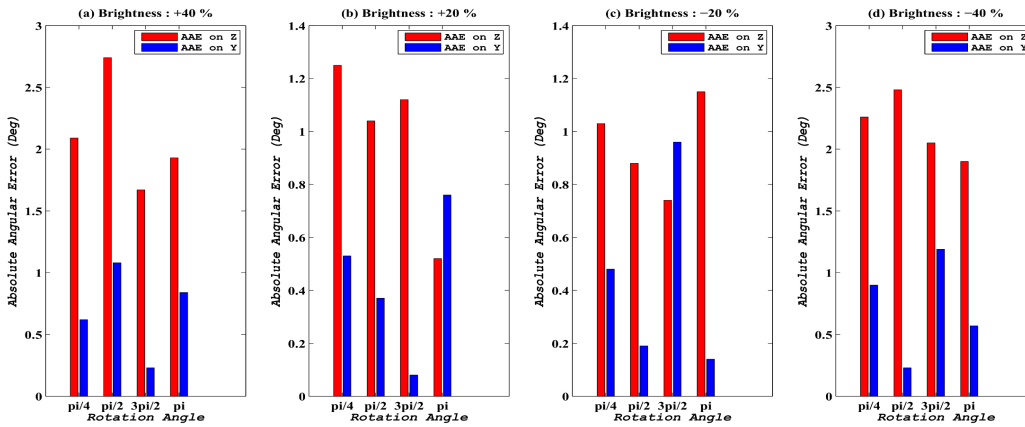


FIGURE 2.20 – Performance en terme d'erreur angulaire absolue pour l'estimation de la rotation autour les axes Z et Y avec différents niveaux de changements de luminosité : (a) +40%, (b) +20%, (c) -20% and (d) -40%.

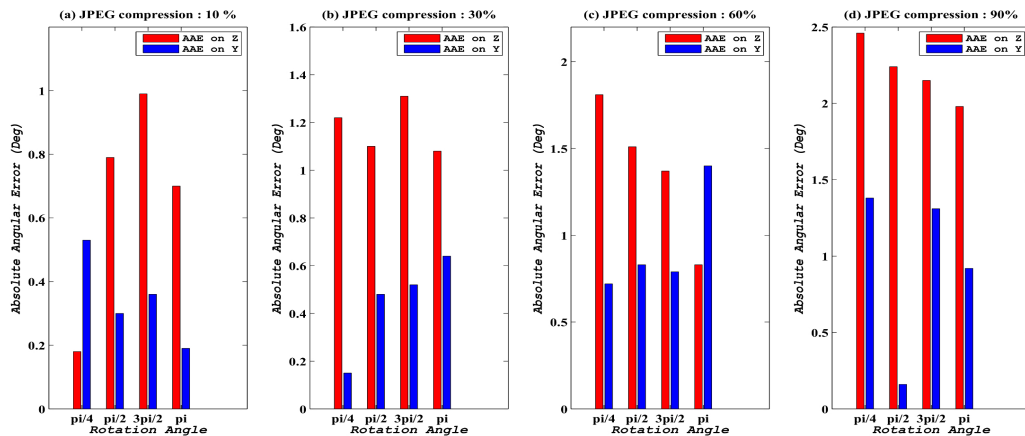


FIGURE 2.21 – Performance en terme d'erreur angulaire absolue pour l'estimation de la rotation autour les axes Z et Y avec différents niveaux de Facteur de Qualité de la compression JPEG : (a) QF = 10, (b) QF = 30, (c) QF = 60 and (d) QF = 90.

En fait, la première approche proposée est DNE (Direct Nonlinear Estimation), où les auteurs ont montré qu'un système non linéaire, dont les inconnues sont les angles d'Euler, peut être minimisé efficacement en utilisant la méthode de quasi-Newton. La deuxième approche proposée est NPM (North Pole Motion), qui consiste à maximiser la corrélation entre les spectres des deux images. Il a été montré qu'à travers la matrice D de Wigner, il est possible d'en extraire le meilleur triplet d'angles d'Euler, qui maximise la corrélation entre les coefficients harmoniques dans l'espace des rotations SO(3).

Comme d'autres solutions, les auteurs ont proposé une approche d'estimation grossière de la rotation appelée NCC (Normalized Cross Correlation), qui se base sur la convolution sphérique des deux images dans le domaine spectral. Elle permet de fournir une localisation des régions dans les images qui constituent un changement de signal important dû à la rotation. Le but de l'estimation grossière est la réduction du coût élevé de la recherche exhaustive de la solution optimale dans tout l'espace des rotations. Cette solution peut fournir une estimation initiale, qu'il est donc possible de raffiner par les deux approches DNE ou NPM. Ce qui a engendré, au total, quatre approches lors de la comparaison, qui sont : DNE, NPM, NCC + DNE et NCC + NPM.

TABLEAU 2.1 – Résumé de la comparaison entre la méthode proposée, DNE, NPM, NCC + DNE et NCC + NPM, pour l'estimation des angles d'Euler. Une valeur de ∞ signifie que la minimisation non linéaire n'a pas convergé.

	Méthodes	α	β	γ
$Is_1(45^\circ, 12^\circ, 45^\circ)$	DNE	29.27°	3.67°	42.8°
	NPM	40.32°	5.12°	32.26°
	NCC+DNE	49.0°	10.2°	43.41°
	NCC+NPM	47.82°	9.52°	46.78°
	Notre méthode	45.78°	11.69°	44.22°
$Is_2(45^\circ, 36^\circ, 45^\circ)$	DNE	–	∞	∞
	NPM	43.88°	13.24°	57.38°
	NCC+DNE	41.19°	30.5°	50.77°
	NCC+NPM	46.28°	33.17°	43.20°
	Notre méthode	44.19°	36.62°	45.85°
$Is_3(45^\circ, 60^\circ, 45^\circ)$	DNE	–	∞	∞
	NPM	38.28°	24.46°	48.36°
	NCC+DNE	43.52°	51.76°	47.34°
	NCC+NPM	44.23°	55.91°	47.82°
	Notre méthode	45.76°	61.47°	44.19°

Il faut noter que, les codes sources des approches proposées dans [124] n'ont pas été fournis. Nous les avons donc programmés par nous-même. La comparaison entre notre approche et ces méthodes compétitives est résumée dans le tableau 2.1. Trois rotations différentes autour de tous les axes (XYZ) ont été appliquées. Le tableau 2.1 montre la supériorité de notre méthode par rapport aux quatre autres méthodes de la littérature.

Nous pouvons voir clairement que l'estimation des angles de rotation obtenus par l'algorithme proposé est plus précise. Cela peut être expliqué par le fait que notre approche possède une étape de raffinement dans l'espace objets, avec une estimation initiale établie dans le domaine spectral. Tandis que les approches compétitives établissent toutes leurs estimations dans le domaine spectral.

Pour conclure, nous avons analysé lors de ces expérimentations la performance et la robustesse de l'algorithme que nous avons proposé pour estimer la rotation 3D entre les images représentées sur la sphère. Cet algorithme présente plusieurs avantages par rapport aux méthodes d'estimation de rotation existantes : 1) La méthode est simple, car elle n'effectue pas de recherche exhaustive sur l'espace de rotation pour trouver le meilleur alignement optimal ; 2) La solution est efficace par rapport aux méthodes directes face aux grandes rotations de la caméra omnidirectionnelle ; 3) Les résultats obtenus sont suffisamment précis (selon l'application) en présence de différents types de perturbations.

2.6.2 Expérimentation sur des images réelles

On présente dans cette partie quelques résultats typiques en conditions réelles afin d'évaluer la pertinence de notre algorithme dans deux cas de figure. Dans ces expériences, nous validons l'algorithme proposé en utilisant des images omnidirectionnelles avec une résolution de 1600 x 1200 pixels, acquises par la caméra catadioptrique centrale (Kumotek VS-C42U-300-TK). Afin de projeter l'image omnidirectionnelle sur la sphère, les paramètres intrinsèques de la caméra ont été récupérés hors ligne, en calibrant la camera par la bibliothèque Omnidirectional Calibration Toolbox proposée dans [125].

2.6.2.1 Première séquence d'images réelles

Afin de valider la précision de l'estimation de l'algorithme proposé sur images réelles d'une part et obtenir une vérité terrain d'autre part, nous avons effectué l'expérience suivante : la camera a été montée rigidement sur un système de trépied (voir Figure. 2.22), on y fait effectuer une rotation sur place et autour de l'axe optique de la caméra (axe Z du trépied) de $[0^\circ \text{ à } 180.0^\circ]$. A partir de l'image de référence captée à l'angle 0° la rotation a été incrémentée d'un pas de 30.0° . Les images ont été prises à chaque position. Les valeurs des angles ont été obtenues à travers les graduations de trépied, alors que les rotations correspondent, par conséquent, à la vérité terrain.

Notons que des petites rotations supplémentaires autour des axes X et Y pourraient être introduites car le système était grossièrement étalonné. De plus, pour assurer une



FIGURE 2.22 – Kumotek VS-C42U-300-TK montée sur un trépied gradué.

grande précision des données de vérité terrain, nous avons opté pour une vérification de la transformation entre une paire d'images omnidirectionnelles captées, par la relation d'homographie [126]. Pour cela, nous avons introduit dans la scène un plan qui apparaît dans les deux prises de vues (voir figure 2.23). Par la suite, la détection des quarts coins du plan ainsi que leur mise en correspondance, ont été établies manuellement. Enfin, le calcul de la transformation qui lie les plans, permet d'extraire avec précision la rotation entre les deux prises de vues.

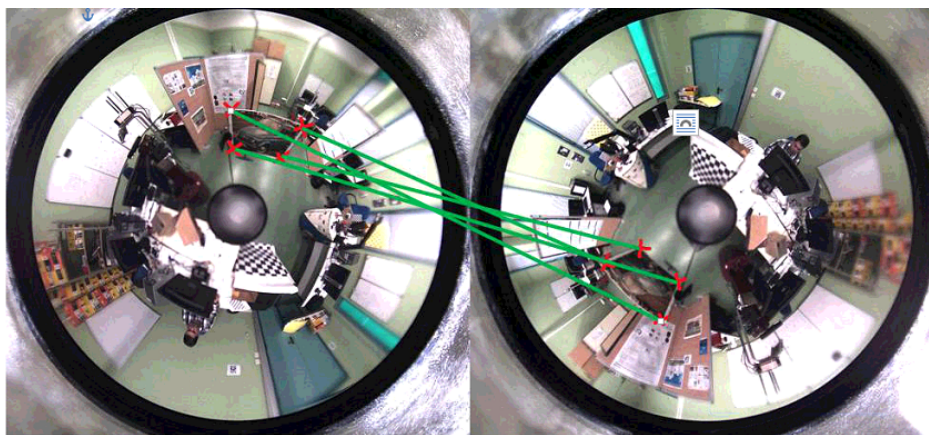


FIGURE 2.23 – Exemple de détection du plan pour le calcul de l'homographie entre les deux images.

Nous évaluons la pertinence de notre algorithme en terme d'erreur angulaire absolue de rotation entre l'estimation de la rotation qui sépare les images successives de la sé-

quence par notre algorithme et les données de la vérité terrain. Les résultats sont illustrés par la figure. 2.24 qui montre que les rotations estimées sont étroitement alignées avec les données de la vérité terrain.

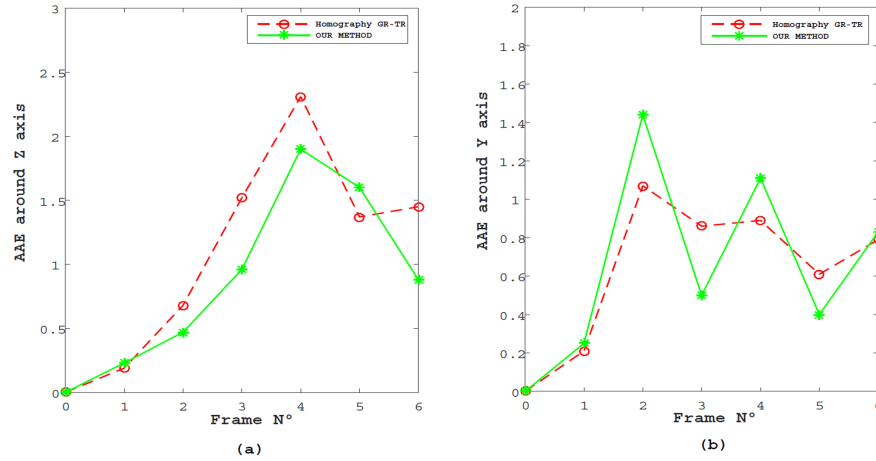


FIGURE 2.24 – Erreur angulaire absolue de l'estimation de la rotation autour des axes : (a) : Z et (b) : Y.

Évidemment, les images réelles peuvent comprendre différent type de perturbations, dont le bruit, les problèmes d'éclairage ou les changements de luminosité d'images et ainsi d'autres perturbations. Par conséquent, les résultats obtenus en cette expérimentation viennent de valider une nouvelle fois l'efficacité de notre approche vis-a-vis le bruit, des changements de luminosité et de la compression JPEG, présenté dans la section précédente.

Dans l'expérience suivante, nous examinons la robustesse de l'approche proposée vis-à-vis des grandes rotations et vis à vis des occultations de l'image.

2.6.2.2 Deuxième séquence d'images réelles

Dans cette expérimentation, nous évaluons la capacité de l'algorithme à estimer la rotation entre les images réelles en présence d'occultations. Deux séquences d'images affectées par trois différents niveaux d'occultation ont été conçus à cet effet. Le niveau d'occultation a été modélisé par le nombre de personnes introduites en scène à chaque acquisition d'image après la rotation de la camera. Notons que l'image à la position initiale sans occultation est considérée comme l'image de référence.

La figure 2.25 illustre en première ligne l'image de référence non-occultée en (a), avec une personne en (b), deux personnes en (c) et trois personnes en (d). Dans la deuxième ligne de cette figure, le niveau d'occultation correspondant à chaque image est présenté et le pourcentage de pixels affectés par l'occultation par rapport au nombre total de pixels

de l'image est donnée (0 %, 4 %, 9 % et 12 %). Ainsi, nous avons testé la précision de notre algorithme en estimant deux différentes rotations de 17° et 105° autour de l'axe Z, qui correspondent consécutivement à de petites et grandes rotations. Les vérités terrain sont calculées avec des mesures physiques à partir du système trépied et confirmées en utilisant la relation d'homographie entre les images omnidirectionnelles.

Le tableau 2.2 montre les erreurs d'estimation pour les différents niveaux d'occultation. On constate que notre algorithme montre toujours une bonne performance lors de l'estimation de la rotation, malgré la perte d'information due à la présence d'occultation dans les images affectées. Cela vient confirmer que les algorithmes denses sont robustes en présence d'occultations d'image. Ces résultats suggèrent que la représentation géométrique de l'intensité lumineuse des images sphériques en l'associant avec la description SPHARM, est adaptée à l'estimation de la rotation des images occultées.

2.6.2.3 Troisième séquence d'images réelles

Cette dernière expérimentation vise à évaluer le comportement de l'algorithme proposé lorsque la transformation entre les images n'est pas une rotation pure, mais en présence de translation aussi. Par conséquent, une séquence d'images omnidirectionnelles réelles a été générée, en appliquant des mouvements de rotation et de translation au système trépied (voir figure 2.22).

Afin d'examiner l'influence des petites et des grandes translations, on déplace le système le long de l'axe X avec les translations de : 10 cm, 30 cm et 60 cm. Pour chaque translation, on applique des rotations autour de l'axe Z (l'axe optique de la camera) de 15° et 45°. Par la suite, nous appliquons notre algorithme pour estimer la rotation entre l'image de référence et les images capturées aux positions données.

Le tableau 2.3 montre les résultats de l'erreur angulaire absolue (AAE) de l'estimation des angles de rotation. Nous pouvons constater que, pour les petites translations de l'ordre de 30 cm, notre algorithme permet d'estimer les angles de rotation avec une er-

TABLEAU 2.2 – Résultats de l'estimation en présence d'occultations. La première colonne correspond aux rotations mesurées à partir de la graduation du trépied. La deuxième colonne correspond aux rotations obtenues à partir de la relation d'homographie (vérité terrain). Les colonnes 3-5 sont les résultats de rotation de chaque image avec clutter (4 %, 9 % et 12 %), respectivement.

Rotation Réelle	Homographie (Vérité terrain) Occultation = 0%		Notre méthode Occultation = 0%		Notre méthode Occultation = 4%		Notre méthode Occultation = 9%		Notre méthode Occultation = 12%	
	R _Z	R _Y	R _Z	R _Y	R _Z	R _Y	R _Z	R _Y	R _Z	R _Y
17°	17.53°	0.35°	17.41°	0.28°	18.16°	0.42°	18.45°	1.13°	15.87°	0.98°
105°	105.66°	0.41°	105.41°	0.45°	105.44°	0.58°	104.64°	1.69°	105.74°	2.31°

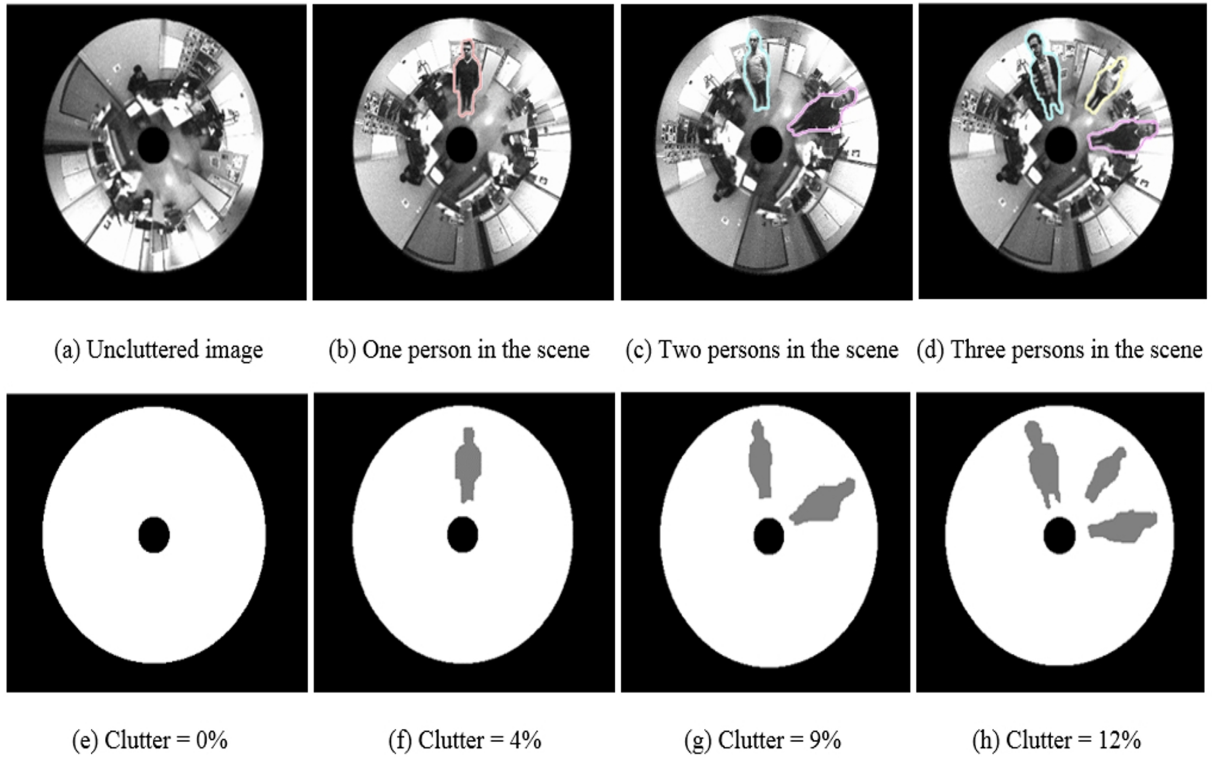


FIGURE 2.25 – Première colonne : les images avec occultation (a) L'image de référence non-occultée, (b) Une personne, (c) Deux personnes et (d) Trois personnes dans la scène. Deuxième rangée : le niveaux d'occultation introduit.

TABEAU 2.3 – AAE de l'estimation de la rotation estimation avec des translations entre les images réelles.

Tran \ Rot	0°		15°		45°	
	R _Z	R _Y	R _Z	R _Y	R _Z	R _Y
0 cm	0	0	0.03	0.98	0.78	0.2
10 cm	0.72	0.91	0.49	0.78	0.89	0.55
30 cm	0.8	1.88	1.57	2.64	1.31	2.59
60 cm	1.78	2.94	1.22	4.55	4.21	2.38

reur absolue moyenne de 2.8° , ce qui est une précision raisonnable, selon la littérature et l'application utilisant la rotation estimée. Alors que l'estimation de la rotation devient moins précise lorsque la translation est supérieure à 30 cm. Par exemple, l'AAE a plus de 4° pour une translation de 60 cm. Ceci vient du fait que l'objet 3D généré à partir de l'intensité lumineuse de l'image sphérique est invariant par rapport au mouvement de rotation. Par contre, il est variant aux translations. En effet, l'approche proposée est basée sur l'hypothèse donnée par l'équation 2.11. Cette dernière n'est pas satisfaite lors de l'existence d'une translation entre les images. En revanche, ces résultats expérimentaux montrent que notre approche reste robuste en présence de faibles mouvements de translation.

2.6.3 Efficacité et coût en temps de calcul

Afin d'évaluer notre algorithme en terme de complexité, nous proposons d'étudier et de comparer ses performances avec une autre méthode similaire et très répandue dans la littérature. Cette méthode, proposée dans [100], se base sur les coefficients harmoniques des images sphériques, une recherche exhaustive dans l'espace des rotations 3D, pour sélectionner la meilleure solution qui aligne les deux images.

Nous rappelons que la complexité d'un algorithme est le nombre d'opérations de base qu'il doit effectuer en fonction de la taille de données d'entrée traitées. Cette complexité est indépendante de la vitesse du processeur sur lequel est exécuté l'algorithme. En effet, la vitesse du processeur peut faire changer le temps d'exécution d'une opération de base, mais ne change pas le nombre d'opérations à effectuer (donc la complexité).

Dans la suite, nous allons détailler davantage les coûts computationnels des deux algorithmes comparés. Commençons par notre algorithme, la première étape de la génération d'objets 3D à partir d'une image sphérique échantillonnée par bande passante B , requiert dans sa forme originale $O(B^2)$ opérations, lorsque la complexité de l'algorithme de la paramétrisation sphérique est de $O(B^2)$. Le coût de la décomposition des harmoniques sphériques implique $O(L^2 \log^2 L)$ opérations [111]. Enfin, le coût de la décomposition SVD et l'algorithme ICP est de $O(L^2)$. L'approche proposée dépend globalement de la bande d'harmoniques sphériques, L_{max} . À partir des coûts de chaque étape de l'algorithme, on constate que $O(L^2 \log^2 L)$ est le coût computationnel de notre algorithme afin d'estimer la rotation optimale.

Les coûts computationnels des étapes, pour estimer la rotation en minimisant la distance L_2 entre deux images sphériques comme présenté dans [100], sont les suivants : la transformation harmonique sphérique pour obtenir les coefficients harmoniques des images sphériques est de l'ordre de $O(L^2 \log^2 L)$, l'inter-multiplication entre les coefficients harmoniques sphériques vaut $O(L^3)$, le calcul de la transformée inverse de Wigner-D est de l'ordre $O(L^4)$ et l'échantillonnage de la fonction inverse obtenue a besoin de $O(L^3)$ opération [106]. Au total, le calcul de l'estimation de la rotation en utilisant l'algorithme proposé dans [100], peut être effectué en $O(L^4)$.

Notre approche permet la réduction de la complexité de calcul du problème de l'estimation de la rotation entre les images sphériques : au lieu d'effectuer le calcul avec un coût computationnel de $O(L^4)$ pour calculer la rotation en utilisant une approche dense, grâce à notre approche, il suffira de faire un calcul avec un coût computationnel de $O(L^2 \log^2 L)$, ce qui permet une réduction considérable du temps de calcul.

2.7 Conclusion

Dans ce chapitre, nous avons présenté une approche originale qui utilise tout le signal image pour estimer la rotation 3D à partir des images sphériques. Cette approche est basée sur la représentation 3D de l'intensité lumineuse de ces images. A travers cette représentation, la transformation géométrique entre deux images dans le domaine sphérique a été reformulée comme étant un problème d'estimation des paramètres de cette transformation entre les deux objets 3D générés. En effet, notre approche est aussi valable pour estimer la rotation relative entre deux objets 3D. En explorant le descripteur global SPHARM (SPHerical HARMonic) des maillages surfaciques 3D de topologie sphérique, une solution directe basée sur les coefficients d'harmoniques sphériques, a été proposée, pour trouver une estimation initiale. La solution finale est obtenue en raffinant l'estimation initiale à l'aide de la technique ICP qui assure une convergence rapide vers la rotation 3D optimale.

Les résultats obtenus ont illustré explicitement les performances de notre approche. La représentation géométrique 3D de l'intensité de l'image ainsi que la description SPHARM, ont permis d'obtenir une approche dense adaptée à l'estimation de grandes rotations entre les images sphériques. En outre, elle permet d'envisager avec efficacité des estimations de la rotation en présence de bruit, de changements de luminosité, de compression JPEG et d'occultations.

Chapitre 3

Estimation de la rotation 3D basée moments sphériques

Sommaire

3.1 Introduction	78
3.2 Les moments géométriques	79
3.3 Matrice d'interaction	83
3.3.1 Matrice d'interaction d'un point	84
3.4 Matrices d'interaction associées aux moments 2D (cas continu)	85
3.5 Matrices d'interaction associées aux moments 2D (cas discret)	87
3.6 Matrice d'interaction des moments sphériques	88
3.6.1 Expression des moments en projection sphérique	89
3.6.2 Matrice d'interaction d'un point en projection sphérique	89
3.6.3 Projection d'un élément de surface sur la sphère unitaire	90
3.6.4 Matrice d'interaction des moments sphériques	91
3.7 Estimation de la rotation basée moments photométriques sphériques	93
3.7.1 Première méthode (M1)	97
3.7.2 Deuxième méthode (M2) :	98
3.7.3 Troisième méthode (M3) :	98
3.8 Expérimentations	99
3.8.1 Données synthétiques	99
3.8.2 Données réelles	104
3.9 Conclusion	114

3.1 Introduction

Le problème de l'estimation de la rotation 3D entre deux images sphériques, basé sur une approche dense, peut être résolu en exploitant les moments de l'image. Les propriétés des moments ont d'abord été introduites dans les communautés de vision par ordinateur et de traitement d'images en 1962 [127], lorsque Hu a proposé les moments géométriques 2-D d'une fonction de distribution structurée (une image) de ce qu'il a appelé "les moments invariants ". Dans ce travail, Hu a utilisé la théorie des invariants algébriques afin de dériver ses sept fameux moments invariants aux transformations linéaires (translation, rotation et changement d'échelle). Dès lors, un grand nombre d'approches dans la théorie des moments et l'invariance des moments ont été présentées. Ce qui a permis de renforcer l'applicabilité des moments dans de nombreuses disciplines. Plusieurs méthodes très diversifiées sont ainsi présentées dans la littérature, dont certaines seront citées dans ce qui suit.

Dans l'analyse d'images, les propriétés importantes des moments, dont la capacité de capturer de manière compacte et avec une redondance minimale l'information contenue dans une image, ont permis de les utiliser avec succès dans de nombreuses applications telles que la segmentation de texture [128], l'enregistrement d'images [129, 130], la détection de bord de sous-pixel [131, 132], la compression d'image [133] et débruitage d'images [134, 135]. Récemment, les moments d'image ont montré une performance prometteuse dans la description des primitives extraites de l'image [136, 137].

Dans le domaine de la reconnaissance et la de classification des formes, les propriétés principales des moments tels que la robustesse par rapport aux transformations géométriques, non géométriques et les invariances du contenu de l'image, permettent de décrire de manière unique la distribution des pixels. La première application des moments invariants remonte aux premiers travaux de [127, 138, 139]. Depuis, ils ont montré une performance remarquable dans la reconnaissance des cibles [140, 141], la reconnaissance de l'iris [142, 143], la reconnaissance des gestes [144, 145], la reconnaissance des expressions faciales [146, 147], la classification des actions humaines [148] et plusieurs moments invariants ont été utilisés comme primitives caractéristiques pour détecter et classer les tumeurs [149].

En robotique, les moments de l'image ont été largement utilisés pour réaliser la tâche de positionnement et l'estimation de déplacement notamment pour l'asservissement visuel [150]. En effet, afin de rendre possible l'utilisation des moments 2D dans un asservissement visuel, il s'agit, de modéliser la matrice d'interaction permettant d'exprimer leur variation en fonction du torseur cinématique de la caméra. Dans un autre travail, Tahri et Al [151] ont exploité les propriétés d'invariance des moments pour obtenir un dé-

couplage entre la commandes de différents degrés de liberté. Ceci a permis d'élargir le domaine de convergence. La méthodologie proposée dans [151] a été exploitée pour des moments calculés à partir d'images binaires ou à partir de nuage de points [152]. Dans [153], il a été démontré que les propriétés de découplage des moments peuvent être également exploité pour le calcul de pose. Dans [154], une nouvelle variété de moments invariants appelés "Shifted moments invariants" a été proposée. L'utilisation de cette nouvelle variété d'invariants permet une nette amélioration du domaine de convergence de la commande, mais aussi l'obtention d'une combinaison unique d'information visuelles indépendante de la forme des objets par auxquels les six degrés de liberté d'une caméra sont asservis. Récemment, Bakthavatchalam et al [155] a exploité les invariants proposés dans [154] pour des moments photométriques calculés directement à partir des niveaux de gris d'une image. Afin de remédier à la problématique des entrées sorties de parties d'image pendant l'asservissement visuel, une fonction de pondération a été introduite dans le calcul des moments dans [156].

3.2 Les moments géométriques

Dans cette section, nous présenterons les définitions relatives aux moments (géométriques et orthogonaux). Aussi, nous détaillerons la nature de l'information qui peut être dégagée des moments géométriques, ainsi que leur signification physique pour décrire les formes des objets.

Les moments ont été proposés initialement en statistiques. Un ensemble de moments caractérisant une entité représente généralement les caractéristiques globales de la forme de l'entité et donne aussi, en général, une large indication sur la position de l'entité. Communément utilisés en mécanique du solide, la formule générale des moments se décompose en deux termes : une fonction de base qui est un polynôme $B(p, q)$ représentant la base dans laquelle sera projetée le deuxième terme, à savoir la fonction image (ou forme) $f(x, y)$.

$$M(p, q) = \int_x \int_y B(p, q) f(x, y) dx dy \quad (3.1)$$

Les indices p et q représentent les degrés des polynômes au position définie par x et y de la fonction f .

Les moments proposés dans la littérature diffèrent par la nature de leurs fonctions de base utilisées. Ainsi, nous retrouvons des fonctions de bases reposant sur des poly-

nômes non-orthogonaux (moments géométriques, moments complexes, ...), et d'autres reposant sur des polynômes orthogonaux (moments de Legendre, moments de Zernike, moments de Tchebychev, ...). Nous nous intéressons dans la suite de ce chapitre, seulement aux moments cartésiens, car les autres moments ne présentent pas un intérêt pour les applications traitées dans ce travail.

• Les moments 2D

La fonction noyau à partir de laquelle sont définis les moments géométriques (appelés aussi moments cartésiens) s'appuie sur les coordonnées (x, y) dans le plan cartésien. Le calcul des moments géométriques est le plus simple à mettre en œuvre comparant aux moments définis par d'autres fonctions.

Les moments géométriques sont définis par l'ensemble des monômes $x^p y^q$; le $(n)^{ième}$ moment géométrique, n représente son ordre avec $n = p + q$, noté m_{pq} est défini par :

$$m_{p,q} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q f(x, y) dx dy \quad (3.2)$$

Il est possible de définir les moments en termes de distribution d'intensité à partir d'une image qui capture une scène. Ces moments d'image définis à partir des intensités sont aussi connus par les moments photométriques, qui peuvent être définis par :

$$m_{p,q} = \int \int_{\xi} x^p y^q I(x, y) dx dy \quad (3.3)$$

Où (x, y) sont les coordonnées des pixels dans l'image, appartenant à la région ξ sur laquelle la fonction de l'intensité de l'image I est définie.

Chacun des moments géométriques (selon son ordre) représente une caractéristique spatiale de la distribution de l'objet dans l'espace. Un ensemble de moments peut donc former un descripteur global de la forme de l'objet. L'interprétation physique de quelques moments est donnée dans cette section.

Par définition, le moment d'ordre 0 (m_{00}) représente l'intensité totale de la fonction f représentant l'entité. Dans le cas d'une entité surfacique, ce moment donne la mesure de son aire.

$$m_{p,q} = \int \int_{\xi} f(x, y) dx dy \quad (3.4)$$

Les moments de premier ordre m_{10} et m_{01} donnent une indication sur la répartition de la fonction f , respectivement, selon l'axe des x et selon l'axe des y . Par la suite, ils

participent à la détermination du centre de masse de l'entité donné par :

$$x_0 = \frac{m_{10}}{m_{00}}, \quad y_0 = \frac{m_{01}}{m_{00}} \quad (3.5)$$

Les moments d'ordre 2 m_{20} , m_{11} et m_{02} sont appelés les moments d'inertie. Avec les coordonnées de centre de gravité, ils permettent de définir un ellipse dont les d'ordre inférieur ou égal à 2 ont la même valeur que ceux de l'objet. Soit "y = c" l'axe de giration (de rayon c) du polygone autour de l'axe des x et parallèle à l'axe des abscisses (voir figure 3.1). On peut donc écrire la relation suivante :

$$m_{p,q} = \int \int_{\xi} c^2 f(x, y) dx dy \quad (3.6)$$

Avec :

$$c = \sqrt{\frac{m_{02}}{m_{00}}}$$

D'une manière analogue, le rayon de giration du polygone autour de l'axe des ordonnées est donné par :

$$c = \sqrt{\frac{m_{20}}{m_{00}}}$$

Afin de rendre les moments invariants au mouvement de translation, il est judicieux d'utiliser les moments centraux. Ces moments sont définis comme suit :

$$\mu_{p,q} = \int \int_{\xi} (x - x_0)^p (y - y_0)^q f(x, y) dx dy \quad (3.7)$$

En utilisant l'équation 3.7, nous obtenons : $\mu_{00} = m_{00}$ et $\mu_{10} = \mu_{01} = 0$

Les moments principaux d'inertie notés $I1$, $I2$ peuvent être exprimés en fonction des moments centraux du second ordre de la manière suivante :

$$I1 = \frac{(\mu_{20} + \mu_{02}) + \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2} \quad I2 = \frac{(\mu_{20} + \mu_{02}) - \sqrt{(\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2}}{2} \quad (3.8)$$

On note que si $\mu_{11} = 0$, on a $I1 = \mu_{20}$ et $I2 = \mu_{02}$.

L'angle d'orientation θ que fait l'axe principal majeur d'inertie avec l'axe des abscisses est donné par l'équation suivante :

$$\mu_{11} \tan^2(\theta) + (\mu_{20} - \mu_{02}) \tan(\theta) - \mu_{11} = 0$$

$$\theta = \frac{1}{2} \arctan\left(\frac{2\mu_{11}}{\mu_{20} - \mu_{02}}\right) \quad (3.9)$$

Les équations 3.8 et 3.9 peuvent être utilisées pour définir une ellipse caractéristique qui aura les mêmes moments d'inertie, ainsi que les mêmes directions des axes principaux du polygone d'origine. Le grand axe a et le petit axe b de l'ellipse sont donnés par les équations suivantes :

$$a = 2\sqrt{\frac{I_1}{\mu_{00}}} \quad b = 2\sqrt{\frac{I_2}{\mu_{00}}} \quad (3.10)$$

L'ellipse caractéristique donne une idée sur la forme fondamentale de l'objet. Les paramètres a , b et θ sont appelés les descripteurs elliptiques de la forme du polygone (voir figure 3.1).

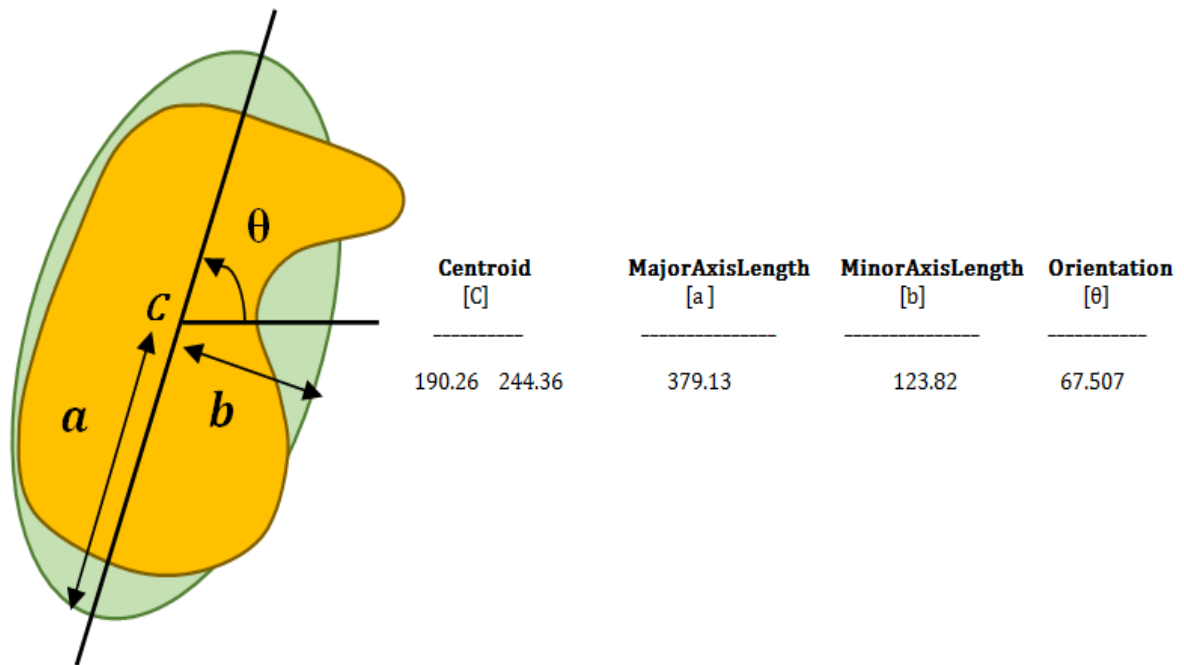


FIGURE 3.1 – Orientation d'un objet et définition des paramètres usuels d'une ellipse.

Les moments du troisième ordre μ_{30} et μ_{03} donnent respectivement une indication sur la symétrie selon l'axe des abscisses et selon l'axe des ordonnées. Si le polygone est symétrique autour de l'axe : $x = x_0$, alors la valeur de μ_{30} est nulle. On peut, en effet, considérer μ_{30} comme une mesure de symétrie autour de l'axe "moyen" $x = x_0$. Étant donnée que la valeur de μ_{20} est toujours positive, on peut diviser le terme μ_{30} par $\mu_{20}^{3/2}$ pour obtenir une quantité sans dimension. Cette quantité est appelée le "coefficient de symétrie" autour de l'axe des abscisses. Le coefficient $\frac{\mu_{03}}{\mu_{02}^{3/2}}$ peut être également calculé pour mesurer la symétrie autour de l'axe des ordonnées.

- **Les moments 3D**

Les moments géométriques d'ordre $(p+q+r)$, notés par m_{pqr} de la fonction de densité $f(x, y, z)$, peuvent être calculés de la même manière que les moments 2D, par l'expression suivante :

$$m_{pqr} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x^p y^q z^r f(x, y, z) dx dy dz \quad (3.11)$$

Ainsi, les moments 3D centrés sont définis par :

$$\mu_{pqr} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} (x - x_g)^p (y - y_g)^q (z - z_g)^r f(x, y, z) dx dy dz \quad (3.12)$$

ou $x_g = \frac{m_{100}}{m_{000}}$, $y_g = \frac{m_{010}}{m_{000}}$ et $z_g = \frac{m_{001}}{m_{000}}$ représentent les coordonnées du centre de gravité de l'objet 3D.

3.3 Matrice d'interaction

Afin de réaliser la tâche de positionnement et d'estimation du mouvement en se base sur les informations visuelles, il est nécessaire de modéliser la variation de ces informations induite par le mouvement du capteur visuel. En effet, la matrice d'interaction, exprime les variations créées dans l'image en fonction du mouvement de la camera. Cela peut être formulé et illustré par ce qui suit. Considérant les deux vecteurs $\mathbf{s}$ et $\mathbf{x}$ qui représentent respectivement les primitives visuelles et l'ensemble des coordonnées des pixels de l'image, et soit le vecteur $\mathbf{q}$ quantifiant le mouvement effectué par la camera. La matrice d'interaction $\mathbf{L}_x$ relie les changements différentiels $\dot{\mathbf{x}}$ de $\mathbf{x}$ aux changements différentiels du mouvement de la camera $\dot{\mathbf{q}}$ de $\mathbf{q}$:

$$\dot{\mathbf{x}} = \mathbf{L}_x \dot{\mathbf{q}}$$

Les changements différentiels $\dot{\mathbf{s}}$ de $\mathbf{s}$ peuvent être écrits comme suit :

$$\begin{aligned} \dot{\mathbf{s}} &= \frac{\partial \mathbf{s}}{\partial \mathbf{x}} \dot{\mathbf{x}} \\ &= \frac{\partial \mathbf{s}}{\partial \mathbf{x}} \mathbf{L}_x \dot{\mathbf{q}} \\ &= \mathbf{L}_s \dot{\mathbf{q}} \end{aligned} \quad (3.13)$$

$\mathbf{L}_s$ est la matrice d'interaction reliant la variation des informations visuelles dans l'image

$\mathbf{s}$ à la vitesse $\mathbf{v}$ de la camera, tel que $\mathbf{v} = (\mathbf{v}, \boldsymbol{\omega})$ est le torseur cinématique du capteur. Il se compose des vitesses de translation $\mathbf{v} = (v_x, v_y, v_z)^\top$ et de rotation $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^\top$.

3.3.1 Matrice d'interaction d'un point

Différents travaux ont permis de déterminer le modèle algébrique de la matrice d'interaction pour différents types de primitives. Dans le cas où le modèle de projection de la caméra est perspective, la matrice d'interaction a été présentée pour des primitives visuelles de type point, droite, cercle et etc ... [157]. De manière très classique, le modèle mathématique d'une caméra est défini alors par une projection perspective telle que tout point $\mathbf{P}$ ayant pour coordonnées $\mathbf{P} = (X, Y, Z)^\top$ dans le repère camera, se projette sur le plan image en un point $\mathbf{p}$ de coordonnées $(x, y)^\top$, comme indiqué sur la figure 3.2, avec :

$$\begin{cases} x = \frac{X}{Z} \\ y = \frac{Y}{Z} \end{cases} \quad (3.14)$$

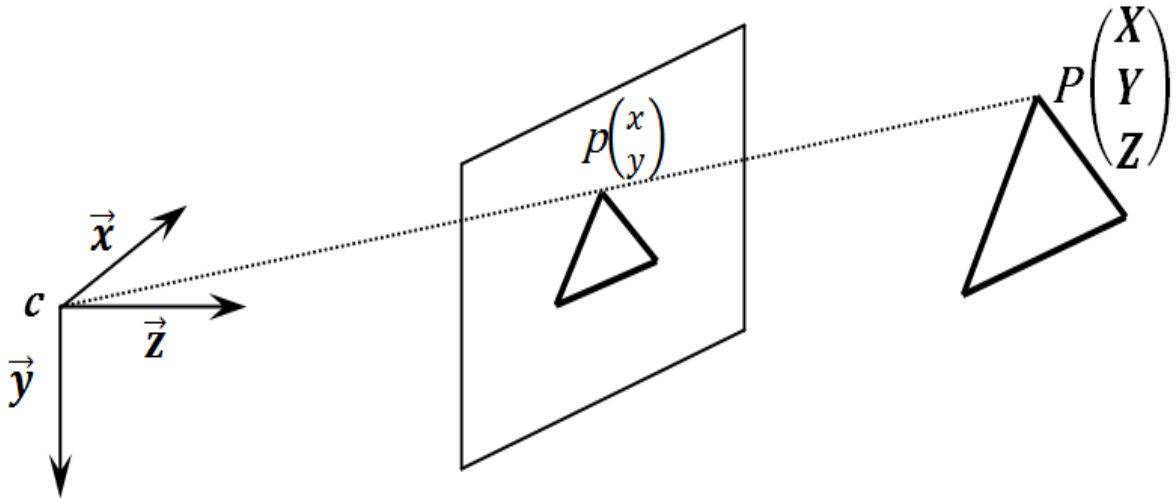


FIGURE 3.2 – Modèle de projection d'un point.

Le vecteur des informations sensorielles $\mathbf{s}$ est alors défini par $\mathbf{s} = (x, y)^\top$. En vision perspective, la matrice d'interaction L_s s'écrit [157] :

En vision perspective, la matrice d'interaction L_s s'écrit [157] :

$$L_s = \begin{pmatrix} \dot{x} \\ \dot{y} \end{pmatrix} = \begin{bmatrix} -\frac{1}{Z} & 0 & \frac{x}{Z} & xy & -(1+x^2) & y \\ 0 & -\frac{1}{Z} & \frac{y}{Z} & 1+y^2 & -xy & -x \end{bmatrix} \quad (3.15)$$

3.4 Matrices d'interaction associées aux moments 2D (cas continu)

Dans ce qui suit, nous rappelons la matrice d'interaction associées au moments 2D. On suppose que l'image est binaire, ou que le niveau de gris est invariant aux mouvements de la caméra. L'objectif est de calculer la variation temporelle de m_{pq} défini par [158] :

$$m_{pq} = \iint_{\Omega_{int}} k(x, y, \Omega_{int}) dx dy \quad (3.16)$$

où le descripteur de la région est défini par l'expression :

$$k(x, y, \Omega_{int}) = x^p y^q f(x, y) \quad (3.17)$$

Pour calculer la variation temporelle de m_{pq} , nous pouvons utiliser le théorème de Green-Riemann [159] ou bien une approche basée sur la dérivée Eulérienne [160]. Le calcul est détaillé en annexe (A.2).

De même, les variations d'un point de l'image par rapport aux mouvements de la caméra sont analytiquement connues (voir l'équation 3.15).

$$\dot{x} = -\frac{v_x}{Z} + x \frac{v_z}{Z} + xy\omega_x - (1+x^2)\omega_y + y\omega_z \quad (3.18a)$$

$$\dot{y} = -\frac{v_y}{Z} + y \frac{v_z}{Z} + (1+y^2)\omega_x - xy\omega_y + x\omega_z \quad (3.18b)$$

Considérons un objet O fixe de la scène qui se projette dans l'image en une région R ayant un contour fermé C (voir figure 3.3).

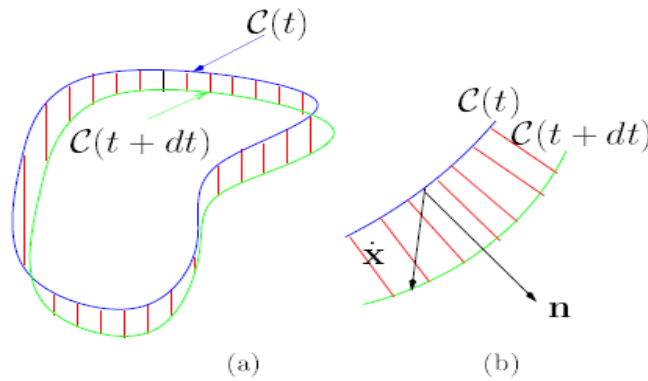


FIGURE 3.3 – (a) Illustration des variations temporelles d'un contour d'image fermé $C(t)$ (b) détails

Nous admettons que cet objet est formé d'une surface continue, c'est-à-dire ne pré-

sentant pas de discontinuité sur la profondeur. Dans ce cas, la profondeur Z de chaque point de l'objet O peut s'exprimer en fonction des coordonnées (x, y) de sa projection dans l'image de la manière suivante [1], [157], [161] :

$$\frac{1}{Z} = \sum_{p \geq 0, q \geq 0} A_{pq} x^p y^q \quad (3.19)$$

En utilisant les équations de projections perspectives 3.14, la profondeur pour un point appartient à un plan, peut s'écrire par la forme suivante [1] :

$$\frac{1}{Z} = Ax + By + C \quad (3.20)$$

A , B et C sont les paramètres liés au plan définie par la surface des limbes plane. En remplaçant l'équation 3.20 dans les équations 3.18a et 3.18b, nous obtenons les dérivées partielles suivantes :

$$\frac{\partial \dot{x}}{\partial x} = -Av_x + (2Ax + By + C)v_z + y\omega_x - 2x\omega_y \quad (3.21a)$$

$$\frac{\partial \dot{y}}{\partial y} = -Bv_y + (Ax + 2By + C)v_z + 2y\omega_x - x\omega_y \quad (3.21b)$$

La matrice d'interaction $\mathbf{L}_{m_{pq}}$ associée au moment m_{pq} , est exprimée sous la forme suivante (pour plus de détail voir [1]) :

$$\mathbf{L}_{m_{pq}} = \left[\mathbf{L}_{m_{pq}}^v \quad \mathbf{L}_{m_{pq}}^\omega \right]^\top = \left[L_{m_{pq}}^{v_x} \quad L_{m_{pq}}^{v_y} \quad L_{m_{pq}}^{v_z} \quad L_{m_{pq}}^{\omega_x} \quad L_{m_{pq}}^{\omega_y} \quad L_{m_{pq}}^{\omega_z} \right]^\top \quad (3.22)$$

Avec :

$$\begin{cases} L_{m_{pq}}^{v_x} = -A(p+1)m_{p,q} - Bp m_{p-1,q+1} - Cp m_{p-1,q} \\ L_{m_{pq}}^{v_y} = -Aq m_{p+1,q-1} - B(q+1)m_{p,q} - Cq m_{p,q-1} \\ L_{m_{pq}}^{v_z} = A(p+q+3)m_{p+1,q} + B(p+q+3)m_{p,q+1} + C(p+q+2)m_{p,q} \\ L_{m_{pq}}^{\omega_x} = q m_{p,q-1} + (p+q+3)m_{p,q+1} \\ L_{m_{pq}}^{\omega_y} = -p m_{p-1,q} - (p+q+3)m_{p+1,q} \\ L_{m_{pq}}^{\omega_z} = p m_{p-1,q+1} - q m_{p+1,q-1} \end{cases} \quad (3.23)$$

Ainsi, une forme analytique de la matrice d'interaction peut être construite pour les moments d'image. A partir de 3.23, en calculant la matrice d'interaction d'un moment d'ordre $p + q$, les moments d'ordre jusqu'au $p + q + 1$ sont nécessaires. La partie de la matrice d'interaction liée aux vitesses de rotation $\mathbf{L}_{m_{pq}}^\omega$ est indépendante des paramètres du plan A, B et C. Ce qu'est tout à fait logique, car les mouvements de rotation ne change pas de la profondeur de la scène.

3.5 Matrices d'interaction associées aux moments 2D (cas discret)

Les moments peuvent également être calculés à partir d'un ensemble de points discrets extraits de l'image. Dans le cas d'un ensemble de points 2D $\{\mathbf{x}\}_i = \{x_i, y_i\}$, nous avons :

$$m_{pq} = \sum_{i=1}^N x_i^p y_i^q \quad (3.24)$$

En dérivant l'équation 3.24, on obtient :

$$\dot{m}_{pq} = \sum_{i=1}^N \left(p x_i^{p-1} y_i^q \dot{x}_i + q x_i^p y_i^{q-1} \dot{y}_i \right) \quad (3.25)$$

En remplaçant les équations 3.18a, 3.18b et 3.20 dans 3.25, on peut obtenir la matrice d'interaction des moments pour le cas discret, qu'est définie par :

$$\mathbf{L}_{m_d pq} = \left[\mathbf{L}_{m_d pq}^v \mathbf{L}_{m_d pq}^\omega \right]^\top = \left[\mathbf{L}_{m_d pq}^{v_x} \mathbf{L}_{m_d pq}^{v_y} \mathbf{L}_{m_d pq}^{v_z} \mathbf{L}_{m_d pq}^{\omega_x} \mathbf{L}_{m_d pq}^{\omega_y} \mathbf{L}_{m_d pq}^{\omega_z} \right]^\top \quad (3.26)$$

avec :

$$\left\{ \begin{array}{l} L_{m_{pq}}^{v_x} = -p(Am_{p,q} + Bm_{p-1,q+1} + Cm_{p-1,q}) - \delta Am_{p,q} \\ L_{m_{pq}}^{v_y} = -q(Am_{p+1,q-1} + Bm_{p,q} - Cm_{p,q-1}) - \delta Bm_{p,q} \\ L_{m_{pq}}^{v_z} = (p + q + 3\delta)(Am_{p+1,q} + Bm_{p,q+1} + Cm_{p,q}) - \delta Cm_{p,q} \\ L_{m_{pq}}^{\omega_x} = qm_{p,q-1} + (p + q + 3\delta)m_{p,q+1} \\ L_{m_{pq}}^{\omega_y} = -pm_{p-1,q} - (p + q + 3\delta)m_{p+1,q} \\ L_{m_{pq}}^{\omega_z} = pm_{p-1,q+1} - qm_{p+1,q-1} \end{array} \right. \quad (3.27)$$

Quand $\delta = 0$, la matrice d'interaction pour les moments calculés à partir d'un ensemble discret de points (cas discrets) est différente de celle des moments calculés à partir d'une distribution continue. Il est à noter que l'équation 3.27 est générale et applicable aux moments calculés à la fois dans le cas d'une distribution continue et discrets. En effet, quand $\delta = 1$, l'ensemble d'équations ci-dessus sont similaires aux équations 3.23 obtenues pour le cas continu.

3.6 Matrice d'interaction des moments sphériques

Récemment, il est apparu de nombreux systèmes de vision qui, contrairement aux caméras perspectives, ont un large champ de vue. La proposition du modèle de projection sphérique a mené à une modélisation unifiée et valable pour tous les systèmes de vision centraux. Ce modèle de projection présente beaucoup d'avantages notamment pour les mouvements de rotation. En fait, une rotation 3D de la caméra se traduit par le même mouvement de la projection sphérique de l'objet présent sur l'image. Cette propriété va permettre à l'objet de conserver sa forme quelque soit le mouvement de rotation appliqué à la caméra. En explorant cette propriété, des nouvelles informations visuelles peuvent être extraites depuis les moments de l'image, pour estimer la rotation entre deux poses de la caméra.

Nous présentons dans ce qui suit, les moments définis à partir d'une projection sphérique, qui sont dénommés : "*moments sphériques*". En effet, ce sont les moments classiques calculés sur la surface de la sphère. Étant les équations sont obtenues, nous abordons ensuite la matrice d'interaction correspondante à ces moments, pour le cas d'une distribution continue et discret.

3.6.1 Expression des moments en projection sphérique

A partir du modèle de projection unifié des caméras centrales, il est possible d'exprimer les coordonnées de points sur la sphère unitaire par :

$$\begin{cases} x_s = \frac{x}{\sqrt{x^2+y^2+1}} = \frac{X}{\sqrt{X^2+Y^2+Z^2}} \\ y_s = \frac{y}{\sqrt{x^2+y^2+1}} = \frac{Y}{\sqrt{X^2+Y^2+Z^2}} \\ z_s = \frac{1}{\sqrt{x^2+y^2+1}} = \frac{Z}{\sqrt{X^2+Y^2+Z^2}} \end{cases} \quad (3.28)$$

De même, les coordonnées dans l'image 2D sont données par :

$$\begin{cases} x = \frac{x_s}{\sqrt{1-(x_s^2+y_s^2)}} \\ y = \frac{y_s}{\sqrt{1-(x_s^2+y_s^2)}} \end{cases} \quad (3.29)$$

Les moments sphériques sont donc définis par :

$$m_{spqr} = \iiint_{\Omega(t)} x_s^p y_s^q z_s^r f(x_s, y_s, z_s) ds \quad (3.30)$$

où $\Omega(t)$ représente la surface de la sphère à un instant t .

3.6.2 Matrice d'interaction d'un point en projection sphérique

En dérivant l'équation 3.28, on obtient [162] :

$$\begin{cases} \dot{x}_s = \frac{\dot{X}}{\sqrt{X^2+Y^2+Z^2}} - \frac{X}{\sqrt{X^2+Y^2+Z^2}} \frac{\dot{X}X+\dot{Y}Y+\dot{Z}Z}{X^2+Y^2+Z^2} \\ \dot{y}_s = \frac{\dot{Y}}{\sqrt{X^2+Y^2+Z^2}} - \frac{Y}{\sqrt{X^2+Y^2+Z^2}} \frac{\dot{X}X+\dot{Y}Y+\dot{Z}Z}{X^2+Y^2+Z^2} \\ \dot{z}_s = \frac{\dot{Z}}{\sqrt{X^2+Y^2+Z^2}} - \frac{Z}{\sqrt{X^2+Y^2+Z^2}} \frac{\dot{X}X+\dot{Y}Y+\dot{Z}Z}{X^2+Y^2+Z^2} \end{cases} \quad (3.31)$$

En remplaçant $\dot{X}$, $\dot{Y}$ et $\dot{Z}$ par l'équation de la cinématique d'un point 3D, la matrice d'interaction d'un point sphérique peut s'obtenir aisément :

$$\mathbf{L}_s = \begin{bmatrix} -\frac{1}{r} + \frac{x_s^2}{r} & \frac{x_s y_s}{r} & \frac{x_s z_s}{r} & 0 & -z_s & y_s \\ \frac{x_s y_s}{r} & -\frac{1}{r} + \frac{y_s^2}{r} & \frac{y_s z_s}{r} & z_s & 0 & -x_s \\ \frac{x_s z_s}{r} & \frac{y_s z_s}{r} & -\frac{1}{r} + \frac{z_s^2}{r} & -y_s & x_s & 0 \end{bmatrix} \quad (3.32)$$

avec $r = \sqrt{X^2 + Y^2 + Z^2}$. Si on considère un objet plan, la valeur de $\frac{1}{r}$ qui apparait dans la matrice 3.32 peut s'écrire, en utilisant 3.20 et 3.28 correspondantes aux équations du plan et les coordonnées spherique respectivement. Plus précisément on a :

$$\begin{aligned} \frac{1}{r} &= \frac{1}{\sqrt{X^2 + Y^2 + Z^2}} = \frac{1}{Z} \frac{1}{\sqrt{x^2 + y^2 + 1}} \\ &= \frac{Ax}{\sqrt{x^2 + y^2 + 1}} + \frac{By}{\sqrt{x^2 + y^2 + 1}} + \frac{C}{\sqrt{x^2 + y^2 + 1}} \end{aligned} \quad (3.33)$$

Finalement on obtient :

$$\frac{1}{r} = Ax_s + By_s + Cz_s \quad (3.34)$$

3.6.3 Projection d'un élément de surface sur la sphère unitaire

Le but ici est d'exprimer la correspondance entre un pixel du plan image et l'élément de surface associé, soit ds'' en fonction de ds . Pour cela, nous utilisons l'intermédiaire ds' (voir figure 3.4).

Pour les éléments de surface infinitésimaux, $ds'\vec{n}'$ est la projection de $ds\vec{n}$ sur la direction du point de vue (voir figure 3.4). On obtient alors [162] :

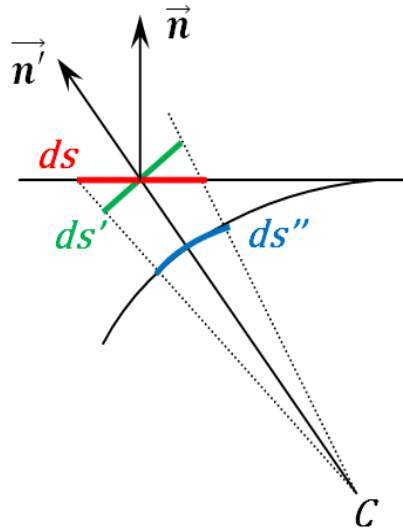


FIGURE 3.4 – Élément de surface en projection spherique.

$$ds' = ds \cdot \begin{pmatrix} 0 & 0 & 1 \end{pmatrix} \cdot \frac{\begin{pmatrix} x \\ y \\ 1 \end{pmatrix}}{\sqrt{x^2 + y^2 + 1}} = \frac{ds}{\sqrt{x^2 + y^2 + 1}} \quad (3.35)$$

D'autre part, toujours sous l'hypothèse d'éléments infinitésimaux :

$$ds'' = \frac{ds'}{x^2 + y^2 + 1}$$

Le changement de variable concernant les éléments de surface est donc donnée par :

$$ds'' = \frac{ds}{(x^2 + y^2 + 1)^{\frac{3}{2}}} \quad (3.36)$$

En combinant 3.36, 3.30 et 3.28, les moments sphériques peuvent être calculés directement à partir de la projection sur le plan image par :

$$m_{spqr} = \iint_{\mathcal{R}} \left(\frac{x}{\sqrt{x^2 + y^2 + 1}} \right)^p \left(\frac{y}{\sqrt{x^2 + y^2 + 1}} \right)^q \left(\frac{1}{\sqrt{x^2 + y^2 + 1}} \right)^r f(x, y) \frac{dx dy}{(x^2 + y^2 + 1)^{\frac{3}{2}}} \quad (3.37)$$

ou encore :

$$m_{spqr} = \iint_{\mathcal{R}} x_s^p y_s^q z_s^{r+3} f(x, y) dx dy \quad (3.38)$$

Dans le cas où on traite un nuage de points $(x_{s_n}, y_{s_n}, z_{s_n})$, les moments sphériques sont donnés par :

$$m_{spqr} = \sum_{n=0}^N x_{s_n}^p y_{s_n}^q z_{s_n}^r \quad (3.39)$$

On peut noter la différence entre la définition des moments dans le cas continu et dans le cas discret : dans le cas continu, le terme z_s^3 apparait pour prendre en compte l'élément de surface, ce qui n'intervient pas dans le cas discret. Comme nous allons le voir à présent, les matrices d'interaction seront donc différentes.

3.6.4 Matrice d'interaction des moments sphériques

3.6.4.1 Cas continu

On considère tout d'abord le cas continu. En dérivant l'équation 3.37 et par analogie avec la méthode utilisée pour l'équation A.21, la variation temporelle $\dot{m}_{pq}$ est définie par :

$$\dot{m}_{pq} = \iint_{\mathcal{R}} \left[\frac{\partial h}{\partial x} \dot{x} + \frac{\partial h}{\partial y} \dot{y} + h(x, y) \left(\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} \right) \right] dx dy \quad (3.40)$$

Avec

$$h(x, y) = \left(\frac{x}{\sqrt{x^2 + y^2 + 1}} \right)^p \left(\frac{y}{\sqrt{x^2 + y^2 + 1}} \right)^q \left(\frac{1}{\sqrt{x^2 + y^2 + 1}} \right)^{r+3} f(x, y)$$

En combinant 3.18 et 3.21 dans 3.40 et afin de la écrire sous cette forme :

$$\mathbf{L}_{m_{spqr}} = \left[\mathbf{L}_{m_{spqr}}^v \mathbf{L}_{m_{spqr}}^w \right]^\top = \left[\mathbf{L}_{m_{spqr}}^{v_x} \mathbf{L}_{m_{spqr}}^{v_y} \mathbf{L}_{m_{spqr}}^{v_z} \mathbf{L}_{m_{spqr}}^{\omega_x} \mathbf{L}_{m_{spqr}}^{\omega_y} \mathbf{L}_{m_{spqr}}^{\omega_z} \right]^\top \quad (3.41)$$

Où :

$$\left\{ \begin{array}{l} \mathbf{L}_{m_{spqr}}^{v_x} = A(\beta m_{sp+2,q,r} - (p+1)m_{sp,q,r}) + B(\beta m_{sp+1,q+1,r} - p m_{sp-1,q+1,r}) \\ \quad + C(\beta m_{sp+1,q,r+1} - p m_{sp-1,q,r+1}) \\ \mathbf{L}_{m_{spqr}}^{v_y} = A(\beta m_{sp+1,q+1,r} - p m_{sp+1,q-1,r}) + B(\beta m_{sp,q+2,r} - (q+1)m_{sp,q,r}) \\ \quad + C(\beta m_{sp,q+1,r+1} - q m_{sp,q-1,r+1}) \\ \mathbf{L}_{m_{spqr}}^{v_z} = A(\beta m_{sp+1,q,r+1} - r m_{sp+1,q,r-1}) + B(\beta m_{sp,q+1,r+1} - r m_{sp,q+1,r-1}) \\ \quad + C(\beta m_{sp,q,r+2} - (r+1)m_{sp,q,r}) \\ \mathbf{L}_{m_{spqr}}^{\omega_x} = q m_{sp,q-1,r+1} - r m_{sp,q+1,r-1} \\ \mathbf{L}_{m_{spqr}}^{\omega_y} = r m_{sp+1,q,r-1} - p m_{sp-1,q,r+1} \\ \mathbf{L}_{m_{spqr}}^{\omega_z} = p m_{sp-1,q+1,r} - q m_{sp+1,q-1,r} \end{array} \right. \quad (3.42)$$

Avec $\beta = p + q + r + 3$.

3.6.4.2 Cas discret

On considère a présent le cas discret. En dérivant l'équation 3.39, on obtient :

$$\dot{m}_{spqr} = \sum_{n=0}^N \left(p x_{s_n}^{p-1} y_{s_n}^q z_{s_n}^r \dot{x}_{s_n} + q x_{s_n}^p y_{s_n}^{q-1} z_{s_n}^r \dot{y}_{s_n} + r x_{s_n}^p y_{s_n}^q z_{s_n}^{r-1} \dot{z}_{s_n} \right) \quad (3.43)$$

En combinant cette équation avec 3.31 et 3.32, on déduit la matrice d'interaction $\mathbf{L}_{m_{sd} pqr}$ comme suit :

$$\mathbf{L}_{m_{sd} pqr} = \left[\mathbf{L}_{m_{sd} pqr}^v \mathbf{L}_{m_{sd} pqr}^w \right]^\top = \left[\mathbf{L}_{m_{sd} pqr}^{v_x} \mathbf{L}_{m_{sd} pqr}^{v_y} \mathbf{L}_{m_{sd} pqr}^{v_z} \mathbf{L}_{m_{sd} pqr}^{\omega_x} \mathbf{L}_{m_{sd} pqr}^{\omega_y} \mathbf{L}_{m_{sd} pqr}^{\omega_z} \right]^\top \quad (3.44)$$

Où :

$$\left\{ \begin{array}{l} \mathbf{L}_{m_{sd} pqr}^{v_x} = A(\beta_d m_{sd p+2,q,r} - p m_{sd p,q,r}) + B(\beta_d m_{sd p+1,q+1,r} - p m_{sd p-1,q+1,r}) \\ \quad + C(\beta_d m_{sd p+1,q,r+1} - p m_{sd p-1,q,r+1}) \\ \mathbf{L}_{m_{sd} pqr}^{v_y} = A(\beta_d m_{sd p+1,q+1,r} - p m_{sd p+1,q-1,r}) + B(\beta_d m_{sd p,q+2,r} - q m_{sd p,q,r}) \\ \quad + C(\beta_d m_{sd p,q+1,r+1} - q m_{sd p,q-1,r+1}) \\ \mathbf{L}_{m_{sd} pqr}^{v_z} = A(\beta_d m_{sd p+1,q,r+1} - r m_{sd p+1,q,r-1}) + B(\beta_d m_{sd p,q+1,r+1} - r m_{sd p,q+1,r-1}) \\ \quad + C(\beta_d m_{sd p,q,r+2} - r m_{sd p,q,r}) \\ \mathbf{L}_{m_{sd} pqr}^{\omega_x} = q m_{sd p,q-1,r+1} - r m_{sd p,q+1,r-1} \\ \mathbf{L}_{m_{sd} pqr}^{\omega_y} = r m_{sd p+1,q,r-1} - p m_{sd p-1,q,r+1} \\ \mathbf{L}_{m_{sd} pqr}^{\omega_z} = p m_{sd p-1,q+1,r} - q m_{sd p+1,q-1,r} \end{array} \right. \quad (3.45)$$

Avec $\beta_d = p + q + r$.

3.7 Estimation de la rotation basée moments photométriques sphériques

Dans le cas où la caméra se déplace avec une vitesse de translation $\mathbf{v} = [v_x \ v_y \ v_z]^\top$ et une vitesse de rotation $\boldsymbol{\omega} = [\omega_x \ \omega_y \ \omega_z]^\top$, par rapport au repère caméra, lorsqu'elle observe un point de la scène $\mathbf{P} = (X, Y, Z)$.

Supposons maintenant que la camera effectue un mouvement de rotation pure $\mathbf{R}$, c-à-d la vitesse de translation est nulle ($\mathbf{v} = 0$), on peut facilement montrer qu'un mouvement de rotation d'un point 3D $\mathbf{P}$ ($\mathbf{P}' = \mathbf{R}\mathbf{P}$, où $\mathbf{P}'$ est le même point 3D après la rotation) induit le même mouvement sur sa projection sur la sphère unité ($\mathbf{P}'_s = \mathbf{R}\mathbf{P}_s$).

Dans l'espace tangent et en reprenant les équations 3.32, l'effet de la vitesse de rotation $\boldsymbol{\omega} = [\omega_x \ \omega_y \ \omega_z]^\top$ sur un point projeté sur la sphère est donné par sa dérivée temporelle $\dot{\mathbf{P}}_s$:

$$\dot{\mathbf{P}}_s = \mathbf{L}_{\mathbf{P}_s} \boldsymbol{\omega} = \begin{bmatrix} 0 & -z_s & y_s \\ z_s & 0 & -x_s \\ -y_s & x_s & 0 \end{bmatrix} \boldsymbol{\omega} \quad (3.46)$$

avec $\mathbf{L}_{\mathbf{P}_s}$ est la matrice d'interaction liée à la vitesse de rotation.

Par ailleurs, en reprenant les équations 3.44 et 3.45, la variation par rapport au temps du moment m_{pqr} est liée à la vitesse de rotation $\boldsymbol{\omega}$ par :

$$\dot{m}_{pqr} = \begin{bmatrix} qm_{p,q-1,r+1} - rm_{p,q+1,r-1} \\ rm_{p+1,q,r-1} - pm_{p-1,q,r+1} \\ pm_{p-1,q+1,r} - qm_{p+1,q-1,r} \end{bmatrix}^\top \boldsymbol{\omega} \quad (3.47)$$

Dans ce qui suit, nous allons montrer comment, à partir de moments, il est possible de définir un triplet $\mathbf{x}_m = [x(m_{pqr}) \ y(m_{pqr}) \ z(m_{pqr})]^\top$ qui se comporte comme un point projeté sur la sphère par rapport à la vitesse de rotation $\boldsymbol{\omega}$. Plus précisément, la matrice d'interaction liée à $\mathbf{x}_m$ a la même forme que (3.46) et donc donnée par :

$$\dot{\mathbf{x}}_m = \begin{bmatrix} 0 & -z(m_{pqr}) & y(m_{pqr}) \\ z(m_{pqr}) & 0 & -x(m_{pqr}) \\ -y(m_{pqr}) & x(m_{pqr}) & 0 \end{bmatrix} \boldsymbol{\omega} \quad (3.48)$$

Dans ce cas, nous avons $\mathbf{x}'_m = \mathbf{R} \mathbf{x}_m$ après un mouvement de rotation.

A partir du moment d'ordre 1, il est facile de montrer que la matrice d'interaction liée au triplet $\mathbf{x}_{m_1} = [m_{100} \ m_{010} \ m_{001}]^\top$ vérifie la contrainte (3.48). En effet, en utilisant (3.47), la dérivée par rapport au temps de $\mathbf{x}_{m_1}$ est liée au $\boldsymbol{\omega}$ par :

$$\dot{\mathbf{x}}_{m_1} = \begin{bmatrix} 0 & -m_{001} & m_{010} \\ m_{001} & 0 & -m_{100} \\ -m_{010} & m_{100} & 0 \end{bmatrix} \boldsymbol{\omega} \quad (3.49)$$

Comme on peut le voir, la matrice d'interaction donnée par (3.49) a une forme identique à celle obtenue pour un point projeté sur la sphère. Ceci est attendu, car ce triplet divisé par le moment m_{000} n'est rien d'autre que le centre de gravité. Notons qu'une rotation 3D entraîne le même mouvement sur le centre de gravité de la projection sphérique.

En effet, une forme générale pour le triplet assurant le même comportement peut être obtenue à partir des moments d'ordre supérieure à 1. Pour cela, considérons les

vecteurs $\mathbf{v}_n$ de moments d'ordre $p + q + r = n$, comme exemple, les vecteurs suivants $\mathbf{v}_3 = [\mathbf{m}_{300} \ \mathbf{m}_{210} \ \mathbf{m}_{201} \ \cdots \ \mathbf{m}_{003}]^\top$ ou $\mathbf{v}_2 = [\mathbf{m}_{200} \ \mathbf{m}_{110} \ \cdots \ \mathbf{m}_{002}]^\top$.

Selon l'équation 3.47, la matrice d'interaction du moment m_{ijk} est une combinaison linéaire des moments de même ordre. Par conséquent, les entrées de la matrice d'interaction $\mathbf{L}_{\mathbf{v}_n}$ qui associent $\dot{\mathbf{v}}_n$ à $\boldsymbol{\omega}$ sont définies comme des combinaisons linéaires de $\mathbf{v}_n$ entrées. Considérons maintenant les vecteurs instantanés $\mathbf{w}$ avec les entrées définies par tous les produits possibles entre les entrées de deux vecteurs $\mathbf{v}_n$.

Comme exemples des vecteurs possibles : $\mathbf{w} = [\mathbf{m}_{300}\mathbf{m}_{200} \ \mathbf{m}_{300}\mathbf{m}_{110} \ \cdots \ \mathbf{m}_{003}\mathbf{m}_{002}]^\top$
ou $\mathbf{w} = [\mathbf{m}_{300}^2 \ \mathbf{m}_{300}\mathbf{m}_{210} \ \cdots \ \mathbf{m}_{003}^2]^\top$.

En ce qui concerne le vecteur de type $\mathbf{v}_n$, on peut montrer que la dérivée temporelle des vecteurs $\mathbf{w}$ peut être exprimée comme combinaisons linéaires de ses entrées :

$$\dot{\mathbf{w}} = \mathbf{L}_{\mathbf{w}} \boldsymbol{\omega} = \begin{bmatrix} l_{1\omega_x} & l_{1\omega_y} & l_{1\omega_z} \\ l_{2\omega_x} & l_{2\omega_y} & l_{2\omega_z} \\ \vdots & \vdots & \vdots \end{bmatrix} \boldsymbol{\omega} \quad (3.50)$$

où $l_{i\omega_x}, l_{i\omega_y}, l_{i\omega_z}$ sont définis comme des combinaisons linéaires des entrées du vecteur $\mathbf{w}$. Le nombre de lignes de $\mathbf{L}_{\mathbf{w}}$ est égal à la taille du vecteur de moment considéré $\mathbf{w}$.

Définissons un triplet comme fonction linéaire des entrées vectorielles du moment $\mathbf{w}$ par :

$$\mathbf{x}_{\mathbf{w}} = [\boldsymbol{\alpha}_x^\top \mathbf{w} \quad \boldsymbol{\alpha}_y^\top \mathbf{w} \quad \boldsymbol{\alpha}_z^\top \mathbf{w}]^\top \quad (3.51)$$

où $\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y$ et $\boldsymbol{\alpha}_z$ sont des vecteurs scalaires à déterminer afin que $\mathbf{x}_{\mathbf{w}}$ se comporte comme un point 3D par rapport aux rotations. Par conséquent, en fonction de (3.51), la variation de temps de $\mathbf{x}_{\mathbf{w}}$ peut être obtenue à partir de $\mathbf{L}_{\mathbf{w}}$ par :

$$\dot{\mathbf{x}}_{\mathbf{w}} = \mathbf{L}_{\mathbf{xw}} \boldsymbol{\omega} = \begin{bmatrix} \boldsymbol{\alpha}_x^\top \mathbf{L}_{\mathbf{w}} \\ \boldsymbol{\alpha}_y^\top \mathbf{L}_{\mathbf{w}} \\ \boldsymbol{\alpha}_z^\top \mathbf{L}_{\mathbf{w}} \end{bmatrix} \boldsymbol{\omega} \quad (3.52)$$

Rappelons que les entrées de la matrice $\mathbf{L}_{\mathbf{w}}$ sont des combinaisons linéaires du composant du vecteur moment $\mathbf{w}$. Ensuite, en fonction de l'équation (3.52), les entrées de $\mathbf{L}_{\mathbf{xw}}$ sont des combinaisons linéaires du vecteur moment $\mathbf{w}$ et aussi du vecteur scalaire $\boldsymbol{\alpha}_x, \boldsymbol{\alpha}_y$

et α_z . Enfin, pour que $\mathbf{x}_w$ se comporte de manière identique à un point 3D dans l'espace tangent par rotation, les vecteurs scalaires α_x , α_y et α_z sont choisis de telle sorte que (3.52) puisse être écrit comme suit :

$$\dot{\mathbf{x}}_w = \begin{bmatrix} 0 & -\alpha_z^\top \mathbf{w} & \alpha_y^\top \mathbf{w} \\ \alpha_z^\top \mathbf{w} & 0 & -\alpha_x^\top \mathbf{w} \\ -\alpha_y^\top \mathbf{w} & \alpha_x^\top \mathbf{w} & 0 \end{bmatrix} \omega \quad (3.53)$$

Ces résultats dans les conditions linéaires sur les inconnues α_x , α_y et α_z qui peuvent être écrits tous ensemble sous la forme suivante :

$$\mathbf{B}(\mathbf{w}) \begin{bmatrix} \alpha_x \\ \alpha_y \\ \alpha_z \end{bmatrix} = \mathbf{0} \quad (3.54)$$

où $\mathbf{B}(\mathbf{w})$ est une fonction du vecteur moment $\mathbf{w}$. Ce système d'équations peut être résolu en utilisant un logiciel de calcul symbolique (par exemple Matlab). Pour le triplet obtenu, la contrainte dans le plan tangent donné par (3.53) est transformé en $\mathbf{x}'_w = \mathbf{R}\mathbf{x}_w$ pour les mouvements de rotation. Résoudre (3.54) pour le cas de

$\mathbf{w} = [\mathbf{m}_{300}\mathbf{m}_{200} \quad \mathbf{m}_{300}\mathbf{m}_{110} \quad \cdots \quad \mathbf{m}_{003}\mathbf{m}_{002}]^\top$ conduit à trois solutions pour le triplet $\mathbf{x}_w$ donné par (3.55), (3.56) et (3.57) :

$$\mathbf{P}_1 = \begin{bmatrix} m_{003}m_{101} + m_{012}m_{110} + m_{021}m_{101} + m_{030}m_{110} + m_{101}m_{201} + m_{102}m_{200} + m_{110}m_{210} + m_{120}m_{200} + m_{200}m_{300} \\ m_{003}m_{011} + m_{011}m_{021} + m_{012}m_{020} + m_{020}m_{030} + m_{011}m_{201} + m_{102}m_{110} + m_{020}m_{210} + m_{110}m_{120} + m_{110}m_{300} \\ m_{002}m_{003} + m_{002}m_{021} + m_{011}m_{012} + m_{011}m_{030} + m_{002}m_{201} + m_{101}m_{102} + m_{011}m_{210} + m_{101}m_{120} + m_{101}m_{300} \end{bmatrix} \quad (3.55)$$

$$\mathbf{P}_2 = \begin{bmatrix} m_{002}m_{120} - 2m_{011}m_{111} + m_{020}m_{102} + m_{002}m_{300} - 2m_{101}m_{201} + m_{102}m_{200} + m_{020}m_{300} - 2m_{110}m_{210} + m_{120}m_{200} \\ m_{002}m_{030} - 2m_{011}m_{021} + m_{012}m_{020} + m_{002}m_{210} + m_{012}m_{200} - 2m_{101}m_{111} + m_{020}m_{210} + m_{030}m_{200} - 2m_{110}m_{120} \\ m_{002}m_{021} + m_{003}m_{020} - 2m_{011}m_{012} + m_{002}m_{201} + m_{003}m_{200} - 2m_{101}m_{102} + m_{020}m_{201} + m_{021}m_{200} - 2m_{110}m_{111} \end{bmatrix} \quad (3.56)$$

$$\mathbf{P}_3 = \begin{bmatrix} m_{002}m_{102} + 2m_{011}m_{111} + m_{020}m_{120} + 2m_{101}m_{201} + 2m_{110}m_{210} + m_{200}m_{300} \\ m_{002}m_{012} + 2m_{011}m_{021} + m_{020}m_{030} + 2m_{101}m_{111} + 2m_{110}m_{120} + m_{200}m_{210} \\ m_{002}m_{003} + 2m_{011}m_{012} + m_{020}m_{021} + 2m_{101}m_{102} + 2m_{110}m_{111} + m_{200}m_{201} \end{bmatrix} \quad (3.57)$$

Dans la pratique, deux triplets différents $\mathbf{x}_w$ sont suffisants pour une solution analytique du mouvement de rotation car ils peuvent définir une base orthogonale. Considérons, par exemple, les deux triplets $\mathbf{P}_1$ et $\mathbf{P}_2$ calculés à partir des moments d'une première

image en utilisant (3.55) et (3.56) et $\mathbf{P}'_1$ et $\mathbf{P}'_2$ l'image correspondante calculée pour l'image ayant subi une rotation. Par conséquent, nous avons $\mathbf{P}'_1 = \mathbf{R}\mathbf{P}_1$ et $\mathbf{P}'_2 = \mathbf{R}\mathbf{P}_2$, où $\mathbf{R}$ est la matrice de rotation liant les deux images I_1 et I_2 .

Définissons la base suivante :

$$\begin{aligned}\mathbf{v}_1 &= \frac{\frac{\mathbf{P}_1}{\|\mathbf{P}_1\|} + \frac{\mathbf{P}_2}{\|\mathbf{P}_2\|}}{\left\| \frac{\mathbf{P}_1}{\|\mathbf{P}_1\|} + \frac{\mathbf{P}_2}{\|\mathbf{P}_2\|} \right\|} \\ \mathbf{v}_2 &= \frac{\frac{\mathbf{P}_1}{\|\mathbf{P}_1\|} - \frac{\mathbf{P}_2}{\|\mathbf{P}_2\|}}{\left\| \frac{\mathbf{P}_1}{\|\mathbf{P}_1\|} - \frac{\mathbf{P}_2}{\|\mathbf{P}_2\|} \right\|} \\ \mathbf{v}_3 &= \mathbf{v}_1 \times \mathbf{v}_2\end{aligned}\tag{3.58}$$

Tout d'abord, on peut facilement montrer que les vecteurs $\mathbf{v}_1$, $\mathbf{v}_2$ et $\mathbf{v}_3$ forment une base orthogonale, c'est-à-dire :

$\|\mathbf{v}_1\| = \|\mathbf{v}_2\| = \|\mathbf{v}_3\| = 1$ et $\mathbf{v}_1 \cdot \mathbf{v}_2 = \mathbf{v}_1 \cdot \mathbf{v}_3 = \mathbf{v}_3 \cdot \mathbf{v}_2 = 0$. Ensuite, considérons maintenant une deuxième base $\mathbf{v}'_1$, $\mathbf{v}'_2$ et $\mathbf{v}'_3$ calculée à partir $\mathbf{P}'_1$ et $\mathbf{P}'_2$ en utilisant (3.58). Comme $\mathbf{P}'_1 = \mathbf{R}\mathbf{P}_1$ et $\mathbf{P}'_2 = \mathbf{R}\mathbf{P}_2$, il peut être prouvé en utilisant (3.58), que $\mathbf{v}'_1 = \mathbf{R}\mathbf{v}_1$, $\mathbf{v}'_2 = \mathbf{R}\mathbf{v}_2$ et $\mathbf{v}'_3 = \mathbf{R}\mathbf{v}_3$. Les dernières équations peuvent être écrites sous la forme suivante :

$$\mathbf{V} = \mathbf{R}\mathbf{V}'\tag{3.59}$$

avec $\mathbf{V} = [\mathbf{v}_1 \ \mathbf{v}_2 \ \mathbf{v}_3]$ et $\mathbf{V}' = [\mathbf{v}'_1 \ \mathbf{v}'_2 \ \mathbf{v}'_3]$. Enfin, la matrice de rotation peut être estimée directement par :

$$\mathbf{R} = \mathbf{V}\mathbf{V}'^T\tag{3.60}$$

Trois schémas de calcul des moments sphériques sont considérés afin de valider la solution proposée pour estimer le mouvement de rotation.

3.7.1 Première méthode (M1)

Appelée dans la suite M1, elle consiste à utiliser tout le signal image sphérique pour le calcul des moments photométriques pour estimer le mouvement de rotation.

3.7.2 Deuxième méthode (M2) :

Dans cette méthode, nous utilisons une fonction de pondération gaussienne appliquée à l'image sphérique. Plus précisément, une première estimation de la rotation, en utilisant la première méthode M1, est utilisée pour maximiser la région de chevauchement des deux images sphériques liées par la rotation. Ensuite, une fonction de pondération gaussienne est utilisée pour attribuer un poids significatif aux régions les plus visibles dans l'image de référence, cette fonction est définie par

La figure 3.5 présente un exemple d'un filtre gaussien sphérique, qu'est défini par [163] :

$$G_s(\theta, \varphi, t) = e^{-(\tan^2(\theta/2))/t}$$

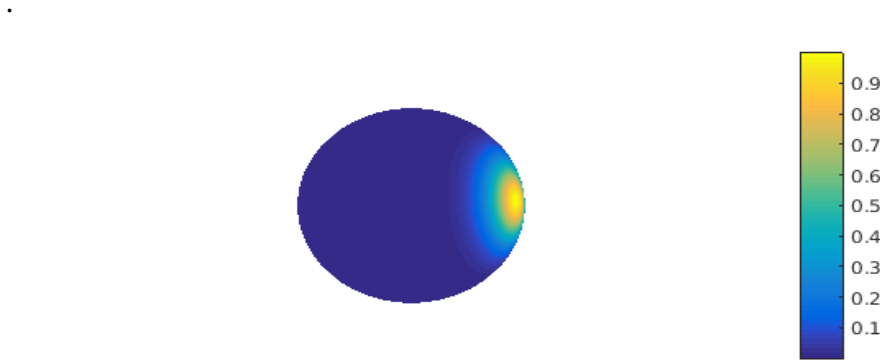


FIGURE 3.5 – Exemple d'un filtre gaussien sphérique.

3.7.3 Troisième méthode (M3) :

Dans cette méthode les moments sphériques photométriques sont calculés en utilisant le voisinage d'un point d'intérêt détecté et mis en correspondance préalablement. En effet, un seul point apparié est suffisant pour estimer la rotation. La figure 3.6 montre la meilleure correspondance des points caractéristiques détectés dans deux images liées par une rotation. Le voisinage sphérique exploité pour le calcul des moments sphériques photométriques est représenté dans la figure 3.7.

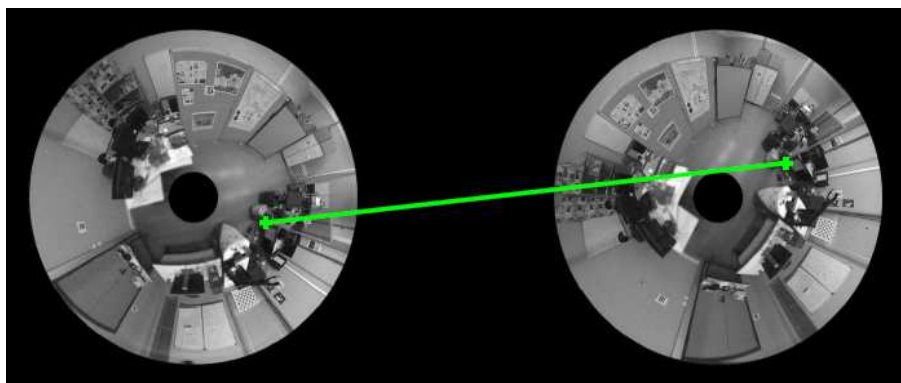


FIGURE 3.6 – La mise en correspondance d’un point d’intérêt pour la méthode M3.



FIGURE 3.7 – Voisinages sphériques des points d’intérêts utilisés pour la méthode M3.

3.8 Expérimentations

Cette section est consacrée à la présentation de différents résultats expérimentaux qualitatifs et quantitatifs, pour évaluer les approches proposées dans ce chapitre. Ils comprennent à la fois une partie des tests effectués sur des données synthétiques et une autre partie effectuée sur des données réelles.

3.8.1 Données synthétiques

Dans un premier temps, nous avons testé les techniques proposées dans ce chapitre, sur une séquence synthétique obtenue en appliquant des mouvements de rotation 3D pures connus, à une image sphérique.

Dans la suite, nous utilisons les angles d’Euler (α, β, γ), correspondant aux angles de rotation autour des axes (Z, Y, X), respectivement pour représenter la rotation 3D.

Concrètement, un ensemble de 13 images sphériques est généré en faisant tourner l’image sphérique de référence (issue d’une image omnidirectionnelle réelle) avec différentes valeurs des angles d’Euler. En effet, l’image sphérique de référence subit à la fois,

une rotation autour des trois axes, sur des valeurs spécifiques dans les intervalles échantillonnés suivants : de 0° à 359° autour de l'axe Z, de 0° à 45° autour de l'axe Y et de -22.5° à 22.5° pour l'axe X. On peut se référer à l'axe des abscisses de la Figure 3.8 pour apercevoir les différentes valeurs des angles d'Euler utilisées afin de générer la séquence de 13 images sphériques synthétiques.

La figure 3.8 montre les résultats obtenus de l'estimation de la rotation 3D en utilisant les trois approches proposées dans ce chapitre. Pour chaque image synthétique générée, nous avons évalué l'estimation entre ces images et l'image de référence. En effet, les données utilisées pour générer les images synthétiques, ont été utilisées comme étant la vérité terrain et les valeurs des estimations des trois angles d'Euler sont donnés sur la Figure 3.8.

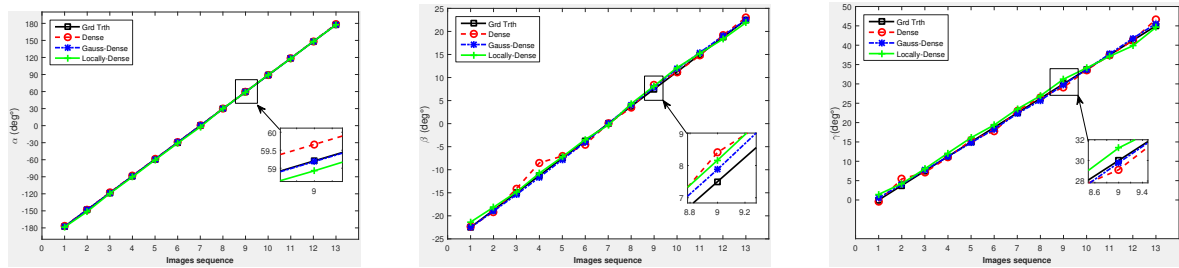


FIGURE 3.8 – Estimation des rotations α autour de l'axe (Z), β autour de l'axe (Y) et γ autour de l'axe (X).

Nous remarquons que l'estimation par les trois approches proposées est très proche à la vérité terrain. En effet, ces approches sont destinées à estimer avec précision, les rotations pures entre les images sphériques. Cela est due au fait que, les méthodes M1, M2 et M3, se basent principalement sur la contrainte d'orthogonalité de la matrice de rotation $\mathbf{R}$ entre les informations visuelles calculées à partir des moments photométriques des images sphériques.

Dans cette partie, nous procédons à l'analyse des trois méthodes basées sur l'approche proposée dans ce chapitre, en présence d'images dégradées par différents types de perturbations. De la même façon qu'au chapitre précédent, sur une base de données des images synthétiques, où les images sont séparées par des rotations pures, nous appliquons trois différentes perturbations, dont le bruit gaussien, changement de luminosité et la compression JPEG. Ainsi, en augmentant l'amplitude de la perturbation, l'analyse a été effectuée pour chaque méthode séparément.

3.8.1.1 Analyse de l'effet du bruit gaussien :

Quatre niveaux de variance du bruit ($\sigma^2=5, 15, 25$ et 50) sont considérés pour évaluer les performances des 3 méthodes. La figure 3.9 illustre l'erreur angulaire absolue de l'esti-

mation de la rotation suivant les trois axes X, Y et Z de chaque méthode. En appliquant les moments photométriques sphériques seulement sur le patch de description locale obtenu de voisinage des points d'intérêts, a permis de réduire l'erreur de l'estimation, par rapport aux deux autres méthodes M1 et M2, qui se basent sur la description globale. L'utilisation de noyau gaussien a apporté des améliorations par rapport à la méthode M1, du fait qu'il se focalise sur les régions des images qui sont liées par la rotation. D'autre part, une variance élevée du bruit ($\sigma^2 = 50$) influence sur l'estimation, car elle intervient directement au niveau photométrique de l'image.

3.8.1.2 Analyse de l'effet du changement de luminosité :

Nous proposons également d'analyser l'influence de perturbations de type changement de luminosité globale des images, sur les performances de notre approche. Pour ce fait, nous avons varié l'intensité des images de la base de données synthétiques, par les valeurs suivantes : +40%, +20%, -20%, -40%. L'estimation de la rotation est effectuée pour chaque image par rapport à l'image de référence. La figure 3.10, présente les différentes erreurs d'estimation de la rotation de chaque méthode et pour chaque changement du niveau de la luminosité. La méthode M3 présente une précision plus élevée notamment pour des forts changements de la luminance d'images (+40% et -40%). En effet, la description locale via les moments sphériques d'une part et la robustesse de détecteur SIFT d'autre part, permet donc d'obtenir une meilleure description locale du voisinage du point pour ce type de transformation. Par conséquent, la vraisemblance entre les zones sélectionnées pour l'estimation de la rotation, joue un rôle important et influence significativement sur la performance de l'estimation.

3.8.1.3 Analyse de l'effet de la compression JPEG :

Nous évaluons maintenant la précision des trois méthodes en présence de la compression JPEG. Nous utilisons la méthodologie présentée dans les analyses précédentes (le cas du bruit et changement de luminosité), afin d'évaluer nos 3 méthodes. En effet, la même base de données synthétique, dont les images sont générées par des rotations pures appliquées sur une image réelle de référence, a été reprise. Ensuite, nous appliquons la compression JPEG sur la base de données en augmentant le facteur de qualité de la compression (avec QF = 10%, 30%, 60% et 90%). Le taux de compression le plus élevées engendre plus de pertes au niveau de la qualité d'images. Une fois cette phase est réalisée, les 3 méthodes sont utilisées afin d'estimer les rotations entre l'image de référence et les images de la base. La figure 3.11 présente les performances des différentes méthodes en termes de l'Erreur Angulaire Absolue de l'estimation.

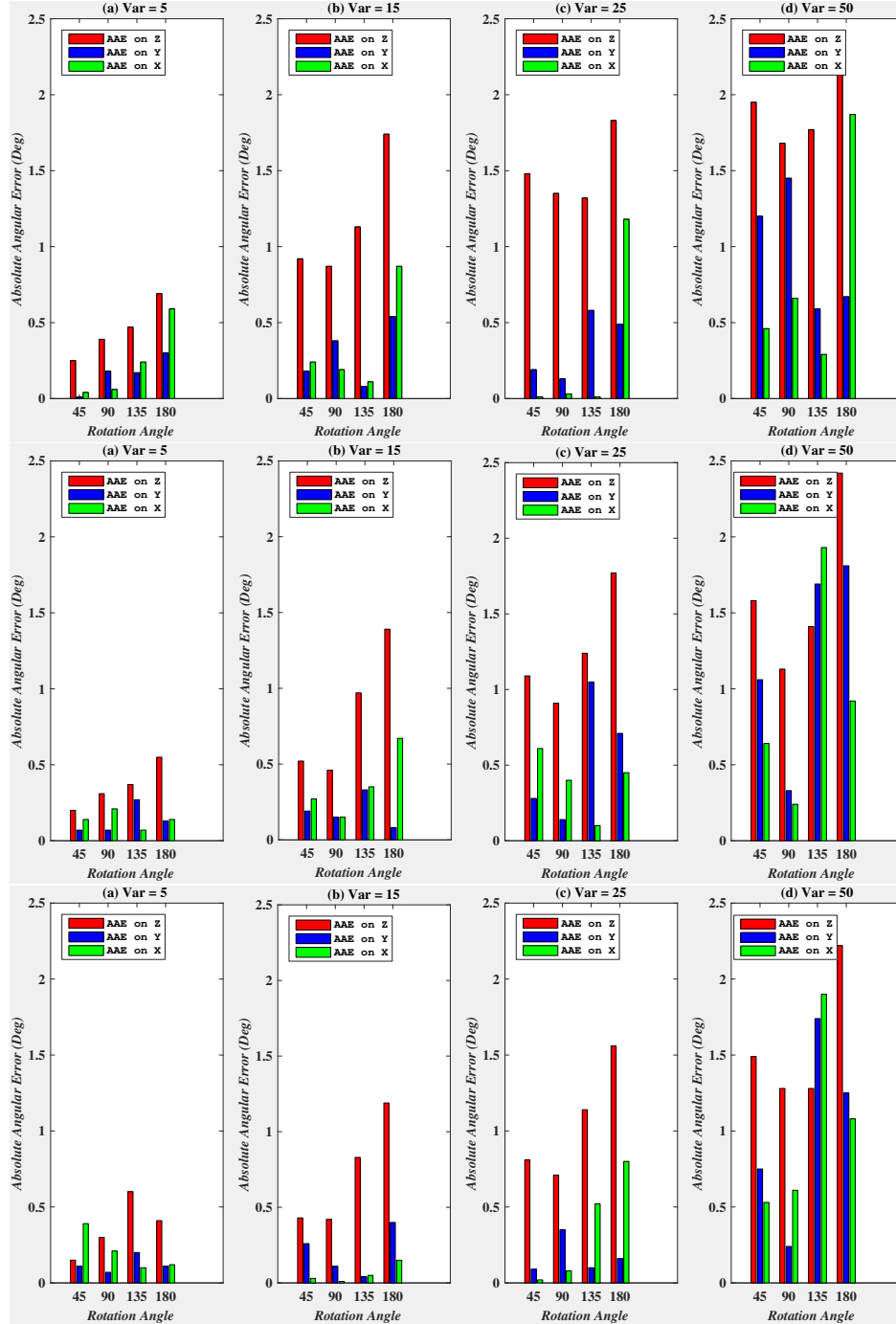


FIGURE 3.9 – Performance en terme de l'erreur angulaire absolue de l'estimation de la rotation autour les axes Z, Y et X avec différents niveaux du bruit gaussien ajouté : (a) $\sigma^2 = 5$, (b) $\sigma^2 = 15$, (c) $\sigma^2 = 25$ and (d) $\sigma^2 = 50$. Première, deuxième et troisième lignes correspond aux méthodes M1, M2 et M3 respectivement

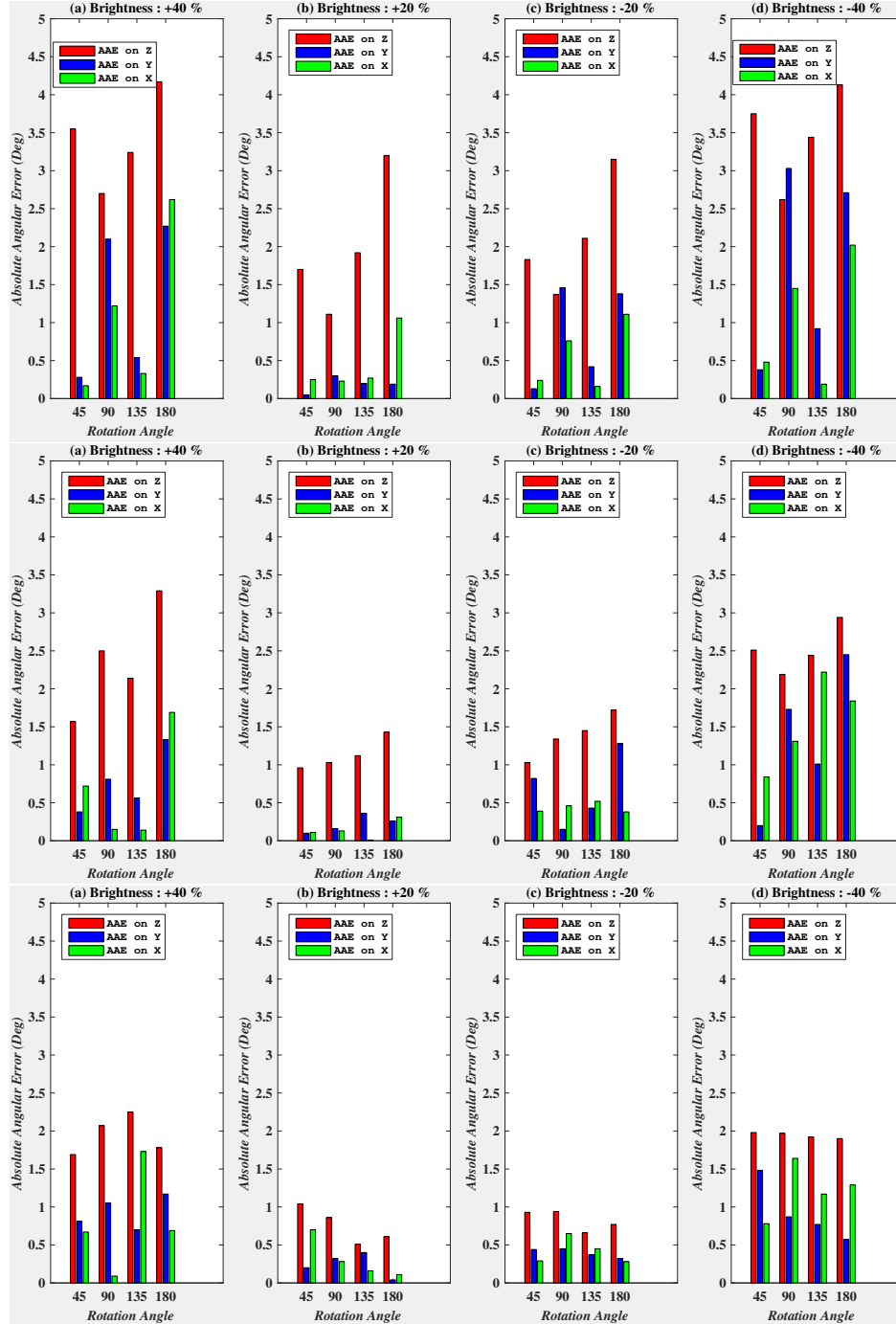


FIGURE 3.10 – Performance en terme de l'erreur angulaire absolue de l'estimation de la rotation autour les axes Z, Y et X avec différents niveaux du changement de luminosité : (a) +40%, (b) +20%, (c) -20% and (d) -40%. Première, deuxième et troisième ligne correspond aux méthodes M1, M2 et M3 respectivement

Les résultats obtenus dans cette partie confirment une fois encore que la méthode M3 a permis d'avoir des estimations plus précises par rapport aux méthodes plutôt globales. Ceci se justifie par le fait que les moments photométriques sont des caractéristiques globales calculées sur l'ensemble de l'image, où tous les pixels contribuent dans l'estimation. Par conséquent, plus la ressemblance entre les zones dans lesquelles on calcule les moments est importante, plus l'estimation de la rotation sera précise.

3.8.2 Données réelles

Dans l'objectif d'évaluer les performances des approches présentées dans ce chapitre, nous avons procédé à des expérimentations et évaluations exhaustives sur les scénarios issues de la base de donnée OVMIS (Omnidirectional Vision, MIS laboratory) [164]. Dans un premier temps, nous présentons cette base de données. Ensuite, nous analysons les résultats obtenus sur les différents jeux de données.

La banque d'images OVMIS est une base de données publique des images omnidirectionnelles. Ce jeu de données a été proposé et précisément décrit dans [164]. Nous n'en rappelons ici que les caractéristiques principales. En effet, la camera catadioptrique utilisée, a été déportée sur deux types de plateformes différentes, celles-ci sont un bras manipulateur industriel et un robot mobile.

1. Le manipulateur *Stäubli* :

Étant donné que les manipulateurs industriels garantissent une haute précision de positionnement, ils sont bien adaptés pour estimer avec une bonne précision le mouvement d'une caméra portée par le robot et donc pour obtenir une vérité terrain précise. Pour qu'elle soit pertinente, la base de données OVMIS a été fournie avec la vérité terrain illustrant les quantités de mouvement entre les images de la base. Pour cela, le système catadioptrique a été monté sur l'effecteur final d'un robot Staubli TX60 (voir figure 3.12) situé dans une pièce de $10.05m \times 7.03m \times 2.70m$.

Pour tenir compte des contraintes géométriques (extension maximale du bras, $670mm$) et des singularités cinématiques du robot, les poses de la caméra ont été limitées dans un disque horizontal de rayon $450mm$ autour du premier axe (vertical) du manipulateur.

2. Le robot Pioneer :

De même, le système catadioptrique a également été monté sur un robot Adept Mobile Robots Pioneer 3-AT. Pour avoir un point de vue élevé, il a été placé sur un sup-

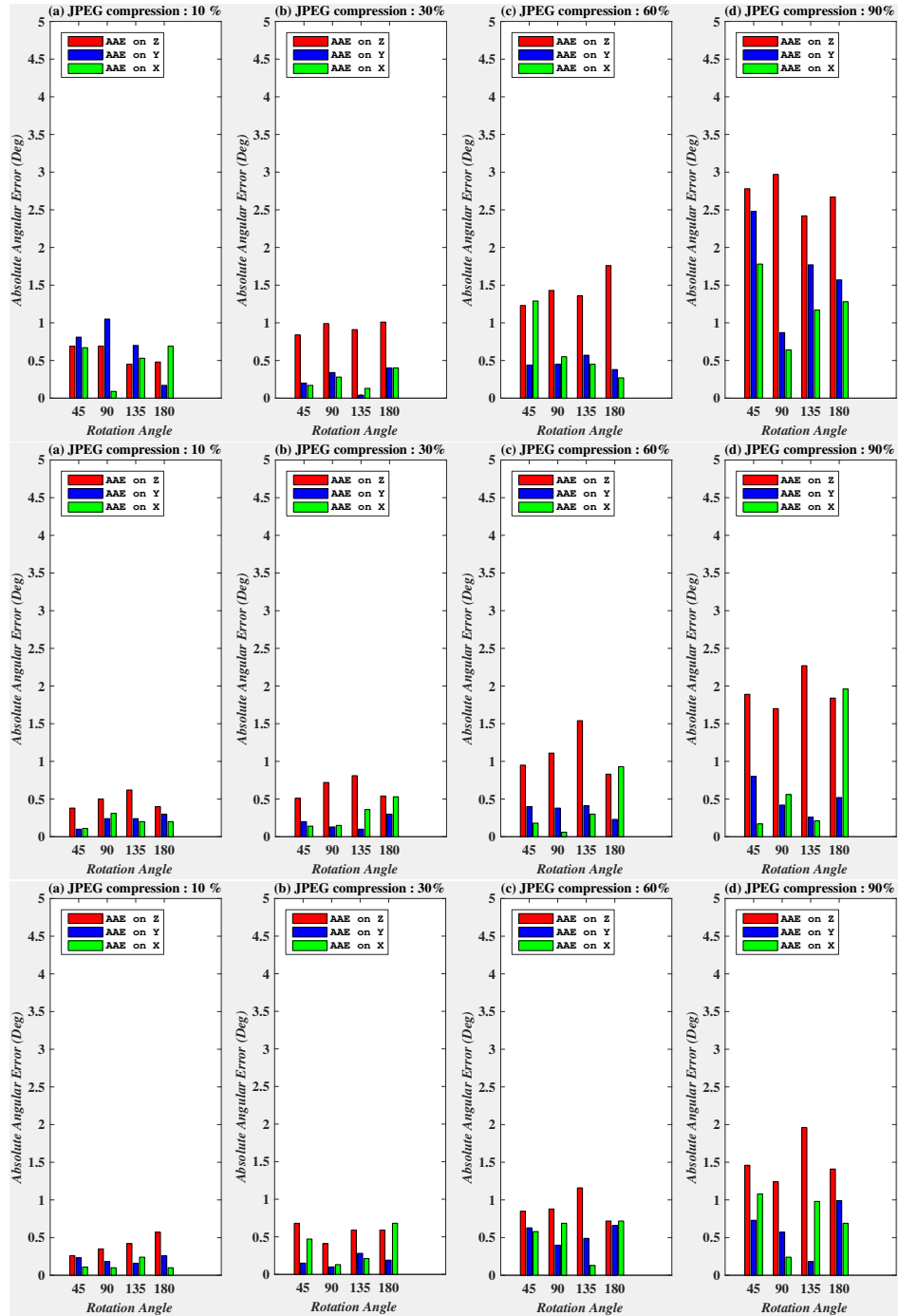


FIGURE 3.11 – Performance en terme de l'erreur angulaire absolue de l'estimation de la rotation autour les axes Z, Y et X avec différents niveaux du Facteur de Qualité de la compression JPEG : (a) QF = 10, (b) QF = 30, (c) QF = 60 and (d) QF = 90. Première, deuxième et troisième lignes correspond aux méthodes M1, M2 et M3 respectivement

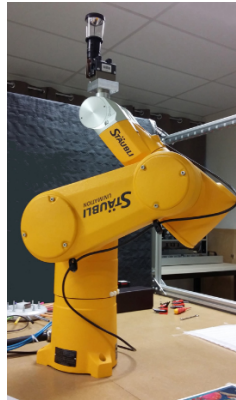


FIGURE 3.12 – Manipulateur *Staubli* TX60.

port vertical, à 134 cm au-dessus du sol (110 cm au-dessus du robot, voir la figure 3.13). L'axe optique de la camera catadioptrique a été aligné avec l'axe verticale de la plateforme mobile. En tournant le robot de 360° autour de son axe vertical, a permis d'avoir des prises d'images omnidirectionnelles, pour des rotations qui allant de 0° à 360°. A l'aide du capteur gyroscope à axe unique qu'en dispose le robot Pioneer, il est capable de détecter avec précision les rotations du robot. Par la synchronisation du système d'acquisition d'images et le module qui enregistre les mouvement du robot, la vérité terrain des déplacements entre les images successives, a été obtenue.



FIGURE 3.13 – Le robot Pioneer 3-AT équipé d'une camera catadioptrique

Dans l'ensemble, quatre scénarios ont été considérés par la base de données OVMIS, dont trois scénarios ont été réalisés par le Pioneer et un autre par le Staubli. C'est quatre scénarios sont proposés pour démontrer l'efficacité et la précision des approches d'estimation de la rotation introduits dans ce chapitre et le prochain.

- Scénario 1 :

Dans ce scénario, nous évaluons la capacité des méthodes proposées pour estimer la rotation entre l'image de référence (la position à 0°) et les autres images de ce scénario. Celui-ci est composé de l'ensemble des images acquises, en appliquant une vitesse angulaire constante au robot, afin d'avoir des pas de rotation régulières (environ 2°). Les images du scénario 1 ont été enregistrées dans un milieu intérieur, dans une salle de laboratoire MIS. Un exemple d'une image de ce scénario est représenté sur la figure 3.14.

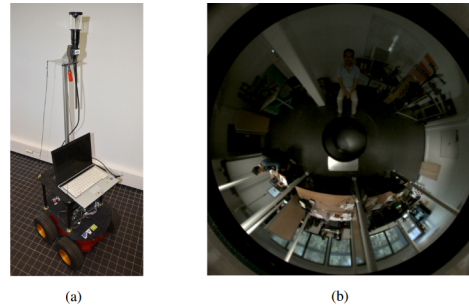


FIGURE 3.14 – Une image catadioptrique du scénario 2 qu'est obtenue en milieu intérieur.

Étant donnée la vérité terrain est fournie, nous avons sélectionné quelques images à des positions où les rotations par rapport à l'image de référence sont bien connues. En effet, nous avons testé nos méthodes sur six différentes rotations autour de l'axe de rotation du robot mobile, en allant de petites rotations à d'autres plus considérables. Les résultats de l'estimation ont été représentés sur la figure 3.15, sous format de l'erreur absolue de l'estimation de la rotation 3D suivant les trois axes X, Y et Z. Or la rotation effectuée autour de l'axe Z, les 2 autres rotations autour de X et Y, sont censées d'être nulles. Nous comparons également les résultats obtenus à celui de l'approche basée "5 primitives + RANSAC". Le choix de cette approche a été motivé principalement par deux raisons : elle a été utilisée avec succès dans de nombreuses applications de l'estimation du mouvement et la disponibilité de la bibliothèque sur matlab (version 2016). L'approche est proposée dans [40] qui combine trois étapes principales : la détection et l'appariement des points d'intérêts, le rejet d'outliers (les mauvaises correspondances) par l'algorithme itérative RANSAC et l'estimation de la transformation géométrique entre les deux images, en utilisant seulement les inliers (les bons appariements).

Les courbes dans la figure 3.15 présentent les erreurs absolues de l'estimation angulaire suivant les trois axes X, Y et Z, respectivement, en fonction de différentes valeurs de la rotation de la vérité terrain. Chaque courbe présente l'erreur de l'estimation de chaque méthode. Nous remarquons que la méthode M3 (approche localement dense) est plus précise par rapport aux autres méthodes proposées, ce qui valide les résultats obtenus par les images synthétiques. Elle est aussi plus précise que l'approche primitives + RANSAC. Ce résultat souligne l'efficacité de la méthode

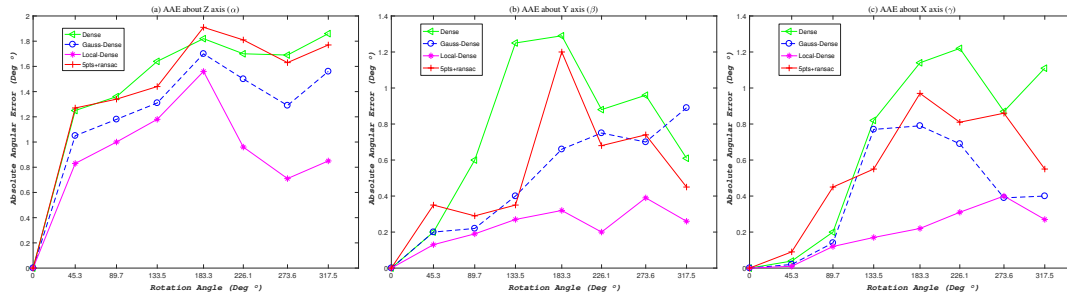


FIGURE 3.15 – Erreur angulaire absolue de l'estimation de la rotation représentée par les angles d'Euler autour des axes : Z, Y et X et comparaison avec la méthode "5 primitives + RANSAC"

M3 pour l'estimation de la rotation dans le cas des images réelles. En effet, cette estimation basée sur le calcul dense des moments sphérique au voisinage d'un seul appariement entre les deux images, est donc plus efficace que d'estimer la rotation en utilisant les positions des primitives obtenues depuis plusieurs appariements.

- Scénario 2 :

Dans ce scénario, la base de données est constituée d'une séquence d'acquisition d'images séparées par des rotations pures, pareille comme le premier scénario, mais cette fois ci dans un milieu extérieur, (un parc public). En gardant la même configuration du robot mobile ainsi que le pas de rotation celui appliqué dans le premier scénario, les images de ce deuxième scénario, ont été obtenues. La figure 3.16 illustre une image de ce scénario.

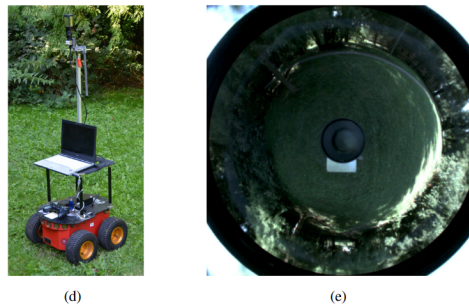


FIGURE 3.16 – Une image catadioptrique du scénario 2 qu'est obtenue en milieu extérieur.

En utilisant ce scénario, nous avons proposé d'évaluer notre approche à travers des 3 méthodes M1, M2, et M3 dans le contexte de la stabilisation vidéo. L'objectif central de stabilisation vidéo est de maintenir la vidéo ou la séquence d'images dans une certaine orientation, en supprimant les fluctuations de mouvement indésirables des données vidéo. En effet, ceci est réalisé en appliquant une certaine quantité de mouvement correctif sur chaque image de la vidéo.

En appliquant notre approche, nous pouvons effectuer la rectification de l'orientation directement à partir des rotations estimées. En effet, la procédure pour rectifier

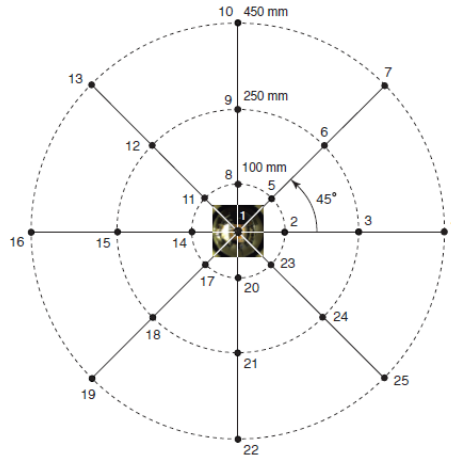


FIGURE 3.17 – Illustration graphique de la structure de l'ensemble de données (les 25 points de collections) généré avec le manipulateur Staubli.

l'orientation d'une image, sous teste, par rapport à l'image de référence, se compose de trois étapes principales. Tout d'abord, nous choisissons une direction sur l'image de référence qu'on doit la représenter par un vecteur, et nous estimons la rotation associée aux deux images par nos 3 méthodes. Deuxièmement, nous appliquons l'inverse de la rotation estimée sur l'image sous test. La troisième et dernière étape consiste à colorer les pixels correspondants au vecteur de direction de l'image sous test. Une bonne estimation de la rotation va mener à une superposition parfaite des deux images.

La figure 3.18 illustre les résultats obtenus de stabilisation vidéo par nos approches sur des images omnidirectionnelles du scénario 2. Nous avons sélectionné trois différentes rotations par rapport à l'image de référence, qui sont (17.8° , 185.3° et 266.3°). Pour chacune de ces rotations, nous appliquons nos trois méthodes, afin d'estimer le vecteur de direction choisi, et tracé en couleur rouge.

Comme expliqué précédemment, nous rectifions l'orientation de chaque image en appliquons l'inverse de la rotation estimée à chaque image sous test. Le résultat de la rectification de chaque approche a été représentée par des couleurs différentes pour la distinction (bleu : l'approche dense (M1), vert : l'approche gaussien dense (M2), magenta : l'approche locale dense (M3)). Nous constatons que, les directions correspondantes aux images sous testes sont parfaitement alignées à celle de l'image de référence, que ce soit pour les petites ou les grandes rotations. Cela prouve que l'effet de rotation contenu dans les images sous test, a été corrigé correctement. Notons que, à l'œil nu, il est difficile de distinguer laquelle de nos 3 méthodes a donné une meilleure rectification. Comme nous disposons de la vérité terrain, nous avons calculé les erreurs absolues de l'estimation de la rotation. Les résultats sont illustrés sur les tableaux 3.1, 3.2 et 3.3. Encore une fois, la méthode M3, a fourni des estimations plus précises comparant aux autres méthodes.

- Scénario 3 : Robustesse au Translation

Dans ce scénario, la base de données est constituée des images captées en utilisant le robot Staubli. A partir de l'axe vertical du manipulateur, les mouvements horizontaux ont été limités dans un cercle de rayon de 450 mm. Ce cercle a été échantillonné en huit raies avec un espacement angulaire de 45° . Suivant chaque raie, trois distances ont été considérées : 100 mm, 250 mm et 450 mm, ce qu'a fait dans l'ensemble 25 points de collections 3.17. En chacun de ces points, des rotations de 0° à 360° autour de l'axe optique qui correspond à l'axe vertical, ont été effectuées.

Pour un pas de rotation de 2.5° de la camera, les images de cette séquence ont été prises. En effet, nous avons utilisé images de cette séquence afin de tester l'impact de la translation sur la précision de l'estimation de la rotation calculée par notre approche. Cependant, nous avons considéré que les déplacements suivant un seul raie choisi. En effet l'utilisation de toutes les images omnidirectionnelles obtenues suivant les huit raies, va mener à une redondance des données, pour effectuer cette évaluation.

Les tests d'évaluations ont concernés le calcul de l'erreur absolue de l'estimation de la rotation relative entre les images, lorsque la translation varie entre les points de collection indiqués sur la figure .

On peut constater à partir des tableaux 3.1, 3.2 et 3.3 que les trois méthodes s'accommodent donc très bien aux translations relativement faibles puisqu'elles donnent une précision bonne . Cependant, la méthode M3 est plus précise pour les translations importants (450 mm). En effet, l'estimation de la rotation entre l'image de référence et l'image de test est réalisée par des approches denses. Pour cela les deux images doivent être proches. Il est donc préférable d'utiliser un algorithme qui privilégie fortement les images de référence dont le contenu est proche de l'image de test. En effet le pouvoir discriminant de l'approche dense-local est plus important que celui des deux autres approches. Ceci provient du fait que le calcul des moments photométriques reposent sur le voisinage du point caractéristique au lieu de l'image entière de et que l'information qu'ils encodent est très caractéristique dans un endroit particulier si les primitives ont été bien mises en correspondances.

TABLEAU 3.1 – Résultats de l'estimation de la rotation suivant les trois axes par l'approche dense en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.

Rotation Translation(mm)		0°	17.5°	125°	285°
0	R _z	0	18.34	126.23	282.23
	R _y	0	0.04	0.393	0.356
	R _x	0	0.005	0.423	0.067
100	R _z	1.08	18.743	127.06	285.32
	R _y	0.76	1.006	0.6	1.389
	R _x	0.32	0.082	0.563	0.32
250	R _z	2.654	19.56	128.23	286.23
	R _y	1.733	2.577	1.18	0.52
	R _x	0.946	0.036	2.29	2.8
450	R _z	5.13	21.09	129.62	287.72
	R _y	2.799	5.352	4.819	5.664
	R _x	1.96	0.17	1.84	0.761

TABLEAU 3.2 – Résultats de l'estimation de la rotation suivant les trois axes par l'approche gaussien-dense en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.

Rotation Translation(mm)		0°	17.5°	125°	285°
0	R _z	0	18.15	126.08	282.53
	R _y	0	0.019	0.171	0.19
	R _x	0	0.014	0.06	0.13
100	R _z	1.05	18.03	126.55	284.62
	R _y	1.069	0.62	0.23	0.28
	R _x	0.64	0.09	0.27	0.8
250	R _z	1.57	19.38	128.07	286.19
	R _y	2.08	2.16	1.63	0.49
	R _x	1.05	0.048	0.34	1.85
450	R _z	4.82	20.94	128.69	287.35
	R _y	2.18	5.05	4.28	5.09
	R _x	1.31	0.11	1.41	0.35

TABLEAU 3.3 – Résultats de l'estimation de la rotation suivant les trois axes par l'approche dense-local en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.

Rotation		0°	17.5°	125°	285°
Translation(mm)					
0	R _z	0	17.26	125.6	283.61
	R _y	0	0.032	0.279	0.52
	R _x	0	0.027	0.05	0.18
100	R _z	0.61	17.72	123.81	285.4
	R _y	0.146	0.61	0.146	0.85
	R _x	0.72	0.74	0.124	-1.14
250	R _z	1.45	19.18	123.38	285.95
	R _y	1.71	1.19	1.49	0.928
	R _x	1.59	0.32	0.48	2.46
450	R _z	2.23	14.29	127.23	281.62
	R _y	-0.48	1.28	0.804	1.28
	R _x	0.151	0.87	1.47	1.19

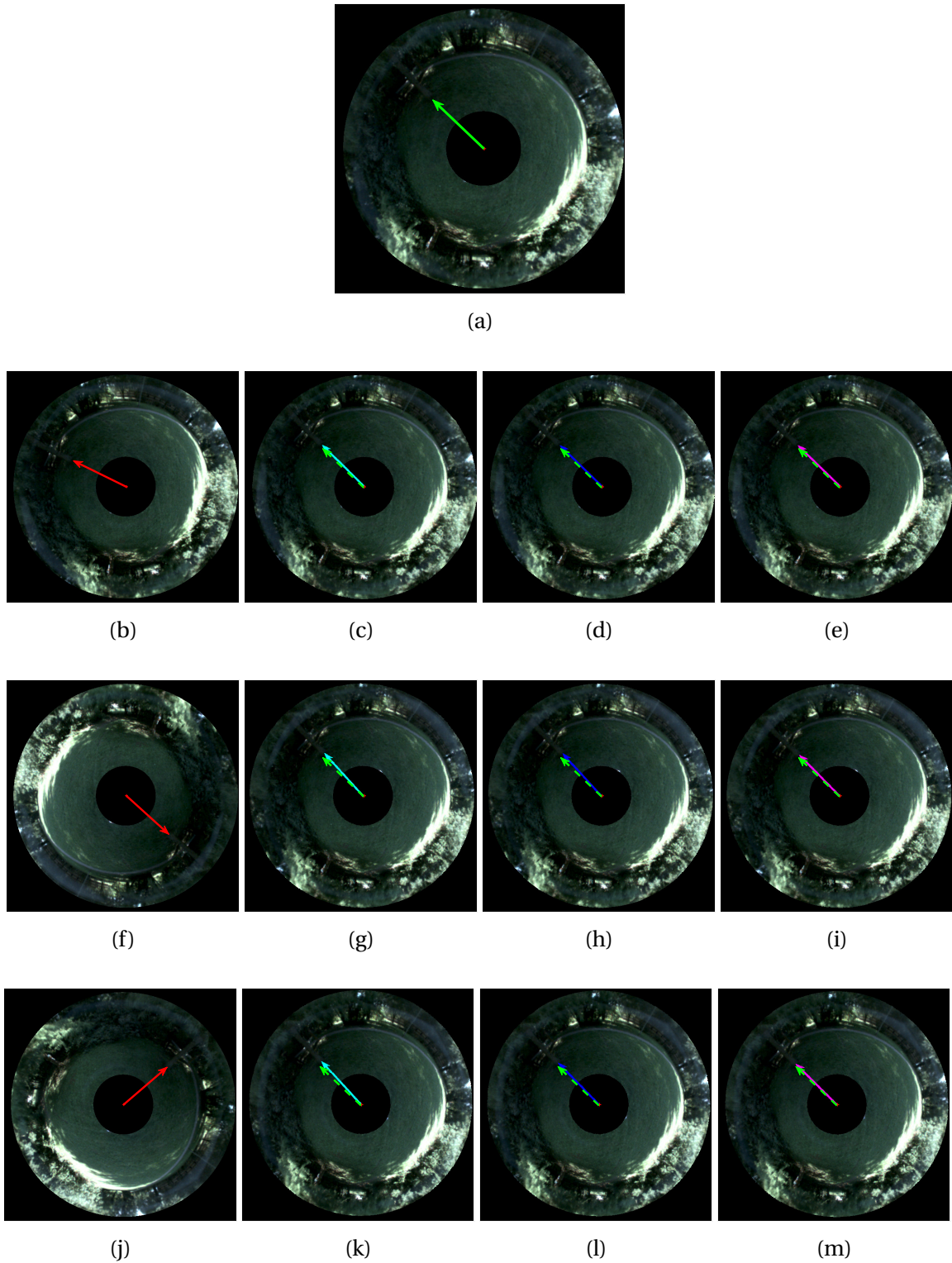


FIGURE 3.18 – Stabilisation de la rotation sur des images catadioptrique par les méthodes proposées. (a) Orientation de référence. (b, f, j) Les images initiales acquises après les rotations suivantes : 17.8° , 185.3° et 266.3° , respectivement. Chaque ligne correspond à l'alignement de rotation par rapport à l'orientation de (a) par les méthodes M1, M2 et M3. Les flèches colorées illustrent une direction initiale (vert) ou alignée M1 (cyan), M2 (bleu) et M3 (magenta). Pour une visualisation plus claire, une seule direction est tracée, mais l'alignement rotatif consiste non seulement en une rotation 1D, mais plutôt en une transformation de rotation 3D complète.

3.9 Conclusion

Ce chapitre a permis d'introduire une nouvelle solution analytique conçue pour l'estimation de la rotation 3D des images sphériques. Trois méthodes ont été introduit à partir de la solution proposée, en s'appuyant sur le calcul des moments photométriques pour offrir une description dense, qu'elle soit globale ou locale. Cette description est construite en calculant les moments de l'intensité de l'image sphérique, qui sert à la conception d'informations visuelles permettant de trouver la matrice orthogonale qui sépare ces informations (la matrice de rotation). En appliquant ce type de solution sur l'image omnidirectionnelle sphérique intégrale ou seulement au voisinage d'une paire des points d'intérêts après la mise en correspondance, ce chapitre a montré la validité de cette solution dense proposée pour les zones locales ainsi que l'image globale.

Nous avons évalué le comportement des méthodes proposées par l'estimation de la rotation à travers des données synthétiques et réelles. Les résultats obtenus ont permis de montrer que les moments photométriques peuvent être utilisés pour estimer la rotation 3D.

Il est à noter que l'approche proposée, basée sur les moments photométriques sphériques, est valable même pour les larges rotations (domaine de convergence grand). Aussi, nous avons constaté que l'approche dense local (M3) est plus précise que les deux approches globales, étant donnée que l'optimisation de l'information visuelle obtenue en local est plus pertinente qu'on globale.

Chapitre 4

Résultats expérimentaux

Sommaire

4.1 Environnement intérieur	115
4.2 Environnement extérieur	119
4.3 Discussion	122

Les expérimentations d'estimation de la rotation 3D à partir d'images sphériques se déroulent en plusieurs étapes. Les expérimentations portent sur les deux algorithmes : les approches basées moments sphériques et l'algorithme utilisant les harmoniques sphériques. Afin de tester les performances de nos approches (basée SPHARM, M1, M2 et M3) vis-a-vis le changement d'environnement, des tests en environnements d'intérieur et d'extérieur ont été menés.

4.1 Environnement intérieur

Afin de valider et comparer les approches développées, il a été nécessaire d'appliquer chacune des approches sur les mêmes données obtenues depuis une expérimentation spécifique. En tout premier lieu, il fallait des images sphériques pour pouvoir d'en valider la performance. A défaut d'images qui couvèrent suffisamment bien la totalité de la sphère, les images obtenues d'une camera catadioptrique centrale ont été utilisées. Ce type d'images obéissant au modèle de projection centrale et contenant donc une information visuelle plus large que l'hémisphère.

En deuxième lieu, on avait préféré que la trajectoire suivie par le robot, doit contenir la fermeture de boucle, c-à-d que le robot revient à son point de départ initial. En effet, une

estimation précise de l'orientation du robot permettant de rapprocher de plus en plus de la trajectoire réelle fournie par la vérité terrain.



FIGURE 4.1 – Robot d'expérimentation en environnement intérieur : Turtlebot 2, équipé de la camera catadioptrique

Pour tout cela, nous avons généré une base de données qui satisfait ces contraintes pour tester les algorithmes dans une telle situation. L'acquisition de la base de données a été réalisée par la camera catadioptrique déportée sur le robot Turtlebot 2 que l'on dispose au niveau du laboratoire IBISC. Cette plateforme robotique est un petit robot mobile qui convient aux environnements d'intérieur. Il s'agit cependant d'une base mobile équipée de différents équipements de mesure et d'acquisition (gyromètre, des capteurs infrarouge, des capteurs chute de roues et odomètre). La figure 4.1 présente le robot Turtlebot 2. Le kit complet est constitué des composants suivants : base mobile, camera rgb-d et laptop, qui sont fournis ensemble et assemblés. Sur le laptop, ROS, ainsi que les packages nécessaires au fonctionnement de robot, sont déjà pré-installés. Il est donc possible d'utiliser certaines fonctionnalités pour mettre le robot en service dès son déballage.

Cette première expérimentation a été effectuée dans la salle robotique de laboratoire IBISC qui se situe au bâtiment Pelvoux de l'université d'Évry. Le plan de la salle affiché sur la figure 4.2 permet de donner une idée de l'environnement dans lequel le robot a évolué durant l'expérimentation, et de la trajectoire réelle parcourue par le robot.

Une fois la trajectoire du robot est tracée au sol de la salle, le robot est téléopéré en suivant le long de la trajectoire de référence en utilisant un joystick. La vérité terrain de la trajectoire réelle parcourue par le robot est supposée donnée par les capteurs d'odométrie et du gyroscope.

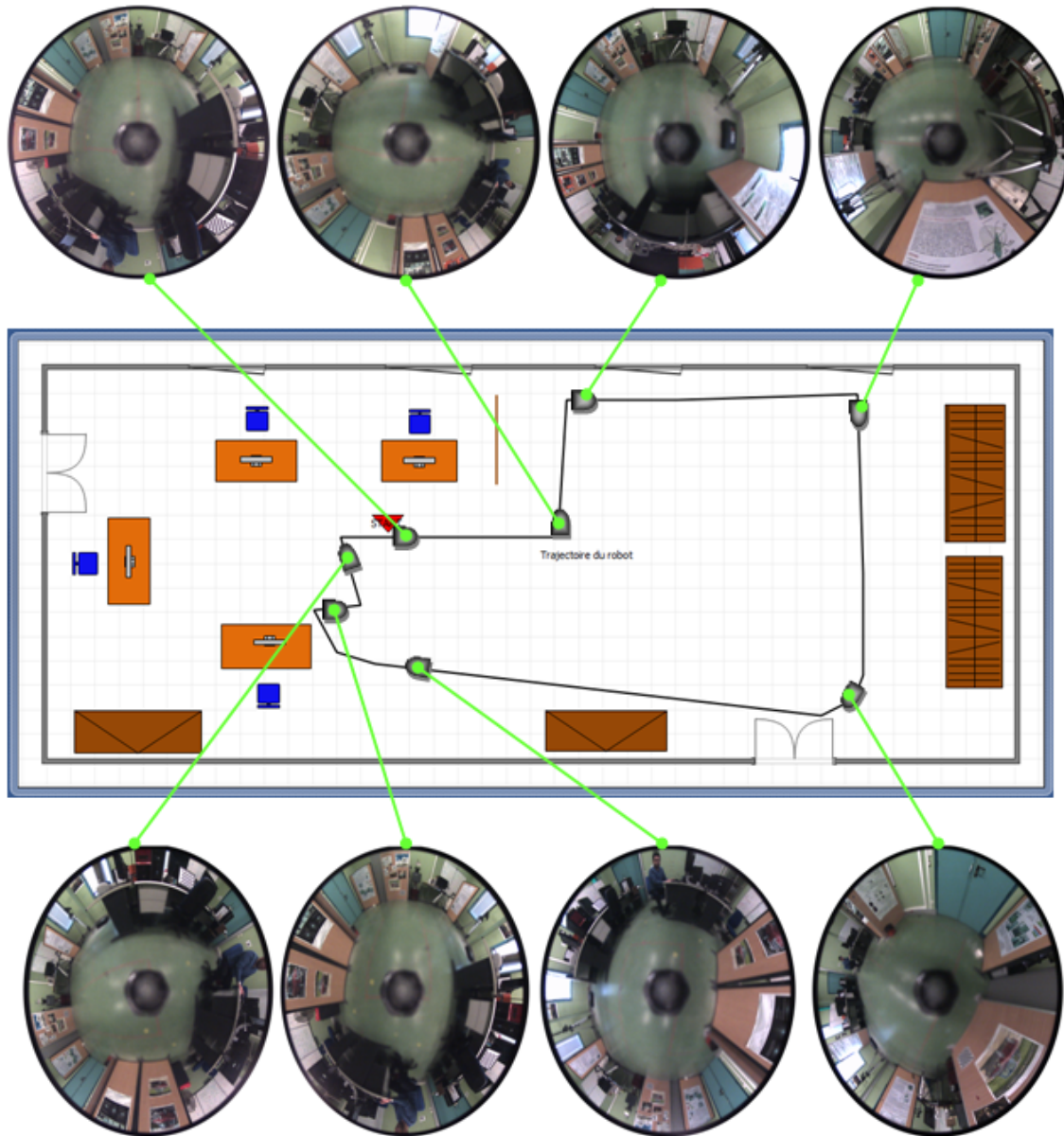


FIGURE 4.2 – Le plan de masse de l’environnement intérieur dont la trajectoire parcourue par le robot a été superposée. La sélection de 8 poses parmi les images de la séquence.

En ce qui concerne la base de données images, la séquence a été acquise avec une cadence de 5 fps (Image par seconde) de taille de 426×128 . Étant donnée que la distance est d’environ 17 mètres et a été parcourue par une vitesse de 0.5 m/s , la base de données est composée de 170 images.

Les résultats de l’expérimentation sont présentés sur la figure 4.3. La trajectoire réelle du robot (obtenue par la vérité terrain) est affichée à chaque fois avec la trajectoire estimée pour chaque approche séparément. Ces trajectoires estimées ont cependant été obtenues en estimant la rotation entre les images de la séquence et en introduisant la translation fournie par les odomètres dans le calcul du déplacement de robot. Une étape de synchronisation a été menée entre les images de la séquence et les mesures de la trans-

lation.

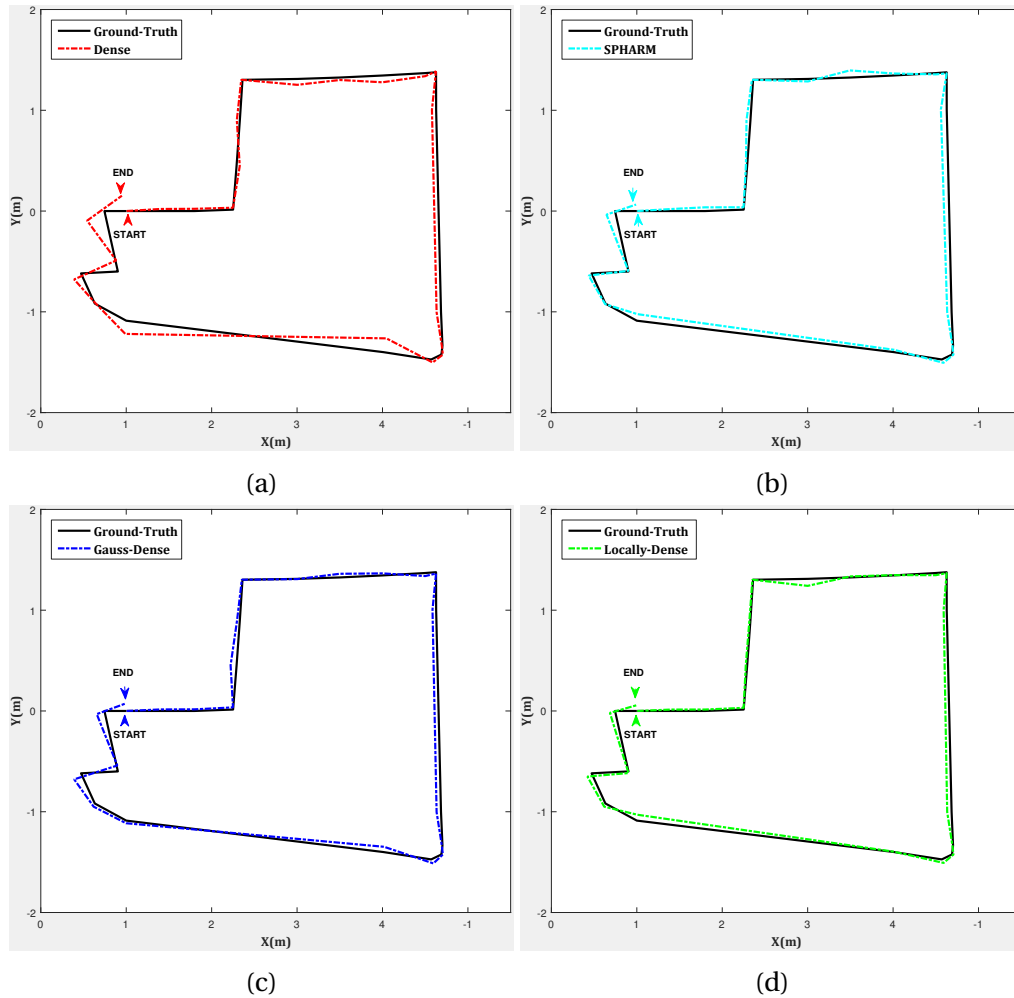


FIGURE 4.3 – Résultats d'estimation de déplacement du robot obtenus par les approches suivantes : (a) SPHARM, (b) M1, (c) M2 et (d) M3.

Bien que les quatre approches sous test ont pu fournir des estimations proches de la trajectoire réelle, nous remarquons que les approches basées moments sont un peu plus précises que celle basée harmoniques sphériques. Étant donné que cette dernière approche utilise des images sous échantillonnées par rapport aux approches basées moments qui utilisent l'image originale et ont donc une influence plus conséquente sur la précision de l'estimation. Toutefois, parmi les approches basées moments, c'est l'approche local dense (M3) qu'a offert l'estimation la plus proche de la vérité terrain. Ce type d'approche et en se basent seulement sur le voisinage d'une seule bonne correspondance permet de trouver la rotation relative effectuée par le robot. A l'opposé des méthodes basées sur un calcul global des moments photométriques (M1 et M2), la méthode M3 exploite l'intensité disponible en zone locale, correspondante au voisinage de primitives. L'utilisation des zones locales au lieu des images intégrales, permettant d'obtenir un recouvrement plus fin et plus efficient. Par conséquent, l'estimation de la rotation obtenue par la méthode M3 est plus précise par rapport à ceux obtenues par les méthodes M1 et

M2.

4.2 Environnement extérieur

Par ailleurs, les environnements d'extérieur sont nettement plus compliqués (niveau de structuration et de texture faible : extraction plus complexe de primitives géométriques), à plus forte raison quand les scènes changent dynamiquement. La différence principale cependant entre un milieu d'extérieur et celui d'intérieur, est que les conditions d'illumination sont non contrôlables et dans certain cas ne sont même pas prédictibles.

Pour ce deuxième test, un environnement d'extérieur a été considéré. Cette expérience est réalisée en utilisant les images de scénario 3 de la base de données OMNIVIS [164], la même base utilisée pour les expérimentations du chapitre 3. Nous rappellerons que, la plateforme robotique est composée d'un robot mobile Pioneer, elle est présentée sur la figure 3.13. En effet, le scénario 3 introduit une séquence d'images avec la vérité terrain, correspondante aux déplacements, présentée sous la forme de translations et rotations, obtenues depuis la centrale inertielle. Selon une trajectoire de forme de huit et de longueur de 25.27 m (voir Figure 4.4-c), le robot a été téléguidé manuellement. Les vitesses pour traverser la trajectoire ont été dotés à 0.187 m/s en vitesse de déplacement moyenne, 0,2 m/s en vitesse de translation et 12 deg/s en vitesse angulaire. Les acquisitions ont été effectuées à Amiens, dans le parc public Bois Bonvallet. Le temps d'exposition de la caméra était limité à 50 ms, ce qui donnait une séquence de 318 d'images.

Le test a donc consisté à estimer les rotations entre les images successives de la séquence. Comme illustré sur la figure 4.5, nous avons testé l'approche basée harmoniques sphériques et les trois approches basées moments sphériques. Celle-ci est rapportée en quatre courbes qu'en ont été tracées en quatre couleurs différentes, chacune correspond à une des approches proposées et la courbe de la vérité terrain est tracée en noire.

Les valeurs de la rotation 3D estimée ont été représentées par les angles d'Euler (yaw, pitch, roll) indiquant les angles de rotation autour des trois axes XYZ. Une illustration de la représentation d'angle d'Euler considérée, est donnée sur la figure 4.6.

Les figures montrent les résultats obtenus des estimations incrémentales de la rotation suivants les trois axes . Notons que, l'axe de rotation du robot c'est l'axe Z et qui s'aligne avec l'axe optique de la camera. D'ailleurs on peut le constater à partir des trois figures, lorsque les valeurs de l'angle yaw apparaissent avec des changements plus significatifs, contrairement aux angles pitch et roll.

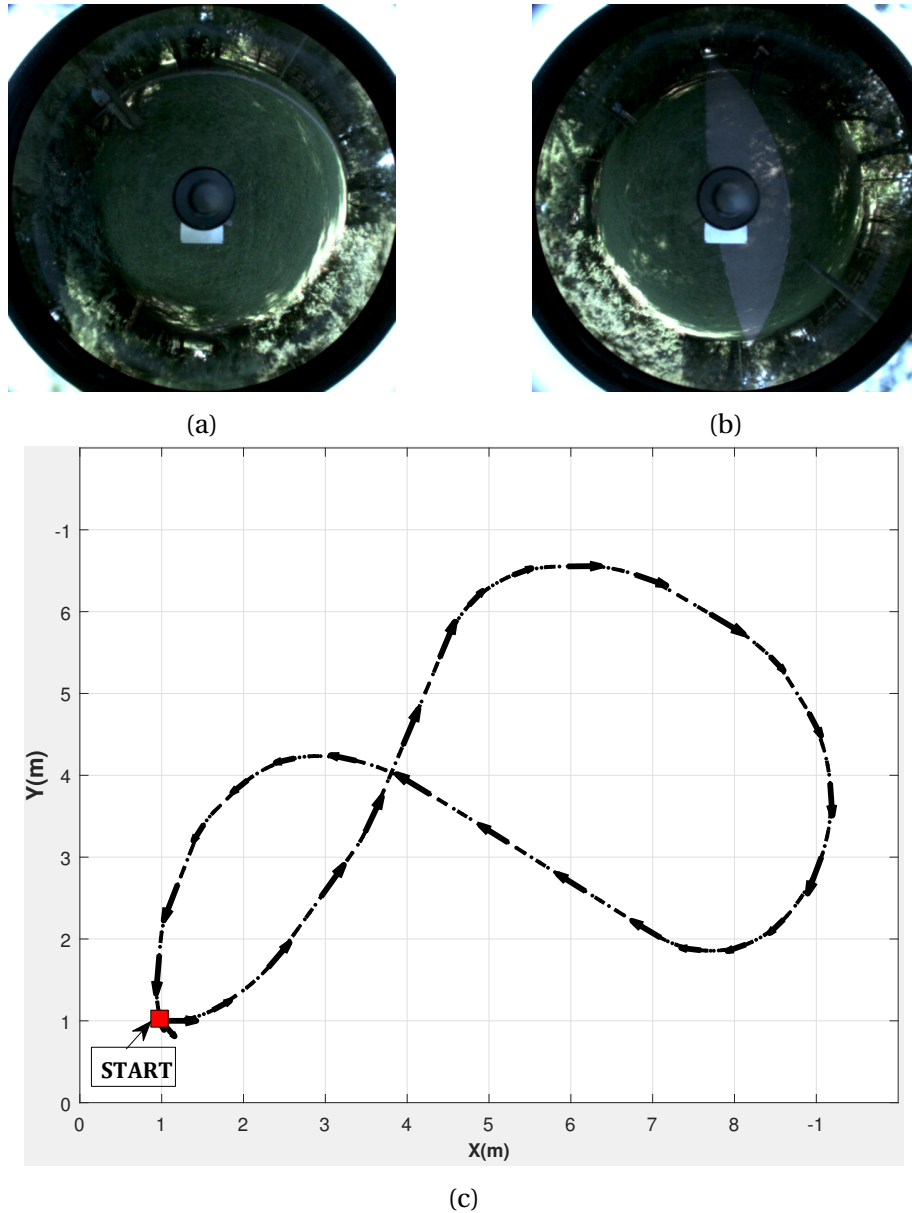


FIGURE 4.4 – (a) L'image initiale de scénario 3 (b) L'image 137 de scénario 3 (c) La trajectoire de forme 8 .

Il a été noté dans [164], que durant les expérimentations, la caméra catadioptrique déportée par le pionner, a subi des oscillations en traversant un terrain accidenté, ce qui engendre un mouvement non planaire de la camera. Ce phénomène peut être constaté sur l'image 137 (voir figure 4.4-b), quand le robot a franchi le sentier en revenant vers la pelouse. De plus, il y avait aussi, des variations d'éclairage dus à l'ombre mobile des arbres. Ce qui valide effectivement ce qu'a été dit précédemment, que les conditions expérimentales d'un milieu extérieure sont souvent plus difficiles qu'en un milieu intérieur. Malgré ces sources d'erreur importantes, nos approches ont montré leurs efficacités d'estimation du mouvement de rotation contre les changements inattendus d'images, avec des précisions plus ou moins hétérogènes.

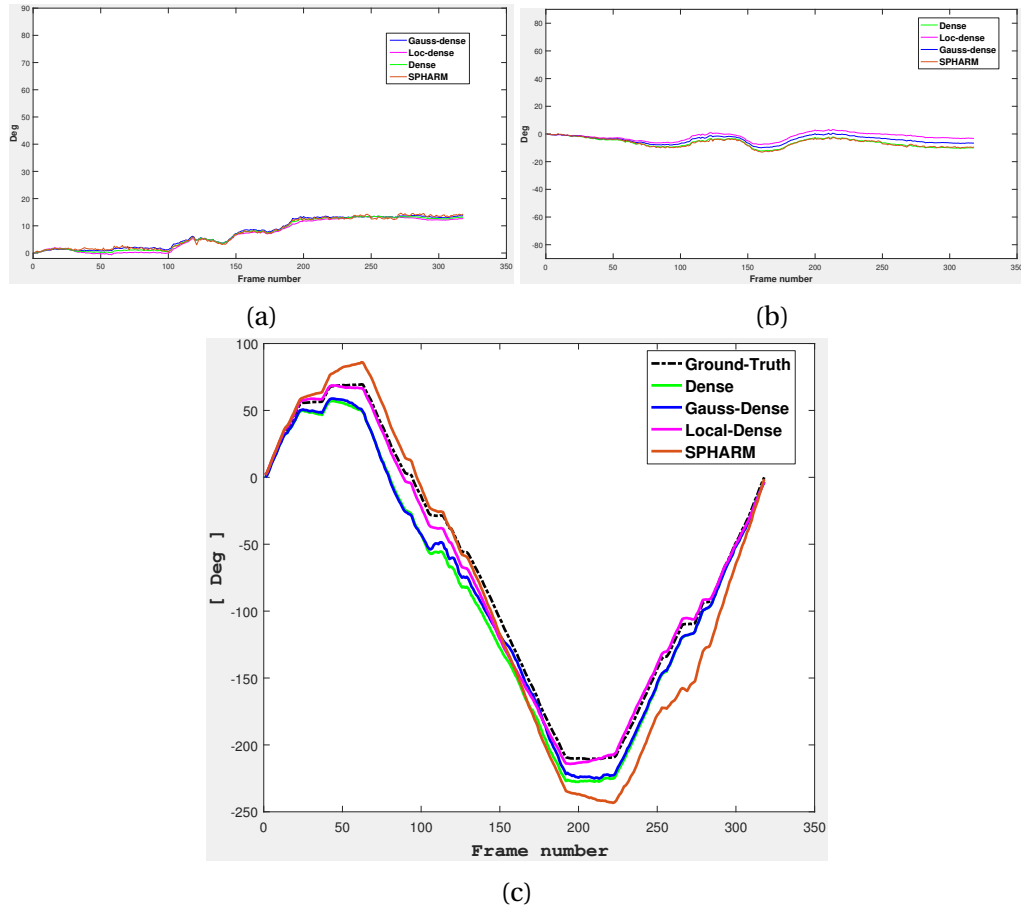


FIGURE 4.5 – L'estimation des angles d'Euler pour la trajectoire de 8 par les approches suivantes : SPHARM (marron), M1 (bleu), M2 (vert) , et M3 (rose), par rapport à la vérité de terrain (noir). (a) roll, (b) pitch et (c) yaw .

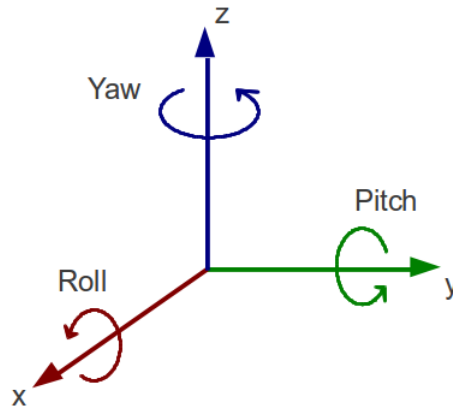


FIGURE 4.6 – La nomenclature des angles d'Euler : Roll, Pitch et Yaw, indiquant les angles de rotation autour des trois axes XYZ.

La qualité des résultats obtenus de chacune de nos approches est mesurée, par la suite, selon différents critères. Le critère de robustesse qui mesure la stabilité des résultats face aux perturbations des données, le critère de fiabilité, mesurant le degré de confiance que l'on peut avoir dans l'affirmation que les résultats soient proches de la vérité terrain.

Une telle analyse, nécessite le calcul de la moyenne, l'écart-type, les valeurs minimales et maximales de l'amplitude des erreurs d'estimation pour les approches proposées. Ces métriques ont été déjà mises en œuvre dans le travail lié à la proposition de la base de données OVMIS, dont une méthode d'estimation de la rotation 2D a été présentée [164]. En calculant les métriques basées sur les données de scénario 3, le tableau comparatif 4.1 a été obtenu. Nous comparons ici les performances de nos approches, d'abord de chacune d'entre elles, puis par rapport aux résultats des approches basées corrélation de phase (COR) et photométrique (PHO), proposées dans [164].

TABLEAU 4.1 – Résultats de l'estimation de la rotation suivant les trois axes par l'approche dense en fonction de la translation. La vérité terrain est la rotation autour de l'axe Z.

	Moyenne	Écart de type	Min	Max
COR	0.60	0.67	0.00	3.47
PHO	0.30	0.26	0.00	1.40
SPHARM	0.56	0.51	0	2.59
M1	0.50	0.44	0	1.98
M2	0.41	0.32	0	1.78
M3	0.19	0.16	0	1.05

Le Tableau 4.1 compare les différentes valeurs de l'erreur d'estimation liées à chaque approche considérée. On voit à partir du tableau que l'approche local-dense (M3) présente les meilleurs résultats de l'estimation par rapport à toutes les autres approches, y compris celle basée corrélation de phase.

De nouveau, le calcul des moments sphériques local, à partir de voisinage d'un seul appariement entre primitives, a montré leur intérêt. Ces résultats démontrent bien la capacité de la solution proposée basée moments sphériques en estimant la rotation entre deux images, notamment lorsqu'il y a du forte similitude entre les deux signaux, comme c'est le cas du voisinage des primitives. On peut également ajouter que nos approches présenterait en outre un important avantage, à savoir la possibilité de fournir l'estimation de la rotation autour des 3 axes. Les figures de 4.5 montrent toutefois l'estimation incrémentale des pitch et roll angles, engendré par le mouvement non planaire du système catadioptrique causé par ses oscillations.

4.3 Discussion

Dans ce chapitre, nous avons exposé nos résultats et analyses pour l'estimation de la rotation 3D. La première approche basée sur la représentation SPHARM montre des

résultats intéressants mais présente divers inconvénients :

- Sous-échantillonnage de l'image sphérique
- Complexité de la paramétrisation sphérique qui fait intervenir des optimisations numériques en calculant la projection bijective
- Raffinement de l'estimation par l'ICP.
- Très couteuse en terme de temps de calcul.

Malgré ces défauts, cet algorithme présente de nombreux avantages dont le plus important est la validation de la preuve de concept quant à l'utilisation pour l'estimation de la rotation, en représentons l'intensité d'images sous forme d'objets 3D. Les différents avantages de ce premier sont alors :

- Estimation de la rotation 3D par approche dense bien établie.
- Preuve de concept de la représentation de l'intensité lumineuse d'images sphériques par objet 3D.
- Calcul du descripteur global complètement adapté à la sphère.
- Résultats prometteurs pour les différents environnements et robuste contre les occultations.
- Utilisation du descripteur global "SPHARM" de très faible dimension.

Suite à cette première approche, nous avons élaboré une seconde solution basée sur le calcul des moments sphériques pour la description de l'information contenue dans le signal d'image.

Cette deuxième méthode est mieux formulée et présente une solution analytique au problème de l'estimation de la rotation 3D entre les images sphériques. Trois variantes approches qui se basent sur cette solution, ont été proposées. La solution prend bien en compte la totalité de l'image sphérique. Les résultats obtenus sont supérieurs à ceux obtenus par la représentation SPHARM, notamment l'approche basée sur le voisinage des points d'intérêts, qui concrétisent l'amélioration de la précision de l'algorithme.

Les résultats obtenus sont significatifs et correspondent efficacement à la vérité terrain fournie par la base de données. Le désavantage central de cette nouvelle solution est :

- La dépendance à la conservation de l'intensité lumineuse entre, les images ou régions, sous testes.

Malgré cette inconvénient, nous obtenons une solution directe et efficace de l'estimation de la rotation 3D. Toutefois, la majorité des problèmes évoqués pour la première approche on été résolus. En ce qui concerne les avantages de cette nouvelle approche, ils sont nombreux :

- Estimation de la rotation 3D par des approches denses et localement dense, bien établie.
- Pas de sous-échantillonnage de l'image.
- Solution directe sans raffinement de l'estimation.
- Détermination de la rotation 3D à partir d'une seule correspondance.
- Descripteur global sphérique bien adapté à la représentation sphérique.
- Algorithme temps-réel.
- Utilisation d'un descripteur global de très faible dimension.
- Résultats d'estimation suffisamment précises pour les environnements intérieur et extérieur.

CONCLUSION GÉNÉRALE

Le travail de cette thèse s'inscrit dans le sillage des nombreuses recherches menées autour du problème d'estimation du mouvement basée sur la vision omnidirectionnelle. L'objectif fixé au début était de proposer des méthodes génériques d'estimation des grands déplacements sans détection ou suivi des primitives visuelles dans l'image, dans le but de s'affranchir des phases de détection, description et la mise en correspondance des primitives. Pour cela, les outils et les procédures de traitement d'image et du signal les plus efficaces, ont été sollicités afin d'assurer une description globale d'images dans le domaine sphérique. Dans cette directive, deux approches d'estimation de la rotation 3D, ont été proposés.

Dans la première approche, nous avons proposé une représentation originale de l'intensité lumineuse de l'image sphérique sous forme d'un objet 3D. Étant donné que l'objet 3D est généré, nous nous sommes focalisés sur le calcul des coefficients des harmoniques sphériques directement sur le maillage de l'objet. Dans le cas où sa topologie n'est pas équivalente à celle de la sphère, nous avons abordé une technique de réparation du maillage d'objet pour en obtenir une sphérique. Une fois les coefficients harmoniques des deux objets sont obtenus, la rotation qui sépare les deux objets a été obtenue par la décomposition SVD. Une étape de raffinement par l'algorithme ICP a été proposée, afin d'extraire la rotation optimale entre les deux images sphériques. D'autre part, cette solution dense de l'estimation de la rotation, repose sur la nouvelle représentation de l'image qu'est explicite et moins compacte. L'avantage d'une telle représentation réside dans la possibilité de faire appel à la théorie de description globale des objets 3D, et d'y généraliser différentes solutions d'estimation de la rotation. Afin de remédier aux éventuelles erreurs d'estimation de la rotation dans le domaine spectral, nous avons pensé à un raffinement de l'estimation par l'ICP. En effet, cette technique itérative assure la convergence vers la solution optimale. Il en résulte un algorithme efficace présentant une estimation de la rotation très satisfaisante.

Une deuxième approche a été proposée, utilisant les moments photométriques sphériques, pour l'estimation de la rotation d'image. Cette approche propose une solution

analytique de calcul de la rotation entre les images sphériques. La problématique initiale d'une estimation dense des larges rotations est respectée par cette solution, tandis que l'intensité globale de l'image est capturée sous la forme des moments photométriques sphériques calculés depuis l'intensité lumineuse de l'image. Le modèle analytique a été en l'occurrence établi pour obtenir la matrice d'interaction des moments photométriques.

A travers ceci, nous avons adopté trois méthodes pour l'estimation de la rotation. Dans la première, les moments sphériques ont été calculés à partir de l'image brute. La deuxième méthode consiste à appliquer une gaussienne, dans le but de se focaliser sur la région contenant plus d'information sur la rotation. En effet, en appliquant la première méthode, le résultat de l'estimation est utilisé pour orienter une gaussienne. Le filtrage d'image par la gaussienne orientée, sert à maximiser le recouvrement entre les deux images sphériques. Cette stratégie de pondération conduit naturellement à un raffinement de l'estimation. La troisième méthode, est une approche hybride combinant des moments photométriques avec des caractéristiques géométriques. A travers cette approche, il est possible d'estimer la rotation 3D, en utilisant seulement un point mis en correspondance entre les deux images. Ce type de solution est localement dense. Les moments photométriques sont calculés depuis le voisinage des points appariés. La méthode hybride présente une solution très efficace, néanmoins les étapes de détection et d'appariement de primitives sont présentes.

Tous les algorithmes proposés dans cette thèse, sont le plus générique possible, et ne nécessitent aucune connaissance a priori sur l'environnement. La validité du modèle sphérique unifié, pour les images conventionnelles ainsi que omnidirectionnelles, permettant l'intégration de ces algorithmes au profit de divers applications en vision par ordinateur. En effet, l'objectif fixé au début a été atteint, étant donné que nos approches sont essentiellement basées sur l'information extraite de l'image entière sans dépendre des processus d'extraction, de description et d'appariement ou de suivi des primitives. Ainsi, les algorithmes fournissent des résultats précis lors de l'estimation de la rotation 3D et une robustesse vis-à-vis des conditions variables telles que le changement de luminosité et le bruit. Les algorithmes présentent également de bonnes propriétés d'estimation de l'orientation en environnement intérieur structuré, et en environnement moins structuré (outdoor).

Perspectives

Parmi les perspectives de cette thèse, nous citons les points suivants :

L'utilisation d'un capteur qui ne fournit pas une image parfaitement sphérique, a certaines limitations en appliquant nos approches adaptées à la sphère. En outre, la validation des approches par un capteur permettant l'acquisition de sphères visuelles, peut être bénéfique en estimant la rotation 3D à travers ces images. Cela nécessite de développer ou d'acquérir ce type de capteur qui est conçu spécialement pour cette tâche.

Dans ce document, nous avons cherché à proposer des solutions denses à l'estimation de la rotation 3D permettant d'éviter la détection et la mise en correspondance de points d'intérêt entre images. Cependant, nous n'avons pas évoqué le problème de l'estimation du facteur d'échelle. Une autre solution possible envisageable est l'utilisation de l'information dense afin d'estimer la translation à un facteur d'échelle près. Le fait qu'on puisse obtenir la rotation et la translation nous permet d'envisager l'estimation de l'orientation relative à 5 degrés de liberté. Cette estimation donnera lieu à l'odométrie visuelle dense à propos de déplacement du robot mobile. Ainsi, une extension à l'espace 6D de la méthode d'estimation de mouvement est possible en ajoutant la valeur de la profondeur fournie soit par un système de stéréovision ou par un télémètre laser.

L'objectif des algorithmes d'estimation de la rotation élaborés, est de fournir une solution permettant de fournir l'orientation 3D du robot. De plus, la solution proposée fournit une méthode générale applicable à n'importe quel type de robot dans n'importe quel type d'environnement. Sans être une amélioration, conduire des expérimentations avec un drone permettrait de compléter l'analyse de l'estimation de l'orientation 3D, que n'était pas possible par un robot mobile effectuant un mouvement planaire. Pour cela, le drone devra effectuer des manœuvres engendrant des rotations suivant les trois axes et en appliquant nos approches, les rotations 3D doivent être obtenues.

L'intégration des algorithmes et méthodes développés sur une plateforme logicielle et matérielle commune constitue l'étape finale de ce travail de thèse. Il est présent possible d'implémenter nos algorithmes, notamment les approches basées moments photométriques, sur une plateforme robotique mobile équipée du système d'exploitation ROS (Robot Operating System). Nous avons déjà commencé le déploiement d'une plateforme robotique dotée du système ROS, pour l'acquisition d'images et le calcul de déplacement relatif. Les contraintes de calcul en temps réel peuvent être respectées par des calculs déportés ou avec une implémentation sur carte graphique à l'aide d'outil tel que CUDA. Le scénario réalisable à l'aide des approches développées est la navigation d'un robot mobile terrestre ou aérien où l'estimation de la rotation peut s'effectuer à l'aide des algorithmes développés.

Annexe A

Annexes

A.1 La transformée de Fourier sphérique

Le traitement d'images planes implique, dans la plupart d'outils de traitement, une étape de filtrage avec un opérateur adapté. Souvent ces opérations s'effectuent mathématiquement par un produit de convolution entre l'image et un opérateur, qui se calcule aisément dans le domaine spectral par la transformée de Fourier. En effet, nous allons présenter la transformée de Fourier dans l'espace sphérique, tout en effectuons une analogie avec le cas du plan.

Soit $f \in L(\mathbb{R}^2)$, sa transformée de Fourier de f est définie par :

$$\hat{f}(\xi) = \int_{\mathbb{R}^2} f(x) e^{-ix \cdot \xi} dx \quad (\text{A.1})$$

Notons que la transformée de Fourier de f est encore une fonction définie sur $\mathbb{R}^2$. Étant donnée la formule de convolution sur le plan :

$$f * g(x) = \int_{\mathbb{R}^2} f(x - y) g(y) dy$$

(avec f et g deux fonctions définies sur $\mathbb{R}^2$), il est facile de montrer que :

$$\widehat{f * g} = \hat{f} \cdot \hat{g}$$

Cette propriété est souvent utilisée pour l'implémentation des outils de traitement d'images.

La transformation de Fourier (TF), est une opération qui transforme une fonction intégrable sur $\mathbb{R}$ en une autre fonction, décrivant le spectre fréquentiel de cette dernière. Si f est une fonction intégrable sur $\mathbb{R}$, sa transformée de Fourier est la fonction $\text{TF}(f) = \hat{f}$ donnée par la formule (A.1), qu'est développée sur une base d'exponentielles. Cette base est la solution de l'équation de Laplace en coordonnées cartésiennes dans un espace euclidien de dimension 3.

Sur la surface de la sphère, pour trouver une base de fonctions sphériques sur laquelle on peut décomposer judicieusement toute fonction définie sur la sphère, on est amené à résoudre l'équation de Laplace en coordonnées sphériques $\Delta\psi = 0$, qui s'écrit sous cette forme :

$$\Delta_{\mathbb{S}^2} = \frac{\partial^2 \psi}{\partial^2 \theta} + \frac{\cos \theta}{\sin \theta} \frac{\partial \psi}{\partial \theta} + \frac{1}{\sin^2 \theta} \frac{\partial^2 \psi}{\partial^2 \varphi} \quad (\text{A.2})$$

Les harmoniques sphériques sont les fonctions propres de l'opérateur de Laplace sphérique :

$$\Delta_{\mathbb{S}^2} Y_l^m = -l(l+1) Y_l^m$$

En effet, ce sont les solutions de l'équation A.2. Ils se sont définies par :

$$Y_l^m(\theta, \varphi) = \sqrt{\frac{2l+1}{4\pi} \frac{(l-m)!}{(l+m)!}} P_l^m(\cos \theta) e^{\pm i m \varphi}$$

pour $m \geq 0$, et

$$Y_l^m(\theta, \varphi) = (-1)^m \overline{Y_l^{-m}(\theta, \varphi)} \quad (\text{A.3})$$

pour $m < 0$,

Avec $l \in \mathbb{N}$, $|m| \leq l$ et P_l^m sont les polynômes de Legendre associés de degré l et d'ordre m , ça formule est donnée par :

$$P_l^m(u) = (-1)^m (1-u^2)^{\frac{m}{2}} \frac{d^m}{du^m} P_l(u)$$

D'une façon plus explicite, il s'écrit :

$$P_l^m(u) = \frac{1}{2^l l!} (1-u^2)^{\frac{m}{2}} \frac{d^{l+m}}{du^{l+m}} (u^2-1)^l$$

Nous listons en ce qui suit, les harmoniques sphériques pour $l = 0, 1, 2$ et les valeurs non négatives de m .

$$\begin{aligned} l=0, \quad & Y_0^0 = \frac{1}{\sqrt{4\pi}} \\ l=1, \quad & \begin{cases} Y_1^1 = -\sqrt{\frac{3}{8\pi}} \sin \theta e^{i\varphi} \\ Y_1^0 = -\sqrt{\frac{3}{4\pi}} \cos \theta \end{cases} \\ l=2, \quad & \begin{cases} Y_2^2 = -\frac{1}{4} \sqrt{\frac{15}{2\pi}} \sin^2 \theta e^{2i\varphi} \\ Y_2^1 = -\sqrt{\frac{15}{8\pi}} \sin \theta \cos \theta e^{i\varphi} \\ Y_2^0 = -\frac{1}{2} \sqrt{\frac{5}{4\pi}} (3 \cos^2 \theta - 1) \end{cases} \end{aligned}$$

Les harmoniques sphériques correspondantes aux valeurs négatives de m sont obtenues en utilisant l'équation A.3.

Étant donnée que les harmoniques sphériques forment une base orthonormale pour $L^2(\mathbb{S}^2)$, chaque fonction $f \in L^2(\mathbb{S}^2)$ s'écrit à l'aide des harmoniques sphériques sous la forme suivante :

$$f = \sum_{l \in \mathbb{N}} \sum_{|m| \leq l} \hat{f}_l^m Y_l^m$$

Avec

$$\hat{f}_l^m = \int_0^{2\pi} \int_0^\pi f(\theta, \varphi) \overline{Y_l^m(\theta, \varphi)} \sin(\theta) d\theta d\varphi$$

L'ensemble des coefficients $\hat{f}_l^m$ est appelé les coefficients de Fourier sphérique ou spectre de f .

A.2 Le calcul de la variation temporelle des moments

Soit $(\mathbf{F}(t, \cdot))_{t \geq 0}$ une famille de transformations bijectives. Pour un point $\mathbf{s} = [x, y]^\top$ on note :

$$\mathbf{s}(t) = \mathbf{F}(t, \mathbf{s}) \quad (\text{A.4})$$

où $\mathbf{s}(0) = \mathbf{s}$ et $\Omega_{int}(t) = \mathbf{F}(t, \Omega_{int})$. Soit $\dot{\mathbf{s}}$ la vitesse d'un point du contour $\mathbf{s} = [x, y]^\top$:

$$\dot{\mathbf{s}}(t, \mathbf{s}(t)) = \frac{\partial \mathbf{F}}{\partial t}(t, \mathbf{s}) \quad (\text{A.5})$$

En s'intéressant aux petites déformations pour le calcul de la dérivée, et en utilisant le théorème d'expansion de Taylor, on a :

$$\mathbf{F}(t, \mathbf{s}) = \mathbf{F}(0, \mathbf{s}) + t \frac{\partial \mathbf{F}}{\partial t}(0, \mathbf{s}) \quad (\text{A.6})$$

La dérivée Eulérienne dans la direction du champ de vecteurs $\dot{\mathbf{s}}$ (vitesse d'un point $\mathbf{s}(x, y)$) nous donne :

$$\dot{m}_{pq} = \lim_{t \rightarrow 0} \frac{(m_{pq}(\Omega_{int}(t)) - m_{pq}(\Omega_{int}))}{t} \quad (\text{A.7})$$

La dérivée de $k(\mathbf{s}, \Omega_{int})$:

$$\dot{k} = \lim_{t \rightarrow 0} \frac{(k(\mathbf{s}(t), \Omega_{int}(t)) - k(\mathbf{s}, \Omega_{int}))}{t} \quad (\text{A.8})$$

La dérivée de domaine de $k(\mathbf{s}, \Omega_{int})$ est définie par :

$$k' = \lim_{t \rightarrow 0} \frac{(k(\mathbf{s}, \Omega_{int}(t)) - k(\mathbf{s}, \Omega_{int}))}{t} \quad (\text{A.9})$$

De toute évidence, en transformant A.7 selon la première formule de Taylor, nous avons :

$$\dot{k}(\mathbf{s}, \Omega_{int}) = k'(\mathbf{s}, \Omega_{int}) + \nabla k(\mathbf{s}, \Omega_{int}) \cdot \dot{\mathbf{s}} \quad (\text{A.10})$$

La transformation de m_{pq} donne :

$$\frac{(m_{pq}(\Omega_{int}(t)) - m_{pq}(\Omega_{int}))}{t} = \frac{1}{t} \left[\int_{\Omega_{int}(t)} k(\mathbf{s}, \Omega_{int}(t)) d\mathbf{s} - \int_{\Omega_{int}} k(\mathbf{s}, \Omega_{int}) d\mathbf{s} \right] \quad (\text{A.11})$$

Dans la première intégrale on fait un changement de variable de telle sorte que : $\mathbf{s}(t) = \mathbf{s} + t\dot{\mathbf{s}}$ où :

$$\dot{\mathbf{s}} = [\dot{\mathbf{s}}_1(x, y), \dot{\mathbf{s}}_2(x, y)]^\top \quad (\text{A.12})$$

Qui donne :

$$\int_{\Omega_{int}(t)} k(\mathbf{s}, \Omega_{int}(t)) d\mathbf{s} = \int_{\Omega_{int}} k(\mathbf{s} + t\mathbf{V}, \Omega_{int} + tV) |\det \mathbf{J}(\mathbf{s})| d\mathbf{s} \quad (\text{A.13})$$

où $\mathbf{J}(\mathbf{s})$ est la matrice jacobienne :

$$\mathbf{J}(\mathbf{s}) = \begin{bmatrix} 1 + t \frac{\partial \dot{\mathbf{s}}_1}{\partial x} & t \frac{\partial \dot{\mathbf{s}}_1}{\partial y} \\ t \frac{\partial \dot{\mathbf{s}}_2}{\partial x} & 1 + t \frac{\partial \dot{\mathbf{s}}_2}{\partial y} \end{bmatrix} \quad (\text{A.14})$$

On a donc :

$$\det \mathbf{J}(\mathbf{s}) = 1 + t \operatorname{div}(\dot{\mathbf{s}}) + \Theta(t) \quad (\text{A.15})$$

Enfin on a :

$$\lim_{t \rightarrow 0} \frac{\det \mathbf{J}(\mathbf{s}) - 1}{t} = \operatorname{div}(\dot{\mathbf{s}}) \quad (\text{A.16})$$

L'équation A.16 devient :

$$\frac{m_{pq}(\Omega_{int}(t)) - m_{pq}(\Omega_{int})}{t} = I_1 + I_2 \quad (\text{A.17})$$

Avec :

$$\lim_{t \rightarrow 0} I_1 = \int_{\Omega_{int}} \dot{k}(\mathbf{s}, \Omega_{int}) d\mathbf{s} = \int_{\Omega_{int}} k'(\mathbf{s}, \Omega_{int}) d\mathbf{s} + \int_{\Omega_{int}} \nabla k(\mathbf{s}, \Omega_{int}) \cdot \dot{\mathbf{s}} d\mathbf{s} \quad (\text{A.18})$$

$$\lim_{t \rightarrow 0} I_2 = \int_{\Omega_{int}} k(\mathbf{s}, \Omega_{int}) \operatorname{div}(\dot{\mathbf{s}}) d\mathbf{s} \quad (\text{A.19})$$

On obtient finalement pour la dérivée Eulérienne :

$$\dot{m}_{pq} = \int_{\Omega_{int}} \operatorname{div}(k(\mathbf{s}, \Omega_{int}) \dot{\mathbf{s}}) d\mathbf{s} \quad (\text{A.20})$$

En développant [A.20](#), nous obtenons finalement

$$\dot{m}_{pq} = \iint_{\Omega_{int}} \left[\frac{\partial k}{\partial x} \dot{x} + \frac{\partial k}{\partial y} \dot{y} + k(x, y) \left(\frac{\partial \dot{x}}{\partial x} + \frac{\partial \dot{y}}{\partial y} \right) \right] dx dy \quad (\text{A.21})$$

Sachant que $k(x, y) = x^p y^q$ et leurs dérivées partielles sont définies par

$$\frac{\partial k}{\partial x} = p x^{p-1} y^q$$

et

$$\frac{\partial k(x, y)}{\partial y} = q x^p y^{q-1}$$

A.3 Étude comparative de performances des détecteurs/descripteurs SIFT et SURF

Dans l'objectif était d'utiliser un ensemble complet de paramètres, permettant de choisir judicieusement parmi les deux détecteurs/descripteurs SIFT et SURF (voir figure A.1), qu'il est le plus approprié pour le cas de l'OV.

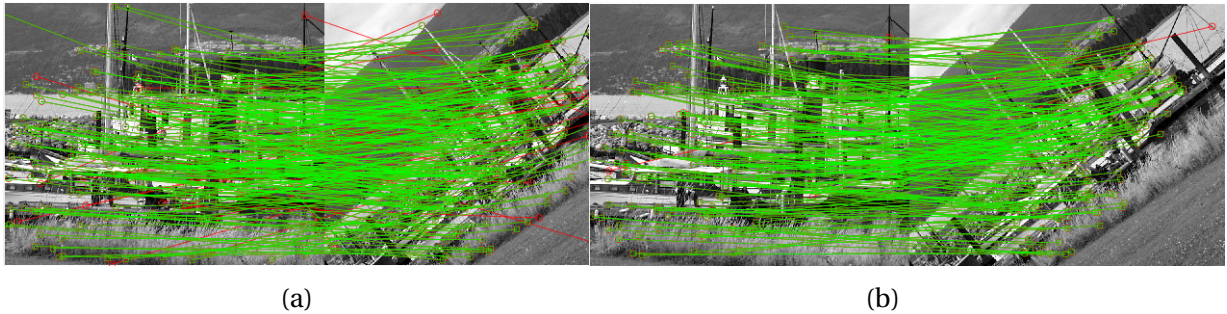


FIGURE A.1 – Detection et description des points d'intérêt par (a) SIFT+RANSAC (b) SURF+RANSAC.

Afin d'évaluer les performances métriques des détecteurs et des descripteurs des points d'intérêt, Mikolajczyk et al. [165] ont proposé un ensemble de trois paramètres à calculer : la répétabilité, le rappel (recall) et la précision. Obtenir des valeurs fortes de ces métriques à une certaine classe de transformations, assure la performance de l'algorithme vis-à-vis les transformations. En effet, ce qui signifie que le détecteur et descripteur contient suffisamment d'information qui lui permette de mettre en correspondance et sans ambiguïté, les bons appariements entre les points d'intérêts détectés.

Les critères d'évaluation dépendent fortement de plusieurs facteurs : la base de donnée d'images, les conditions de visibilité, la relation géométrique entre les scènes à mettre en correspondance. De plus, les critères sont influencés par l'algorithme de mise en correspondance. En se basant sur les évaluations des détecteurs et descripteurs de points d'intérêt dans la littérature, nous avons sélectionné les critères suivantes pour évaluer les détecteurs et descripteurs sélectionnés :

- Nombre de points d'intérêt initialement détectés (après l'application du détecteur sur les deux images).
- Nombre de bons appariements : il s'agit du nombre de points considérés comme bons appariements (inliers), en les vérifiant en présence de la vérité terrain de la transformation présente entre les deux images.
- Répétabilité : elle se calcule par la formule A.22 et elle représente la capacité de

détecter le point d'intérêt dans le cas d'un changement de point de vue dû à un mouvement de la caméra.

- Rappel (Recall) : il se calcule par la formule A.23 en divisant le nombre des bons appariements par le nombre des bonnes correspondances qui existe entre les deux prises de vue total de points d'intérêt mis en correspondance.
- Précision : elle se calcule par la formule A.24 en divisant cette fois ci le nombre des bons appariements par le nombre de mis en correspondance des points d'intérêt retrouvé.

$$\text{Répétabilité} = \frac{\text{Nombre de correspondance}}{\text{Nombre minimum des points d'intérêts détectés (img}_1, \text{img}_2)} \quad (\text{A.22})$$

$$\text{Rappel} = \frac{\text{Nombre des vraies correspondances retrouvées}}{\text{Nombre total des vraies correspondances existantes}} \quad (\text{A.23})$$

$$\text{Precision} = \frac{\text{Nombre des vraies correspondances retrouvées}}{\text{Nombre de mis en correspondance des points d'intérêt retrouvé}} \quad (\text{A.24})$$

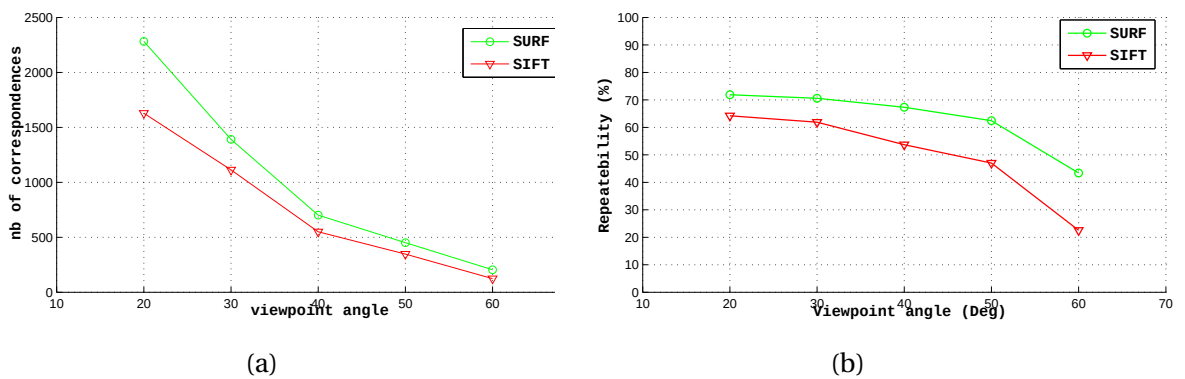


FIGURE A.2 – Évaluation des performances de SIFT et SURF par rapport au changement de l'angle de vue et l'échelle. (a) Nombre de correspondance (b) Répétabilité.

La base de donnée de la scène de bateau a été sollicitée afin d'évaluer le changement d'échelle et les angles de vue, ce qui a produit à chaque image de la base de donnée un changement de nombre de correspondance et les vraies correspondances. Par conséquent on aura différentes valeurs de la précision et du rappel. Le nombre des vraies correspon-

dances retrouvées se vérifie via la vérité terrain, qu'a été fournit sous format de l'homo-graphie qui projette chaque prise de vue à l'image de référence.

Nous avons évalué les performances des deux détecteurs SIFT et SURF. La figure A.2 illustre le nombre de bon appariements et la répétabilité en fonction de changements d'angles de vue et l'échelle par rapport à l'image de référence. Nous en déduisons que plus le changement entre les images est grand, plus les performances de l'extraction et des points d'intérêt sont dégradées. Cependant, le détecteur SURF a montrée des bonnes résultats par rapport à SIFT. Ces résultats confirment que le détecteur SURF est plus stable en changeant l'angle de vue et l'échelle.

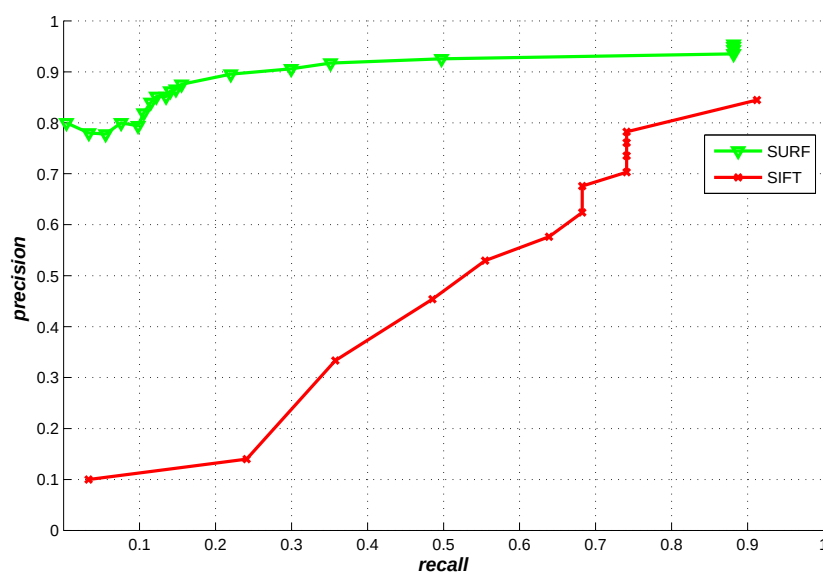


FIGURE A.3 – Les courbes de Recall en fonction de la precision .

Par la suite, les deux descripteurs SIFT et SURF ont été évalué à travers la précision et le recall. Nous rappelons que, une valeur élevée de la précision avec une faible valeur de recall cela peut être expliqué par le faite que le descripteur arrive de trouver des bonnes correspondances, mais beaucoup d'autres n'ont été pas obtenues. Alors que une valeur élevée de recall avec une faible valeur de la précision, peut être expliqué que la plupart des bonnes correspondances ont été trouvées, mais avec une forte présence des mauvais appariements. Pour cela, la situation idéale est que le descripteur fournissait des bonnes valeurs pour les deux métriques simultanément. Or, en traçant les valeurs de la précision en fonction de recall pour les deux descripteurs sur la figue A.3, on peut constater que la description SURF est meilleur que celle de SIFT en présence de changement d'angle de vues et l'échelle.

Nous n'évaluons pas les performances des détecteurs et des descripteurs en terme de temps, car nos implémentations évaluent plutôt la performance des algorithmes et non

leurs rapidités. Cependant, nous prenons en considération cette contrainte (la rapidité) dans le choix des algorithmes. En effet, nous nous sommes basé sur ce critère (la rapidité) qui a été déjà évalué dans l'état de l'art. Les implémentations des deux détecteur et descripteurs SIFT et SURF utilisées ne sont pas optimisées mais restent proches du temps réel. Les tableaux A.1 et A.2 résument le recensement des performances des détecteurs et descripteurs locaux SIFT et SURF. Nous remarquons également que le descripteur SURF dépasse le descripteurs SIFT en termes de temps de calculs et restent comparables par rapport au SIFT en terme de répétabilité, précision et recall.

TABLEAU A.1 – Performances des détecteurs

Descripteurs	Nombre de correspondances	Répétabilité	Temps de calcul
SIFT	++	++	++
SURF	+++	+++	+++

TABLEAU A.2 – Performances des descripteurs

Détecteurs	Précision	Recall	Temps de calcul
SIFT	++	++	+
SURF	++	+++	++

Bibliographie

- [1] Francois Chaumette. Image moments : a general and useful set of features for visual servoing. *IEEE Transactions on Robotics*, 20(4) :713–723, 2004. [xiii](#), [86](#)
- [2] Y. Onoe, N. Yokoya, K. Yamazawa, and H. Takemura. Visual surveillance and monitoring system using an omnidirectional video camera. In *Proceedings. Fourteenth International Conference on Pattern Recognition*, volume 1, pages 588–592 vol.1, 1998. [6](#)
- [3] Multi-view structure-from-motion for hybrid camera scenarios. *Image and Vision Computing*, 30(8) :557 – 572, 2012. [6](#)
- [4] D. Scaramuzza and R. Siegwart. Appearance-guided monocular omnidirectional visual odometry for outdoor ground vehicles. *Robotics, IEEE Transactions*, 2008. [6](#), [25](#)
- [5] A. Alahi, Y. Boursier, L. Jacques, and P. Vanderghelynst. Sport players detection and tracking with a mixed network of planar and omnidirectional cameras. In *2009 Third ACM/IEEE International Conference on Distributed Smart Cameras (ICDSC)*, pages 1–8, 2009. [6](#)
- [6] Y. Yagi, S. Kawato, and S. Tsuji. Real-time omnidirectional image sensor (copis) for vision-guided navigation. *IEEE Transactions on Robotics and Automation*, 10(1) :11–22, 1994. [6](#)
- [7] Tomokazu Sato, Sei Ikeda, and Naokazu Yokoya. Extrinsic camera parameter recovery from multiple image sequences captured by an omni-directional multi-camera system. *Computer Vision-ECCV 2004*, pages 326–340, 2004. [7](#)
- [8] Maxime Meilland, Andrew Ian Comport, and Patrick Rives. Dense visual mapping of large scale environments for real-time localisation. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 4242–4248. IEEE, 2011. [7](#)

- [9] Philipp Koschorrek, Tommaso Piccini, Per Oberg, Michael Felsberg, Lars Nielsen, and Rudolf Mester. A multi-sensor traffic scene dataset with omnidirectional video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 727–734, 2013. [7](#)
- [10] Simon Baker and Shree K Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision*, 35(2) :175–196, 1999. [8](#), [10](#)
- [11] Xianghua Ying and Zhanyi Hu. Can we consider central catadioptric cameras and fisheye cameras within a unified imaging model. *Computer Vision-ECCV 2004*, pages 442–455, 2004. [8](#)
- [12] Christopher Mei and Patrick Rives. Single view point omnidirectional camera calibration from planar grids. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3945–3950. IEEE, 2007. [8](#)
- [13] *Panoramic vision : sensors, theory, and applications*. Springer Verlag, 2001. [8](#)
- [14] R. Descartes and D.T. Smith. The geometry of rené descartes. *Dover Publication*, 1954. [8](#)
- [15] I. GUELZIM. *Contributions aux traitements d'images perspectives et omnidirectionnelles par des outils statiques*. PhD thesis, Université Mohammed V-Agdal, 2012. [8](#), [9](#)
- [16] V. Hlavac, T. Svoboda, and T. Pajdla. Motion estimation using central panoramic cameras. In *International Conference on Intelligent Vehicles*. [9](#)
- [17] S. Bigot-Marchand. *Outils de traitement d'images adaptes au traitement d'images omnidirectionnelles*. PhD thesis, l'Université de Picardie Jules Verne Amiens, 2008. [9](#)
- [18] S. Baker and S. K. Nayar. A theory of single-viewpoint catadioptric image formation. *International Journal of Computer Vision (IJCV)*, 1999. [9](#), [10](#)
- [19] Gregory D Hager, Wen-Chung Chang, and A Stephen Morse. Robot hand-eye coordination based on stereo vision. *IEEE Control Systems Magazine*, 15(1) :30–39, 1995. [9](#)
- [20] S. Baker and S. K. Nayar. A theory of catadioptric image formation. In *International Conference on Computer Vision (ICCV)*. [10](#)
- [21] S.K. Nayar. Catadioptric omnidirectional camera. In *In Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*. [10](#)

- [22] E.M. Mouaddib. Introduction à la vision panoramique catadioptrique. *Traitement du signal*, 2005. [10](#)
- [23] S. Derrien and K. Konolige. Approximating a single viewpoint in panoramic imaging devices. In *IEEE Workshop on Omnidirectional Vision*. [11](#)
- [24] S.Thompson, T. Matsui, and H. Asoh. Mobile robot localization using circular correlations of panoramic images. In *International Conference on Intelligent Robots and Systems IROS2000*. [11](#)
- [25] N. Winters and J. Santos-Victor. Omni-directional visual navigation. In *7th International Symposium on Intelligent Robotic Systems SIRS99*. [11](#)
- [26] Javier Cruz-Mota, Iva Bogdanova, Benoit Paquier, Michel Bierlaire, and Jean-Philippe Thiran. Scale invariant feature transform on the sphere : Theory and applications. *International journal of computer vision*, 98(2) :217–241, 2012. [13](#)
- [27] Cedric Demonceaux, Pascal Vasseur, and Yohan Fougerolle. Central catadioptric image processing with geodesic metric. *Image and Vision Computing*, 29(12) :840–849, 2011. [13](#)
- [28] T. Sato, T. Pajdla, and N. Yokoya. Epipolar geometry estimation for wide-baseline omnidirectional street view images. In *2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 56–63, Nov 2011. [13](#)
- [29] K. Daniilidis and C. Geyer. Omnidirectional vision : Theory and applications. In *Proceedings International Conference on Pattern Recognition*. [13](#)
- [30] E.M. Mouaddib. La vision omnidirectionnelle. In *Journées Nationale de la Recherche en Robotique*. [13](#)
- [31] B. Vandeportaele. *Contributions à la vision omnidirectionnelle : Etude, Conception et Etalonnage de capteurs pour l'acquisition d'images et la modélisation 3D*. PhD thesis, 2006. [13](#)
- [32] Roger Tsai and T Huang. Estimating three-dimensional motion parameters of a rigid planar patch. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 29(6) :1147–1152, 1981. [16](#)
- [33] Zhongfei Zhang and Allen R Hanson. Scaled euclidean 3d reconstruction based on externally uncalibrated cameras. In *Computer Vision, 1995. Proceedings., International Symposium on*, pages 37–42. IEEE, 1995. [16](#)

- [34] Zhengyou Zhang, Rachid Deriche, Olivier Faugeras, and Quang-Tuan Luong. A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry. *Artificial intelligence*, 78(1-2) :87–119, 1995. [17](#)
- [35] Richard Hartley and Rajiv Gupta. Computing matched-epipolar projections. In *Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on*, pages 549–555. IEEE, 1993. [17](#)
- [36] Youcef MEZOUAR. *Habilitation à Diriger des Recherches*. PhD thesis, Université Blaise Pascal, 2009. [18](#)
- [37] Tomáš Svoboda, Tomáš Pajdla, and Václav Hlaváč. Epipolar geometry for panoramic cameras. *Computer Vision—ECCV'98*, pages 218–231, 1998. [18](#)
- [38] O Faugeras. *Panoramic vision : sensors, theory, and applications*. Springer Science & Business Media, 2013. [18](#)
- [39] Tomáš Svoboda and Tomáš Pajdla. Epipolar geometry for central catadioptric cameras. *International Journal of Computer Vision*, 49(1) :23–37, 2002. [18](#)
- [40] D. Nister and al. Visual odometry. In *in Computer Vision and Pattern Recognition, CVPR 2004*. [19](#), [24](#), [25](#), [107](#)
- [41] D. Nister. An efficient solution to the five-point relative pose problem. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2003. [20](#), [21](#)
- [42] H. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. In *Nature*. [20](#), [23](#)
- [43] D. Scaramuzza and F. Fraundorfer. Visual odometry : Part i : The first 30 years and fundamentals. *IEEE ROBOTICS AND AUTOMATION MAGAZINE*, 2011. [21](#), [23](#)
- [44] M. A. Fischler and R. C. Bolles. Random sample consensus : A paradigm for model fitting with applications to image analysis and automated cartography. *Comm. of the ACM*, 1981. [21](#)
- [45] Phil Torr and Andrew Zisserman. Robust computation and parametrization of multiple view relations. In *Computer Vision, 1998. Sixth International Conference on*, pages 727–732. IEEE, 1998. [22](#)
- [46] C. Harris and J. Pike. 3d positional integration from image sequences. In *in Proc. Alvey Vision Conf*. [23](#)

- [47] P. M. Bill Triggs, Richard Hartley, and Andrew Fitzgibbon. Bundle adjustment. In *A Modern Synthesis*. [23](#)
- [48] R. Hartley and A. Zisserman. Multiple view geometry in computer vision. In *Cambridge University Press*. [23](#), [25](#)
- [49] H. Moravec. *Obstacle avoidance and navigation in the real world by a seeing robot rover*. PhD thesis, Stanford Univ, CA, 1980. [23](#), [24](#)
- [50] C. Olson and al. Robust stereo ego-motion for long distance navigation. In *IEEE Conf.Computer Vision and Pattern Recognition*. [23](#), [24](#)
- [51] L. Matthies and S. Shafer. Error modeling in stereo navigation. *IEEE J. Robot. Automat*, 1987. [24](#)
- [52] L. Matthies. *Dynamic stereo vision*. PhD thesis, CarnegieMellon Univ, Pittsburgh, 1989. [24](#)
- [53] C. Olson and al. Rover navigation using stereo ego-motion. In *Robot. Autonom. Syst*, 2003. [24](#)
- [54] A. Milella and R. Siegwart. Stereo-based ego-motion estimation using pixel tracking and iterative closest point. In *IEEE International Conference Computer Vision Systems, ICVS '06*, 2006. [24](#)
- [55] C. T. Jianbo Shi. Good features to track. In *In IEEE International Conference on Computer Vision and Pattern Recognition*. [24](#)
- [56] P. J. Besl and H. D. McKay. A method for registration of 3-d shapes. *IEEE Transactions Pattern Analysis and Machine Intelligence*, 1992. [24](#)
- [57] A. I. Comport and al. Accurate quadrifocal tracking for robust 3d visual odometry. In *IEEE International Conference Robotics and Automation*, 2007. [24](#)
- [58] D. Nister and al. Visual odometry for ground vehicle applications. *J. Field Robot*, 2006. [25](#)
- [59] P. Corke and al. Omnidirectional visual odometry for a planetary rover. In *in Intelligent Robots and Systems, (IROS 2004)*, 2004. [25](#)
- [60] T. K. Bruce and D. Lucas. An iterative image registration technique with an application to stereo vision. In *In International joint conference on Artificial intelligence*. [25](#), [32](#)
- [61] B. Liang and N. Pears. Visual navigation using planar homographies. In *International. Conf. Robotics and Automation (ICRA'02)*, 2002. [25](#)

- [62] Q. Ke and T. Kanade. Transforming camera geometry to a virtual downward-looking camera : Robust ego-motion estimation and groundlayer detection. In *Computer Vision and Pattern Recognition*, 2003. [25](#)
- [63] Z. Zhang. Motion and structure from two perspective views : from essential parameters to euclidean motion through the fundamental matrix. *Journal of the Optical Society of America*, 1997. [25](#)
- [64] T. Svoboda, T. Pajdla, and V. Hlavac. Motion estimation using central panoramic cameras. In *IEEE International Conference on Intelligent Vehicles*. [25](#)
- [65] A. Levin and R. Szeliski. Visual odometry and map correlation. In *CVPR (1)*. [25](#)
- [66] H. Tamimi, H. Andreasson, A. Treptow, T. Duckett, and Zell. Localization of mobile robots with omnidirectional vision using particle filter and iterative sift. In *Selected the 2nd papers from European Conference on Mobile Robots (ECMR '05)*. [25](#)
- [67] H. Hadj-Abdelkader, E. Malis, and P. Rives. Spherical image processing for accurate visual odometry with omnidirectional cameras. In *The 8th Workshop on omnidirectional vision, camera networks and non-classical cameras—OMNIVIS*. [25](#)
- [68] Bogdanova, Bresson, Thiran, and Vanderghelynst. Scale-space analysis and active contours for omnidirectional images. *IEEE Transactions on Image Processing*, 2007. [25](#)
- [69] Michael Bloesch, Sammy Omari, Marco Hutter, and Roland Siegwart. Robust visual inertial odometry using a direct ekf-based approach. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 298–304. IEEE, 2015. [25](#)
- [70] A. Makadia and K. Daniilidis. Rotation recovery from spherical images without correspondences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 2006. [26](#)
- [71] Peter Liepa. Filling holes in meshes. In *Proceedings of the 2003 Eurographics/ACM SIGGRAPH Symposium on Geometry Processing*, pages 200–205. Eurographics Association, 2003. [26](#), [34](#)
- [72] Ameesh Makadia, Lorenzo Sorgi, and Kostas Daniilidis. Rotation estimation from spherical images. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2004. [26](#), [34](#)
- [73] A. Makadia, C. Geyer, and K. Daniilidis. Correspondence-free structure from motion. *International Journal of Computer Vision*, 2007. [26](#)

- [74] Ameesh Makadia, Christopher Geyer, Shankar Sastry, and Kostas Daniilidis. Radon-based structure from motion without correspondences. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2005. [26](#), [34](#)
- [75] Mo Shan, Yingcai Bi, Hailong Qin, Jiaxin Li, Zhi Gao, Feng Lin, and Ben M Chen. A brief survey of visual odometry for micro aerial vehicles. In *Industrial Electronics Society, IECON 2016-42nd Annual Conference of the IEEE*, pages 6049–6054. IEEE, 2016. [26](#)
- [76] Albert S Huang, Abraham Bachrach, Peter Henry, Michael Krainin, Daniel Maturationa, Dieter Fox, and Nicholas Roy. Visual odometry and mapping for autonomous flight using an rgb-d camera. In *International Symposium on Robotics Research (ISRR)*, pages 1–16, 2011. [26](#)
- [77] Christian Kerl, Jürgen Sturm, and Daniel Cremers. Robust odometry estimation for rgb-d cameras. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 3748–3754. IEEE, 2013. [27](#)
- [78] Sergey Alexandrov and Rainer Herpers. Evaluation of recent approaches to visual odometry from rgb-d images. In *Robot Soccer World Cup*, pages 444–455. Springer, 2013. [27](#)
- [79] Ji Zhang, Michael Kaess, and Sanjiv Singh. A real-time method for depth enhanced visual odometry. *Autonomous Robots*, 41(1) :31–43, 2017. [27](#)
- [80] Ji Zhang and Sanjiv Singh. Visual-lidar odometry and mapping : Low-drift, robust, and fast. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 2174–2181. IEEE, 2015. [27](#)
- [81] Ji Zhang and Sanjiv Singh. Visual-inertial combined odometry system for aerial vehicles. *Journal of Field Robotics*, 32(8) :1043–1055, 2015. [27](#)
- [82] Ji Zhang, Michael Kaess, and Sanjiv Singh. Real-time depth enhanced monocular odometry. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 4973–4980. IEEE, 2014. [27](#)
- [83] Zheng Fang and Sebastian Scherer. Experimental study of odometry estimation methods using rgb-d cameras. In *Intelligent Robots and Systems (IROS 2014), 2014 IEEE/RSJ International Conference on*, pages 680–687. IEEE, 2014. [27](#)
- [84] Zheng Fang and Sebastian Scherer. Real-time onboard 6dof localization of an indoor mav in degraded visual environments using a rgb-d camera. In *Robotics and Automation (ICRA), 2015 IEEE International Conference on*, pages 5253–5259. IEEE, 2015. [27](#)

- [85] M. MEILLAND. *Cartographie RGB-D dense pour la localisation visuelle temps-réel et la navigation autonome*. PhD thesis, Paris Tech, École nationale supérieure des mines de Paris, 2012. [28](#)
- [86] M. S. Chris Harris. A combined corner and edge detector. In *In Proceedings of the 4th Alvey Vision Conference*. [28](#)
- [87] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 2004. [29](#)
- [88] T. T. Herbert Bay and L. Van Gool. Surf : Speeded up robust features. In *presented at the Proceedings of the ninth European Conference on Computer Vision*. [29](#)
- [89] R. C. B. Martin and A. Fischler. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. In *ACM Communications*. [29](#)
- [90] Houssem Eddine Benseddik, Oualid Djekoune, and Mahmoud Belhocine. Sift and surf performance evaluation for mobile robot-monocular visual odometry. *Journal of Image and Graphics*, 2(1) :70–76, 2014. [31](#)
- [91] Alexandre Alahi, Raphael Ortiz, and Pierre Vanderghenst. Freak : Fast retina keypoint. In *Computer vision and pattern recognition (CVPR), 2012 IEEE conference on*, pages 510–517. Ieee, 2012. [31](#)
- [92] Stefan Leutenegger, Margarita Chli, and Roland Y Siegwart. Brisk : Binary robust invariant scalable keypoints. In *Computer Vision (ICCV), 2011 IEEE International Conference on*, pages 2548–2555. IEEE, 2011. [31](#)
- [93] Josef Maier, Martin Humenberger, Oliver Zendel, and Markus Vincze. Ground truth accuracy and performance of the matching pipeline. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 29–39, 2017. [31](#)
- [94] S. B. Christopher Mei, Ezio Malis, and Patrick Rives. Constrained multiple planar template tracking for central catadioptric cameras. In *In British Machine Vision Conference*. [32](#)
- [95] C. T. Jianbo Shi. Good features to track. In *In IEEE International Conference on Computer Vision and Pattern Recognition*. [32](#)
- [96] E. M. Guillaume Caron and E.M Mouaddib. Tracking planes in omnidirectional stereovision. In *In IEEE International Conference on Robotics and Automation*. [32](#)

- [97] C. Geyer, S. Sastry, and R. Bajcsy. Euclid meets fourier. In *In Workshop on Omnidirectional Vision, Prague*. [33](#)
- [98] E. De Castro and C. Morandi. Registration of translated and rotated images using finite fourier transforms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, PAMI-9(5) :700–703, 1987. [38](#)
- [99] Ffts on the rotation group], author = Kostelec, PeterJ. and Rockmore, DanielN., pages = 145–179. *Journal of Fourier Analysis and Applications*, 14(2), 2008. [38](#)
- [100] A. Makadia and K. Daniilidis. Rotation recovery from spherical images without correspondences. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(7) :1170–1175, 2006. [38](#), [75](#)
- [101] L. E. Falcon and E. Bayro-Corrochano. Radon transform and harmonical analysis using lines for 3d rotation estimation without correspondences from omnidirectional vision. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–6, Oct 2007. [38](#)
- [102] G. Burel and H. Henoco. Determination of the orientation of 3d objects using spherical harmonics. *Graphical Models and Image Processing*, 57(5) :400–408, 1995. [39](#), [52](#)
- [103] R.C. Hoover, A.A. Maciejewski, and R.G. Roberts. Eigendecomposition of images correlated on $s(1)$, $s(2)$, and $so(3)$ using spectral theory. *Image Processing, IEEE Transactions on*, 18(11) :2562–2571, Nov 2009. [39](#)
- [104] Li Shen and Fillia Makedon. Spherical mapping for processing of 3d closed surfaces. *Image and Vision Computing*, 24(7) :743–761, 2006. [39](#), [53](#), [54](#)
- [105] Heng Huang, Li Shen, Rong Zhang, F. Makedon, A.. Saykin, and J. Pearlman. A novel surface registration algorithm with biomedical modeling applications. *Information Technology in Biomedicine, IEEE Transactions on*, 11(4) :474–482, July 2007. [39](#), [54](#)
- [106] M. Kazhdan. An approximate and efficient method for optimal rotation alignment of 3d models. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 29(7) :1221–1229, 2007. [39](#), [75](#)
- [107] Ch. Brechbuhler, G. Gerig, and O. Kubler. Parametrization of closed surfaces for 3-d shape description. *Comput. Vis. Image Underst.*, 61(2), 1995. [39](#), [53](#)
- [108] M. Mousa, R. Chaine, and S. Akkouche. Frequency-based representation of 3d point-based surfaces using spherical harmonics. *Machine Graphics and Vision International Journ*, 15(3) :537–546, jan 2006. [39](#), [54](#)

- [109] Li Shen, Heng Huang, F. Makedon, and A.J. Saykin. Efficient registration of 3d spharm surfaces. In *Computer and Robot Vision, 2007. CRV '07. Fourth Canadian Conference on*, pages 81–88, May 2007. [39](#)
- [110] P. R. Osteen, J. L. Owens, and C. C. Kessens. Online egomotion estimation of rgb-d sensors using spherical harmonics. In *Robotics and Automation (ICRA), 2012 IEEE International Conference on*, pages 1679–1684, May 2012. [39](#)
- [111] S. Althloothi, M.H. Mahoor, and R.M. Voyles. A robust method for rotation estimation using spherical harmonics representation. *Image Processing, IEEE Transactions on*, 22 : "2306–2316, 2013. [39](#), [75](#)
- [112] MiguelAngel Garcia, BorisXavier Vintimilla, and AngelDomingo Sappa. Approximation and processing of intensity images with discontinuity-preserving adaptive triangular meshes. In *Computer Vision - ECCV 2000*, volume 1842, pages 844–855. 2000. [39](#)
- [113] Atul Sajjanhar, Guojun Lu, Dengsheng Zhang, Jingyu Hou, and Yi-PingPhoebe Chen. Spherical harmonics and distance transform for image representation and retrieval. In *Intelligent Data Engineering and Automated Learning - IDEAL 2009*, volume 5788, pages 309–316. Springer Berlin Heidelberg, 2009. [39](#), [40](#)
- [114] Mathieu Gauthier. Images géométriques de genre arbitraire dans le domaine sphérique. [47](#)
- [115] Paulette Lieby, Nick Barnes, and Brendan D. McKay. Topological repair on voxel-based quadrangular meshes. In *Workshop on Mathematical Foundations of Computational Anatomy : Geometrical, Statistical and Registration Methods for Modeling Biological Shape Variability*, pages 146–155, Oct 2006. [51](#)
- [116] Qianqian Fang and David A. Boas. Tetrahedral mesh generation from volumetric binary and gray-scale images. In *Proceedings of the Sixth IEEE International Conference on Symposium on Biomedical Imaging : From Nano to Macro*, pages 1142–1145, 2009. [51](#)
- [117] Kun Zhou, Hujun Bao, and Jiaoying Shi. 3d surface filtering using spherical harmonics. *Computer-Aided Design*, 36(4) :363 – 375, 2004. [53](#)
- [118] A. Sheffer, C. Gotsman, and N. Dyn. *Robust Spherical Parameterization of Triangular Meshes*, pages 185–193. Springer Vienna, 2004. [53](#)
- [119] Emil Praun and Hugues Hoppe. Spherical parametrization and remeshing. *ACM Trans. Graph.*, 22(3) :340–349, jul 2003. [53](#)

- [120] Martin Styner, Ipek Oguz, Shun Xu, Christian Brechbuhler, Dimitrios Pantazis, James Levitt, MarthaE Shenton, and Guido Gerig. Framework for the statistical shape analysis of brain structures using spharm-pdm. 1071 :242–50, 10 2006. [54](#)
- [121] M. Irani, B. Rousso, and S. Peleg. Recovery of ego-motion using region alignment. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1997. [60](#)
- [122] P.J. Besl and Neil D. McKay. A method for registration of 3-d shapes. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 14(2) :239–256, 1992. [62](#)
- [123] Christopher Geyer and Kostas Daniilidis. Catadioptric projective geometry. *International Journal of Computer Vision*, 45(3) :223–243, 2001. [64](#)
- [124] A. Makadia, L. Sorgi, and K. Daniilidis. Rotation estimation from spherical images. In *Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on*, volume 3, pages 590–593 Vol.3, 2004. [67](#), [69](#)
- [125] C. Mei and Patrick Rives. Single view point omnidirectional camera calibration from planar grids. In *Robotics and Automation, 2007 IEEE International Conference on*, pages 3945–3950, April 2007. [70](#)
- [126] H.H. Abdelkader, Y. Mezouar, N. Andreff, and P. Martinet. 2 1/2 d visual servoing with central catadioptric cameras. In *Intelligent Robots and Systems, 2005. (IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 3572–3577, 2005. [71](#)
- [127] Ming-Kuei Hu. Visual pattern recognition by moment invariants. *Information Theory, IRE Transactions on*, 8(2) :179–187, February 1962. [78](#)
- [128] Mihran Tuceryan. Moment-based texture segmentation. *Pattern Recognition Letters*, 15(7) :659 – 668, 1994. [78](#)
- [129] J. Flusser and T. Suk. A moment-based approach to registration of images with affine geometric distortion. *IEEE Transactions on Geoscience and Remote Sensing*, 32(2) :382–387, Mar 1994. [78](#)
- [130] Xiaolong Dai and Siamak Khorram. A feature-based image registration algorithm using improved chain-code representation combined with invariant moments. *IEEE Transactions on Geoscience and Remote Sensing*, 37(5) :2351–2362, 1999. [78](#)
- [131] Sugata Ghosal and Rajiv Mehrotra. Orthogonal moment operators for subpixel edge detection. *Pattern recognition*, 26(2) :295–306, 1993. [78](#)
- [132] TJ Bin, Ao Lei, Cui Jiwen, Kang Wenjing, and Liu Dandan. Subpixel edge location based on orthogonal fourier–mellin moments. *Image and Vision Computing*, 26(4) :563–569, 2008. [78](#)

- [133] W. S. Lang, N. A. Abu, and H. Rahmalan. Fast 4x4 tchebichef moment image compression. In *2009 International Conference of Soft Computing and Pattern Recognition*, pages 295–300. IEEE, 2009. [78](#)
- [134] Zexuan Ji, Qiang Chen, Quan-Sen Sun, and De-Shen Xia. A moment-based nonlocal-means algorithm for image denoising. *Information Processing Letters*, 109(23-24) :1238–1244, 2009. [78](#)
- [135] Jiahong Yin, Alvaro Rodolfo De Pierro, and Musheng Wei. Analysis for the reconstruction of a noisy signal based on orthogonal moments. *Applied Mathematics and Computation*, 132(2) :249–263, 2002. [78](#)
- [136] Chong-Yaw Wee, Raveendran Paramesran, Ramakrishnan Mukundan, and Xudong Jiang. Image quality assessment by discrete orthogonal moments. *Pattern Recognition*, 43(12) :4055–4068, 2010. [78](#)
- [137] Kim-Han Thung, Raveendran Paramesran, and Chern-Loon Lim. Content-based image quality metric using similarity measure of moment vectors. *Pattern Recognition*, 45(6) :2193–2204, 2012. [78](#)
- [138] Sahibsingh A Dudani, Kenneth J Breeding, and Robert B McGhee. Aircraft identification by moment invariants. *IEEE transactions on computers*, 100(1) :39–46, 1977. [78](#)
- [139] Yaser S Abu-Mostafa and Demetri Psaltis. Recognitive aspects of moment invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6) :698–706, 1984. [78](#)
- [140] Zheng-Jun Liu, Qi Li, Zhi-Wei Xia, and Qi Wang. Target recognition of ladar range images using even-order zernike moments. *Applied optics*, 51(31) :7529–7536, 2012. [78](#)
- [141] Feng Zhang, Shang-qian Liu, Da-bao Wang, and Wei Guan. Aircraft recognition in infrared image using wavelet moment invariants. *Image and Vision Computing*, 27(4) :313–318, 2009. [78](#)
- [142] Li Ma, Tieniu Tan, Yunhong Wang, and Dexin Zhang. Local intensity variation analysis for iris recognition. *Pattern recognition*, 37(6) :1287–1298, 2004. [78](#)
- [143] Seyed Jabbar Hosaini, Shahpour Alirezaee, Majid Ahmadi, and Seyed Vahab-Al Din Makki. Comparison of the legendre, zernike and pseudo-zernike moments for feature extraction in iris recognition. In *Computational Intelligence and Communication Networks (CICN), 2013 5th International Conference on*, pages 225–228. IEEE, 2013. [78](#)

- [144] S Padam Priyal and Prabin Kumar Bora. A robust static hand gesture recognition system using geometry based normalizations and krawtchouk moments. *Pattern Recognition*, 46(8) :2202–2219, 2013. [78](#)
- [145] Gholamreza Amayeh, George Bebis, Ali Erol, and Mircea Nicolescu. Peg-free hand shape verification using high order zernike moments. In *Computer Vision and Pattern Recognition Workshop, 2006. CVPRW'06. Conference on*, pages 40–40. IEEE, 2006. [78](#)
- [146] Seyed Mehdi Lajevardi and Zahir M Hussain. Higher order orthogonal moments for invariant facial expression recognition. *Digital Signal Processing*, 20(6) :1771–1779, 2010. [78](#)
- [147] Vassilis G Kaburlasos, Stelios E Papadakis, and George A Papakostas. Lattice computing extension of the fam neural classifier for human facial expression recognition. *IEEE Transactions on Neural Networks and Learning Systems*, 24(10) :1526–1538, 2013. [78](#)
- [148] Hao Yan, Zhu Zhenwen, and Lu Jianfeng. Human action classification based on the 3d zernike moment. In *Measuring Technology and Mechatronics Automation (ICMTMA), 2014 Sixth International Conference on*, pages 305–308. IEEE, 2014. [78](#)
- [149] Zafer Iscan, Zümray Dokur, and Tamer Ölmez. Tumor detection by using zernike moments on segmented magnetic resonance brain images. *Expert Systems with Applications*, 37(3) :2540–2549, 2010. [78](#)
- [150] François Chaumette. Image moments : a general and useful set of features for visual servoing. *IEEE Transactions on Robotics*, 20(4) :713–723, 2004. [78](#)
- [151] Omar Tahri and Francois Chaumette. Point-based and region-based image moments for visual servoing of planar objects. *IEEE Transactions on Robotics*, 21(6) :1116–1127, 2005. [78](#), [79](#)
- [152] O. Tahri, Y. Mezouar, F. Chaumette, and P. Corke. Decoupled image-based visual servoing for cameras obeying the unified projection model. *IEEE Transactions on Robotics*, 26(4) :684–697, Aug 2010. [79](#)
- [153] O. Tahri and F. Chaumette. Complex objects pose estimation based on image moment invariants. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 436–441, April 2005. [79](#)
- [154] O. Tahri, A. Y. Tamtsia, Y. Mezouar, and C. Demonceaux. Visual servoing based on shifted moments. *IEEE Transactions on Robotics*, 31(3) :798–804, June 2015. [79](#)

- [155] M. Bakthavatchalam, O. Tahri, and F. Chaumette. Improving moments-based visual servoing with tunable visual features. In *2014 IEEE International Conference on Robotics and Automation (ICRA)*, pages 6186–6191, May 2014. [79](#)
- [156] Manikandan Bakthavatchalam, François Chaumette, and Eric Marchand. Photometric moments : New promising candidates for visual servoing. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 5241–5246. IEEE, 2013. [79](#)
- [157] François Chaumette. La relation vision-commande : théorie et application à des tâches robotiques. 01 1990. [84](#), [86](#)
- [158] Aurélien Yeremou Tamtsia. Nouvelles contributions à l’application des moments en asservissement visuel, 2013. [85](#)
- [159] François Chaumette. Avancées récentes en asservissement visuel. In *Journée Nationales de la Recherche en Robotique, JNRR’03*, pages 103–108, Clermont-Ferrand, France, France, 2003. [85](#)
- [160] Stéphanie Jehan-Besson, Michel Barlaud, and Gilles Aubert. Dream2s : Deformable regions driven by an eulerian accurate minimization method for image and video segmentation. (53) :45–70, 2003. [85](#)
- [161] Bernard Espiau, François Chaumette, and Patrick Rives. A new approach to visual servoing in robotics. 8 :313–326, 07 1992. [86](#)
- [162] Omar Tahri. Utilisation des moments invariants en asservissement visuel et en calcul de pose, 03 2004. [89](#), [90](#)
- [163] K. Daniilidis, A. Makadia, and T. Bulow. Image processing in catadioptric planes : spatiotemporal derivatives and optical flow computation. In *Omnidirectional Vision, 2002. Proceedings. Third Workshop on*, pages 3–10, 2002. [98](#)
- [164] F. Morbidi and G. Caron. Phase correlation for dense visual compass from omnidirectional camera-robot images. *IEEE Robotics and Automation Letters*, 2(2) :688–695, April 2017. [104](#), [119](#), [120](#), [122](#)
- [165] Krystian Mikolajczyk, Tinne Tuytelaars, Cordelia Schmid, Andrew Zisserman, Jiri Matas, Frederik Schaffalitzky, Timor Kadir, and Luc Van Gool. A comparison of affine region detectors. *International journal of computer vision*, 65(1-2) :43–72, 2005. [134](#)

Titre : Estimation de la rotation 3D : Approches denses et génériques basées images sphériques

Mots clés : Approches denses, estimation de la rotation 3D, image sphérique, représentation SPHARM, harmoniques sphériques, moments sphériques.

Résumé : Cette thèse traite du problème de la localisation et de l'estimation du mouvement de rotation dans le contexte de l'analyse d'images sphériques. Plus précisément, le but de ce manuscrit est le développement d'algorithmes génériques d'estimation de la rotation 3D entre les images issues de caméras monoculaires conventionnelles et non-conventionnelles (catadioptrique, fisheye, ...). Les contributions principales de la thèse reposent sur l'exploitation de la riche information d'intensité lumineuse dans l'image, notamment omnidirectionnelle, pour accomplir l'estimation des rotations 3D, y compris les larges. Une première contribution de la thèse est la proposition d'une nouvelle représentation de l'intensité lumineuse de l'image sous forme d'un objet 3D de topologie sphérique. Le calcul de coefficients de la transformée en harmoniques sphériques directement à partir du maillage de

l'objet, est accompagné de la décomposition SVD, permettant d'estimer la rotation qui lie les deux images. La deuxième contribution de la thèse est une solution analytique d'estimation de la rotation reposant sur une stratégie dense et indirecte de capture de l'intensité lumineuse de l'image, par le biais de moments photométriques sphériques. Le formalisme théorique établi vise à définir une modélisation analytique de la matrice d'interaction relative aux moments photométriques. En utilisant les propriétés de cette solution, nous proposons trois approches de l'estimation de la rotation 3D entre deux images sphériques. L'étude de performances menée montrent que nos approches permettant d'estimer efficacement la rotation 3D de l'image sphérique. Pour finir, plusieurs scénarios d'expérimentations sont proposés et effectués. Ces derniers incluent les environnements intérieurs et extérieurs.

Title : 3D rotation estimation: Dense and generic approaches based on spherical images

Keywords : Dense approaches, 3D rotation estimation, spherical image, SPHARM representation, spherical harmonics, spherical moments.

Abstract : This thesis deals with the problem of location and motion estimation in the context of spherical image analysis. More specifically, the purpose of this manuscript is the development of generic algorithms for estimating 3D rotation between images captured by conventional and non-conventional (catadioptric, fisheye, ...) monocular camera. The main contributions of the thesis are based on the exploitation of the rich intensity information available in the image, notably omnidirectional, to accomplish the estimation of the 3D rotation, including the large ones. A first contribution is to propose a new representation for image intensity as a 3D object of spherical topology. The coefficients of the spherical harmonics transform computation, directly from the object mesh, is accompanied by the SVD

decomposition, allowing to estimate the rotation which links the two images. The second contribution of the thesis is a closed form solution for estimating the rotation based on a dense and indirect strategy of perceiving the image intensity, which may be achieved by spherical photometric moments. The established theoretical formalism opts to define an analytical modeling of the interaction matrix relative to the photometric moments. Using the properties of this solution, we propose three approaches to estimate the 3D rotation between two spherical images. The conducted performance study show that our approaches allow to efficiently determine 3D rotation of the spherical image. Finally, several experimental scenarios are proposed and evaluated. This included the indoor and outdoor environments.