



Contributions à la Vision par Ordinateur pour les Systèmes en Lumière Structurée et les Systèmes Catadioptriques

David Fofi

► To cite this version:

David Fofi. Contributions à la Vision par Ordinateur pour les Systèmes en Lumière Structurée et les Systèmes Catadioptriques. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université de Bourgogne, 2008. tel-00950264v2

HAL Id: tel-00950264

<https://theses.hal.science/tel-00950264v2>

Submitted on 3 Mar 2014

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.

Synthèse des travaux scientifiques présentée par

David FOFI

pour obtenir le diplôme d'

Habilitation à Diriger des Recherches

Discipline Sciences et Techniques

**CONTRIBUTIONS À LA VISION PAR ORDINATEUR
POUR LES SYSTÈMES EN LUMIÈRE STRUCTURÉE
ET LES SYSTÈMES CATADIOPTRIQUES**

Soutenue publiquement le 4 décembre 2008

JURY

Michel DEVY	Directeur de Recherche	LAAS CNRS	Rapporteur
Olivier STRAUSS	Maître de Conférences HDR	Université de Montpellier II	Rapporteur
Peter STURM	Directeur de Recherche	INRIA Rhône-Alpes	Rapporteur
El Mustapha MOUADDIB	Professeur des Universités	Université de Picardie Jules Verne	Examineur
Fabrice MÉRIAudeau	Professeur des Universités	Université de Bourgogne	Examineur

Avant-Propos

Ce document présente le bilan des travaux de recherche que j’ai effectués de septembre 1997 à l’été 2008, d’abord au sein des équipes *Perception en Robotique* du CREA (Centre de Robotique, Electrotechnique et Automatique¹ – Université de Picardie Jules Verne) et *Mr3D* du VICOROB (Computer Vision and Robotics Group – Universitat de Girona, Espagne) en tant que doctorant, puis au sein de la cellule *Vision 3D* du Le2i (Laboratoire Electronique, Informatique et Image UMR CNRS 5158 – Université de Bourgogne), depuis mon recrutement en tant que Maître de Conférences en septembre 2002.

Ce mémoire est divisé en deux parties. La première décrit les activités de recherche que j’ai menées et tente une synthèse de celles-ci : elle est divisée en trois chapitres thématiques. La seconde apporte des informations plus administratives, telles que le nombre et la nature des encadrements, les activités annexes à la recherche et le transfert de technologies, les publications.

Mes travaux de recherche concernent essentiellement la vision par ordinateur, ou vision artificielle. Basiquement, je me suis efforcé d’imaginer des dispositifs, d’étudier des algorithmes, d’intégrer des méthodes et techniques connues dans des méthodologies nouvelles, de développer çà et là des aspects théoriques originaux. On se rendra compte à la lecture du document que je me suis beaucoup intéressé à des systèmes de vision alternatifs comme les systèmes en lumière structurée ou catadioptriques. La plus grande partie de ce mémoire leur sera dédiée. Ces systèmes permettent d’étudier les techniques usuelles de vision par ordinateur sous un éclairage différent, ils nous obligent à ajuster le problème aux caractéristiques qui leur sont propres ; ils permettent, en quelque sorte, d’appréhender la vision par ordinateur “de biais”. J’évoquerai également des travaux plus marginaux ou plus récents sur la chirurgie virtuelle et la reconstruction d’objets transparents qui, chacun à leur manière, viendront compléter le cadre de ce que sont mes activités de recherche. Ce qui les ont animées tient en une phrase : comment passer d’une image à sa représentation tridimensionnelle ? – et en corollaire : quelles sont les informations dont j’ai besoin pour y parvenir ? comment adapter le capteur, le principe ou la méthode à l’application et comment adapter les traitements au capteur ? Ceci m’a conduit à étudier, en amont, le traitement des images et à m’aventurer parfois dans des domaines qui vont au-delà de mon champ de compétence, comme celui de l’imagerie polarimétrique ou de la physique.

Le premier chapitre traitera des activités de recherche que j’ai menées et encadrées sur la vision en lumière structurée (systèmes projecteur-caméra et laser-caméra). Les thématiques qui y sont abordées sont essentiellement liées à la géométrie du système : modélisation, calibrage, auto-calibrage et recalage. Des aspects propres à la lumière structurée, comme la codification ou l’étude des techniques de lumière structurée imperceptible, sont également évoqués.

Le deuxième chapitre décrira les recherches que nous avons conduites autour de la vision catadioptrique, dans le cadre du projet CaViAR (*Catadioptric Vision for Aerial Robots*) soutenu par l’ANR. Nous décrirons une chaîne de traitements et propositions, encore en cours d’étude et de développement, allant du traitement des images à la reconstruction 3D. L’apport de l’imagerie polarimétrique y est abondamment souligné.

Le troisième chapitre présentera une application et thématique de recherche n’entrant pas dans

1. Devenu MIS - Modélisation, Information et Systèmes - depuis janvier 2008.

les deux axes de recherche donnant son titre au mémoire. J'y décris une application de la reconstruction 3D à partir de radiographies pour l'assistance à la chirurgie dans le cadre d'une opération de placement de prothèse pour l'épaule. Je clorai cette partie scientifique par une conclusion générale, comprenant quelques pistes de réflexion sur nos travaux futurs et une brève description de nos travaux nouveaux, notamment sur la reconstruction d'objets transparents.

Le métier d'enseignant-chercheur n'étant pas fait que de recherche, j'ai essayé, en quelques pages, de donner un aperçu des activités d'enseignement, d'administration, de valorisation qui m'ont tout autant occupées au cours de ces années. Parmi elles, j'aimerais souligner tout particulièrement notre implication dans le master VIBOT (*Erasmus Mundus Masters in Vision and roBOTics*), qui est à l'interface de nos activités d'enseignement et de recherche.

Table des matières

Table des matières	3
Table des figures	6
Liste des tableaux	9
I Synthèse des Travaux Scientifiques	11
1 Vision en Lumière Structurée	13
1.1 Lumière Structurée et Codée	14
1.2 Auto-Calibrage d'un Système Projecteur-Caméra	15
1.2.1 Etat de l'Art	15
1.2.2 Reconstruction Projective	17
1.2.2.1 Méthode Linéaire	18
1.2.2.2 Méthode Itérative	18
1.2.3 Reconstruction Métrique par Génération de Contraintes	19
1.2.3.1 Détection de Plans dans l'Espace	19
1.2.3.2 Contraintes du Parallélogramme	20
1.2.3.3 Contraintes d'Alignement	21
1.2.3.4 Contraintes d'Orthogonalité	21
1.2.3.5 Résultats Expérimentaux	22
1.2.4 Conclusions	24
1.3 Scanner 3D à Main	24
1.3.1 Classification et Nouvelle Méthode de Minimisation	25
1.3.1.1 Classification et Comparaison des Méthodes de Recalage	25
1.3.1.2 Stratégie pour Minimiser la Propagation de l'Erreur	25
1.3.2 Calibrage du Système Laser-Caméra	34
1.3.3 Segmentation et Identification des Lignes Laser	35
1.3.4 Résultats Expérimentaux	36
1.4 Dispositif de Texturation Virtuelle	40
1.4.1 Système Projecteur-Caméra	40
1.4.1.1 Géométrie du Système	44
1.4.1.2 Conception d'une Mire Codée	46
1.4.1.3 Mise en Correspondance	46
1.4.1.4 Lumière Structurée Imperceptible	47
1.4.2 Recalage et Synthèse de Vue	49
1.4.2.1 Recalage par Analyse en Composantes Principales	49
1.4.2.2 Recalage par ICP	50
1.4.2.3 Synthèse de Vue	51
1.4.3 Résultats Expérimentaux	51
1.5 Conclusions	55

2	Vision Catadioptrique	57
2.1	Introduction à la Vision Catadioptrique	57
2.2	Calibrage par Imagerie Polarimétrique	59
2.2.1	Modèle Générique de Caméra	60
2.2.2	Notions d'Imagerie Polarimétrique	60
2.2.3	Calibrage et Reconstruction	62
2.2.3.1	Estimation des Lignes de Vue	62
2.2.3.2	Triangulation	63
2.2.4	Résultats Expérimentaux	65
2.3	Traitement Adapté des Images Catadioptriques	70
2.3.1	Définition du Voisinage : Cas Paramétrique	70
2.3.2	Définition du Voisinage : Cas Non-Paramétrique	71
2.3.3	Dérivation d'Opérateur de Traitement des Images	72
2.3.4	Résultats Expérimentaux	73
2.4	Mise en Correspondance de Droites Catadioptriques	74
2.4.1	Relation entre les Droites	75
2.4.2	Algorithme de Mise en Correspondance	76
2.4.2.1	Extraction des Lignes, Intersections et Angles	77
2.4.2.2	Corrélation au Voisinage des Intersections	77
2.4.2.3	Calcul de la Rotation	78
2.4.3	Rotation et Translation Longue	79
2.4.4	Résultats Expérimentaux	79
2.4.4.1	Données Synthétiques	80
2.4.4.2	Données Réelles	81
2.5	Conclusions	82
3	Une Application à la Chirurgie Virtuelle	85
3.1	Contexte	85
3.2	De l'Epaule au Modèle	85
3.2.1	Anatomie et Biomécanique de l'Epaule	86
3.2.2	La Prothèse de l'Epaule	86
3.2.3	Technique Opératoire (Prothèse Inversée)	87
3.3	Quel Modèle ?	90
3.4	Maquette Numérique et Chirurgie Virtuelle	92
3.4.1	Modélisation Osseuse	92
3.4.2	Extraction des Paramètres	93
3.4.3	Simulation Chirurgicale Pré-Opératoire	96
3.5	Simulation de Débâtements	98
3.6	Conclusions	101
4	Conclusions et Perspectives	105
4.1	Contributions	105
4.2	Développement des Thématiques	105
4.2.1	Vision en Lumière Structurée	105
4.2.2	Vision Catadioptrique	106
4.2.3	Reconstruction d'Objets Transparents	107
4.3	Collaborations Présentes et Futures	107
4.4	De la Recherche à l'Enseignement	108
II	Curriculum Vitae	109
5	Encadrement de Thèses et de Stage	111
5.1	Post-doctorat	111

5.2	Doctorat	111
5.3	Master	111
6	Responsabilités, Enseignement, Animation et Valorisation	115
6.1	Responsabilités Scientifiques	115
6.2	Autres Responsabilités	116
6.3	Relectures et Evaluations	117
6.4	Activités Pédagogiques	118
6.5	Valorisation et Transfert	119
7	Bibliographie de l'Auteur	121
7.1	Articles de Revue	121
7.2	Communications Internationales	122
7.3	Communications Nationales	124
7.4	Editions	124
	Bibliographie	125

Table des figures

1.1	Schéma général pour la vision en lumière structurée	13
1.2	La lumière structurée	14
1.3	Classification de la codification des lumières structurées	16
1.4	Scène éclairée par le projecteur	17
1.5	Retroprojection de la reconstruction obtenue par méthode linéaire. A gauche : le plan image, à droite : le plan projecteur. Les cercles désignent les points originaux et les croix, les points retroprojetés.	18
1.6	Retroprojection de la reconstruction obtenue par méthode itérative. A gauche : le plan image, à droite : le plan projecteur. Les cercles désignent les points originaux et les croix, les points retroprojetés.	19
1.7	Algorithme général de la méthode	20
1.8	Test de coplanarité. A gauche : plan image. A droite : patron de lumière	20
1.9	Contraintes du parallélogramme	21
1.10	Contrainte d'orthogonalité	22
1.11	Reconstruction 3D - Scène I	23
1.12	Scène II. Les contraintes choisies sont en surbrillance	23
1.13	Reconstruction 3D - Scène II	24
1.14	Classification des méthodes de recalage	26
1.15	Recalage par paire de vues sur des données réelles	27
1.16	Recalage peu précis sur des données réelles présentant peu de détails de forme	27
1.17	Recalage grossier par la méthode RANSAC	28
1.18	Influence sur l'erreur du nombre de points considérés par la méthode de <i>Spin Image</i>	28
1.19	Diagramme de flux de l'approche de minimisation	30
1.20	Exemples d'échantillonnage. (a) Surface initiale. (b) Echantillonnage aléatoire. (c) Echantillonnage normal	31
1.21	Calcul de l'intersection entre le plan tangent et la surface par la méthode de PARK	32
1.22	Recouvrement et boîtes englobantes	32
1.23	Modèle de Chen & Kak	34
1.24	Birapport et quadrangle complet utilisés pour faciliter la mise en correspondance 2D/3D	35
1.25	Génération de points supplémentaires pour améliorer la précision du calibrage	35
1.26	Identification des lignes laser. A gauche : image originale. A droite : image étiquetée	37
1.27	Exemple de d'algorithme d'identification	37
1.28	Premier prototype du scanner à main	37
1.29	Précision en fonction de la distance	38
1.30	Exemples d'indexation de lignes laser	39
1.31	Principe et résultats sur le modèle simulé de l'INRIA	39
1.32	Recalage complet du modèle Beethoven	40
1.33	Exemple de résultats	41
1.34	Résultats de recalage. (a) Notre méthode. (b) Méthode de SHARP. (c) Alignement mécanique. (d) Objet	41
1.35	Séquence de vues partielles	42
1.36	Polychromie de la Cathédrale d'Amiens	43

1.37	Texturation du Taj-Mahal par RASKAR	43
1.38	Création de la mire binaire	46
1.39	Mire pseudo-aléatoire et vérification de l'unicité	47
1.40	Exemple de mises en correspondance	47
1.41	Principe de la lumière structurée imperceptible	48
1.42	Séquence pour obtenir une valeur RVB de 223, 47 et 128. Le miroir peut-être actif (gris clair), inactif (noir) ou indéterminé 'gris foncé)	48
1.43	Séquences obtenues pour les 256 valeurs de rouge	49
1.44	Ajustement des pixels pendant la BIEP	49
1.45	Les modèles tridimensionnels utilisés	51
1.46	Objet avant et après texturation	52
1.47	Changement de texture	52
1.48	Précision de la texturation	53
1.49	Autres modèles texturés	53
1.50	Séquence de texturation	54
2.1	Principe de la vision omnidirectionnelle catadioptrique	58
2.2	Modèle de la sphère équivalente	58
2.3	Principe de la projection stéréographique. De haut en bas : miroir parabolique, miroir hyperboloïde et miroir elliptique	59
2.4	Principe de l'imagerie polarimétrique et configuration	61
2.5	Image des paramètres de polarisation. (a) Degré de polarisation. (b) Angle de polarisation	61
2.6	Variation de l'intensité lumineuse en fonction de l'angle du polariseur	61
2.7	Ambiguïté sur l'angle de polarisation. (a) Masque utilisé. (b) Image résultat	62
2.8	Principe de la triangulation	63
2.9	Carte d'erreur de la reconstruction d'un miroir hyperbolique	65
2.10	Erreurs de mesure des paramètres 3D. (a) L'angle θ . (b) L'angle ϕ . (c) Carte d'erreur	66
2.11	Reconstruction 3D en simulation	66
2.12	Erreur de reconstruction induite par des mesures bruitées de la normale. (a) L'angle θ . (b) L'angle ϕ	67
2.13	Paramètres pris en compte. (a) La normale et la profondeur du miroir. (b) La normale uniquement	67
2.14	Mise en correspondance manuelle	68
2.15	Reconstruction à partir d'un miroir hyperbolique	68
2.16	Reconstruction à partir d'un miroir sphérique	69
2.17	Reconstruction à partir de trois vues d'un miroir sphérique	69
2.18	Exemple de paramétrisation dans le cas orthographique	71
2.19	Exemple de voisinage projeté sur l'image	71
2.20	Equivalence entre le voisinage et les paramètres de polarisation. (a) Le cercle représente le degré de polarisation et le rayon, l'angle. (b) Le voisinage 7×7 du pixel considéré	72
2.21	Détection de contours. (a) Par la méthode perspective. (b) Par la méthode adaptée	74
2.22	Détection de points d'intérêt par la méthode de Harris. (a) Sur un voisinage classique. (b) Sur un voisinage adapté	74
2.23	Mise en correspondance. (En Haut) Sur un voisinage classique. (En Bas) Sur un voisinage adapté	75
2.24	Droites 3D dans l'espace et projection sur la sphère équivalente. \mathbf{n}_1 et \mathbf{n}_2 sont les vecteurs normaux	76
2.25	Projection d'un pixel sur la sphère équivalente, grands cercles et points antipodaux	77
2.26	Consistance des angles sous une rotation pure	77
2.27	Extraction des <i>patches</i>	78
2.28	Exemple de mise en correspondance des intersections	78
2.29	Translation longue. La première et la deuxième colonne sont des images prises à différentes positions. La troisième colonne est l'image dont on a compensé la rotation	79
2.30	Surimpression de l'image initiale et de l'image compensée : les lignes sont parallèles	80

2.31	Mise en correspondance sous un déplacement nul	80
2.32	Images synthétiques séparées par une faible rotation	81
2.33	Compensation de la rotation sur l'image 2	81
2.34	Surimpression de l'image initiale et de l'image compensée	82
2.35	Couple d'images et mise en correspondance	82
3.1	(A gauche) Partie osseuse de l'épaule. (A droite) Les deux axes principaux de l'humérus	86
3.2	Les mouvements de l'épaule	87
3.3	(En haut) Prothèse anatomique. (En bas) Prothèse inversée	87
3.4	(A gauche) Position du patient pendant l'intervention. (A droite) Emplacement en bleu de l'incision	88
3.5	Préparation de l'humérus (gauche) tige centro-médullaire, guide de coupe et broche lisse de la duocentric (centre) coupe de la tête humérale (à droite)	88
3.6	Râpe, embout porte râpe et broche lisse	88
3.7	Ecarteur de glène	89
3.8	Fantôme 1	89
3.9	Fraisage	89
3.10	Fantôme 2	90
3.11	Montage de la duoglène et du support de glène	90
3.12	Platine-insert d'essai	90
3.13	Démarche suivie pour la réalisation du dispositif d'aide à la chirurgie	92
3.14	Humérus réel (à gauche) et humérus reconstruit avec la méthode des formes simples (à droite)	93
3.15	Principales formes utilisées pour modéliser l'omoplate et les paramètres correspondants	94
3.16	D'autres paramètres principaux et secondaires de l'omoplate	94
3.17	Espace sous acromial (flèche rouge)	95
3.18	Extraction du rayon de la tête humérale à partir de radios	95
3.19	Mesure du demi-rayon du cylindre métaphysaire	95
3.20	Extraction du décalage entre l'axe de la tête humérale et l'axe de la diaphyse humérale	96
3.21	Mesure de l'espace sous acromial	96
3.22	Position 3D de l'extrémité de la coracoïde	96
3.23	Modèle CAO assemblé de la prothèse	97
3.24	Interface utilisateur pour l'opération de découpe de la tête humérale	97
3.25	Angle entre l'axe de la tige humérale et l'axe du cône métaphysaire	97
3.26	Résultat obtenu lors d'une découpe de la tête humérale suivant le plan en rouge et l'axe de perçage en jaune	98
3.27	Placement des trous pour le pin's 1 et pin's 2 suivant le fantôme 1	98
3.28	Utilisation du fantôme 2 pour le placement du support glène après avoir travaillé le bas de la glène	99
3.29	Travail du bas de glène et placement du support glène avec un choix entre les tailles S et M	99
3.30	Simulation de l'offset de l'insert huméral	99
3.31	Différents débattements d'une épaule saine	100
3.32	Mouvement d'abduction avec une épaule prothésée	100
3.33	Mouvement d'ante pulsion d'une épaule prothésée	100
3.34	Mouvement de rétropulsion d'une épaule prothésée	101
3.35	Mouvement de rotation interne d'une épaule prothésée	101
3.36	Mouvement de rotation externe d'une épaule prothésée	102
3.37	Comparaison des différents mouvements de l'épaule avec ou sans prothèse	102

Liste des tableaux

1.1	Erreurs de reconstruction	23
1.2	Comparaison des techniques de recalage . (a) Sur des données synthétiques. (b) Sur des données réelles	28
1.3	Comparaison des techniques de recalage fin. (a) Sur des données synthétiques. (b) Sur des données réelles	29
1.4	Pourcentage de points indexés	38
1.5	Comparaison des résultats du recalage multivue	39
1.6	Erreur de recalage	53
1.7	Erreurs de recalage en fonction de l'orientation et des modèles	54
3.1	Formes géométriques utilisées pour modéliser l'humérus	93
5.1	Encadrements post-doctoraux	111
5.2	Encadrements doctoraux	112
5.3	Jury et mention des thèses soutenues	112
5.4	Encadrements de masters	113
6.1	Bilan des publications de la cellule	115
6.2	Responsabilités dans des conférences	118
6.3	Charge d'enseignement	119
6.4	Prestations, Valorisation, Transfert	120

Première partie

Synthèse des Travaux Scientifiques

Chapitre 1

Vision en Lumière Structurée

La vision en lumière structurée est un terme générique qui regroupe à la fois des systèmes projecteur-caméra (pas tous : il en existe qui ne projettent pas de lumière structurée) et les systèmes laser-caméra. Il s'agit de systèmes composés d'au moins une caméra et d'au moins une source lumineuse projetant un motif connu et... structuré sur la scène à analyser. Ce sont des capteurs à triangulation et comme tels, ils sont utilisés principalement pour reconstruire tridimensionnellement des objets ou, plus généralement, des surfaces. Ils sont une alternative efficace aux capteurs stéréoscopiques dans les cas où : la mise en correspondance est en défaut, l'illumination ambiante est faible. Ce chapitre regroupe les travaux que j'ai menés autour cette technique. La première section fait un rapide état de l'art des techniques de lumière structurée *et codée*. Je décrirai ensuite les travaux sur l'auto-calibrage d'un système projecteur-caméra que j'ai effectué en thèse de doctorat. Les deux sections qui suivront détailleront les travaux que j'ai encadrés sur la conception d'un scanner 3D à main (thèse de C. MATABOSCH) et d'un dispositif de texturation virtuelle (thèse de T. MOLINIER). On pourra en déduire une règle générale sur la conception d'un capteur de vision en lumière structurée. De nombreux outils et de nombreuses méthodes sont communs à ces trois thèses : la conception d'un motif adapté à l'application, le calibrage, la reconstruction et le recalage. La figure 1.1 en donne un synoptique très général. Chaque aspect sur lequel nous avons travaillé est annoté et sera détaillé dans la suite du chapitre.

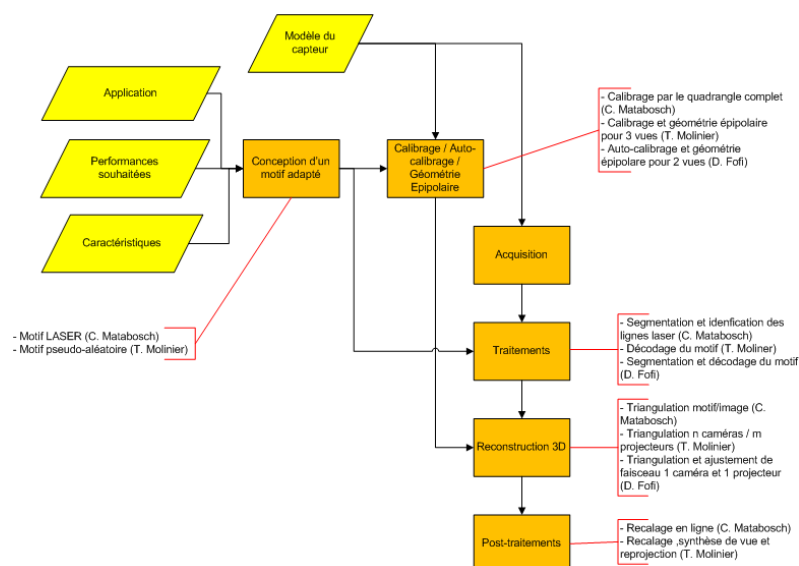


FIGURE 1.1: Schéma général pour la vision en lumière structurée

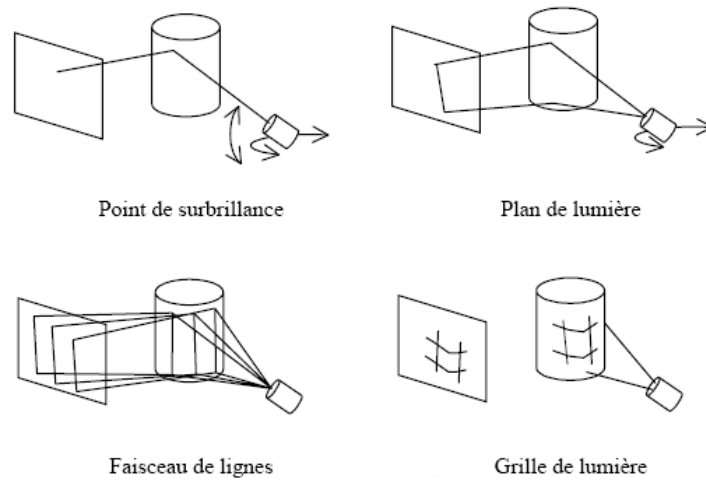


FIGURE 1.2: La lumière structurée

1.1 Lumière Structurée et Codée

Une lumière structurée est une source lumineuse modulée dans l'espace, le temps, l'intensité et/ou la couleur. Dans sa forme la plus simple, la lumière structurée est un faisceau laser projetant un point de surbrillance sur l'environnement (figure 1.2). Puisqu'un seul point est projeté, la mise en correspondance est immédiate. En revanche, un balayage de la scène suivant les deux axes x et y est nécessaire si l'on désire obtenir une mesure globale de celle-ci. Une deuxième solution consiste à projeter un plan de lumière formant un segment sur les objets illuminés. Le plan de lumière est obtenu en "déviant" un faisceau laser à l'aide d'une lentille cylindrique. Le balayage suivant l'axe horizontal demeure nécessaire. Pour éviter toute action mécanique, coûteuse en temps, rendue nécessaire par le balayage, des motifs bi-dimensionnels (ou patrons ou motifs structurants) couvrant le champ de vision de la caméra, ont été proposés : faisceau de lignes, grille ou matrice de points. Les premiers travaux en ce domaine datent de 1971, et sont l'œuvre de WILL et PENNINGTON [Will 1971] qui baptisèrent leur technique *grid coding*. L'inconvénient de ces systèmes vient du fait que la mise en correspondance n'est plus univoque. En effet, dans le cas d'une grille par exemple, puisque rien ne permet de distinguer dans l'image un noeud d'un autre noeud, il est difficile de trouver un unique homologue aux points capturés par la caméra.

Pour contourner ce problème de mise en correspondance, des techniques de codage de la lumière structurée ont été développées au cours de ces trente dernières années. Le codage du patron permet de résoudre de manière efficace le problème de la mise en correspondance. Chaque élément (lignes, points, etc.) de la lumière projetée porte un code permettant de déterminer sa position au sein du patron. La reconnaissance de ce code dans l'image garantit une mise en correspondance unique. Trois coordonnées sont suffisantes pour reconstruire un point ; en conséquence, le patron de lumière peut être codé uniquement le long de ses lignes ou de ses colonnes, étant bien entendu qu'un codage le long des deux axes garantit une codification plus robuste. Une revue exhaustive et une classification de ces différentes techniques a paru voici quelques années [Batlle 1998], puis mise à jour dans un article ultérieur [Salvi 2004]. Elle classe le codage d'une lumière structurée selon trois critères.

Le premier critère est lié à la *dépendance temporelle* du codage :

- **Statique** : la projection du patron est limitée à des scènes statiques. Cette contrainte apparaît quand plusieurs patrons doivent être projetés sur la scène (le code est une résultante des valeurs obtenues sous les différentes illuminations). Un quelconque mouvement dans l'intervalle

de temps séparant deux projections met inévitablement en échec la mise en correspondance.

- **Dynamique** : le codage ne nécessite qu'une seule projection et le patron peut être utilisé pour l'analyse de scènes mobiles.

Le deuxième est lié à la *nature de la lumière émise* :

- **Binaire** : chaque point du patron prend la valeur 0 ou 1 ; ces valeurs correspondent à l'opacité ou à la transparence du patron en ce point.
- **En niveaux de gris** : un niveau de gris est associé à chaque point du patron (il représente son degré d'opacité). Une prise d'image sans projection de lumière, permettant d'évaluer et ainsi d'atténuer l'effet de réflexion des surfaces, est souvent requise avec un tel codage. Les codages en niveaux de gris sont, de ce fait, des codages "statiques".
- **En couleurs** : une couleur est associée à chaque point du patron. Ce codage est plutôt réservé à des scènes composées d'objets de couleurs neutres : des surfaces de couleurs fortement saturées étant susceptibles d'absorber les couleurs projetées.

Le troisième est lié à la *dépendance aux discontinuités en profondeur* :

- **Périodique** : les codes sont répétés périodiquement tout au long du patron. Cette technique permet de réduire le nombre de bits de codification. Cependant, les discontinuités des surfaces doivent être de taille supérieures à la moitié de la période de codage, par respect du théorème de Shannon.
- **Absolu** : le code est unique pour chaque élément du patron.

Une autre clef de classification est donnée par le type de codage utilisé : le **codage par multiplexage temporel**, quand une séquence de motifs est projetée sur la scène ; le **codage par voisinage spatial**, quand le code est donné par un voisinage ; le **codage direct**, quand chaque pixel porte son propre code. La figure 1.3, tirée de [Salvi 2004], synthétise cette classification.

1.2 Auto-Calibrage d'un Système Projecteur-Caméra

L'auto-calibrage consiste à estimer les paramètres intrinsèques et extrinsèques du capteur, en ligne, sans autre connaissance que les points images et leur mise en correspondance. Comment auto-calibrer un système projecteur-caméra ? Est-il possible d'appliquer, ou pour le moins d'adapter, une technique usuelle d'auto-calibrage à un tel système – et si oui, dans quelles mesures ? Ces questions sont loin d'être triviales. Il existe une différence fondamentale entre un système multivues (capteur n -oculaire ou caméra en mouvement) et un système projecteur-caméra : c'est ce que j'avais défini comme étant le *problème du mouvement* pour les systèmes projecteur-caméra et c'est sans doute ce qui différencie le plus la stéréovision de la vision en lumière structurée. Un mouvement du capteur, plus particulièrement du projecteur, provoque un glissement des points projetés sur les surfaces. En d'autres termes, les points 3D illuminés *avant* le mouvement sont différents des points 3D illuminés *après* le mouvement. D'autre part, les paramètres intrinsèques du projecteur et de la caméra sont nécessairement différents, ce qui empêche de considérer, même si le même modèle est utilisé pour les deux appareils, que les paramètres intrinsèques restent constants au cours de la prise de vue.

1.2.1 Etat de l'Art

Il n'existe pas, à ma connaissance, de méthode d'auto-calibrage explicitement développée ou pensée pour les systèmes projecteur-caméra. Longtemps, les méthodes de calibrage furent de simples adaptations des méthodes existant en stéréovision ; à ce titre, elles ne tirèrent pas avantage des particularités de la vision en lumière structurée et n'en évitèrent pas non plus les écueils. On sait pourtant deux ou trois choses des systèmes projecteur-caméra : leur comportement vis-à-vis du mouvement diffère de celui des capteurs stéréoscopiques, un projecteur ne "voit" pas la scène et donc les points 3D, le point principal d'un projecteur est souvent désaxé. Ajoutons ceci : quand on projette une lumière structurée, on a une connaissance parfaite *a priori* de l'image projetée, de sa structure, des relations internes qui lient les éléments du motif. C'est comme si on connaissait





Time-multiplexing	Binary codes 	Posdamer et al. Inokuchi et al. Minou et al. Trobin Valkenburg and Melvor Skocaj and Leonardis Rocchini et al.	✓ ✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓
	n-ary codes	Caspi et al. Horn and Kiryati	✓ ✓	✓ ✓	✓ ✓
	Gray code + Phase shifting	Bergmann Sansoni et al. Wiora Güthring	✓ ✓ ✓ ✓	✓ ✓ ✓ ✓	✓ ✓ ✓ ✓
	Hybrid methods	Kosuke Sato Hall-Holt and Rusinkiewicz	✓ ✓	✓ ✓	✓ ✓
	Non-formal codification	Maruyama and Abe Durdle et al. Ito and Ishii Boyer and Kak Chen et al.	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓
Spatial Neighborhood	De Bruijn sequences 	Hügli and Maître Monks et al. Vuytsteke and Oosterlinck Salvi et al. Lavoie et al. Zhang et al.	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓
	M-arrays 	Morita et al. Petriu et al. Kiyasu et al. Spoelder et al. Griffin and Yee Davies and Nixon Morano et al.	✓ ✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓ ✓ ✓
	Grey levels	Carrhill and Hummel Chazan and Kiryati Hung	✓ ✓ ✓	✓ ✓ ✓	✓ ✓ ✓
Direct coding	Colour 	Tajima and Iwakawa Smutny and Pajdla Geig Wust and Capson Tatsuo Sato	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓	✓ ✓ ✓ ✓ ✓
Scene applicability		Static Moving	✓ ✓	✓ ✓	✓ ✓
Pixel depth		Binary Grey levels Colour	✓ ✓ ✓	✓ ✓ ✓	✓ ✓ ✓
Coding strategy		Periodical Absolute	✓ ✓	✓ ✓	✓ ✓

FIGURE 1.3: Classification de la codification des lumières structurées

parfaitement, précisément et toujours l'une des deux images d'une paire stéréoscopique. C'est une information riche et encore sous-exploitée.

On s'est longtemps contenté d'appliquer plus ou moins finement les méthodes de calibrage développées pour les caméras. Le problème n'est pas théorique, puisque géométriquement un projecteur peut être vue comme une caméra travaillant "à l'envers" (la lumière allant de l'image à la scène), mais bien pratique : si l'image projetée est parfaitement connue (et donc parfaitement "segmentable"), il est plus difficile d'obtenir les coordonnées des points 3D pour établir la correspondance et identifier la matrice de passage. Faut-il viser, par la projection, des points 3D préalablement connus ou mesurés (sur une mire de calibrage par exemple) ? Faut-il mesurer les points projetés – et comment ? Les premières méthodes pratiques de calibrage, un peu frustes, proposaient de couvrir une mire d'un papier millimétré pour y mesurer les coordonnées des points projetés ; ce qui, on l'imagine, posait de sérieux problèmes de précision. En 1999, HUYNH et al. [Huynh 1999] ont proposé une méthode généralisant celle de CHEN et KAK [CHEN 1987] pour les systèmes laser-caméra. Elle est plutôt dédiée aux motifs "lignes" (projection de plans lumineux), mais peut être généralisée à tout motif dont les éléments de base sont alignés. Elle utilise astucieusement le birapport et le calcul d'homographie pour recouvrer les coordonnées des points 3D sur une mire conçue *ad hoc*. En 2001, RASKAR et BEARDSLEY [Raskar 2001] ont proposé une méthode simple pour calibrer un système projecteur-caméra. Ils eurent en outre le mérite de préciser, dans le même article, les différences essentielles qui existent entre une caméra et un projecteur. La méthode est une extension du calibrage de caméra par homographies aux systèmes comprenant un projecteur. Un motif est projeté sur un plan homogène blanc et l'homographie calculée entre le motif et l'image du motif permet d'extraire les paramètres de calibrage du projecteur, une fois que la caméra a été elle-même calibrée. Un peu plus récemment, LI et CHEN [Li 2003, Chen 2002] ont proposé une

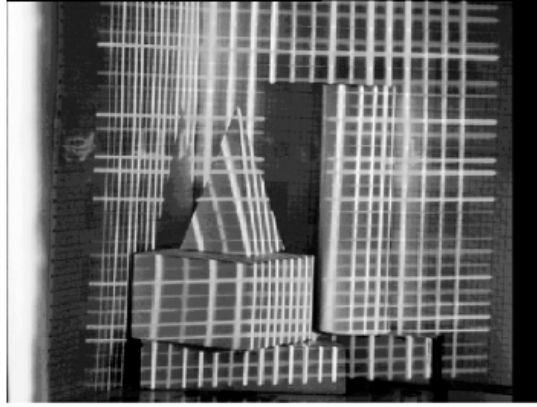


FIGURE 1.4: Scène éclairée par le projecteur

méthode de recalibrage d'un système projecteur-caméra, c'est-à-dire de calcul en ligne de la pose relative entre le projecteur et la caméra à partir d'un système initialement calibré, via encore une fois un calcul d'homographies. Bien que les auteurs indiquent que cette méthode peut être adaptée pour l'auto-calibrage (à condition que le nombre d'inconnues n'excède pas 6), aucune démonstration ni expérimentation n'est proposée. Il est *a priori* possible d'estimer 6 paramètres intrinsèques (les focales et les points principaux par exemple), mais à partir d'un système initialement calibré.

Une méthode d'auto-calibrage, oui, mais *quelle* méthode d'auto-calibrage? Faut-il fixer les paramètres intrinsèques? tous? quelques-uns seulement? Que cherche-t-on en auto-calibrant un système de vision? A se défaire d'une intervention humaine, hors-ligne et pas toujours évidente à réaliser? Ou bien à conférer au capteur d'utiles degrés de liberté supplémentaires? Tous les cas de figure sont possibles, mais je crois pourtant que, plus encore que pour une caméra, il est utile de laisser au projecteur le plus grand nombre de degrés de liberté possible (pour, notamment, assurer une mise au point correcte tout au long de la prise de vue). Nous avons, dans ces travaux, choisis de développer une méthode d'auto-calibrage à partir de deux "vues" seulement, c'est-à-dire une image de la scène et un patron de lumière structurée. Cette méthode se subdivise en deux étapes : la première consiste à reconstruire projectivement la scène ; la seconde à la redresser dans un espace métrique par la génération de contraintes Euclidiennes. Nous détaillons ces deux étapes dans la suite de la section.

1.2.2 Reconstruction Projective

On peut extraire les matrices de projection d'une paire stéréoscopique à partir de la seule connaissance de la géométrie épipolaire. Dans ce cas, ces matrices seront bien évidemment exprimées dans un espace projectif :

$$\mathbf{P}_{proj} \simeq [\mathbf{M} \quad \mathbf{e}] \text{ et } \mathbf{P}'_{proj} \simeq [\mathbf{I} \quad \mathbf{0}] \quad (1.1)$$

avec $\mathbf{M} = -\frac{1}{\|\mathbf{e}\|^2} [\mathbf{e}]_{\times} \mathbf{F}$. Où \mathbf{F} représente la matrice fondamentale et \mathbf{e} l'épipole dans la première image ; \mathbf{P} est la matrice de projection de la caméra et \mathbf{P}' , la matrice de projection du projecteur. L'indice *proj* désigne les vecteurs et matrices exprimés dans un espace projectif ; à l'inverse, l'indice *eucl* désignera les vecteurs et matrices exprimés dans un espace métrique¹. Le symbole \simeq représente l'égalité projective, ou égalité à un facteur d'échelle près.

1. Métrique désignera l'espace Euclidien à un facteur d'échelle près.

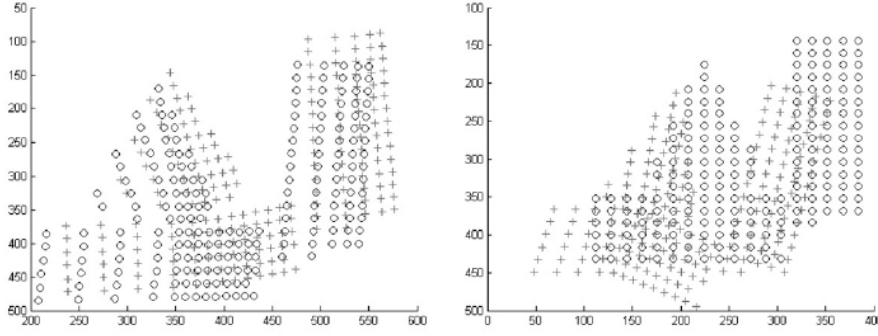


FIGURE 1.5: Retroprojection de la reconstruction obtenue par méthode linéaire. A gauche : le plan image, à droite : le plan projecteur. Les cercles désignent les points originaux et les croix, les points retroprojetés.

1.2.2.1 Méthode Linéaire

Etant donnée une paire de points homologues $\mathbf{m} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T$ et $\mathbf{m}' = \begin{bmatrix} u' & v' & 1 \end{bmatrix}^T$ et le point 3D dont ils sont les images $\mathbf{M}_{proj} = \begin{bmatrix} x & y & z & 1 \end{bmatrix}^T$, on peut écrire :

$$\begin{aligned} \lambda \begin{bmatrix} u & v & 1 \end{bmatrix}^T &= \mathbf{P}_{proj} \begin{bmatrix} x & y & z & 1 \end{bmatrix}^T \\ \lambda' \begin{bmatrix} u' & v' & 1 \end{bmatrix}^T &= \mathbf{P}'_{proj} \begin{bmatrix} x & y & z & 1 \end{bmatrix}^T \end{aligned} \quad (1.2)$$

λ et λ' sont deux facteurs d'échelle non-nuls. En les éliminant et en ré-arrangeant les équations ci-dessus, il vient :

$$\mathbf{Q}\mathbf{M}_{proj} = 0 \quad (1.3)$$

où \mathbf{Q} est une matrice 4×4 donnée par :

$$\mathbf{Q} = \begin{bmatrix} \mathbf{p}_1 - u\mathbf{p}_3 & \mathbf{p}_2 - v\mathbf{p}_3 & \mathbf{p}'_1 - u'\mathbf{p}'_3 & \mathbf{p}'_2 - v'\mathbf{p}'_3 \end{bmatrix} \quad (1.4)$$

Les \mathbf{p}_i et \mathbf{p}'_i sont les vecteurs correspondant à la i^e ligne de \mathbf{P}_{proj} et \mathbf{P}'_{proj} respectivement. Comme \mathbf{M}_{proj} est définie à un facteur d'échelle près, on peut imposer la contrainte $\|\mathbf{M}_{proj}\| = 1$. La solution est donnée par le vecteur propre de la matrice $\mathbf{Q}^T\mathbf{Q}$ associé à la plus petite valeur propre. C'est une méthode de résolution très courante, parfois appelée *Direct Linear Transformation* (DLT).

La figure 1.5 montre les résultats de la retroprojection des points 3D reconstruits par la méthode linéaire et la comparaison avec les points initiaux. L'erreur est importante et montre combien il est essentiel de procéder à un raffinement itératif de la reconstruction.

1.2.2.2 Méthode Itérative

Il s'agit ici de minimiser l'erreur résiduelle séparant les points retroprojetés et les points initiaux. C'est un algorithme d'ajustement de faisceaux projectif, exprimé par :

$$\left(u - \frac{\mathbf{p}_1^T \mathbf{M}_{proj}}{\mathbf{p}_3^T \mathbf{M}_{proj}}\right)^2 + \left(v - \frac{\mathbf{p}_2^T \mathbf{M}_{proj}}{\mathbf{p}_3^T \mathbf{M}_{proj}}\right)^2 + \left(u' - \frac{\mathbf{p}'_1^T \mathbf{M}_{proj}}{\mathbf{p}'_3^T \mathbf{M}_{proj}}\right)^2 + \left(v' - \frac{\mathbf{p}'_2^T \mathbf{M}_{proj}}{\mathbf{p}'_3^T \mathbf{M}_{proj}}\right)^2 \quad (1.5)$$

En pratique, on utilisera un algorithme de minimisation classique de type Levenberg-Marquardt [Marquardt 1963]. Les résultats obtenus par la méthode linéaire serviront à initialiser la méthode itérative. La figure 1.6 montre les résultats de retroprojection : on voit qu'ils sont considérablement améliorés.

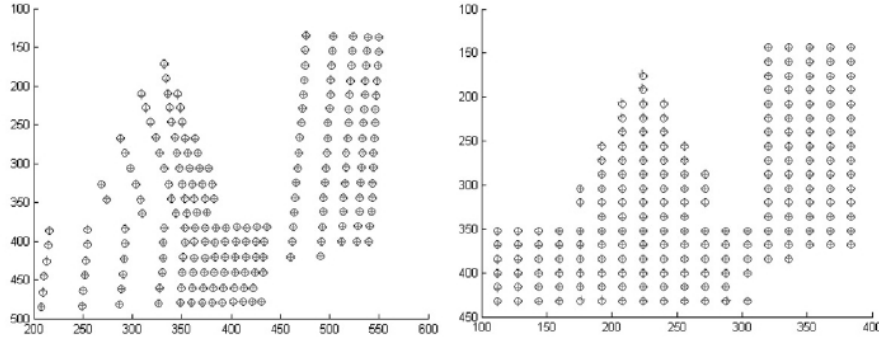


FIGURE 1.6: Retroprojection de la reconstruction obtenue par méthode itérative. À gauche : le plan image, à droite : le plan projecteur. Les cercles désignent les points originaux et les croix, les points retroprojetés.

1.2.3 Reconstruction Métrique par Génération de Contraintes

On sait que la géométrie Euclidienne est un cas particulier de la géométrie projective ; en d'autres termes, il existe une transformation projective permettant de passer de l'une à l'autre. Nous chercherons donc à identifier cette transformation de manière à redresser la reconstruction projective obtenue en reconstruction métrique (puisque le facteur d'échelle ne peut être obtenu sans la connaissance *a priori* d'une distance). Le problème peut se formaliser ainsi :

$$\mathbf{M}_{eucl} = \mathbf{W}\mathbf{M}_{proj} \quad (1.6)$$

où \mathbf{W} est une matrice 4×4 , représentant une transformation projective. Des connaissances géométriques sur la scène observée sont nécessaires pour contraindre et identifier cette matrice. De plus, si nous voulons obtenir une reconstruction métrique, il est nécessaire que ces connaissances reflètent des informations Euclidiennes, telles que le parallélisme, l'orthogonalité, etc. La matrice \mathbf{W} possède 16 éléments et est définie à un facteur d'échelle près : elle a donc 15 degrés de liberté. Ce sont 15 contraintes, cohérentes et indépendantes, qui doivent être générées. L'idée que nous avons défendue est la suivante : ne peut-on pas utiliser la structure de la lumière projetée pour générer ces contraintes, en tirant des informations sur la scène 3D elle-même ? Nous y avons apporté une réponse partielle, puisque la génération de contraintes n'a été validée que pour un projecteur adoptant un comportement affine (c'est-à-dire dans les conditions de prise de vue pour lesquelles un modèle affine est valide). Nous pensons toutefois qu'il est possible de généraliser la méthodologie aux projecteurs "projectifs". La figure 1.7 propose un synoptique général de la méthode ; les contraintes sont détaillées dans les sous-sections qui suivent.

1.2.3.1 Détection de Plans dans l'Espace

Nous ferons l'hypothèse, sans perte de généralité, que le motif projeté forme une grille régulière. La cellule élémentaire du motif est donc un carré. Chaque carré, projeté sur une surface plane, sera vu par la caméra comme un quadrilatère. En théorie, ces quadrilatères peuvent être des carrés, des rectangles, des losanges, des parallélogrammes ou des trapèzes, selon la position relative de la caméra, du projecteur et de la surface considérée. Bien que la détection d'un quadrilatère dans l'image capturée par la caméra ne signifie pas nécessairement la détection d'un plan dans l'espace (il existe des configurations de surfaces courbes donnant un quadrilatère), la probabilité est forte. Nous avons proposé dans [Fof 2001] une méthode de vérification de la planarité de telles surfaces par calcul de la conservation du birapport.

Considérons la configuration de points de la figure 1.8. Si les points o_i, p_i, q_i, r_i et s_i ($i = 1$ ou 2) sont projetés sur un plan, alors le birapport au sein du patron est égal au birapport des cinq points

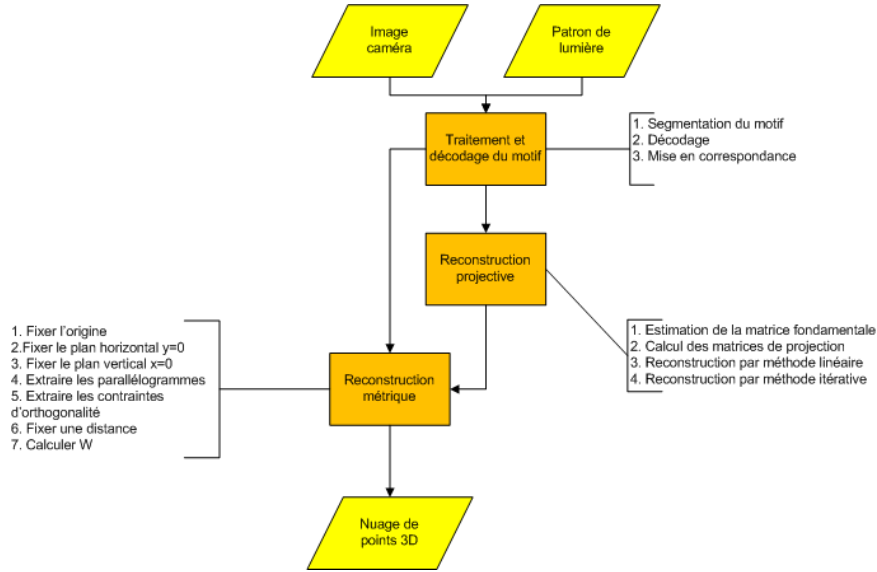


FIGURE 1.7: Algorithme général de la méthode

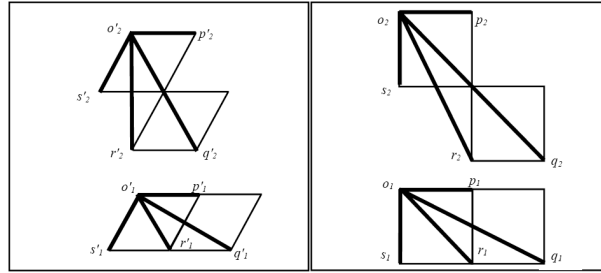


FIGURE 1.8: Test de coplanarité. A gauche : plan image. A droite : patron de lumière

formés sur ce plan ; aussi, le birapport que l'on retrouve dans l'image est égal aux deux autres. La transformation des points du projecteur vers la caméra est obtenue par deux homographies successives. On en déduit que si l'équation 1.7 est vérifiée, alors les points 3D correspondant O , P , Q , R et S sont coplanaires.

$$\{o_i; p_i, q_i, r_i, s_i\} = \{o'_i; p'_i, q'_i, r'_i, s'_i\} \text{ avec } i = 1 \text{ ou } 2 \quad (1.7)$$

où $\{o; \dots\}$ représente le birapport d'un faisceau de quatre droites concourant en o .

1.2.3.2 Contraintes du Parallélogramme

Puisque que nous posons l'hypothèse que notre capteur de vision a un comportement affine, nous pouvons en déduire que chaque carré du motif projeté sur une surface plane de la scène apparaîtra dans l'image sous la forme d'un parallélogramme (figure 1.9). Ce qui est remarquable ici, c'est que les points 3D correspondant sont les sommets d'un parallélogramme dans l'espace. Les propriétés du parallélogramme nous permettent d'en déduire des contraintes de distance entre ces points (eq. 1.8) ainsi que des contraintes de parallélisme (eq. 1.9). Il faut préalablement vérifier que les points considérés appartiennent bien à un même plan de l'espace. On sait que certaines configurations de points peuvent donner naissance à un parallélogramme dans l'image sans qu'elles ne

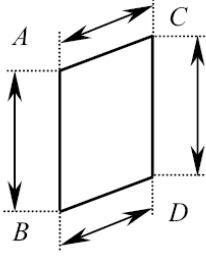


FIGURE 1.9: Contraintes du parallélogramme

correspondent à un parallélogramme dans l'espace ; si l'on teste la coplanarité des quatre sommets du parallélogramme (section 1.2.3.1), cette ambiguïté est levée.

$$d(A, B) = d(C, D), \quad d(A, C) = d(B, D) \quad (1.8)$$

$$(AB) \parallel (CD), \quad (AC) \parallel (BD) \quad (1.9)$$

où $d(.,.)$ représente la distance Euclidienne.

On peut constater que ces deux jeux de contraintes sont redondants. Puisque la géométrie projective préserve l'alignement et la coplanarité, les équations 1.8 et 1.9 décrivent la même configuration de points. Il faut noter également qu'un seul parallélogramme détermine complètement le plan 3D sur lequel il repose. De ce fait, il suffira d'implémenter les contraintes d'un seul parallélogramme pour chaque surface plane de la scène.

Les contraintes du parallélogramme nous permettent d'injecter dans le système une information de distance et une information d'angle.

1.2.3.3 Contraintes d'Alignement

Chaque ligne horizontale du patron forme un plan de lumière dans l'espace qui peut être considéré comme un plan 3D horizontal dans le repère centré sur le projecteur ; similairement, chaque ligne verticale du patron forme un plan de lumière qui peut être considéré comme un plan 3D vertical dans ce même repère. Le motif perçu par la caméra correspond à l'intersection des plans de lumière générés par la projection avec les surfaces qui composent la scène. On peut en déduire que les points qui composent le motif perçu appartiennent à des plans horizontaux et/ou verticaux de l'espace et inférer des contraintes d'alignement. De ce fait, deux points appartenant au même plan horizontal ou vertical ont une coordonnée en commun.

On peut dériver deux nouvelles contraintes de ces contraintes d'alignement : choisir un point d'intersection du patron de lumière comme "origine" en fixant deux de ses coordonnées à 0 (x et y) ; fixer une distance Euclidienne entre deux plans horizontaux et verticaux.

1.2.3.4 Contraintes d'Orthogonalité

L'orthogonalité est une caractéristique importante de la structure Euclidienne. La détection de plans orthogonaux permet, au moins partiellement, de définir un repère Euclidien de l'espace 3D. On constate que la projection de droites orthogonales (OM) et (ON) produit deux plans orthogonaux dans l'espace (figure 1.10). Quand ces plans coupent des surfaces planes de l'environnement, ils produisent un segment lumineux en leur superficie qui sera capturé par la caméra. Nous avons donc deux droites ($O'M'$) et ($O'N'$) dans l'espace qui appartiennent à des plans orthogonaux. Puisque O' et M' appartiennent au même plan vertical et O' et N' appartiennent au même plan horizontal, on a :

$$x_{O'} = x_{M'} \text{ et } y_{O'} = y_{N'} \quad (1.10)$$

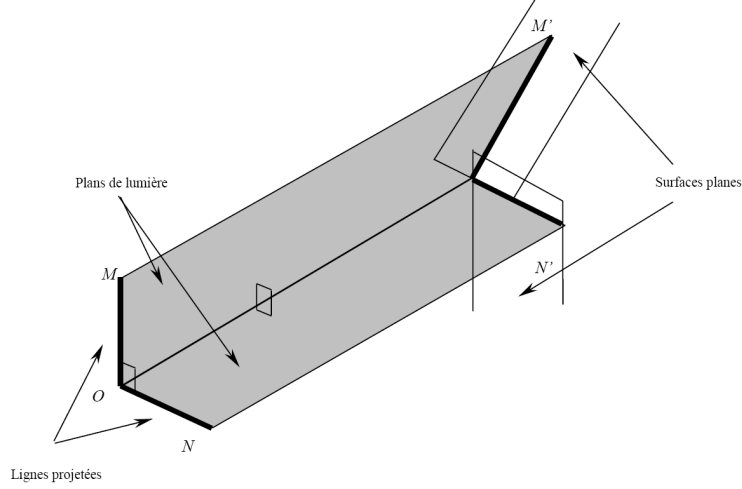


FIGURE 1.10: Contrainte d'orthogonalité

Deux vecteurs sont orthogonaux si leur produit scalaire est nul, donc :

$$(O'M') \perp (O'N') \Leftrightarrow \overrightarrow{O'M'} \cdot \overrightarrow{O'N'} = 0 \quad (1.11)$$

En développant le second membre, on obtient :

$$(x_{M'} - x_{O'}) (x_{N'} - x_{O'}) + (y_{M'} - y_{O'}) (y_{N'} - y_{O'}) + (z_{M'} - z_{O'}) (z_{N'} - z_{O'}) = 0 \quad (1.12)$$

En combinant avec l'équation 1.10, on obtient finalement :

$$(O'M') \perp (O'N') \Leftrightarrow z_{O'} = z_{M'} \text{ ou } z_{O'} = z_{N'} \quad (1.13)$$

Si les conditions imposées par l'équation 1.13 sont vérifiées, nous obtenons une contrainte d'orthogonalité ; dans le cas contraire, nous obtenons une *contrainte d'orthogonalité réduite* :

$$(x_{O'} - x_{M'}) (x_{O'} - x_{N'}) + (y_{O'} - y_{M'}) (y_{O'} - y_{N'}) = 0 \quad (1.14)$$

1.2.3.5 Résultats Expérimentaux

Nous présentons quelques résultats expérimentaux illustrant la méthode. Nous avons tout d'abord reconstruit les éléments de la scène présentée en figure 1.4 : il s'agit de trois objets géométriques et achromatiques. De nombreuses contraintes ont pu être générées : contraintes du parallélogramme sur toutes les facettes des objets, contraintes d'orthogonalité, distance, origine. La figure 1.11 montre que la reconstruction est *qualitativement* acceptable : les proportions semblent préservées, le parallélisme et l'orthogonalité sont visuellement respectées, etc. Nous avons accompagné ces tests d'une étude menée en simulation, dont les résultats sont consignés dans le tableau 1.1. L'erreur relative reste relativement faible sous un bruit uniforme de ± 1 pixel sur les points images.

La figure 1.12 représente une scène un peu plus réaliste. Elle a le mérite de ne pas être capturée sous illumination contrôlée, ni sur des objets uniformément blancs. Les lignes apparaissant en surbrillance représentent les contraintes que nous avons générées. La figure 1.13 représente la reconstruction 3D correspondante. Les résultats corroborent ceux obtenus pour la scène précédente et ceux obtenus en simulation : parallélisme, orthogonalité, proportions sont correctement préservés.

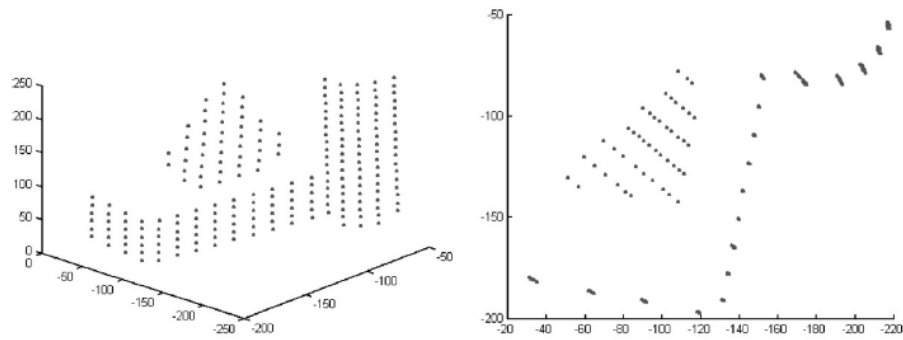


FIGURE 1.11: Reconstruction 3D - Scène I

Errors on reconstruction with uniform noise ± 1

Real co-ordinates			Errors on estimate co-ordinates		
X	Y	Z	ΔX	ΔY	ΔZ
100	-50	4000	0.518	-0.267	3.95
300	-50	2000	-0.65	-0.242	-1.5
700	-50	4000	0.614	-0.33	6.43
500	-400	4020	-1.132	-1.768	-4.332
300	50	4000	0.091	0.397	2.597
500	50	2000	0.076	-0.119	0.449
900	50	4000	0.13	0.171	2.007
300	-430	3000	0.505	-0.911	5.079
450	75	2500	0.76	-1.154	4.016
705	-120	1000	0.603	-0.827	0.829
Mean relative error (%)			0.518	1.539	0.169
Standard deviation			0.610	0.655	3.222

TABLE 1.1: Erreurs de reconstruction

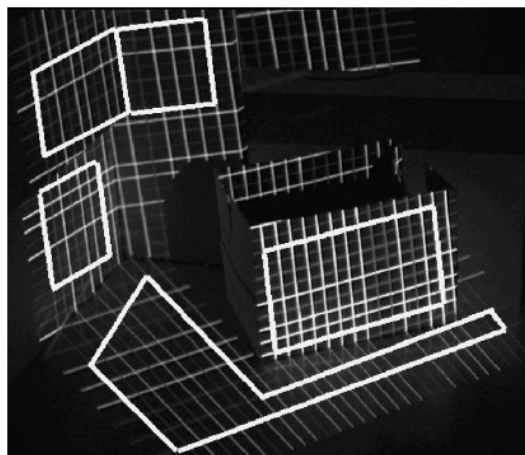


FIGURE 1.12: Scène II. Les contraintes choisies sont en surbrillance

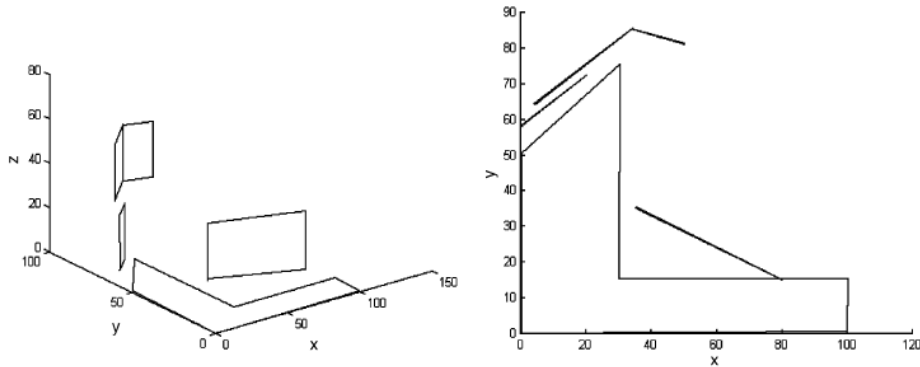


FIGURE 1.13: Reconstruction 3D - Scène II

1.2.4 Conclusions

Ces travaux datent de l'année 1999/2000 (la thèse fut soutenue en 2001). Le formalisme utilisé, l'approche en elle-même, la méthode de résolution s'en ressentent forcément. Si l'auto-calibrage d'un système projecteur-caméra est *encore* un problème qui, à mon sens, est ouvert, je ne l'aborderais sans doute pas de la même manière aujourd'hui. L'automatisation de la génération des contraintes reste difficile à implémenter et le grand nombre d'équations qu'elle produit rend la résolution lourde en charge de calcul et parfois même incertaine (problème de convergence). En ce sens, la méthode proposée aurait besoin d'amendements conséquents. L'utilisation du formalisme de la géométrie projective dans l'expression des contraintes (projection d'une ligne de l'image dans l'espace via les coordonnées de Plücker et leur loi de transformation, par exemple) en est un. Elle aurait l'avantage de poser le problème de manière plus compacte et lisible et, pourquoi pas, de révéler des vices cachés, des propriétés ou caractéristiques permettant de rendre la résolution plus efficiente et de diriger la génération des contraintes. En second lieu, l'état de l'art sur l'auto-calibrage a quelque peu avancé depuis lors. Les contributions nouvelles, essentiellement développées pour des caméras, pourraient être adaptées aux projecteurs et contraintes par les caractéristiques qui lui sont inhérentes (connaissance parfaite et a priori de l'image qu'ils projettent et relations internes des éléments qui la composent). Dès 1999, POLLEFEYS [Pollefeys 1999] proposait une méthode flexible d'auto-calibrage utilisable avec deux vues au moins et permettant de laisser libre la distance focale. Une tentative pour adapter cette approche (et les variantes qui ont été publiées depuis) au problème de l'auto-calibrage d'un système projecteur-caméra (avec ou sans génération de contraintes Euclidiennes) serait sans doute à entreprendre.

1.3 Scanner 3D à Main

Ces travaux sont le fruit d'une collaboration, qui a pris la forme d'un co-encadrement de thèse, entre le Le2i et le VICOROB de l'Université de Gérone en Espagne. L'objectif était de concevoir un scanner 3D à triangulation laser, portable, compact et maniable, capable de reconstruire un objet complet (c'est-à-dire, "sous toutes ses facettes"), même de grande dimension, sans avoir recours à aucun système de localisation de la tête laser (pour précisément en assurer la compacité et la maniabilité). Nous avons donc été tout naturellement conduit à employer une méthode de recalage en ligne pour re-positionner chaque nouveau morceau reconstruit à ceux qui l'ont été précédemment. Il est entendu qu'en utilisant une source laser de type point ou ligne, il est impossible d'appliquer un algorithme de recalage car l'information fournie (un point 3D, dans le premier cas, un profil 3D dans le second) est trop pauvre géométriquement. Nous avons donc choisi de projeter un faisceau de 19 lignes laser et réfléchi à un algorithme permettant de les segmenter et de les identifier. Le principal verrou scientifique du projet était ailleurs cependant ; posons-le ainsi : comment assurer

un recalage précis sur un grand nombre de vues, autrement dit comment minimiser la propagation de l'erreur de recalage ? En effet, si l'on s'en tient aux techniques de recalage (en laissant de côté les techniques de SLAM visuel, par exemple), on sait que les méthodes dites "multivues" offrent une meilleure précision ; or, il est impossible de les utiliser pour un recalage à la volée. Les méthodes de recalage "par paire de vues" possèdent elles des caractéristiques inverses. La solution a donc consisté à combiner les deux approches en les intégrant à une stratégie originale de recalage de vues.

1.3.1 Classification et Nouvelle Méthode de Minimisation

Nous présentons ici les deux contributions majeures de ces travaux : l'état de l'art comparatif des méthodes de recalage et la nouvelle stratégie de minimisation de l'erreur sur une séquence de vues, le premier ayant fourni les éléments d'analyse utiles au second.

1.3.1.1 Classification et Comparaison des Méthodes de Recalage

La figure 1.14, tirée de la thèse de doctorat de C. MATABOSCH [Matabosch 2007], synthétise l'état de l'art et la classification proposés des méthodes de recalage rigide tri-dimensionnel. Nous distinguons dans un premier temps les techniques dites "grossières" (*coarse registration*) permettant le recalage sans initialisation préalable de l'algorithme (une connaissance a priori et approximative de la pose relative des surfaces à recaler n'est pas nécessaire), des techniques dites "fines" (*fine registration*) garantissant une meilleure précision mais nécessitant une initialisation. Nous proposons ensuite des critères, que nous avons jugé pertinents, pour la classification telles que la méthode de résolution (linéaire ou itérative), les primitives considérées (points, surfaces, courbes), l'approche utilisée (par paire de vues ou multivues), etc.

Cette étude bibliographique nous a permis d'acquérir une expertise fine sur les méthodes de recalage rigide, leurs performances, la précision attendue, leur charge de calcul et leur applicabilité. La plupart ont été implémentées, testées et comparées ; une toolbox Matlab est disponible au téléchargement². Nous présentons dans les figures 1.15, 1.16, 1.17 et 1.18 des résultats obtenus avec différentes méthodes et dans différentes conditions.

Nous avons également fourni une comparaison détaillée des méthodes de recalage par évaluation quantitative de leurs performances (tableaux 1.2 et 1.3). Il en ressort que l'Analyse en Composantes Principales [Chung 1998] est la méthode qui offre le meilleur compromis précision/temps de calcul. Ses résultats sont excellents sur des données synthétiques, un peu moins sur des données réelles pour lesquelles elles semblent souffrir du bruit. Parmi les méthodes de recalage fin, c'est l'ICP (*Iterative Closest Point*) et ses nombreuses variantes et extensions qui donnent les meilleurs résultats [Chen 1991] [Zinsser 2003].

1.3.1.2 Stratégie pour Minimiser la Propagation de l'Erreur

La méthode que nous avons proposée peut être vue comme une variante robuste de la méthode de SHARP et al. [Sharp 2004], nous allons donc en donner les grandes lignes afin de faciliter la compréhension de la section. Elle est basée sur la minimisation d'un ensemble de vues formant un cycle. Un cycle est défini comme étant un ensemble de vues dont la combinaison des déplacements est nulle – c'est-à-dire, dont la matrice combinée des déplacements est égale à l'identité. La différence entre la matrice combinée et la matrice identité est appelée erreur de déplacement. SHARP propose de distribuer cette erreur sur toutes les vues composant le cycle. Pour distribuer l'erreur correctement, la translation et la rotation sont tout d'abord découplées.

La stratégie que nous avons développée reprend partiellement les idées émises par SHARP. Elle se décompose en deux étapes principales (le détail de l'algorithme est donné en figure 1.19) :

1. Deux vues consécutives sont recalées en utilisant la méthode de type *point-à-plan* proposée par PARK et SUBBARAO [Park 2003].

2. <http://eia.udg.es/%7Ecatabos/ToolboxRegistration.rar>

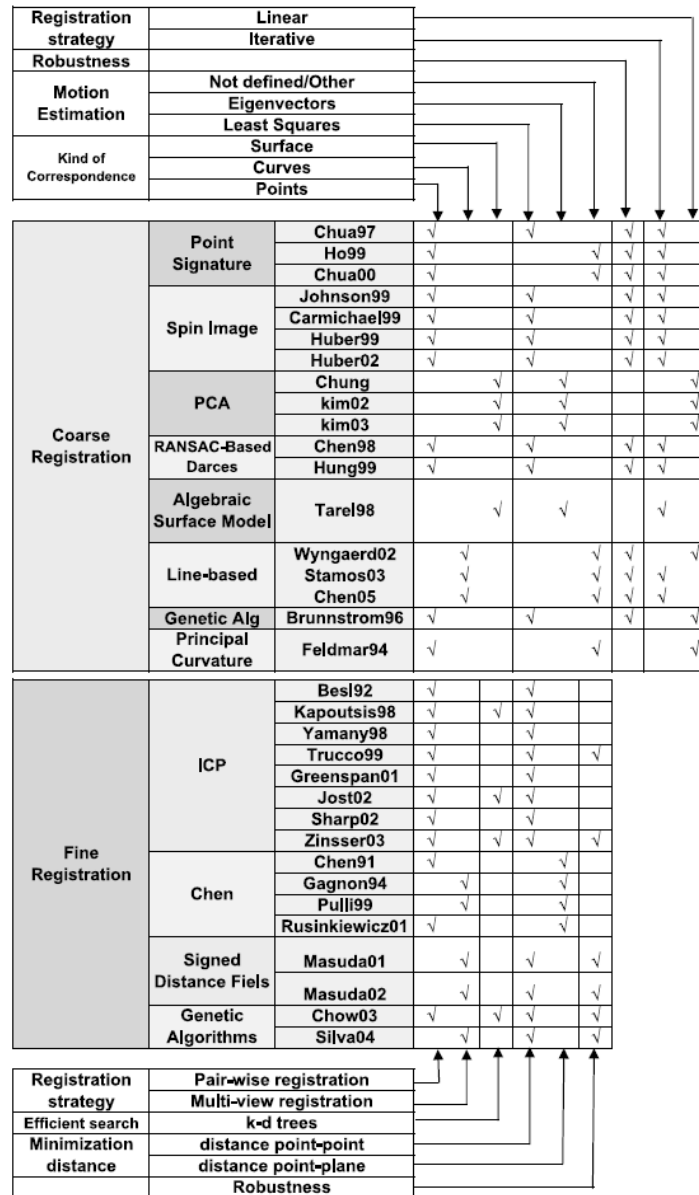


FIGURE 1.14: Classification des méthodes de recalage

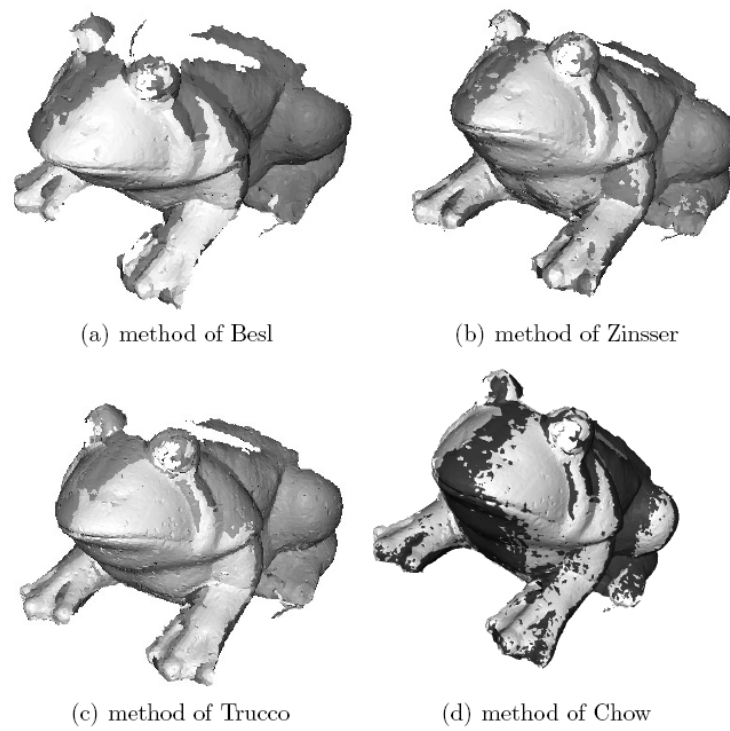


FIGURE 1.15: Recalage par paire de vues sur des données réelles

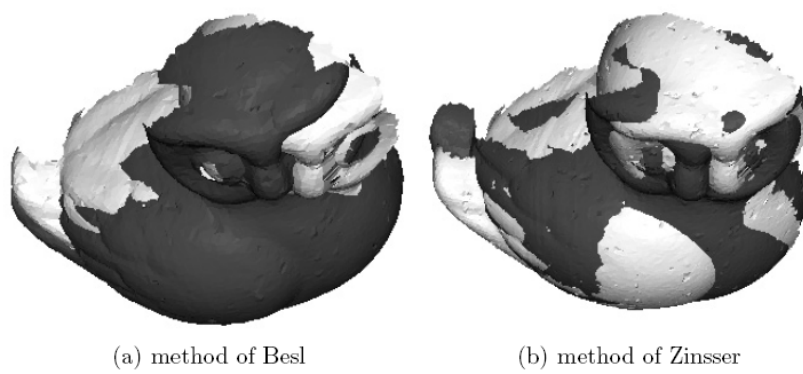


FIGURE 1.16: Recalage peu précis sur des données réelles présentant peu de détails de forme

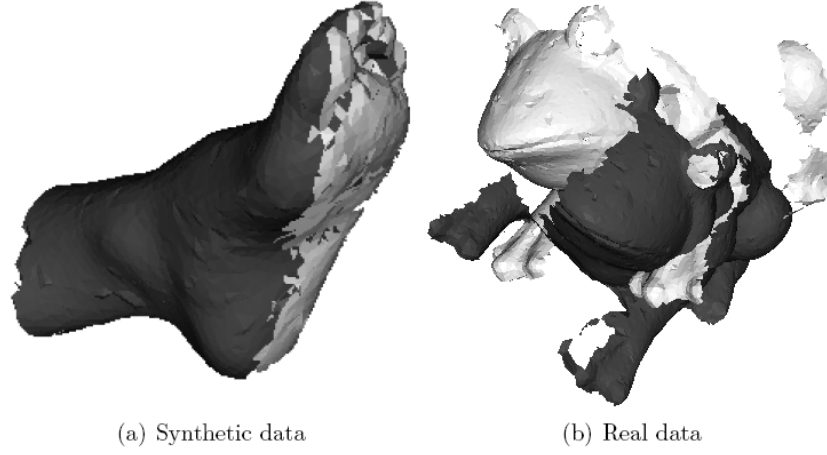
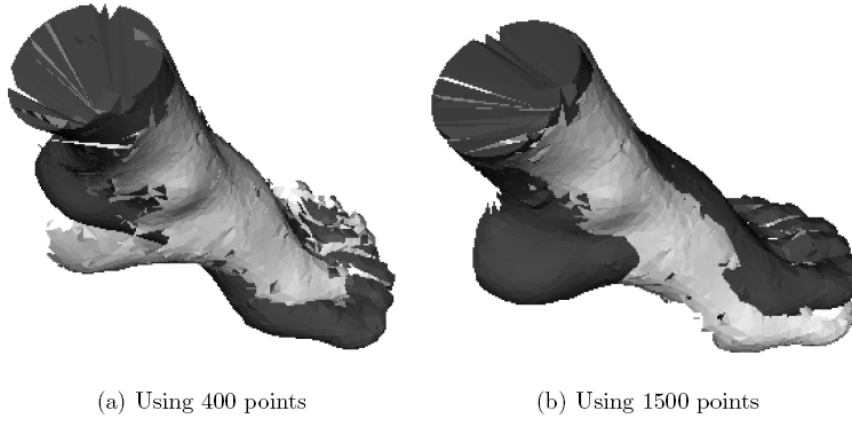


FIGURE 1.17: Recalage grossier par la méthode RANSAC

FIGURE 1.18: Influence sur l'erreur du nombre de points considérés par la méthode de *Spin Image*

Experimental results using synthetic data obtained by coarse registration methods					Experimental results using real range images obtained by coarse registration methods					
Points	Method	Translation error (mm)	$\alpha-\beta$ (rad)	Time (s)	Points	Method	Translation error (mm)	$\alpha-\beta$ (rad)	γ (rad)	Time (s)
200	Ransac-based	50.613	0.98573	1.062	250	Ransac-based	69.406	2.3562	1.3855	2.86
	PCA	21.229	2.9533	0.016		PCA	67.464	1.9923	1.2985	0.016
	Spin image	149.84	2.8911	7.969		Spin image	53.861	0.91301	1.9844	24.156
	GenAlg	81.359	0	4.281		GenAlg	72.094	2.1475	1.5708	44.766
400	Ransac-based	21.394	0.30569	106.08	500	Ransac-based	127.03	2.3562	1.5708	2.453
	PCA	18.777	3.0037	0.016		PCA	69.87	1.9703	1.3306	0.016
	Spin image	81.19	2.3465	50.703		Spin image	40.147	0.54215	0.68824	101.59
	GenAlg	33.537	1.6721	122.56		GenAlg	82.698	2.1988	1.5708	243.94
700	Ransac-based	27.557	0.096427	32820	1000	Ransac-based	40.53	0.17204	1.5236	1657.7
	PCA	16.566	0.21871	0.016		PCA	69.893	1.9236	1.3262	0.203
	Spin image	64.254	0.90339	169.76		Spin image	23.133	0.54339	1.5708	565.45
	GenAlg	18.017	2.9687	249.89		GenAlg	38.059	1.7571	0.85619	1051.6

TABLE 1.2: Comparaison des techniques de recalage . (a) Sur des données synthétiques. (b) Sur des données réelles

Experimental results using synthetic data obtained by fine registration methods						Experimental results using real range images obtained by fine registration methods					
Points	Method	Translation error (mm)	$\alpha-\beta$ (rad)	RMS error (mm)	Time (s)	Points	Method	Translation error (mm)	$\alpha-\beta$ (rad)	RMS error (mm)	Time (s)
700	Besl	0.37	0.01	1.59	0.91	500	Besl	3.34	-0.09	4.18	1.41
	Zinsser	0.25	0.003	2.64	3.28		Zinsser	0.68	-0.06	2.35	1.95
	Jost	0.82	0.006	4.85	0.38		Jost	2.01	-0.07	4.54	0.36
	Trucco	0.87	0.011	2.27	4.13		Trucco	2.87	-0.05	3.93	19.33
	Chen	0	0	1.49	25.03		Chen	1.3373	-0.008	2.1608	18.391
1500	Besl	0.41	0.007	1.48	4.77	1000	Chow	0.27	0.0	4.60	154.05
	Zinsser	0.093	0.002	1.75	9.28		Besl	3.47	-0.09	3.78	5.22
	Jost	0.257	0.009	4.22	1.30		Zinsser	0.47	-0.02	1.42	7.97
	Trucco	0.768	0.016	5.44	112.8		Jost	2.38	-0.08	3.70	1.38
	Chen	0	0	1.68	145.06		Trucco	2.71	-0.05	5.13	33.95
6000	Besl	0.166	0.002	0.71	47.86	5000	Chen	0.29957	0.003	1.7305	67.954
	Zinsser	0.169	0.002	1.86	406.05		Chow	0.12	-0.01	4.51	281.61
	Jost	0.801	0.019	3.01	8.51		Besl	3.12	-0.08	3.20	57.36
	Trucco	0.458	0.015	2.91	198.28		Zinsser	0.26	-0.02	0.85	281.58
	Chen	0	0	1.36	1217.3		Jost	2.24	-0.09	2.83	9.14
							Trucco	2.57	-0.05	1.77	149.59
							Chen	1.2535	-0.015	1.2543	993.09
							Chow	0.06	0.0	3.57	1776.00

TABLE 1.3: Comparaison des techniques de recalage fin. (a) Sur des données synthétiques. (b) Sur des données réelles

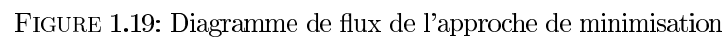
2. Quand un cycle est détecté, une technique de recalage multivue est appliquée aux vues composant le cycle.

Recalage de deux vues consécutives. Généralement, un recalage fin doit être précédé d'un recalage grossier afin d'obtenir une initialisation du déplacement. Nous considérerons ici que les images sont prises à fréquence suffisamment haute, que les surfaces reconstruites sont donc suffisamment proches pour que le déplacement initial soit considéré comme nul. En pratique, cette condition est toujours vérifiée.

La technique que nous proposons est basée sur une variante robuste du recalage point-à-plan de CHEN et MEDIONI [CHEN 1991] publiée par PARK et SUBBARAO en 2003 [Park 2003]. Quelques modifications y ont été apportées pour accroître la précision.

Tout d'abord, la zone de *non*-recouvrement des deux vues est éliminée de la vue courante avant le recalage avec la vue précédente. En théorie, cette zone est inconnue puisque le déplacement n'a pas encore été calculé. Cependant, comme l'hypothèse d'un faible déplacement d'une vue à l'autre a été faite, nous considérons les points placés au centre de la vue comme appartenant à la zone de recouvrement. Qui plus est, la plupart des points situés en périphérie de la surface auraient été difficilement mis en correspondance. La boîte englobante est définie comme étant la projection de la vue 3D sur le plan XY de la caméra orthogonal à l'axe focal. En pratique, un rectangle dont la dimension est égale à 80% de la boîte englobante, centré sur l'image, représentera l'image de la zone de recouvrement. Tous les points qui lui sont extérieurs seront éliminés.

Seul un échantillon des points appartenant à cette boîte englobante seront conservés pour le recalage. Il existe différents types d'échantillonnage : uniforme [Turk 1996, Matsuda 2001], aléatoire [Matsuda 1996], normal [Rusinkiewicz 2001], notamment. L'échantillonnage est souvent utilisé pour accélérer le processus de recalage, mais il permet également d'améliorer la précision de celui-ci en sélectionnant judicieusement les points à conserver. En effet, considérons une surface plate aux bords perpendiculaires (un créneau). Si tous les points sont aveuglément considérés, le résultat du recalage sera peu précis du fait de l'influence minimale des points aux bords par rapport aux points (beaucoup plus nombreux) qui sont sur la surface. L'objectif visé par l'échantillonnage est donc de sélectionner des points représentatifs de la surface : nous avons opté pour l'échantillonnage normal. Tous les points sont préalablement transformés dans un espace bidimensionnel défini par α et β comme suit :



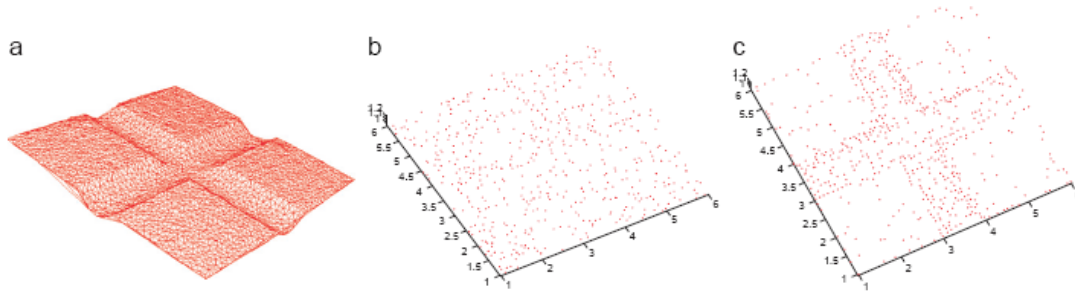


FIGURE 1.20: Exemples d'échantillonnage. (a) Surface initiale. (b) Echantillonnage aléatoire. (c) Echantillonnage normal

$$\begin{aligned}\alpha &= \arctan \left(n_x, \sqrt{n_z^2 + n_y^2} \right) \\ \beta &= \arctan 2(n_y, n_z)\end{aligned}\tag{1.15}$$

où α et β sont les coordonnées dans l'espace normal, et n_x , n_y et n_z sont les trois composantes du vecteur normal de chaque point. Tous les points sont ainsi placés sur une grille 2D. Un unique point par case de la grille est arbitrairement conservé. Un exemple d'échantillonnage est donné en figure 1.20.

Nous utilisons une méthode récursive pour calculer l'intersection entre les lignes et les surfaces. En premier lieu, les points sélectionnés sont projetés orthographiquement sur le plan image. Une grille de 50×50 cellules est dimensionnée de manière à ce qu'elle puisse recueillir tous les points projetés. Un point p_0 de la vue courante est projeté dans une cellule de la grille dans laquelle nous chercherons le point le plus proche q_{p_0} . La projection de q_{p_0} sur le vecteur normal à p_0 donne un nouveau point p_1 qui est une approximation de l'intersection recherchée. Nous chercherons récursivement des nouveaux points p_i jusqu'à ce que la norme $|p_i - q_{p_i}|$ soit plus petite qu'un seuil donné (figure 1.21). Le processus est répété pour tous les points de la vue courante.

Une fois la mise en correspondance ainsi établie, on minimise l'expression suivante pour calculer le déplacement séparant les deux surfaces :

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \|m_i - R p_i - t\|^2\tag{1.16}$$

où N_p est le nombre de correspondances ; m_i , l'ensemble des points sélectionnés de la vue courante ayant un correspondant dans la vue précédente ; p_i , les correspondants des m_i ; R et t , la matrice de rotation et le vecteur de translation respectivement. L'algorithme prend fin quand la moyenne de l'erreur quadratique est inférieure à un seuil donné. Les vues sont recalées consécutivement. L'erreur de recalage est donc propagée tout au long de la séquence.

Détection de cycle. Nous cherchons à détecter les cycles provoqués par le passage du scanner sur une même portion de l'objet. A chaque fois que deux vues sont recalées, la transformation Euclidienne (le déplacement) qui les sépare est calculée et le lien établi. Si nous représentons notre surface reconstruite sous forme d'un graphe, avec les portions de surface aux sommets et les déplacements aux arêtes, il est possible de considérer les liens entre les vues comme des chemins dans le graphe. Le déplacement du scanner entre une vue et une autre sera donnée par le produit des transformations Euclidiennes successives. Ainsi, si la translation entre deux vues non-voisines est quasi nulle, alors un cycle est potentiellement détecté. Toutefois, ce critère n'est pas suffisant, puisque, notamment quand la rotation est importante, une séquence de vues donnant une translation proche de zéro ne forme pas toujours un cycle. De ce fait, un second critère sera utilisé : il faudra en plus, que les deux vues aient une zone de recouvrement assez importante.

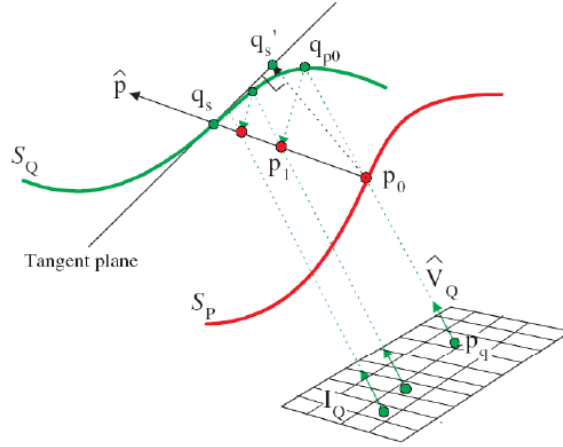


FIGURE 1.21: Calcul de l'intersection entre le plan tangent et la surface par la méthode de PARK

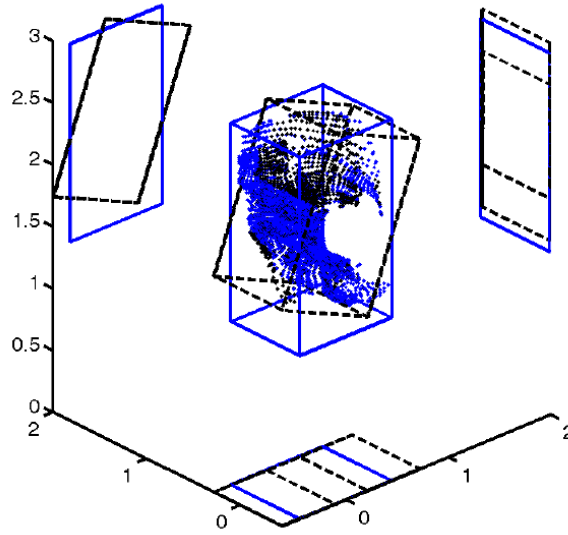


FIGURE 1.22: Recouvrement et boîtes englobantes

Afin d'éviter d'avoir à recalculer finement deux vues pour estimer leur zone de recouvrement, nous avons proposé une méthode rapide basée sur le recouvrement dans une boîte englobante, représentant ici une approximation de l'enveloppe convexe des deux surfaces. La boîte englobante d'une surface donnée est définie comme étant le plus petit parallélépipède contenant tous les points de la surface. Ces boîtes sont projetées sur les plans XY , XZ et YZ (figure 1.22) : chaque plan contient donc deux projections des boîtes englobantes (une par surface à comparer) ce qui permet de calculer trois zones de recouvrement bidimensionnelles. Si la plus grande de ces zones dépassent un seuil donné et si la matrice de déplacement entre les deux vues extrêmes est proche de l'identité, alors un cycle est détecté. Il est à noter que nous avons choisi la plus grande des zones et pas le cumul des trois zones afin de pouvoir estimer la zone de recouvrement même pour une surface "plate".

Minimisation de cycle. La dernière étape consiste à minimiser l'erreur de recalage à l'intérieur d'un cycle. Nous devons considérer que la première et la dernière vue coïncident. En pratique, cela n'arrive jamais, c'est pourquoi nous ajoutons une vue virtuelle au cycle qui n'est autre que la première vue recalée avec la dernière. C'est par un algorithme de minimisation du type Levenberg-Marquardt que la minimisation globale est effectuée. Les paramètres de minimisation sont les matrices de rotation (représentées par des quaternions) et les vecteurs de translation entre deux vues consécutives. La fonction à minimiser est la somme des distances entre les points de correspondance :

$$\min \left\{ \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{k=1}^{N_p} \left(P_i(k) - T_j^i \times P_j(k) + T_i^j \times P_i(k) - P_j(k) \right) \right\} \quad (1.17)$$

où $P_i(k)$ et $P_j(k)$ sont les points des vues i et j en correspondance; N_p est le nombre de correspondances; N est le nombre de vues; T_j^i et T_i^j sont les déplacements qui transforment les points de i à j et de j à i respectivement :

$$T_j^i = \prod_{k=i+1}^j T_k^{k-1} \quad (1.18)$$

et

$$T_i^j = \left(\prod_{k=j}^{N-1} T_{k+1}^k \right) T_1^N \left(\prod_{k=2}^i T_k^{k-1} \right) \quad (1.19)$$

avec $j > i$.

La minimisation est effectuée sous la contrainte du produit des matrices du cycle à l'identité, ainsi exprimée :

$$\varepsilon_{cr} = \varepsilon_R + s_f \varepsilon_T \quad (1.20)$$

où ε_R est la contrainte de rotation; ε_T , la contrainte de translation et s_f le facteur d'échelle permettant d'exprimer la translation sur la même gamme que les paramètres de rotation.

$$\varepsilon_R = \sum (\text{abs}(R_{accum} - I_{3 \times 3})) \quad (1.21)$$

où R_{accum} est le produit de toutes les matrices de rotation.

La contrainte de translation est définie par :

$$\varepsilon_T = |t| \quad (1.22)$$

où t est le vecteur de translation séparant la première et la dernière vue du cycle.

Bien que cette approche soit rapide, comparée à beaucoup d'autres méthodes, si le cycle contient beaucoup de vues, la charge de calcul pour établir les correspondances peut être très élevée. Nous avons donc proposé une approche alternative, plus rapide. Pendant le recalage des paires de vue, les correspondances sont mémorisées. Quand un cycle est détecté, les paramètres de déplacement sont minimisés en prenant en compte uniquement les vues consécutives dont les correspondances ont été mémorisées. Une équation simplifiée peut être obtenue :

$$\min \left\{ \sum_{i=1}^{N-1} \sum_{k=1}^{N_p} \left(P_i(k) - T_{i+1}^i \times P_{i+1}(k) + T_i^{i+1} \times P_i(k) - P_{i+1}(k) \right) \right\} \quad (1.23)$$

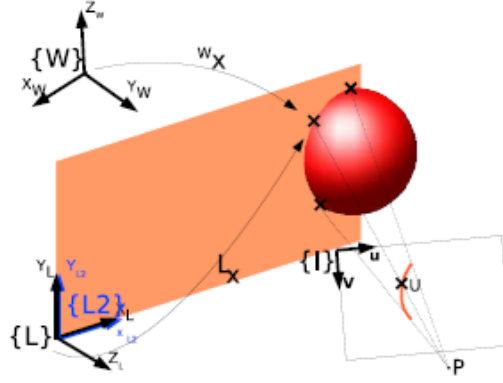


FIGURE 1.23: Modèle de Chen & Kak

1.3.2 Calibrage du Système Laser-Caméra

De nombreuses méthodes ont été proposées pour le calibrage d'une caméra [Salvi 2002] ou pour celui d'un système laser-caméra [Forest 2004] ; c'est une étape essentielle de la mesure 3D par capteur visuel. Pour les systèmes rigides (c'est-à-dire dont la pose relative de la source lumineuse et de la caméra est invariante dans le temps), on a coutume d'utiliser le modèle proposé par CHEN et KAK [CHEN 1987]. La figure 1.23 illustre l'idée qui sous-tend ce modèle. On part de l'image de la ligne laser dans le repère $\{I\}$. Cette image peut être projetée par homographie dans le système de coordonnées propre au plan laser $\{L2\}$, lui-même positionné dans l'espace à trois dimensions $\{L\}$. Le tout peut être maintenant exprimé dans le repère-monde $\{W\}$. Ainsi, un système laser-caméra peut être représenté par une unique matrice 4×3 (éq. 1.24).

$$\begin{aligned} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} &= {}^W \mathbf{T}_L \cdot {}^L \mathbf{T}_{L2} \cdot {}^{L2} \mathbf{H}_I \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = {}^W \mathbf{T}_L \cdot \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} e_{11} & e_{12} & e_{13} \\ e_{21} & e_{22} & e_{23} \\ e_{31} & e_{32} & e_{33} \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \dots \\ &\dots = \begin{bmatrix} t_{11} & t_{12} & t_{13} \\ t_{21} & t_{22} & t_{23} \\ t_{31} & t_{32} & t_{33} \\ t_{41} & t_{42} & t_{43} \end{bmatrix} \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = {}^W \mathbf{T}_I \cdot \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \end{aligned} \quad (1.24)$$

En lumière structurée, la source lumineuse ne "voit" pas la scène. Elle illumine au contraire des points 3D dont il est difficile de mesurer précisément les coordonnées. C'est pourquoi nous avons opté pour la méthode du quadrangle complet, initialement proposée par [Forest 2004] et généralisée par nos soins pour prendre en compte la projection de plusieurs lignes laser [Matabosch 2008].

Le principe est simple. Le système laser-caméra observe un quadrangle complet, dont les coordonnées des sommets A , B , C et D sont connues et dont découlent les coordonnées du point de fuite G (voir figure 1.24). La ligne laser vient couper le quadrangle en deux nouveaux points P_A et P_B . Sachant que le birapport de quatre points est invariant aux transformations projectives, on en déduit que :

$$\{A, B, G, P_A\} = \{A', B', G', P'_A\} \text{ et } \{D, C, G, P_B\} = \{D', C', G', P'_B\} \quad (1.25)$$

Les seules inconnues de ces équations sont les points P_A et P_B , qu'il est facile de calculer. On peut désormais aisément faire la mise en correspondance 2D (un ensemble de points P'_A et P'_B dans l'image)/3D (l'ensemble correspondant des points P_A et P_B calculés). Le quadrangle est translaté le long de l'axe z et un jeu de points est calculé pour chaque nouvelle position. Il est également

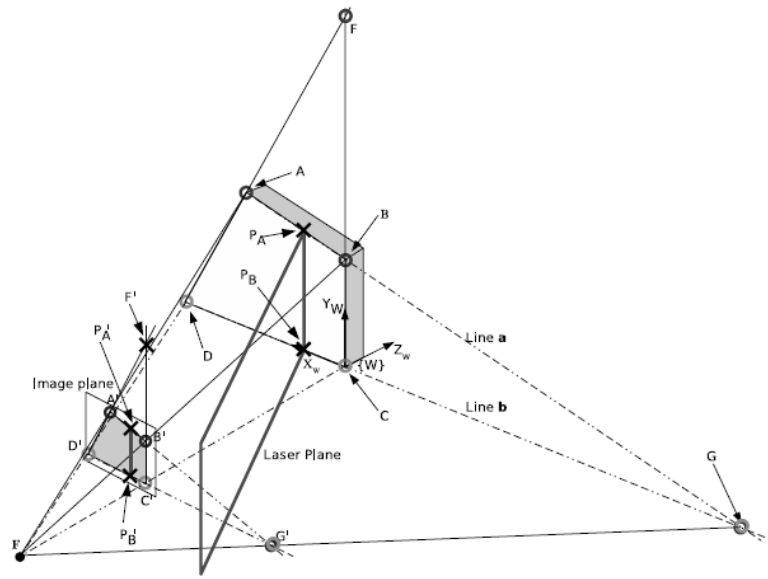


FIGURE 1.24: Birapport et quadrangle complet utilisés pour faciliter la mise en correspondance 2D/3D

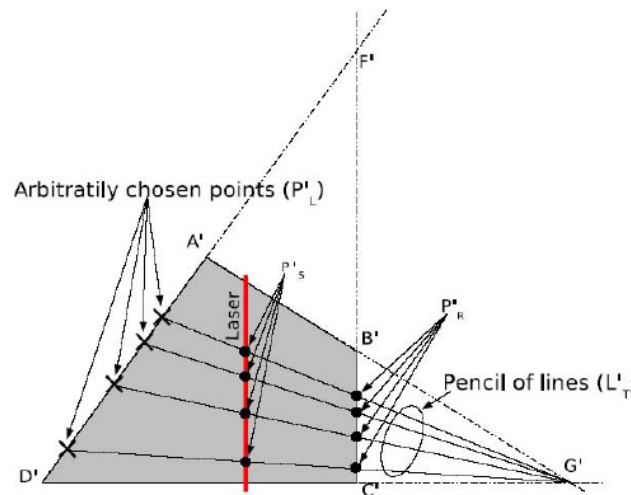


FIGURE 1.25: Génération de points supplémentaires pour améliorer la précision du calibrage

possible, comme illustré sur la figure 1.25, de générer facilement des points supplémentaires de manière à rendre l'estimation de la matrice de calibrage plus robuste.

1.3.3 Segmentation et Identification des Lignes Laser

La plupart des motifs structurés composés de plusieurs lignes sont codés; l'identification ou l'indexation des lignes pour la mise en correspondance est donc directement donnée par leur décodage. Ici au contraire, puisque nous projetons 19 lignes laser de couleur rouge, nous avons à faire face à un problème de lumière structurée *non codée* pour lequel il faut trouver une stratégie d'indexation des lignes afin de rendre efficace la mise en correspondance. Très peu de travaux sont

publiés sur le sujet. Il est clair qu'il est nécessaire de poser des hypothèses, éventuellement fortes mais raisonnables, sur les conditions de prise de vue. En premier lieu, nous supposons qu'une ligne laser ne peut pas apparaître deux fois dans une même ligne de l'image, ce que ROBINSON et al. [Robinson 2004] formule comme étant la contrainte d'inclinaison commune (*common inclination constraint*) – cette hypothèse est vérifiée quand la ligne projetée et réfléchie du faisceau partagent toujours une "inclinaison commune". Nous posons également la contrainte locale de douceur. De ce fait, s'il est possible de détecter moins de ligne laser que de lignes projetées, il devient impossible de les identifier de manière désordonnée. L'algorithme que nous avons proposé est une variante de celui de ROBINSON [Robinson 2004] adapté au cas où le nombre de plans laser projeté est connu.

L'image doit être préalablement segmentée, c'est-à-dire que les lignes laser, formant crêtes, doivent être précisément détectées. On utilise pour cela un algorithme classique de détection de pics sub-pixel. Celui que nous avons implémenté reprend les grandes lignes de celui décrit dans [Forest 2004]. Au final, nous obtenons une image contenant des -1 à la position des pics et des 0 partout ailleurs. Ainsi, si l'on détecte 19 pics sur une même ligne de l'image, on peut affecter de manière ordonnée l'étiquette i aux au i^e pic. Au contraire, si aucune ligne de l'image ne contient 19 pics, l'algorithme ne peut pas s'y appliquer et l'image devient tout simplement inutilisable. Quand toutes les lignes contenant 19 pics sont détectées, on en choisit une pour lancer l'algorithme. On suit (*tracking*) les 19 lignes de façon à étiqueter tous les pics d'un même profil, tout en respectant les règles d'indexation. Si un pic est détecté à moins de 1,5 pixels de distance par rapport au pic de la ligne-image précédent, alors les deux porteront la même étiquette. Le processus prend fin quand il ne reste plus aucun pic à étiqueter ou quand les contraintes ne sont plus respectées. Quand tous les pics d'une même ligne ont été identifiés, on passe aléatoirement à une autre ligne. On ré-itére le processus jusqu'à ce que l'étiquetage ne soit plus possible ou que tous les pics soient étiquetés. Les règles et contraintes sont données ci-dessous (on dira qu'un pic est classé "comme plan i " s'il est identifié comme appartenant au i^e plan laser du motif) :

Contraintes d'indexation :

- Tous les pics i peuvent être classés comme plans j , avec $j < i$
- Tous les pics i dans une ligne de n pics peuvent être classés comme plans j , avec $n - i > 19 - j$
- Le pic i peut être classé comme plan laser j ssi tous les pics k (de 1 à $i-1$) sont classés comme plan l , avec $l < j$
- Le pic i dans une ligne de n pics peut seulement être classé comme plan j si tous les pics k (de $i-1$ à n) sont classés comme plan l , avec $l > j$

Règles d'indexation :

- Le pic i est classé comme plan i quand le pic j est classé comme plan j , avec $j > i$
- Le pic i , dans une ligne de n pics, est classé comme plan $k = 19 + i - n$ si le pic j est classé comme plan l , $n - j = 19 - l$ et $i > j$
- Le pic i est classé comme plan $l = m + i - j$ si les pics j et k sont classés comme plans m et n respectivement et $n - m = k - j$

La figure 1.26 présente les résultats d'un processus d'indexation et la figure 1.27 illustre, par l'exemple, l'algorithme proposé.

1.3.4 Résultats Expérimentaux

Le scanner 3D que nous avons développé est composé d'une caméra CCD, d'un objectif de 6mm et d'un émetteur laser LASIRIS de 635nm auquel on a adjoint une tête optique permettant la projection de 19 lignes. Le système capture des images d'une résolution de 1200×900 pixels. Les axes optiques du laser et de la caméra forment un angle de 60° et la distance qui sépare les deux éléments est d'environ 20cm. Dans cette configuration, la distance d'acquisition idéale se situe entre 10 et 30cm de la caméra (figure 1.28).

Pour évaluer la précision du scanner, nous avons placé la tête de mesure sur une table XY et capturé des images du quadrangle utilisé pour le calibrage à différentes profondeurs par incréments

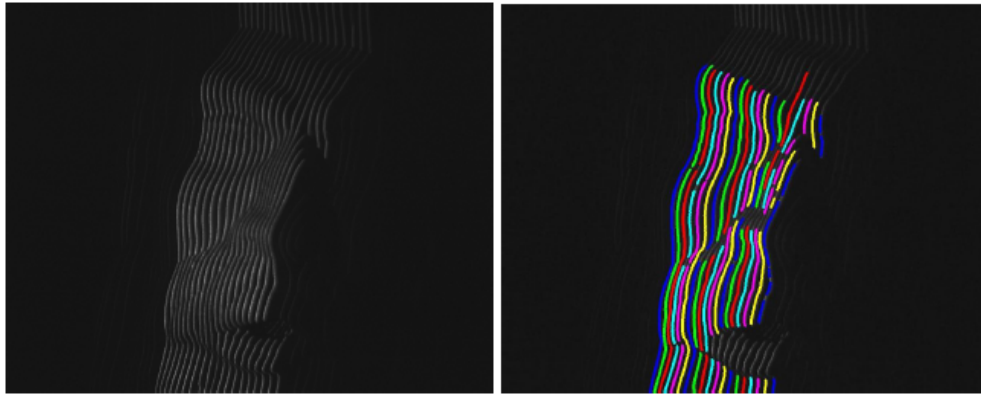


FIGURE 1.26: Identification des lignes laser. A gauche : image originale. A droite : image étiquetée

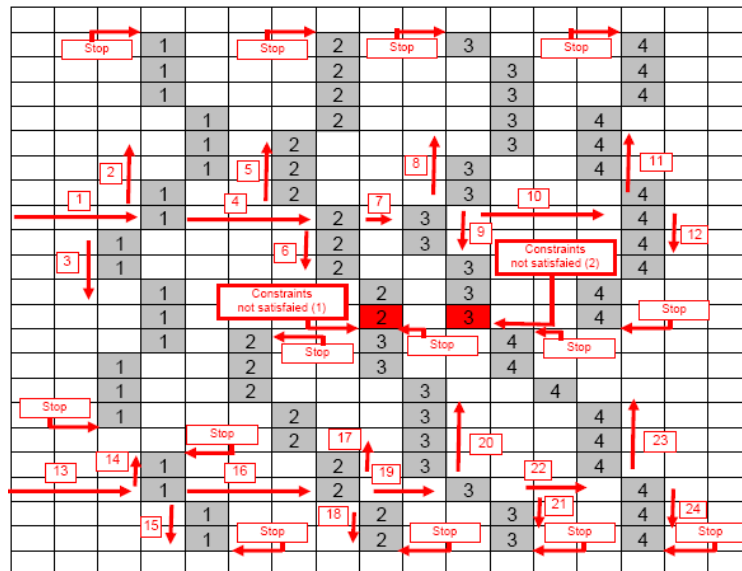


FIGURE 1.27: Exemple de d'algorithme d'identification

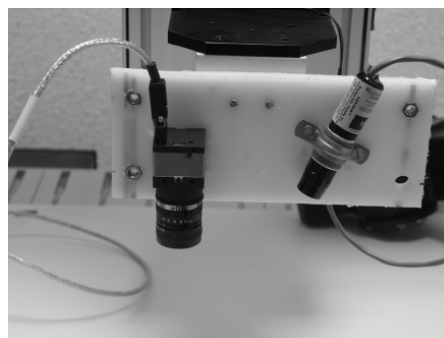


FIGURE 1.28: Premier prototype du scanner à main

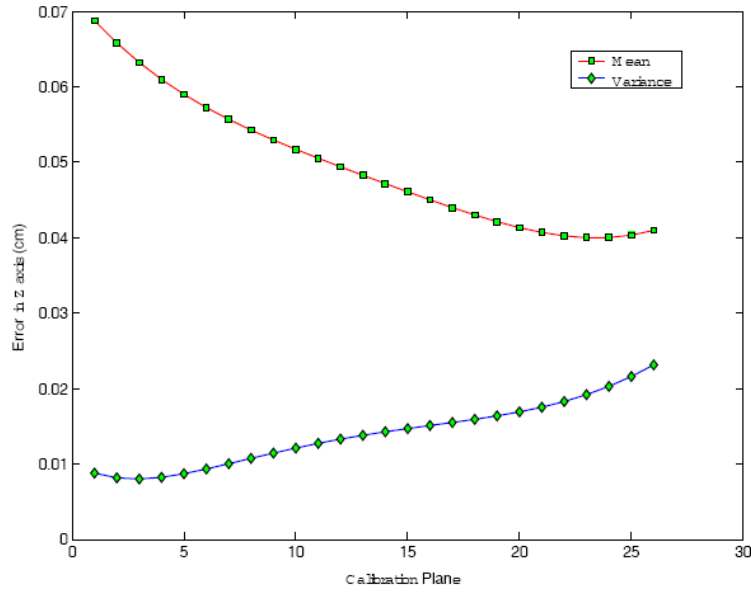


FIGURE 1.29: Précision en fonction de la distance

View	Detected peaks	indexed peaks	%Indexing
1	13628	12746	93.53
2	14936	14714	98.51
3	15061	14893	98.88
4	15007	14556	96.99
5	15548	15142	97.39
6	13670	11138	81.48
7	16386	11758	71.76
8	12846	10288	80.09
9	11850	8495	71.69
10	12874	10744	83.46

TABLE 1.4: Pourcentage de points indexés

de 2mm. Nous avons estimé l'erreur en comparant le z donné par la reconstruction et le z de référence (figure 1.29). Si l'objet est trop proche du capteur, le flou rend la mesure moins précise. L'erreur minimale est de 0,04cm environ.

Nous avons également évalué les performances de notre algorithme d'indexation des lignes laser (tableau 1.4). Le pourcentage est très élevé ; mieux, 100% des points indexés le sont correctement. Un jeu d'images utilisées pour ces tests est présenté figure 1.30. Au regard des applications visées, les conditions de prise de vue et les surfaces à numériser sont réalistes. La figure 1.33 présente d'autres résultats obtenus.

Nous avons finalement comparé notre méthode de recalage multivue avec la méthode de SHARP. Pour cela, nous avons utilisé le modèle Beethoven de l'INRIA (figure 1.31) et simulé son *scanning*. Les résultats sont compilés dans le tableau 1.5. Nos performances, en termes de précision, sont sensiblement meilleures, même pour ceux obtenus avec l'approche simplifiée. Nous présentons en figure 1.32 le recalage complet du modèle synthétique de l'INRIA.

Nous présentons d'autres résultats en figures 1.33, 1.34 et 1.35. La deuxième de ces figures

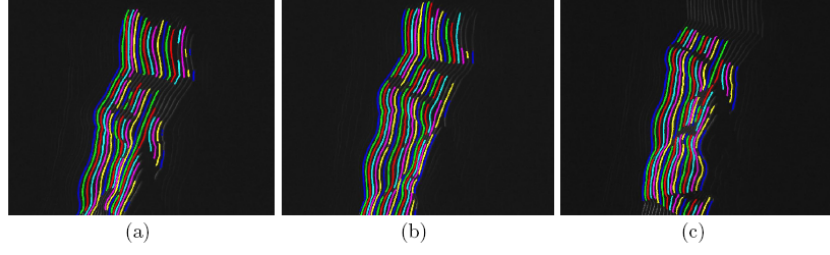


FIGURE 1.30: Exemples d'indexation de lignes laser

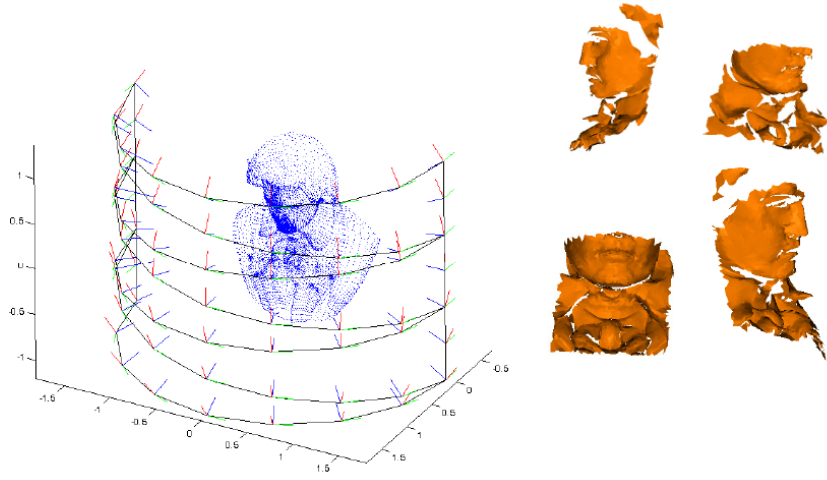


FIGURE 1.31: Principe et résultats sur le modèle simulé de l'INRIA

Scene	Our Method			Fast Variant			Sharp's Method		
	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE
$\sigma=0$	0.516	0.008	0.003	0.339	0.079	0.001	0.511	0.074	0.002
	1.120	0.004	0.003	0.867	0.191	0.006	1.006	0.034	0.002
$\sigma=1.25\%$	0.675	0.154	0.004	1.177	0.459	0.006	2.225	4.403	0.026
	1.115	0.305	0.005	1.388	0.265	0.006	1.385	1.196	0.051
$\sigma=2.5\%$	1.1286	0.4698	0.005	1.202	0.316	0.005	1.472	1.367	0.001
	1.1905	0.2149	0.005	1.410	0.217	0.005	1.202	0.704	0.002
$\sigma=3.75\%$	0.246	0.056	0.003	1.552	0.875	0.007	2.601	3.485	0.026
	0.732	0.024	0.002	1.169	0.425	0.006	1.134	2.106	0.046
$\sigma=5.0\%$	1.570	0.890	0.005	1.533	0.828	0.007	2.753	3.126	0.017
	1.284	0.682	0.005	1.144	0.497	0.006	1.212	2.225	0.020

TABLE 1.5: Comparaison des résultats du recalage multivue

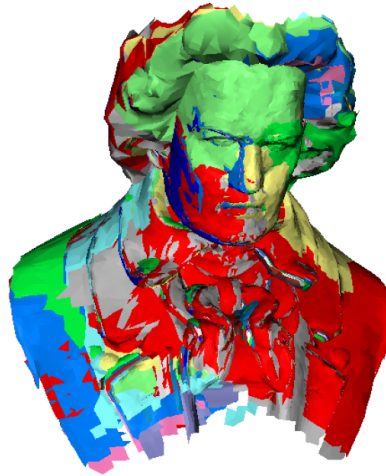


FIGURE 1.32: Recalage complet du modèle Beethoven

illustre qualitativement l'apport de notre méthode. La dernière figure montre une séquence complète de vues partielles obtenues avec notre prototype.

1.4 Dispositif de Texturation Virtuelle

Ces travaux avaient pour objectif la conception d'un dispositif pour la texturation virtuelle d'objets réels. Les applications visées sont l'architecture, la muséographie ou le prototypage augmenté. Nous nous sommes tout naturellement orienté vers un système projecteur-caméra permettant à la fois la capture de l'objet et la projection de texture sur sa surface. Nous verrons qu'il est possible d'effectuer ces deux étapes simultanément et de manière imperceptible.

Comment appliquer correctement une texture virtuelle sur un objet tridimensionnel réel ? Comment calculer l'image à projeter pour qu'elle "colle" parfaitement à la surface à colorer ? Nous sommes ici sur une problématique qui se trouve à la frontière entre vision par ordinateur et réalité augmentée. Notre ambition n'était pas de rivaliser avec les salles d'immersion virtuelle, mais de proposer un dispositif léger et des outils adaptés de vision par ordinateur pour répondre simplement et automatiquement au problème. Pour la polychromie de la Cathédrale d'Amiens (figure 1.36), deux gros projecteurs fixés au sol sont utilisés pour colorer la façade. Les images ont été calculées hors-ligne et le dispositif – comme, bien évidemment, la Cathédrale – est fixe. C'est cette méthodologie que nous souhaitons étudier, en y apportant plus de flexibilité et en laissant plus de degrés de liberté au système. En 2001, RASKAR et al. ont proposé un système de texturation virtuelle [Raskar 2001]. L'image la plus célèbre issue de ces travaux est sans doute celle du Taj-Mahal (figure 1.37). Le système qu'il propose n'est malheureusement pas encore complètement automatique. C'est sur ce point bien précis que nous avons concentré nos efforts. De plus, pour permettre au système de capturer l'information tridimensionnelle nécessaire au positionnement relatif de l'objet par rapport au projecteur et de projeter simultanément la texture virtuelle, nous avons été amené à étudier les techniques dites de lumière structurée imperceptible, notamment celles permettant la projection intégrée d'une mire codée (*embedded pattern*) [Cotting 2004]. Nous introduirons ainsi le concept de projection simultanée de *lumière structurée* et de *lumière texturante*.

1.4.1 Système Projecteur-Caméra

Le système projecteur-caméra est composé d'un unique projecteur de lumière et de deux caméras, le projecteur étant placé entre les deux caméras. Les principes géométriques que nous allons

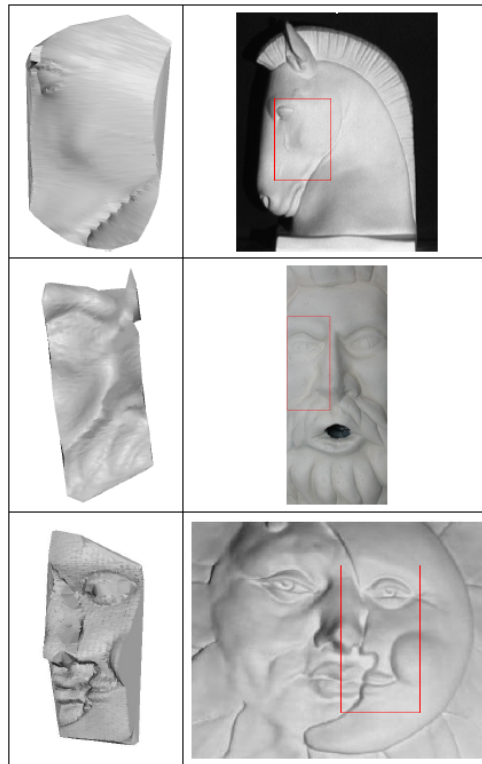


FIGURE 1.33: Exemple de résultats

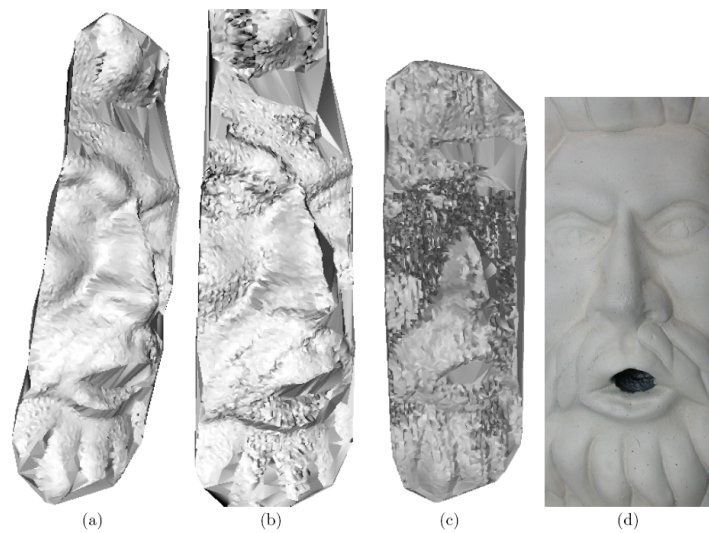


FIGURE 1.34: Résultats de recalage. (a) Notre méthode. (b) Méthode de SHARP. (c) Alignement mécanique. (d) Objet

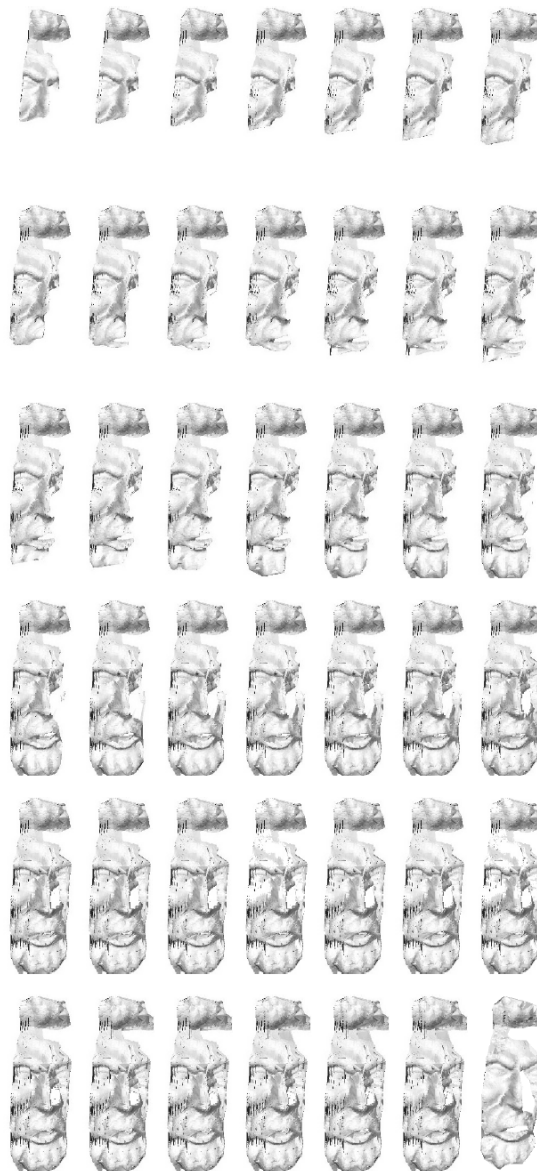


FIGURE 1.35: Séquence de vues partielles



FIGURE 1.36: Polychromie de la Cathédrale d'Amiens

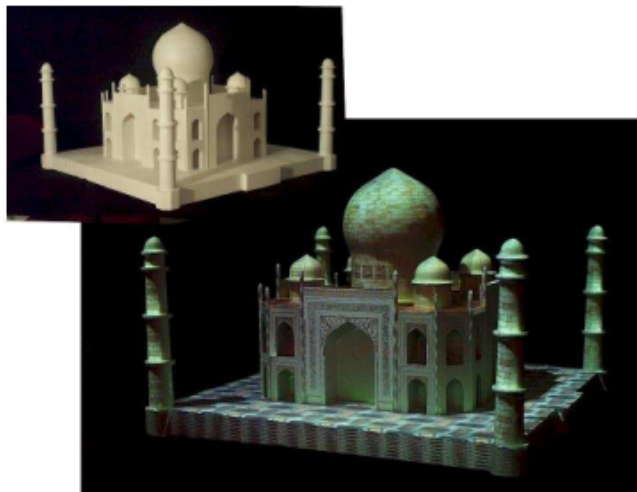


FIGURE 1.37: Texturation du Taj-Mahal par RASKAR

évoqués permettent de généraliser à n projecteurs et m caméras et à tout positionnement relatif. Nous supposons que les objets à texturer sont blancs. Il est entendu que dans de telles conditions, il est difficile d'extraire des points d'intérêt et d'effectuer une mise en correspondance suffisamment dense pour que la projection d'un motif colore assez finement l'objet. Il est donc nécessaire d'employer une méthode de stéréovision active, basée sur la projection de lumière structurée, pour reconstruire ou localiser l'objet.

1.4.1.1 Géométrie du Système

Notre dispositif est composé d'un seul projecteur et de deux caméras. Nous avons souhaité cependant rester assez général pour éviter la dépendance au matériel et prendre en compte des systèmes multi-projecteur (au moins géométriquement et même si nous n'avons pas encore expérimenté ce cas de figure). Le projecteur et les caméras sont modélisés par le sténopé. Nous proposons une méthode simple de calibrage et avons étudié la géométrie épipolaire des couples projecteur-caméra et la géométrie trifocale du système expérimental.

Calibrage du système. Nous avons choisi une méthode de calibrage par calcul d'homographies pour le couple de caméras. Pour le projecteur, nous avons contourné la difficulté en utilisant les données de calibrage des caméras, au risque de propager l'erreur d'un élément du système à l'autre. Cinq étapes sont nécessaires, dont seulement une nécessite l'intervention d'un opérateur :

1. calibrage par calcul d'homographies des deux caméras ;
2. projection d'un motif sur une mire blanche ;
3. extraction des points d'intérêt du motif et triangulation à l'aide des deux caméras calibrées ;
4. utilisation des points 3D calculés pour la mise en correspondance 2D/3D du projecteur ;
5. estimation des paramètres du projecteur par la méthode de Faugeras.

La méthode de calibrage utilisée ne requiert qu'une mire plane et le calcul des homographies permettant de passer du plan image aux plans de la mire dans différentes positions et orientations pour estimer les paramètres intrinsèques [Sturm 1999, Zhang 1999]. La projection d'un point $\mathbf{M} = [x \ y \ z \ 1]^T$ de l'espace sur le plan image est donnée par :

$$\mathbf{m} \approx \mathbf{K} [\mathbf{R} \ \mathbf{t}] \mathbf{M} \quad (1.26)$$

où $\mathbf{m} = [u \ v \ 1]^T$; $\mathbf{K} = \begin{bmatrix} f_x & s & u_0 \\ 0 & f_y & v_0 \\ 0 & 0 & 1 \end{bmatrix}$ est la matrice des paramètres intrinsèques ; \mathbf{R} et

\mathbf{t} forment la matrice de pose de la caméra dans un repère de référence ; le symbole \approx représente l'égalité projective. En vue de simplifier la notation, on considère que la mire de calibration plane est telle que $z = 0$:

$$\mathbf{m} \approx \mathbf{K} [\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3 \ \mathbf{t}] \begin{bmatrix} x \\ y \\ 0 \\ 1 \end{bmatrix} \approx \mathbf{K} [\mathbf{r}_1 \ \mathbf{r}_2 \ t] \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (1.27)$$

où les \mathbf{r}_i correspondent aux colonnes de \mathbf{R} . Cette équation exprime la relation homographique reliant les points de la mire aux points images :

$$\mathbf{m} \approx \mathbf{H} \bar{\mathbf{M}} \quad (1.28)$$

où $\bar{\mathbf{M}}$ est le vecteur des coordonnées du point \mathbf{M} sans la troisième coordonnée et $\mathbf{H} = [h_1 \ h_2 \ h_3]$ est la matrice d'homographie 3×3 . Cette matrice peut s'exprimer en fonction des paramètres intrinsèques de la caméra :

$$[h_1 \ h_2 \ h_3] \approx \mathbf{K} [\mathbf{r}_1 \ \mathbf{r}_2 \ t] \quad (1.29)$$

Comme les vecteurs \mathbf{r}_i sont orthonormés, on en déduit que :

$$\mathbf{h}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_2 = 0 \quad (1.30)$$

et :

$$\mathbf{h}_1^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_1 = \mathbf{h}_2^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{h}_2 \quad (1.31)$$

Une homographie donne deux contraintes sur les paramètres intrinsèques. On remarque que $\omega \approx \mathbf{K}^{-T} \mathbf{K}^{-1} = \lambda \mathbf{K}^{-T} \mathbf{K}^{-1}$ est l'image de la conique absolue. En posant :

$$\omega \approx \begin{bmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{21} & \omega_{22} & \omega_{123} \\ \omega_{31} & \omega_{32} & \omega_{33} \end{bmatrix} = \dots \quad (1.32)$$

$$\dots = \begin{bmatrix} \frac{1}{f_x^2} & \frac{-s}{f_x^2 f_y} & \frac{sv_0 - u_0 f_y}{f_x^2 f_y} \\ \frac{-s}{f_x^2 f_y} & \frac{s^2}{f_x^2 f_y^2} + \frac{1}{f_y^2} & \frac{-s(sv_0 - u_0 f_y)}{f_x^2 f_y^2} - \frac{v_0}{f_y^2} \\ \frac{sv_0 - u_0 f_y}{f_x^2 f_y} & \frac{-s(sv_0 - u_0 f_y)}{f_x^2 f_y^2} - \frac{v_0}{f_y^2} & \frac{(sv_0 - u_0 f_y)^2}{f_x^2 f_y^2} + \frac{v_0^2}{f_y^2} + 1 \end{bmatrix}$$

Les paramètres intrinsèques (et le facteur d'échelle λ) peuvent donc être extraits de ω . Les deux caméras sont ainsi calibrées et utilisées pour fournir les points 3D nécessaires à la calibration du projecteur, par la méthode de FAUGERAS-TOSCANI [Faugeras 1986] que nous ne rappellerons pas ici.

Géométrie épipolaire et trifocale. Nous avons étudié la géométrie épipolaire reliant les paires de vues (les deux images caméras et le motif structuré), ainsi que la géométrie trifocale, ceci pour à la fois réduire la zone de recherche lors de la mise en correspondance et faciliter la synthèse de vue par le transfert de point. Il n'est pas utile de décrire la géométrie épipolaire [Zhang 1998, Armangué 2003] ou trifocale [Ressl 2003] ni de détailler l'estimation de la matrice fondamentale ou du tenseur trifocal ; nous allons en rappeler quelques grandes lignes.

Soient deux points de correspondance \mathbf{m} et \mathbf{m}' (ces points peuvent appartenir aux images ou au motif) ; la relation suivante exprime la contrainte épipolaire :

$$\mathbf{m}'^T \mathbf{F} \mathbf{m} = 0 \quad (1.33)$$

où \mathbf{F} est la matrice fondamentale, de rang 2, définie à un coefficient multiplicatif près et dépendant à la fois de la pose relative de la vue 1 par rapport à la vue 2 et des paramètres intrinsèques des caméras. Cette matrice possède 7 degrés de liberté et contient toute la géométrie épipolaire. Elle est indépendante de la scène observée. Elle relie un point de la première image à sa ligne épipolaire dans la seconde image :

$$\mathbf{l}'_m = \mathbf{F} \mathbf{m} \quad (1.34)$$

Cette contrainte peut être généralisée à un dispositif comportant plus de deux vues. On parle alors de tenseur trifocal, quadrifocal, n -focal. Nous disposons de deux images et d'un motif structuré, donc de trois vues différentes. Il est tout naturel de chercher à modéliser le système grâce au tenseur trifocal. Imaginons les trois matrices de projection suivantes :

$$\begin{aligned} \mathbf{P} &= \begin{bmatrix} \mathbf{I}_{3 \times 3} & \mathbf{0} \end{bmatrix} \\ \mathbf{P}' &= \mathbf{K}' \begin{bmatrix} \mathbf{R}' & \mathbf{t}' \end{bmatrix} = \begin{bmatrix} \mathbf{a}'_i & \mathbf{v}' \end{bmatrix} \\ \mathbf{P}'' &= \mathbf{K}'' \begin{bmatrix} \mathbf{R}'' & \mathbf{t}'' \end{bmatrix} = \begin{bmatrix} \mathbf{b}''_i & \mathbf{v}'' \end{bmatrix} \end{aligned} \quad (1.35)$$

Le tenseur trifocal, tableau de $3 \times 3 \times 3$ éléments, peut s'exprimer ainsi :

$$\mathbf{T}_i^{kj} = \mathbf{v}'_j \mathbf{b}''_i^k - \mathbf{v}''_k \mathbf{a}'_i^j \quad (1.36)$$

\mathbf{T}_i^{kj} possède 27 éléments et seulement 18 degrés de liberté. Le tenseur peut également être calculé à partir des trois matrices fondamentales reliant les caméras et le projecteur. Nous avons

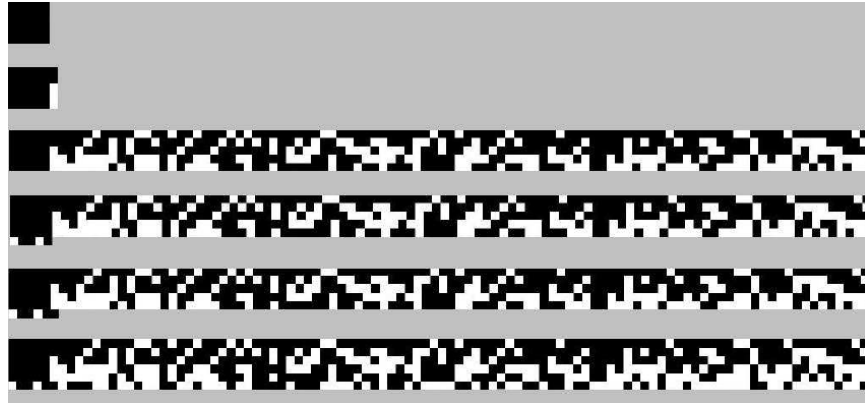


FIGURE 1.38: Création de la mire binaire

donc deux possibilités : soit estimer les matrices fondamentales et les compiler dans le tenseur ; soit estimer directement le tenseur. Pour des raisons pratiques, nous avons opté pour la première de ces possibilités tout en utilisant le formalisme tensoriel, qui permet plus directement d'envisager une généralisation à n vues.

1.4.1.2 Conception d'une Mire Codée

Parmi toutes les méthodes de codification d'une lumière structurée que nous avons rappelée précédemment, nous avons identifié deux critères pertinents :

1. la codification doit être directe, afin d'envisager une texturation dynamique de l'objet ;
2. la mire doit être binaire afin de répondre aux critères nécessaires à son intégration dans une stratégie de lumière structurée imperceptible.

Nous avons choisi de concevoir une mire pseudo-aléatoire, de type *perfect sub-map* [Griffin 1992], composée de motifs élémentaires binaire de 5×5 pixels (appelés cellules). Des règles simples, plutôt qu'une recherche exhaustive dans un graphe comme on le fait habituellement, nous ont permis d'en élaborer un :

1. Une cellule entièrement noire est choisie pour le coin supérieur gauche.
2. Une liste aléatoire de toutes les combinaisons possibles pour une colonne de 5 pixels est créée. On ajoute la première combinaison et on vérifie que tout carré de 5×5 est unique. Si l'unicité n'est pas respectée, on vérifie avec la deuxième combinaison et ainsi de suite. On ajoute des colonnes jusqu'à atteindre la limite en résolution.
3. On effectue la même procédure qu'en 2, mais pour une ligne de 5 pixels.
4. On complète la ligne en ajoutant pixel par pixel, la liste se limitant à 2 choix (blanc ou noir).
5. Une fois la ligne complétée, on commence une nouvelle ligne jusqu'à remplir intégralement l'image.

La figure 1.38 illustre ces différentes étapes. La figure 1.39 donne le résultat final et montre, sur un exemple, que l'unicité des cellules est vérifiée (les cellules rouge et bleue se chevauchent mais portent un code différent).

1.4.1.3 Mise en Correspondance

Deux approches sont envisageables pour la mise en correspondance : (i) le décodage du motif dans l'image, c'est-à-dire la reconnaissance des cellules après projection sur la surface de l'objet et projection sur le plan image ; (ii) l'utilisation directe des algorithmes de mise en correspondance



FIGURE 1.39: Mire pseudo-aléatoire et vérification de l'unicité

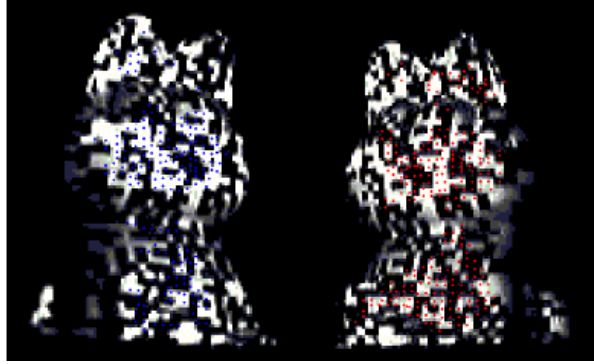


FIGURE 1.40: Exemple de mises en correspondance

classique de type corrélation [Devernay 2002]. C'est cette deuxième approche qui nous a, en l'occurrence, semblée plus directe et efficace. Pour chaque point d'intérêt du motif, nous calculons la ligne épipolaire correspondante dans les images caméra. Nous définissons un voisinage 5×5 autour des points des lignes épipolaires (une dans l'image de gauche et l'autre dans l'image de droite) sur lequel nous calculons la corrélation :

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A}) (B_{mn} - \bar{B})}{\left(\sqrt{\sum_m (A_{mn} - \bar{A})^2} \sqrt{\sum_n (B_{mn} - \bar{B})^2} \right)} \quad (1.37)$$

où \bar{A} et \bar{B} sont les moyennes des niveaux de gris sur le voisinage. Nous avons décidé d'être très sélectif sur la mise en correspondance afin de réduire au maximum le nombre de points mal appariés (qui viendront fausser la texturation ultérieure). La figure 1.40 présente un résultat de mise en correspondance

1.4.1.4 Lumière Structurée Imperceptible

Nous avons évoqué la possibilité de rendre une lumière structurée *imperceptible*, c'est-à-dire capturable par la caméra mais invisible à l'œil humain. Les premiers travaux en ce domaine datent de 1998 [Raskar 1998]. Le principe consistait à projeter un motif suivi de son motif complémentaire à assez haute fréquence de manière à ce que l'œil intègre les deux illuminations en une illumination uniforme (figure 1.41). Si la caméra est synchronisée sur le temps de projection d'un motif, elle

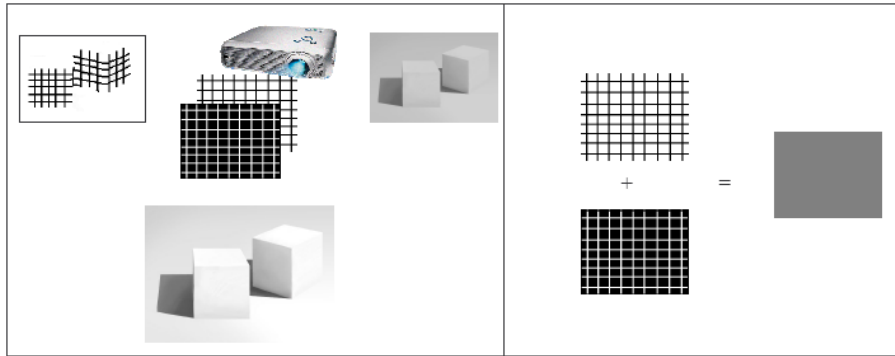


FIGURE 1.41: Principe de la lumière structurée imperceptible

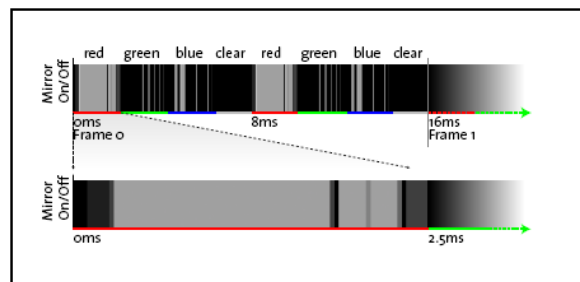


FIGURE 1.42: Séquence pour obtenir une valeur RVB de 223, 47 et 128. Le miroir peut-être actif (gris clair), inactif (noir) ou indéterminé (gris foncé)

capturera l'objet illuminé comme un système de vision en lumière structurée classique ; si, au contraire, elle est configurée avec un temps d'exposition plus long, elle capturera la scène sous illumination uniforme. Rien n'empêche évidemment d'adjoindre deux caméras au projecteur et de profiter des deux types d'information différents. La fréquence de projection est donnée par un indice perceptuel appelé *Fréquence Critique de Fusion* (FCF), représentant la limite de résolution temporelle de l'œil humain.

Plus tard, COTTING et al [Cotting 2004] ont proposé un système beaucoup plus élaboré dont la portée est beaucoup plus vaste. Ils ont proposé de “cacher” le motif structuré dans une image à projeter en exploitant les caractéristiques des vidéoprojecteurs de technologie DLP (*Digital Light Processing*). Ces projecteurs sont composés d'une matrice dense de micro-miroirs (DMD – *Digital Micromirror Device*) correspondant chacun à un pixel. L'intensité du pixel est déterminée par le pourcentage de temps pendant lequel le miroir est tourné vers la scène : “blanc” si le miroir est fixement tourné vers la scène et “noir” si le miroir est en orientation inverse. Les micro-miroirs peuvent passer d'une orientation à l'autre très rapidement (au-delà de la FCF) : approximativement en $10\mu s$. Il est donc envisageable, par un détournement rapide de certains miroirs, de coder imperceptiblement une information dans une image projetée. Pour obtenir une image couleur, chaque composante est traitée successivement via un filtre de couleur (figure 1.42, tirée, comme les deux suivantes, de [Cotting 2004]). Les auteurs définissent un laps de temps, appelé *Binary Image Exposure Period* (BIEP), pendant lequel le motif structuré peut-être projeté – et donc couvert par le temps d'exposition de la caméra (figure 1.43). Chaque pixel est réajusté (du blanc vers le noir ou du noir vers le blanc) de manière à obtenir le motif souhaité (figure 1.44). Ce réajustement altère évidemment l'image originale à projeter : un algorithme de diffusion de l'erreur permet de compenser partiellement cette altération.

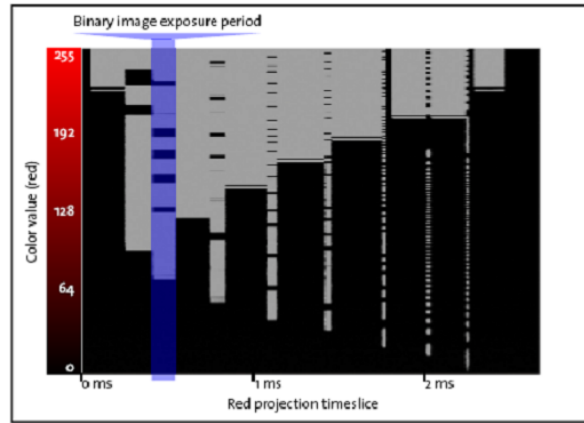


FIGURE 1.43: Séquences obtenues pour les 256 valeurs de rouge

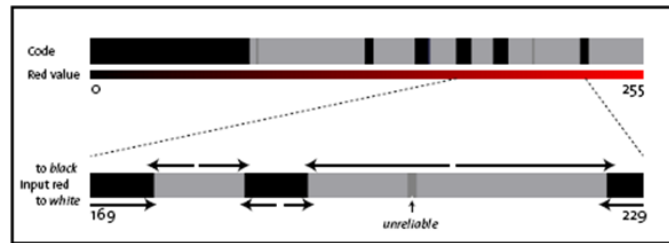


FIGURE 1.44: Ajustement des pixels pendant la BIEP

1.4.2 Recalage et Synthèse de Vue

Il est difficile d'obtenir un modèle dense ou suffisamment dense pour garantir une texturation précise en utilisant la lumière structurée codée (qui plus est quand la codification est directe!) Nous avons pris l'hypothèse que les objets à texturer sont soit *a priori* connus (on peut accéder à leur modèle CAO), c'est le cas pour le prototypage augmenté, la muséographie; soit numérisable hors-ligne à l'aide d'un capteur de vision en lumière structurée (notamment à partir du système projecteur-caméra que nous décrivons). On part donc du principe qu'il est toujours possible d'avoir un modèle 3D complet et suffisamment dense de l'objet à texturer – que nous appellerons modèle de référence. Quand on pose un objet en face du capteur, on dispose d'une reconstruction partielle de celui-ci. En recalant, cette reconstruction partielle au modèle de référence et en exprimant les coordonnées des points qui le compose dans le repère du projecteur, il nous est donc possible de synthétiser une carte de texture à projeter ajustée à l'objet.

1.4.2.1 Recalage par Analyse en Composantes Principales

La méthode de recalage par Analyse en Composantes Principales (ACP) utilise la direction des axes principaux du volume créé par un nuage de points pour les recalcr. Si la zone de recouvrement est suffisante (50% est le seuil critique pour obtenir un résultat), les deux axes doivent presque coïncider et le recalage sera précis. Néanmoins cette méthode simple, rapide par rapport à d'autres méthodes qui demandent l'estimation de la courbure ou le calcul d'une signature particulière, peut servir d'initialisation à une méthode fine. CHUNG et LEE [CHUNG 1998] proposent une méthode qui utilise les vecteurs direction du nuage de points. La matrice de covariance est estimée pour chaque nuage :

$$Cov = \frac{1}{N} \sum_{i=0}^{N-1} (p_i - \bar{p}) (p_i - \bar{p})^T \quad (1.38)$$

où N est le nombre de points, \bar{p} le centre de masse du nuage de points et p_i , le i^e point. La direction U_i de l'axe principale est déduite de la décomposition en valeurs singulières suivantes :

$$Cov_i = U_i D_i U_i^T \quad (1.39)$$

La matrice de rotation permettant de passer d'un nuage à l'autre peut être dès lors extraite :

$$R = U_1 U_2^{-1} \quad (1.40)$$

Le vecteur de translation est donné par la distance séparant les deux centres de masse $\bar{\mu}_i$:

$$t = \bar{\mu}_2 - R \bar{\mu}_1 \quad (1.41)$$

1.4.2.2 Recalage par ICP

La méthode ICP a été initialement introduite par BESL et MCKAY [BESL 1992]; elle consiste à minimiser la distance entre les points de pseudo-correspondance, soit le point le plus proche. L'estimation initiale sert à déplacer le nuage de point vers sa nouvelle position. On cherche pour chaque point de la première vue p_i son plus proche voisin m_i dans la seconde vue de manière à minimiser leur distance. Ce processus est répété jusqu'à convergence :

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \left\| \vec{m}_i - R(\vec{q}_R) \vec{p}_i - \vec{t} \right\| \quad (1.42)$$

Pour minimiser cette équation, on construit d'abord la matrice symétrique $Q(\sum_{pm})$:

$$Q(\sum_{pm}) = \begin{bmatrix} tr(\sum_{pm}) & \Delta^T \\ \Delta & \sum_{pm} + \sum_{pm}^T - tr(\sum_{pm}) I_3 \end{bmatrix} \quad (1.43)$$

avec tr , la trace de la matrice; $\Delta = [A_{23} \ A_{31} \ A_{12}]^T$, extraite de la matrice anti-symétrique $A_{ij} = (\sum_{pm} - \sum_{pm}^T)$; I_3 , la matrice identité; \sum_{pm} , la matrice de co-variance des points m et p :

$$\sum_{pm} = \frac{1}{N_p} \sum_{i=1}^{N_p} [\vec{p}_i \vec{m}_i] - \vec{\mu}_p \vec{\mu}_m \quad (1.44)$$

Le vecteur absolu unitaire $\vec{q}_R = [q_0 \ q_1 \ q_2 \ q_3]^T$ correspondant à la plus grande valeur propre de la matrice Q est sélectionné comme étant la rotation optimale exprimée par des quaternions. Une fois que R a été calculée, on peut estimer le vecteur translation :

$$R = \begin{bmatrix} q_0^2 + q_1^2 - q_2^2 - q_3^2 & 2(q_1 q_2 - q_0 q_3) & 2(q_1 q_3 + q_0 q_2) \\ 2(q_1 q_2 + q_0 q_3) & q_0^2 - q_1^2 + q_2^2 - q_3^2 & 2(q_2 q_3 - q_0 q_1) \\ 2(q_1 q_3 - q_0 q_2) & 2(q_2 q_3 + q_0 q_1) & q_0^2 - q_1^2 - q_2^2 + q_3^2 \end{bmatrix} \quad (1.45)$$

$$t = \bar{\mu}_m - R \bar{\mu}_p \quad (1.46)$$



FIGURE 1.45: Les modèles tridimensionnels utilisés

1.4.2.3 Synthèse de Vue

Une fois la portion reconstruite et le modèle de référence recalés, il est assez direct de synthétiser la vue à projeter. On projette le nuage de points sur le plan du projecteur :

$$\begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} = \mathbf{K} \begin{bmatrix} R & T \end{bmatrix} \begin{bmatrix} x_{ref} \\ y_{ref} \\ z_{ref} \\ 1 \end{bmatrix} \quad (1.47)$$

où $\begin{bmatrix} u_p & v_p & 1 \end{bmatrix}^T$ sont les points de la diapositive et $\begin{bmatrix} x_{ref} & y_{ref} & z_{ref} & 1 \end{bmatrix}^T$ sont les points “visibles” (c’est-à-dire non occultés) du modèle de référence recalé. On peut aisément définir une texture ou à coloriage à appliquer à chaque maille de la surface triangulée projetée.

1.4.3 Résultats Expérimentaux

Nous présentons quelques résultats de texturation virtuelle accompagnés d’une évaluation quantitative du recalage. Nous n’avons pas jugé utile d’insérer ici des résultats expérimentaux sur la précision du calibrage ou de la géométrie épipolaire : on trouve, dans la littérature, suffisamment de rapports, d’articles ou d’états de l’art, traitant du sujet [Salvi 2002, Armangué 2003]. La figure 1.45 présente les modèles que nous avons utilisés pour nos expérimentations. Les figures 1.46, 1.43 et 1.48 montrent qu’il est possible de changer la texture et ainsi d’habiller le modèle de manière automatique et flexible. Visuellement, la précision obtenue est correcte ; les résultats consignés dans le tableau 1.6 montrent que l’erreur moyenne de recalage (de laquelle dépend fortement l’erreur de texturation) reste raisonnablement basse – une partie de la séquence sur laquelle ces erreurs ont été calculées se trouve en figure 1.50. La vue 10 donne une erreur plus importante : elle correspond au recalage du modèle sur une partie douce, sans points de saillance. On sait que le recalage est souvent mis en défaut dans de telles situations. Le fait que cette zone ne présente pas d’aspérité ou de détails fait que, cependant, le rendu de la texturation ne semble pas plus mauvais ou aberrant.

La figure 1.49 illustre d’autres résultats de texturation. Les erreurs de recalage présentées dans le tableau 1.7 montrent que l’erreur est peu dépendante de l’orientation de l’objet ou de l’objet lui-même – il est vrai que tous les modèles testés ont une morphologie proche et que les effets de perspective qui pourraient être produits par des objets plus allongés n’ont pas été suffisamment éprouvés.

La contribution de ces travaux est essentiellement géométrique. Dans une phase plus avancée, il conviendrait d’étendre la méthodologie à un système composé de plusieurs projecteurs et d’étudier – d’autres l’ont fait avant nous – les problèmes liés au mosaïchage des vues projetées et de la compensation radiométrique (pour que deux vues issues de deux projecteurs différents projettent



FIGURE 1.46: Objet avant et après texturation



FIGURE 1.47: Changement de texture



FIGURE 1.48: Précision de la texturation

Numéro de vue	erreur minimale	erreur maximale	erreur moyenne
1	0.28	8.66	3.5
2	0.14	10.09	1.90
3	0.14	8.6	1.94
4	0.26	23.9	6.8
5	0.07	11.2	1.8
6	0.13	6.4	1.7
7	0.09	9.8	1.9
8	0.34	13.59	5.26
9	0.15	5.87	1.86
10	0.44	39.27	17.42
11	0.08	35.16	2.5
12	0.07	25.66	2.37
13	0.044	35.91	3.19

TABLE 1.6: Erreur de recalage

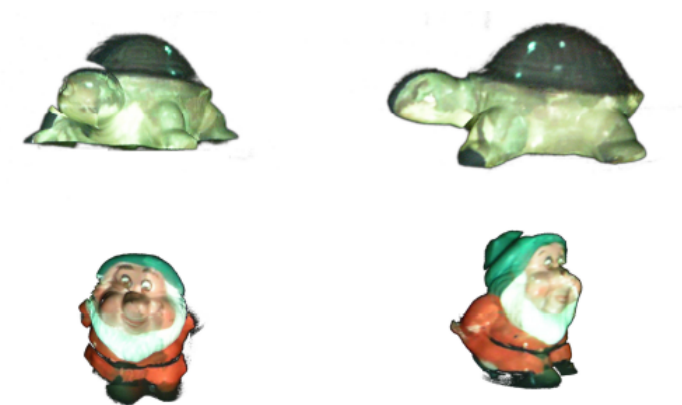


FIGURE 1.49: Autres modèles texturés

objet	erreur minimale	erreur maximale	erreur moyenne
tortue de face	0.15	11.87	2.96
tortue de 3 quart	0.12	15.83	2.69
nain de jardin de face	0.15	14.40	2.75
nain de jardin de 3 quart	0.11	10.00	2.34
taupe de face	0.14	8.67	2.18
taupe de 3 quart	0.10	12.24	2.7

TABLE 1.7: Erreurs de recalage en fonction de l'orientation et des modèles

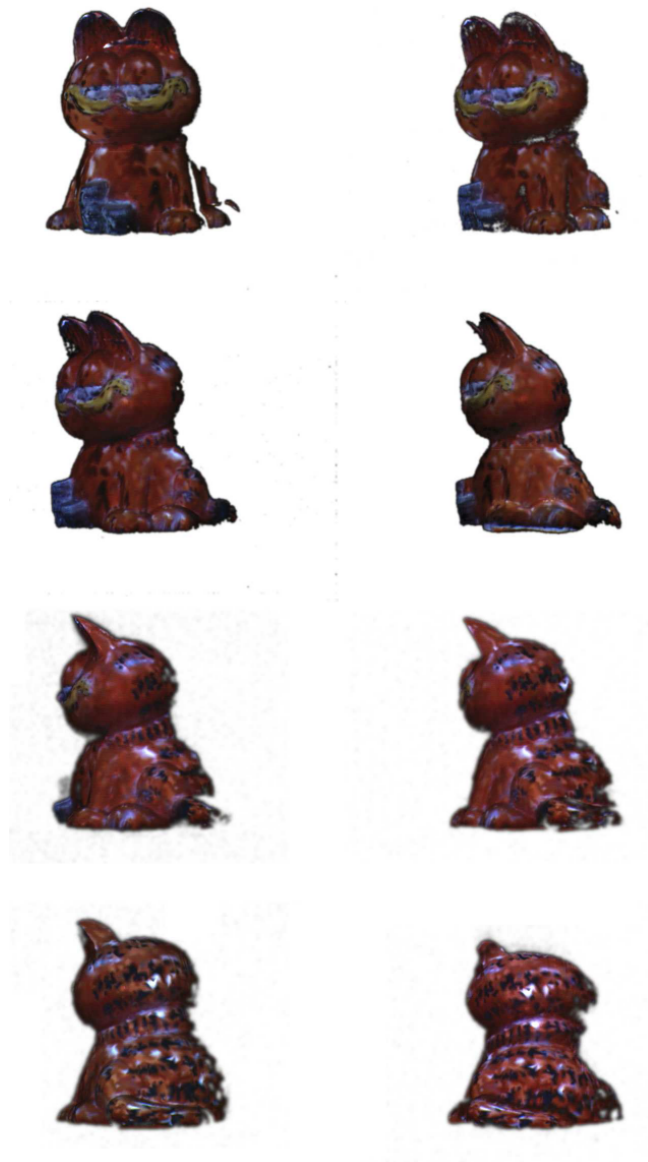


FIGURE 1.50: Séquence de texturation

bien la même couleur pour une consigne donnée). De même, le calibrage photométrique des projecteurs devra être pris en compte pour contrôler le rendu (c'est-à-dire l'effet de la projection sur le matériau de l'objet) par rapport à la consigne (c'est-à-dire la valeur du pixel projeté).

1.5 Conclusions

La géométrie des systèmes projecteur-caméra est bien connue et il est souvent possible d'appliquer les outils de la stéréovision au couple projecteur-caméra pour achever la transposition d'un système à l'autre. Le traitement des images en lumière structurée, utile à la segmentation et au décodage éventuel du motif, relève de la théorie générale du traitement des images. Il est fortement contraint, au sens où l'on sait *a priori* ce que l'on cherche dans les images, et donc plus simple en général que le traitement des images en lumière ambiante. Il reste toutefois au moins quatre thématiques d'intérêt peu étudiées ou dont l'état de l'art est encore susceptible de progresser considérablement : (i) la conception d'un motif structuré codé adaptatif; (ii) l'étude géométrique et photométrique des réseaux de projecteur; (iii) l'auto-calibrage du système; (iv) l'analyse du mouvement, notamment du projecteur.

Pour le point (i), des travaux proposant un système complet et performant ont été publiés par KONINCKX et al en 2005 [Koninckx 2005] (après ceux de CASPI et al en 1998 [Caspi 1998]). On pourrait leur objecter que l'adaptation qu'ils effectuent est globale et pas ajustée plus précisément aux différentes zones de l'objet ou de la scène, ni en couleur ni en géométrie. Il y a sans doute quelque chose à creuser dans cette direction. Nous y reviendrons dans la conclusion générale de ce mémoire.

Pour le point (ii), la vitalité du *workshop* sur les systèmes projecteur-caméra³ et le nombre d'articles publiés traitant du sujet atteste des travaux qu'il reste à mener sur cet aspect.

Nous avons rapidement évoqué le problème de l'auto-calibrage en début de chapitre. La difficulté ici est d'obtenir une estimation des paramètres intrinsèques assez précise et robuste à partir de deux vues uniquement (le motif et l'image), tout en laissant suffisamment de degrés de liberté au projecteur pour la mise au point. Les informations internes au motif permettront peut-être de lever certaines ambiguïtés et d'adapter plus rationnellement les techniques d'auto-calibrage aux systèmes projecteur-caméra.

L'analyse du mouvement reste un problème ouvert et très peu abordé. On sait qu'un déplacement du projecteur provoquera un déplacement (et donc une perte, une inconsistance) des points 3D visés par la projection. Comment alors établir le lien entre mouvement apparent et mouvement réel? Comment définir et interpréter le flot optique? Si ces questions ont été si peu posées, c'est sans doute parce que l'encombrement du projecteur empêchait d'envisager sérieusement l'embarquement d'un tel capteur. Avec l'apparition prochaine des *pico-projecteurs*⁴ sur le marché (technologie développée par Texas Instruments et dont la commercialisation est promise par Microvision), il paraît évident qu'un regain d'intérêt pour les systèmes projecteur-caméra en contrôle ou vision embarquée rendra cette thématique plus attrayante.

Pour ce qui est des systèmes laser-caméra, on peut considérer que la thématique de recherche est mûre déjà depuis plusieurs années. On peut envisager des avancées technologiques permettant d'obtenir une plus grande précision ou d'appliquer la reconstruction sur des surfaces ou des matériaux différents. Mais les principales contributions à nos yeux seront sur les à-côtés, sur l'utilisation du scanner en configuration particulière : le recalage de vues, le *best-next-view*, etc.

3. ProCams, www.procams.org

4. Projecteur de la taille d'un téléphone portable permettant de projeter des images de bonne résolution. Microvision annonce 848×480.

Chapitre 2

Vision Catadioptrique

La totalité des travaux exposés dans ce chapitre ont été effectués dans le cadre du projet Ca.Vi.A.R. (Catadioptric Vision for Aerial Robots) sélectionné dans le programme blanc de l'ANR. Ce projet, dont le financement court sur 3 ans, de décembre 2005 à décembre 2008, est le fruit d'une collaboration entre 5 laboratoires nationaux : le MIS de l'Université de Picardie Jules Verne, le LAAS CNRS de Toulouse, l'équipe AROBAS de l'INRIA Sophia-Antipolis, l'équipe PERCEPTION de l'INRIA Rhône-Alpes et le Le2i. Trois thèmes ont été développés par la cellule Vision 3D du Le2i : le calibrage des capteurs catadioptriques par imagerie polarimétrique, le traitement des images catadioptriques adapté à la forme du miroir et la mise en correspondance de droites catadioptriques. Nous les détaillerons dans les différentes sections à venir. Au préalable, nous ferons une brève introduction à la vision catadioptrique.

2.1 Introduction à la Vision Catadioptrique

Un système est dit catadioptrique quand il a trait à la fois à la catoptrique (lois de la réflexion de la lumière) et à la dioptrique (lois de la réfraction). Un objectif catadioptrique est donc composé à la fois de miroirs et de lentilles.

Le système catadioptrique le plus célèbre est sans doute le périscope qui permet depuis un sous-marin d'observer au dessus du niveau de la mer grâce à une double réflexion sur des miroirs plans. On rencontre aussi fréquemment des systèmes composés d'un objectif et d'un ou plusieurs miroirs plans permettant de faire de la stéréovision à partir d'une seule caméra – la seconde vue est créée virtuellement par réflexion sur le miroir [Mathieu 1995]. En utilisant un miroir de révolution (sphérique, conique, parabolöide ou hyperbolöide), il est possible d'obtenir une vue à 360° à partir d'une seule image. On dit un peu abusivement de ces capteurs qu'ils sont omnidirectionnels (figure 2.1, tirée de [Mouaddib 2003]).

En 1999, BAKER et NAYAR [Baker 1999] proposèrent une théorie de la formation des images propre à ces capteurs. Ils en tirèrent une contrainte, connue sous le nom de *contrainte du point de vue unique* (bien qu'ils ne soient pas les premiers à avoir exprimé ces relations, voir [Mouaddib 2003]). Si cette contrainte est respectée, alors le capteur catadioptrique doit échantillonner la fonction plénoptique 5D, en d'autres termes, à une direction unique donnée dans l'espace 3D, on associe un seul point dans l'image. Les capteurs respectant cette contrainte, sont dit *centraux*. Une conséquence remarquable est qu'il est théoriquement possible de générer une image perspective de la scène observée à partir de tels capteurs. De l'étude menée par BAKER et NAYAR, il ressort que seules quatre formes de miroir permettent le respect de cette contrainte :

1. le miroir parabolöide, quand il est associé à une caméra orthographique (les lignes de vue passent par le foyer de la parabole et sont rejetés à l'infini – c'est-à-dire au point focal de la caméra) ;
2. le miroir hyperbolöide, quand une caméra perspective est placée précisément sur le second foyer de l'hyperbole ;

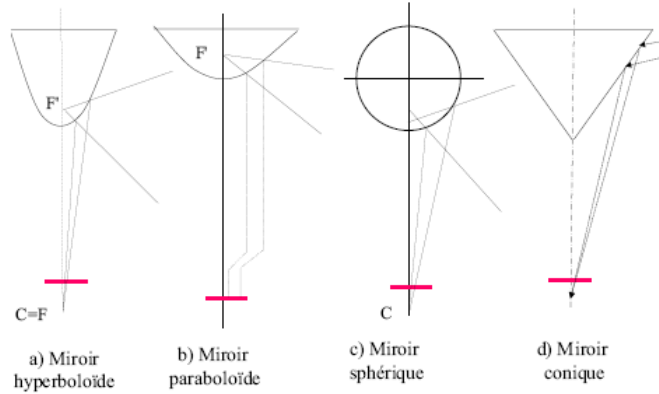


FIGURE 2.1: Principe de la vision omnidirectionnelle catadioptrique

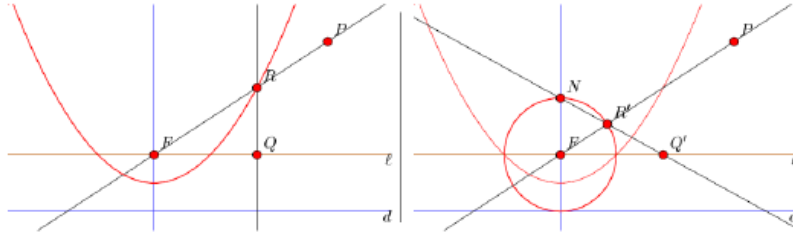


FIGURE 2.2: Modèle de la sphère équivalente

3. le miroir elliptique, quand une caméra perspective est placée sur le second foyer de l'ellipse – cette configuration apporte peu, puisque c'est l'intérieur de l'ellipse qui réfléchira la scène ;
4. le miroir plan, quand il respecte une configuration très particulière de sorte que le point de vue est sur la bissectrice perpendiculaire à la droite reliant le point focal à son image virtuelle.

En théorie, il est possible également de construire un capteur central à partir d'un miroir sphérique ou conique : dans le premier cas, la caméra devrait être placée au centre de la sphère, dans le second cas, la caméra devrait être placée au sommet de la conique, ce qui est impossible pratiquement.

GEYER et DANILIDIS [Geyer 2000, Geyer 2001] ont proposé un modèle unifié pour les capteurs catadioptriques centraux. Ils ont démontré qu'il était possible de les modéliser en utilisant d'abord une projection sur une sphère centrée sur le point de vue unique du capteur, suivie d'une autre projection sur le plan image. La figure 2.2 illustre cette double projection dans le cas d'un miroir paraboloïde. La sphère équivalente est centrée en F . Le point P est d'abord projeté en R , puis stéréographiquement en Q' . Il est démontré que Q et Q' sont confondus. Dans le cas d'un miroir hyperboloïde, la projection stéréographique ne se fera pas à partir du pôle nord de la sphère, mais d'un point situé au dessous du pôle nord et défini par l'excentricité de la conique – il en sera de même pour le miroir elliptique (figure 2.3, tirée de [Geyer 2001]). Ainsi, tous les capteurs catadioptriques centraux peuvent être représentés par une projection sur la sphère équivalente. Corollairement, cela signifie que les traitements ou les propriétés géométriques du capteur peuvent être développés ou établis sur cette sphère et restés valides quel que soit le capteur central considéré.

Théoriquement, l'inconvénient d'un tel modèle est qu'il ne prend en compte "que" les capteurs centraux. Pratiquement, on sait qu'il est parfois difficile d'obtenir un capteur suffisamment bien

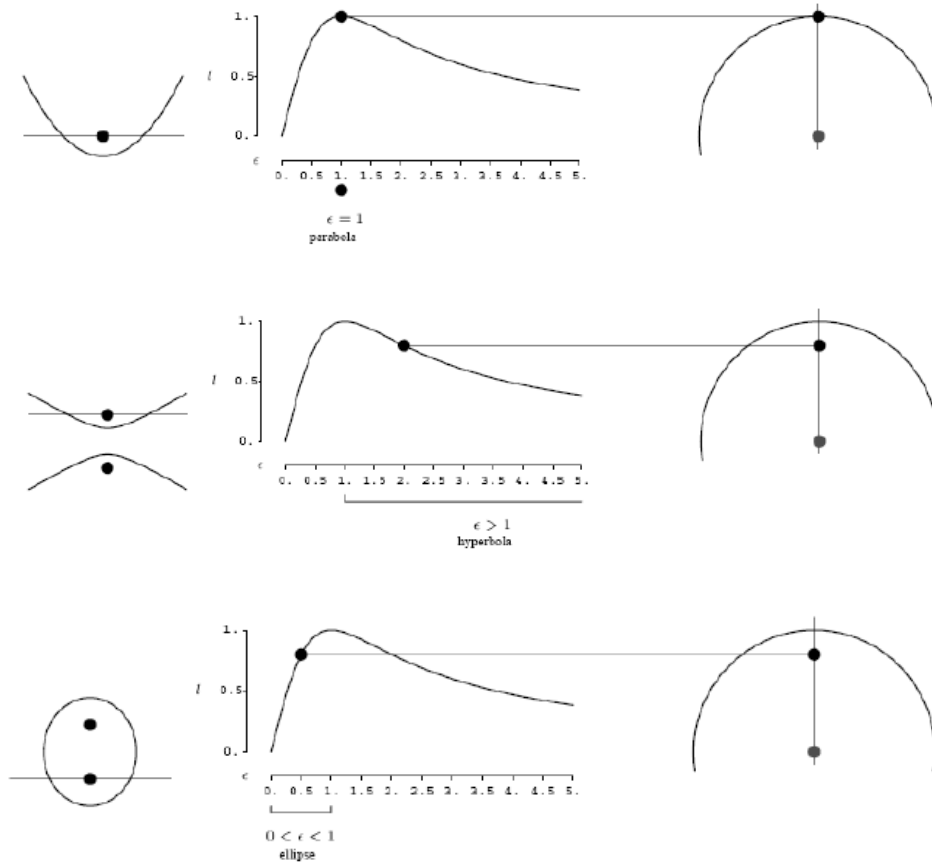


FIGURE 2.3: Principe de la projection stéréographique. De haut en bas : miroir parabolöide, miroir hyperboloöide et miroir elliptique

aligné pour que la contrainte du point de vue unique puisse être satisfaite. Quand bien même, il arrive, notamment au cours de mission de navigation en robotique, que les chocs et les vibrations provoqués par le déplacement désaligne le capteur.

Les capteurs ne satisfaisant pas ou plus la contrainte du point de vue unique peuvent être modélisés par des caustiques (lieu des rayons lumineux subissant une réflexion ou une réfraction sur une surface ou une courbe – quand un capteur est central les rayons convergent tous en un unique point). Ces modèles restent relativement compliqués à définir, identifier et mettre en oeuvre [Swaminathan 2001], c’est pourquoi nous leur avons préféré le modèle générique, dont la simplicité et le formalisme emprunté à la géométrie projective est moins déroutant et qui trouve dans l’imagerie polarimétrique une méthode de calibrage particulièrement bien adaptée.

2.2 Calibrage par Imagerie Polarimétrique

Ces travaux ont été inspirés par les articles de Peter STURM sur le modèle générique de caméra [Sturm 2004, Sturm 2005, Ramalingam 2006] et par ceux d’Olivier MOREL sur la reconstruction de surfaces métalliques par imagerie polarimétrique [Morel 2006b]. L’objectif était de combiner les deux approches pour proposer une nouvelle méthode de calibrage dédiée aux capteurs catadioptriques basée sur le modèle générique.

2.2.1 Modèle Générique de Caméra

La plupart des modèles de caméra sont paramétriques, c'est-à-dire définis par un jeu de paramètres intrinsèques. On dispose donc d'une fonction permettant de passer d'un point 3D au pixel correspondant, et, en l'inversant, du pixel à une ligne de vue dans l'espace. C'est vrai pour le modèle orthographique, pour le modèle perspective, pour le modèle unifié pour les capteurs catadioptriques centraux, etc. Le problème vient du fait que l'on a besoin d'un modèle différent pour chaque type ou famille de caméra, que l'on traite aussi différemment les capteurs centraux et non-centraux et que, de ce fait, on manipule les capteurs mixtes (composés de caméras de types différents) de manière plutôt inélégante.

En vision catadioptrique, trois hypothèses sont couramment admises pour le calibrage : (i) la forme du miroir est parfaitement connue ; (ii) l'alignement du capteur est parfait, de sorte que la contrainte du point de vue unique est respectée ; (iii) le modèle de projection peut être paramétré. Il existe des méthodes permettant de relaxer la deuxième contrainte et quelques-unes permettent de relaxer la première, cependant, jusqu'il y a peu [Grossberg 2001, Pless 2003, Sturm 2004], toutes sous-tendaient un modèle paramétrique de projection explicite. Comment décrire le fonctionnement projectif d'une caméra, le plus généralement possible ? Une caméra capture la lumière qui transite le long d'une ligne de vue. Chaque ligne de vue est intégrée dans un pixel du plan image. Ainsi, une caméra est totalement décrite par :

- les coordonnées 3D des lignes de vue ;
- la correspondance entre ces lignes de vue et les pixels.

L'avantage d'un tel modèle vient du fait qu'il décrit correctement tout type de caméra (orthographique, perspective, *fish-eye*, catadioptrique, etc.), qu'il unifie capteurs centraux et non-centraux et qu'il permet de traiter indifféremment tous les éléments d'un capteur mixte.

2.2.2 Notions d'Imagerie Polarimétrique

L'imagerie polarimétrique permet d'étudier l'état de polarisation d'une onde lumineuse. L'application la plus courante en vision par ordinateur est celle permettant de distinguer des objets diélectriques et métalliques [Wolff 1990] et de détecter les surfaces transparentes. L'imagerie polarimétrique permet également de mesurer tridimensionnellement un objet : c'est la méthode dite de *Shape from Polarization*. Le principe physique est le suivant : après avoir été réfléchie, une onde lumineuse non-polarisée devient partiellement linéairement polarisée, en fonction de la normale à la surface et de l'indice de réfraction du matériau qu'elle vient illuminer (figure 2.4). Une lumière partiellement linéairement polarisée a trois paramètres : l'amplitude lumineuse I , le degré de polarisation ρ et l'angle de polarisation φ (figure 2.5). Il ne s'agit pas ici de proposer un cours sur la polarisation de la lumière, mais seulement d'en donner les caractéristiques et propriétés qui permettent de bien comprendre les sous-sections qui suivent.

L'état de polarisation de la lumière peut être calculée. L'intensité de chaque pixel est liée aux paramètres de polarisation par la fonction suivante :

$$I_p(\alpha) = \frac{I}{2} (\rho \cos(2\alpha - 2\varphi) + 1) \quad (2.1)$$

où α représente l'angle du polariseur, $I = I_{max} + I_{min}$ et $\rho = \frac{I_{max} - I_{min}}{I_{max} + I_{min}}$ (figure 2.6).

WOLFF et BOULT ont démontré comment déterminer des contraintes sur les normales des surfaces en utilisant le modèle de réflectance de Fresnel [Wolff 1991]. La surface du miroir est supposée continue et décrite par l'expression cartésienne $z = f(x, y)$. Ainsi, chaque normale à la surface est donnée par l'expression non-normalisée suivante :

$$\vec{n} = \begin{bmatrix} -\frac{\partial f(x,y)}{\partial x} \\ -\frac{\partial f(x,y)}{\partial y} \\ 1 \end{bmatrix} = \begin{bmatrix} p = \tan \theta \cos \phi \\ q = \tan \theta \sin \phi \\ 1 \end{bmatrix} \quad (2.2)$$

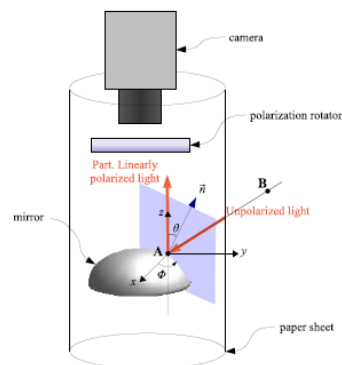
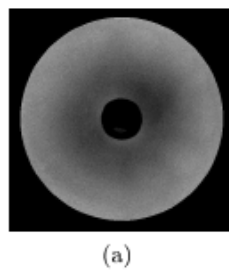
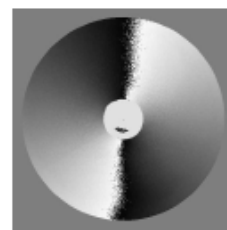


FIGURE 2.4: Principe de l'imagerie polarimétrique et configuration



(a)



(b)

FIGURE 2.5: Image des paramètres de polarisation. (a) Degré de polarisation. (b) Angle de polarisation

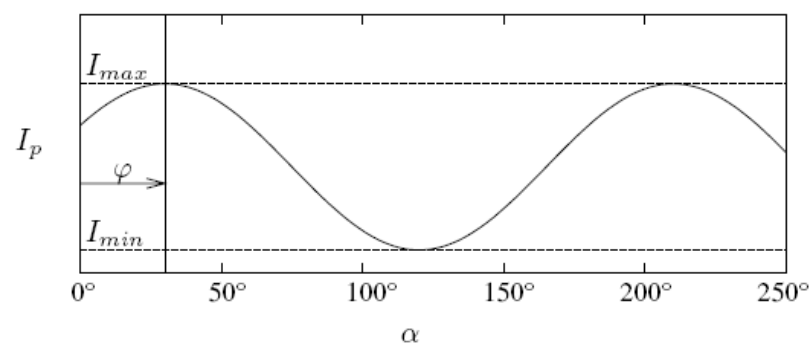


FIGURE 2.6: Variation de l'intensité lumineuse en fonction de l'angle du polariseur



FIGURE 2.7: Ambiguïté sur l'angle de polarisation. (a) Masque utilisé. (b) Image résultat

Le but du *shape from polarization* est de calculer ces normales à partir des angles θ et ϕ . En combinant les formules de Fresnel et la loi de Snell-Descartes, on peut trouver une relation entre le degré de polarisation ρ et l'angle zenithal θ [Miyazaki 2004]. Pour des surfaces métalliques, la formule suivante peut être appliquée :

$$\rho(\theta) = \frac{2n \tan \theta \sin \theta}{\tan^2 \theta \sin^2 \theta + |\hat{n}|^2} \quad (2.3)$$

où $\hat{n} = n(1 + i\kappa)$ est l'indice complexe de réfraction du miroir.

L'angle azimuthal ϕ est lié à l'angle de polarisation φ puisque la lumière réfléchie linéairement partiellement polarisée en fonction de la normale au plan d'incidence. En prenant l'hypothèse que la caméra est orthographe (nous avons ajouté une lentille télécentrique à l'objectif), il est très simple d'établir la correspondance entre les paramètres de polarisation et les pixels. Ainsi, l'angle ϕ peut être déduit de l'angle de polarisation comme suit :

$$\phi = \varphi \pm \frac{\pi}{2} \quad (2.4)$$

L'ambiguïté sur l'angle peut être levée via l'algorithme présenté dans [Morel 2006a] – figure 2.7.

2.2.3 Calibrage et Reconstruction

2.2.3.1 Estimation des Lignes de Vue

Le modèle générique considère qu'une caméra est complètement déterminée par un faisceau de lignes de vue dans l'espace et leur correspondance aux pixels. Une ligne de vue peut être représentée par un couple de points :

$$\mathbf{A} = \begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix}, \mathbf{B} = \begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix} \quad (2.5)$$

Ces points peuvent être obtenus à partir de la reconstruction de la surface du miroir et de la connaissance de la normale au point considéré sur le miroir. Il est possible, une fois connu le champ de normales, de l'intégrer afin d'estimer la surface du miroir. L'algorithme proposé par FRANKOT et CHELLAPPA [Frankot 1988] permet un tel calcul. En considérant que le point \mathbf{A} est sur la surface du miroir, on peut facilement estimer un point \mathbf{B} par :

$$\mathbf{B} = \mathbf{A} + k \begin{bmatrix} \tan 2\theta \cos \phi \\ \tan 2\theta \sin \phi \\ 1 \end{bmatrix} \quad (2.6)$$

où k est une constante non-nulle. La correspondance entre les pixels de l'image et les lignes de vue est assurée par le fait que la caméra est orthographe.

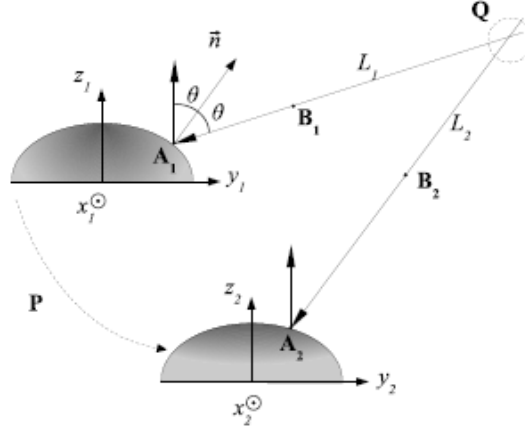


FIGURE 2.8: Principe de la triangulation

2.2.3.2 Triangulation

Deux méthodes de triangulation ont été adaptées au modèle générique. La première est une méthode linéaire basée sur l'extraction des valeurs propres souvent appelée *Direct Linear Transformation* [Hartley 2000]. La seconde est la méthode du point milieu, ré-écrite à partir de la connaissance de deux points appartenant aux lignes de vue.

Méthode linéaire par les valeurs propres. Soient les deux points A_1 et B_1 appartenant à une ligne de vue du miroir liée à l'image 1 et A_2 et B_2 deux points appartenant à une ligne de vue du miroir liée à l'image 2. Si ces deux lignes de vue sont en correspondance, c'est-à-dire si les points images auxquels elles sont liées sont des points homologues, alors elles se coupent dans l'espace en un point Q (figure 2.8).

Un point appartenant à une ligne dans l'espace peut être exprimé par une combinaison linéaire de deux autres points de cette ligne. Si la caméra 1 est placée à l'origine, alors :

$$Q = \lambda_1 A_1 + \mu_1 B_1 \quad (2.7)$$

Si $P = \begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix}$ est la matrice de pose relative de la caméra 2 par rapport à la caméra 1, alors :

$$PQ = \lambda_2 A_2 + \mu_2 B_2 \quad (2.8)$$

En combinant les deux dernières équations, on obtient :

$$\forall i = 1, 2 \quad \lambda_i A_i + \mu_i B_i = P_i Q \quad (2.9)$$

avec $P_1 = I_{4 \times 4}$ et $P_2 = P$. Le système d'équations peut être exprimé sous forme matricielle :

$$\begin{bmatrix} I & A_1 & B_1 & 0 & 0 \\ P & 0 & 0 & A_2 & B_2 \end{bmatrix} \begin{bmatrix} Q \\ -\lambda_1 \\ -\mu_1 \\ -\lambda_2 \\ -\mu_2 \end{bmatrix} = 0 \quad (2.10)$$

Ce système peut être facilement généralisé :

$$\mathbf{H}\mathbf{x} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad (2.11)$$

avec $\mathbf{x} = [\mathbf{Q} \quad -\lambda_1 \quad -\mu_1 \quad \cdots \quad -\lambda_n \quad -\mu_n]^T$ et

$$\mathbf{H} = \begin{bmatrix} \mathbf{P}_1 & \mathbf{A}_1 & \mathbf{B}_1 & 0 & 0 & \cdots & 0 & 0 \\ \mathbf{P}_2 & 0 & 0 & \mathbf{A}_2 & \mathbf{B}_2 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \mathbf{P}_n & 0 & 0 & 0 & 0 & \cdots & \mathbf{A}_n & \mathbf{B}_n \end{bmatrix}.$$

La résolution peut être obtenue en minimisant $\|\mathbf{H}\mathbf{x}\|$ sous la contrainte que $\|\mathbf{x}\| = 1$. La solution est donnée par le vecteur propre unitaire correspondant à la plus petite valeur propre de $\mathbf{H}^T\mathbf{H}$.

Méthode du point milieu. Soient $(\mathbf{A}_1, \mathbf{B}_1)$ et $(\mathbf{A}_2, \mathbf{B}_2)$, deux points appartenant respectivement à la ligne de vue \mathcal{L}_1 et \mathcal{L}_2 . Comme précédemment, \mathcal{L}_1 est attachée à la caméra 1, \mathcal{L}_2 est attachée à la caméra 2 et \mathbf{P} est la matrice de pose de la caméra 2 par rapport à la caméra 1. Dans la méthode du point milieu, \mathbf{Q} est le barycentre de \mathbf{M}_1 et \mathbf{M}_2 , projections orthogonales de \mathbf{Q} sur \mathcal{L}_1 et \mathcal{L}_2 respectivement. Les relations suivantes peuvent être établies :

$$\begin{aligned} \overrightarrow{\mathbf{M}_1\mathbf{Q}} \cdot \overrightarrow{\mathbf{A}_1\mathbf{B}_1} &= 0 \\ \overrightarrow{\mathbf{M}_2\mathbf{Q}} \cdot \overrightarrow{\mathbf{A}_2\mathbf{B}_2} &= 0 \\ \overrightarrow{\mathbf{Q}\mathbf{M}_1} \cdot \overrightarrow{\mathbf{Q}\mathbf{M}_2} &= 0 \end{aligned} \quad (2.12)$$

Puisque \mathbf{M}_1 et \mathbf{M}_2 appartiennent à \mathcal{L}_1 et \mathcal{L}_2 respectivement, on peut écrire :

$$\begin{aligned} \mathbf{M}_1 &= \mathbf{A}_1 + \mu_1 (\mathbf{B}_1 - \mathbf{A}_1) \\ \mathbf{M}_2 &= \mathbf{A}_2 + \mu_2 (\mathbf{B}_2 - \mathbf{A}_2) \end{aligned} \quad (2.13)$$

Les équations 2.12 peuvent être ré-écrites :

$$\begin{aligned} (\mathbf{Q} - \mathbf{M}_1)^T (\mathbf{B}_1 - \mathbf{A}_1) &= 0 \\ (\mathbf{P}\mathbf{Q} - \mathbf{M}_2)^T (\mathbf{B}_2 - \mathbf{A}_2) &= 0 \\ \mathbf{M}_1 + \mathbf{P}^{-1}\mathbf{M}_2 &= 2\mathbf{Q} \end{aligned} \quad (2.14)$$

En injectant 2.14 dans 2.15, il vient :

$$\begin{aligned} \mathbf{A}_1^T (\mathbf{B}_1 - \mathbf{A}_1) &= (\mathbf{B}_1 - \mathbf{A}_1)^T \mathbf{Q} - \mu_1 \|\mathbf{B}_1 - \mathbf{A}_1\|^2 \\ \mathbf{A}_2^T (\mathbf{B}_2 - \mathbf{A}_2) &= (\mathbf{B}_2 - \mathbf{A}_2)^T \mathbf{P}\mathbf{Q} - \mu_2 \|\mathbf{B}_2 - \mathbf{A}_2\|^2 \\ \mathbf{A}_1 + \mathbf{P}^{-1}\mathbf{A}_2 &= 2\mathbf{Q} - \mu_1 (\mathbf{B}_1 - \mathbf{A}_1) - \mu_2 \mathbf{P}^{-1} (\mathbf{B}_2 - \mathbf{A}_2) \end{aligned} \quad (2.15)$$

Ou sous forme matricielle :

$$\begin{bmatrix} \mathbf{A}_1^T (\mathbf{B}_1 - \mathbf{A}_1) \\ \mathbf{A}_2^T (\mathbf{B}_2 - \mathbf{A}_2) \\ \mathbf{A}_1 + \mathbf{P}^{-1}\mathbf{A}_2 \end{bmatrix} = \mathbf{K} \begin{bmatrix} \mathbf{Q} \\ \mu_1 \\ \mu_2 \end{bmatrix} \quad (2.16)$$

avec $\mathbf{K} = \begin{bmatrix} (\mathbf{B}_1 - \mathbf{A}_1)^T & -\|\mathbf{B}_1 - \mathbf{A}_1\|^2 & 0 \\ (\mathbf{B}_2 - \mathbf{A}_2)^T & 0 & -\|\mathbf{B}_2 - \mathbf{A}_2\|^2 \\ 2\mathbf{I} & -(\mathbf{B}_1 - \mathbf{A}_1) & -\mathbf{P}^{-1}(\mathbf{B}_2 - \mathbf{A}_2) \end{bmatrix}.$

La méthode peut également être généralisée à n vues :

$$\begin{aligned} \forall i \in [1, n], (\mathbf{P}_i\mathbf{Q} - \mathbf{M}_i)^T (\mathbf{B}_i - \mathbf{A}_i) &= 0 \\ \forall i \in [1, n], \mathbf{M}_i &= \mathbf{A}_i + \mu_i (\mathbf{B}_i - \mathbf{A}_i) \\ \sum \mathbf{P}_i^{-1}\mathbf{M}_i &= n\mathbf{Q} \end{aligned} \quad (2.17)$$

Soit, sous forme matricielle :

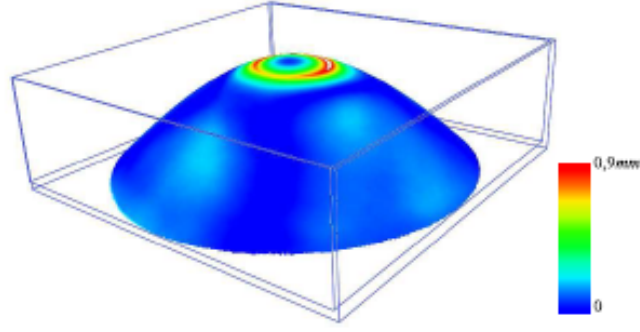


FIGURE 2.9: Carte d'erreur de la reconstruction d'un miroir hyperbolique

$$\mathbf{y} = \mathbf{K}\mathbf{x} \quad (2.18)$$

avec $\mathbf{y} = [\mathbf{Q}, \mu_1, \dots, \mu_n]^T$ et $\mathbf{x} = [\mathbf{A}_1^T (\mathbf{B}_1 - \mathbf{A}_1), \dots, \mathbf{A}_n^T (\mathbf{B}_n - \mathbf{A}_n), \sum \mathbf{P}_i^{-1} \mathbf{A}_i]^T$

et la matrice $\mathbf{K} = \begin{bmatrix} (\mathbf{B}_1 - \mathbf{A}_1)^T \mathbf{P}_1 & -\|\mathbf{B}_1 - \mathbf{A}_1\|^2 & 0 & \dots \\ \vdots & 0 & \ddots & 0 \\ (\mathbf{B}_n - \mathbf{A}_n)^T \mathbf{P}_n & \vdots & 0 & -\|\mathbf{B}_n - \mathbf{A}_n\|^2 \\ n\mathbf{I} & -\mathbf{P}_1^{-1} (\mathbf{B}_1 - \mathbf{A}_1) & \dots & -\mathbf{P}_n^{-1} (\mathbf{B}_n - \mathbf{A}_n) \end{bmatrix}$.

Le système d'équations peut être résolu au sens des moindres carrés :

$$\mathbf{x} = (\mathbf{K}^T \mathbf{K})^{-1} \mathbf{K}^T \mathbf{y} \quad (2.19)$$

2.2.4 Résultats Expérimentaux

Une évaluation expérimentale de la méthode a été effectuée en simulation et sur données réelles, avec différents miroirs. La qualité du calibrage est estimée par la mesure d'erreur sur la reconstruction de la surface du miroir. Le premier miroir hyperbolique que nous avons fait usiner avait pour équation :

$$\frac{z^2}{789.3274} - \frac{x^2 + y^2}{548.1440} = 1 \quad (2.20)$$

Nous avons superposé le modèle idéal et la reconstruction obtenue par imagerie polarimétrique (figure 2.9). L'erreur est minime. En supposant que le miroir a été parfaitement construit, on peut en déduire que le calibrage est correct : l'évaluation n'est pas suffisante et l'explication un peu courte, mais elle permet au moins de valider la méthode de reconstruction et d'avoir une première idée sur la viabilité de notre proposition. Nous avons ré-itéré l'opération avec un miroir sphérique et en comparant les paramètres d'angle avec les paramètres théoriques du miroir (figure 2.10). L'erreur quadratique moyenne sur les angles θ et ϕ est de 0.49° et 1.02° respectivement. Précisons que pour obtenir l'estimation des normales à la surface du miroir, nous avons utilisé un objectif à lentille télécentrique, de ce fait, le capteur formé avec le miroir hyperbolique ou sphérique est dans les deux cas non-central.

La figure 2.11 donne un exemple de reconstruction d'une scène après calibrage du capteur.

Nous avons ensuite évalué l'erreur de reconstruction au regard du bruit de mesure (figure 2.12) sur les deux angles de la normale calculée. Nous l'avons fait dans deux cas de figure : le premier quand la méthode est calculée de bout en bout, c'est-à-dire jusqu'à la reconstruction de la surface du miroir (coordonnée z), la seconde en utilisant uniquement l'orientation de la normale (et donc en

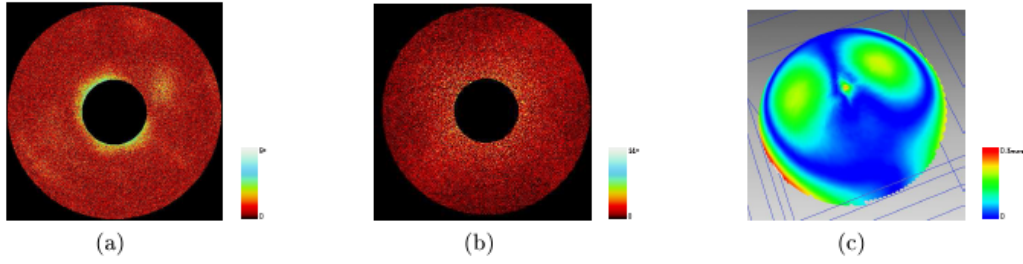
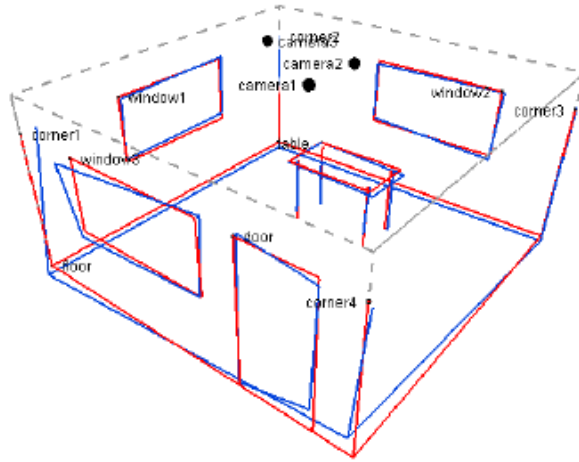
FIGURE 2.10: Erreurs de mesure des paramètres 3D. (a) L'angle θ . (b) L'angle ϕ . (c) Carte d'erreur

FIGURE 2.11: Reconstruction 3D en simulation

considérant que tous les points du miroir sont sur un même plan). Etant donné la faible profondeur du miroir par rapport à la distance qui le sépare de la caméra et à l'échelle de la scène entière, on voit que la reconstruction de la surface du miroir est facultative. En pratique, quand les conditions de prise de vue et de rapport d'échelle sont respectés, nous abandonnerons cette étape.

Quelques résultats ont été obtenus sur des images réelles. Nous présentons en figure 2.14 une mise en correspondance manuelle et en figures 2.15 et 2.16 les résultats de la reconstruction à partir, respectivement, d'un miroir hyperbolique et sphérique. Les parties rouges ont été obtenues par la méthode DLT et les parties bleues par la méthode du point milieu. La figure 2.17 présente un résultat de reconstruction à partir de trois vues cette fois (avec un miroir sphérique : le capteur est donc non-central). Les points bleutés représentent la position du capteur dans la scène. On peut ainsi se rendre compte de l'étendu de la reconstruction au regard du faible déplacement de la caméra. Ces résultats méritent évidemment d'être complétés : (i) en évaluant la qualité de la reconstruction (estimation de l'erreur 3D et de l'erreur sur le déplacement) et (ii) en automatisant et augmentant le nombre de points de correspondance. C'est l'objet de notre prochaine section.

Le dispositif expérimental mis en place permet de reconstruire, à l'aide d'un scanner à temps de vol, la scène dans laquelle évolue le capteur et de recalculer l'information reconstruite par le capteur catadioptrique et celle reconstruite par le capteur à temps de vol. Une vidéo de présentation est disponible sur le site du projet CaViAR¹.

1. <http://www.anr-caviar.org>

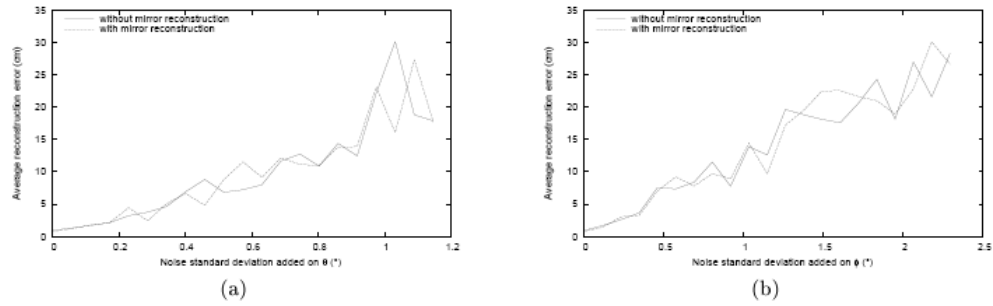


FIGURE 2.12: Erreur de reconstruction induite par des mesures bruitées de la normale. (a) L'angle θ . (b) L'angle ϕ

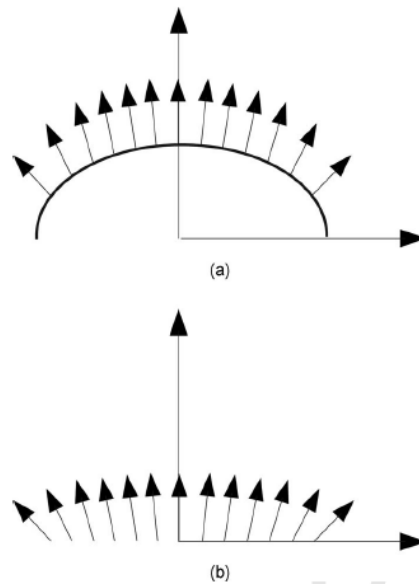


FIGURE 2.13: Paramètres pris en compte. (a) La normale et la profondeur du miroir. (b) La normale uniquement

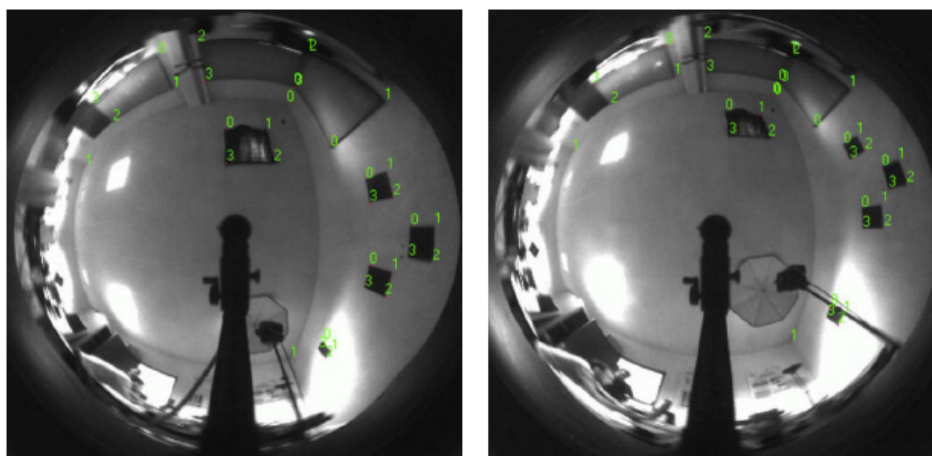


FIGURE 2.14: Mise en correspondance manuelle

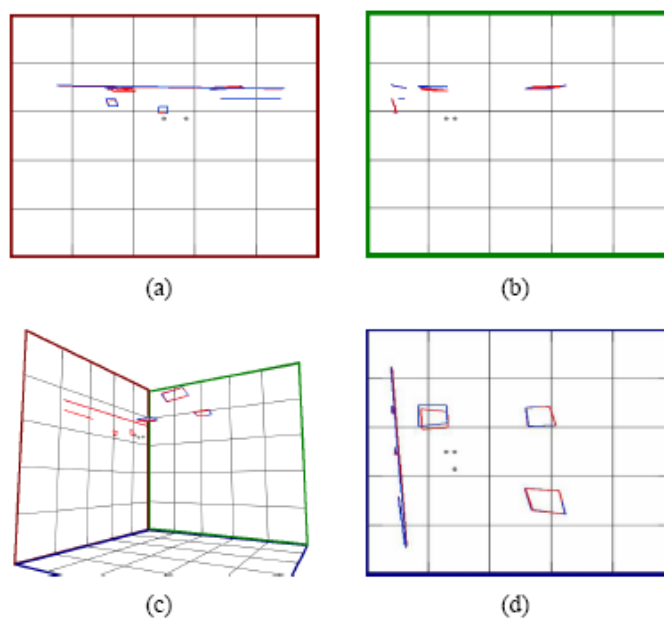


FIGURE 2.15: Reconstruction à partir d'un miroir hyperbolique

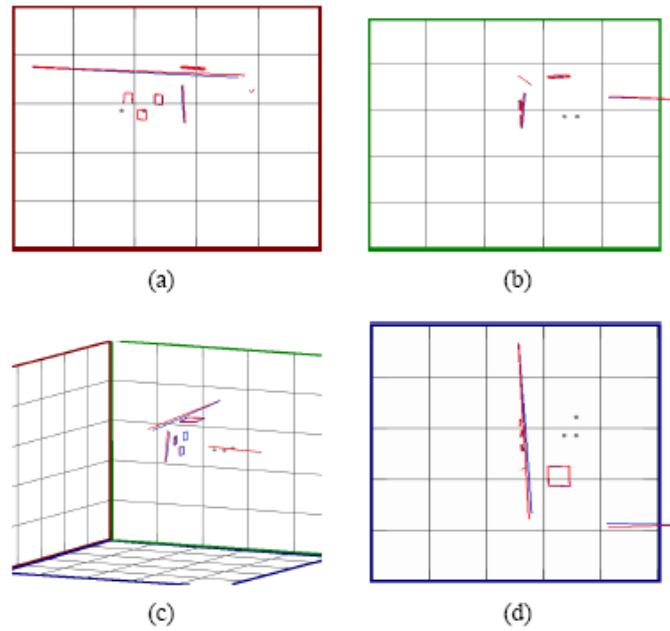


FIGURE 2.16: Reconstruction à partir d'un miroir sphérique

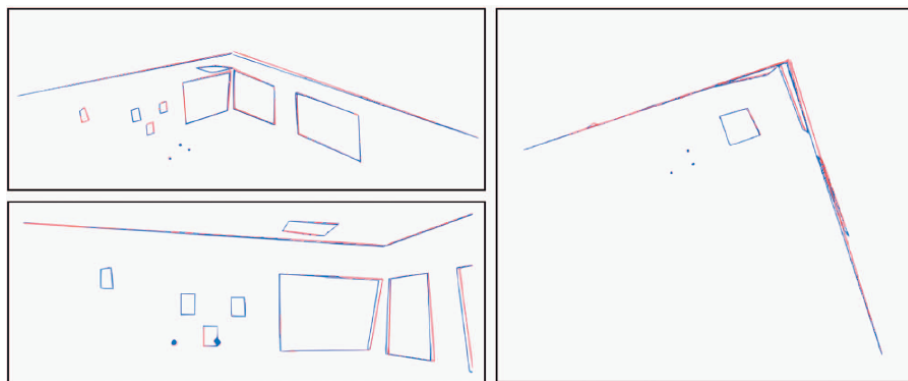


FIGURE 2.17: Reconstruction à partir de trois vues d'un miroir sphérique

2.3 Traitement Adapté des Images Catadioptriques

Les images catadioptriques sont anamorphosées : elles ont une résolution non-uniforme et un voisinage variant avec sa position dans l'image. Il est donc peu pertinent d'y appliquer les opérateurs de convolution classique, par exemple, ou tout autre opérateur de traitement des images "à noyau" (convolution, noyau gaussien, morphologie mathématique). Une première idée serait de redresser les images, puis d'y appliquer les traitements. Or, cette technique fait intervenir des interpolations et du lissage qui modifient l'information contenue dans l'image. Il faudrait à l'inverse distordre le noyau pour l'adapter en taille à la résolution et en forme au voisinage. Nous proposons une méthode permettant de définir un voisinage adapté quel que soit le miroir utilisé, nous montrons qu'il est également possible de définir ce voisinage à partir des paramètres de polarisation du capteur (degré et angle) et sans calibrage – que le capteur soit central ou non.

L'adaptation du voisinage et des traitements est un problème récurrent en vision omnidirectionnelle [Daniilidis 2002, Strauss 2005]. Nous l'avons abordé de manière directe (paramétrisation du miroir et ré-écriture *ad hoc* du voisinage et des opérateurs), sans perdre toutefois de vue que nous voulions rester assez général pour prendre en compte à la fois les capteurs centraux et non-centraux. L'objectif, *in fine*, est de proposer un paradigme générique pour le traitement des images anamorphosées, un peu à la manière des modèles génériques de caméra pour la géométrie.

2.3.1 Définition du Voisinage : Cas Paramétrique

L'idée est de paramétrer chaque point de la surface du miroir. Nous devons supposer que le capteur a été initialement calibré. Après projection du point de la surface du miroir sur le plan image, les paramètres du point du miroir apparaîtront comme argument du pixel correspondant. Ceci conduit à une relation non-linéaire entre un pixel et son voisinage.

Prenons l'exemple du miroir parabolique – sans perte de généralité. L'équation de la paraboloïde est donnée par :

$$z = -f + \frac{x^2 + y^2}{4f} \quad (2.21)$$

où $\mathbf{X} = [x \ y \ z]^T$ sont les coordonnées d'un point à la surface du miroir et f est la focale du miroir obtenue par calibrage [Mei 2007]. Une paramétrisation possible du point \mathbf{X} est donnée par :

$$x = f\sqrt{2t} \cos \tau \quad y = f\sqrt{2t} \sin \tau \quad z = 2f(t - 1) \quad (2.22)$$

avec $\tau \in [0, \pi]$ et $t \in [0, h]$; h étant la distance entre le sommet et le contour du miroir.

La relation liant un point de la surface du miroir et son image est donnée par :

$$\begin{bmatrix} u \\ v \end{bmatrix} = \mathbf{KR}(\mathbf{X} - \mathbf{t}) \quad (2.23)$$

où \mathbf{K} , \mathbf{R} et \mathbf{t} sont respectivement la matrice des paramètres intrinsèques de la caméra, la rotation et la translation entre les systèmes de coordonnées du miroir et de la caméra. Le vecteur de translation est défini par $\mathbf{t} = [0 \ 0 \ -2\epsilon]^T$ où ϵ est l'excentricité du miroir; $\mathbf{t} = \mathbf{0}_{3 \times 1}$ quand on considère un miroir parabolique et une caméra orthographique. On en déduit donc que :

$$\begin{cases} u(t, \tau) = u_0 + \alpha_u \sqrt{2t} \cos(\tau + \frac{\pi}{2}) \\ v(t, \tau) = v_0 + \alpha_v \sqrt{2t} \sin(\tau + \frac{\pi}{2}) \end{cases} \quad (2.24)$$

où u_0, v_0, α_u et α_v sont les paramètres intrinsèques de la caméra, respectivement les coordonnées du point principal et la distance focal en pixels. Dans le cas d'une matrice de pixels orthogonale, un déplacement infinitésimal (h, k) dans l'image est défini par $(u + h, v + k)$. En considérant le déplacement dans l'image comme la projection du déplacement sur le miroir, le déplacement infinitésimal sera cette fois défini par $(u(t(i + h, j + k), \tau(i + h, j + k)), v(t(i + j, j + k), \tau(i + h, j + k)))$ où i

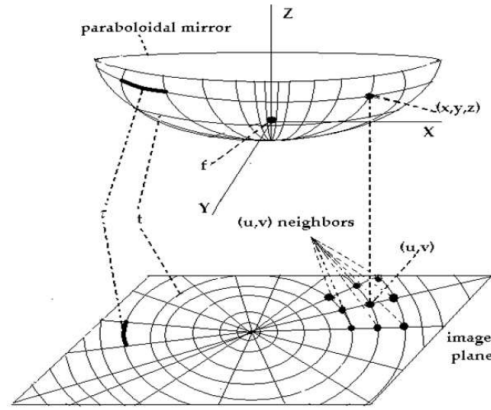


FIGURE 2.18: Exemple de paramétrisation dans le cas orthographique

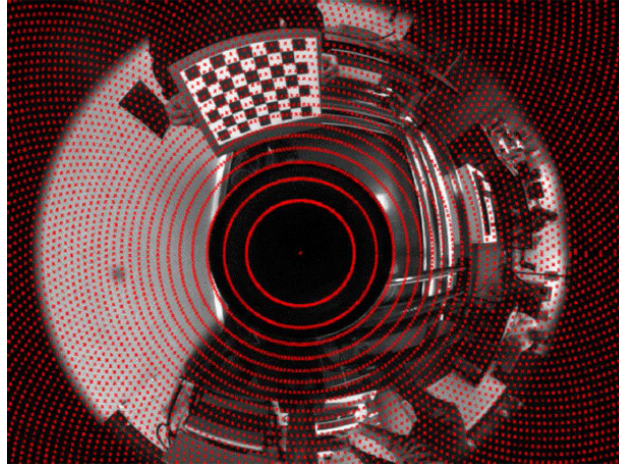


FIGURE 2.19: Exemple de voisinage projeté sur l'image

et j sont les indices échantillonnés des paramètres. La figure 2.18 illustre la projection miroir/image et le voisinage estimé.

En conséquence, deux pixels voisins ne partagent plus la même ligne ou la même colonne de l'image, mais reposent sur le même cercle de rayon $t = \frac{u-u_0^2}{2\alpha_u} + \frac{v+v_0^2}{2\alpha_v}$ ou sur le même rayon d'angle $\tau = \arctan\left(\frac{\alpha_u}{\alpha_v} \frac{v-v_0}{u-u_0}\right)$. La figure 2.19 présente la projection du voisinage sur une image réelle. Il faut noter que, par souci de lisibilité, un échantillon seulement des pixels a été projeté. Les masques que l'on peut former à partir de ce voisinage balayent l'image en spirale du centre vers la périphérie. Il est possible de calculer des dérivées, de détecter des contours ou des points d'intérêt, d'appliquer des opérateurs de morphologie mathématique (même si cette dernière possibilité n'a pas été expérimentée).

2.3.2 Définition du Voisinage : Cas Non-Paramétrique

Nous montrons maintenant qu'il est possible d'obtenir de manière directe ce voisinage, sans calibrage préalable du capteur, en utilisant judicieusement les paramètres de polarisation du miroir. Nous avons vu dans la section 2.2 que la polarisation nous permettait de calibrer de manière dense et sans traitement d'image une caméra catadioptrique. En partant des mêmes présupposés et en

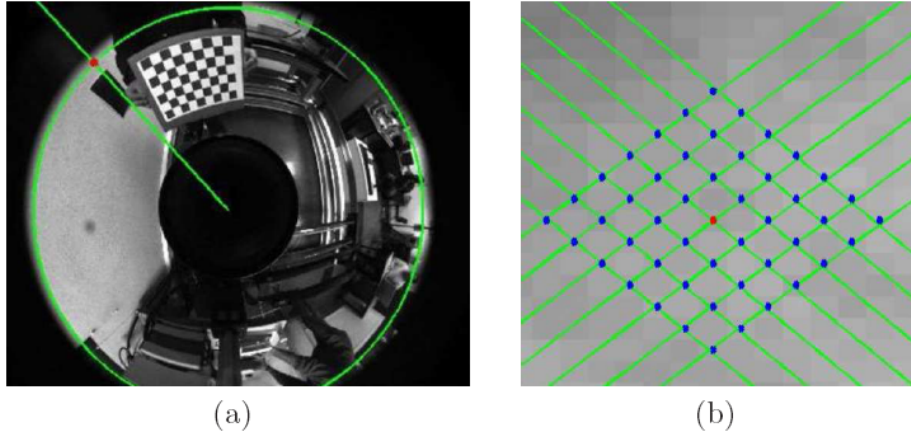


FIGURE 2.20: Equivalence entre le voisinage et les paramètres de polarisation. (a) Le cercle représente le degré de polarisation et le rayon, l'angle. (b) Le voisinage 7×7 du pixel considéré

appliquant les mêmes principes, nous allons montrer qu'il est directement possible de passer de l'angle et du degré de polarisation au voisinage adapté, ce, même pour un capteur non-central.

La normale en tout point du miroir est définie par :

$$\mathbf{n} = \begin{bmatrix} \tan \theta \cos \phi \\ \tan \theta \sin \phi \\ 1 \end{bmatrix} \quad (2.25)$$

La loi de SNELL-DESCARTES nous permet d'établir le lien entre les coordonnées de la normale et les paramètres de polarisation :

$$\begin{cases} \varphi = \phi \pm \frac{\pi}{2} \\ \rho(\theta) = \frac{2n \tan \theta \sin \theta}{\tan^2 \theta \sin^2 \theta + |\hat{n}|} \end{cases} \quad (2.26)$$

avec $\hat{n} = n + ik$, l'indice complexe de réfraction du matériau. La projection des angles θ et ϕ est directement liée aux paramètres t et τ : nous considérons que deux voisins dans l'image partagent le même degré et le même angle de polarisation. Les pixels peuvent donc être exprimés par $(u(\rho, \varphi), v(\rho, \varphi))$. La figure 2.20 illustre cette équivalence.

2.3.3 Dérivation d'Opérateur de Traitement des Images

De nombreux opérateurs de traitement des images peuvent être dérivés du voisinage adapté. En fait, tous les opérateurs de convolution et les opérateurs basés sur un masque. Nous donnerons quatre exemples : la détection de contours, la dérivée de l'image, la détection de points d'intérêt et l'appariement de points.

Détection de contours. Considérons que E est l'ensemble des contours d'une image I , détecté par la formule suivante :

$$E(i, j) = \sum_k \sum_l G(k, l) |I(i + k, j + l) - I(i, j)| \quad (2.27)$$

où G est le noyau Gaussien. Dans le plan catadioptrique, la formule devient :

$$E(i, j) = \sum_k \sum_l G(\rho(k, l), \phi(k, l)) |I(\rho(i + k, j + l), \phi(i + k, j + l)) - I(\rho(i, j), \phi(i, j))| \quad (2.28)$$

Le déplacement dans l'image est bien porté par les paramètres (ρ, θ) et pas par les coordonnées (u, v) .

Dérivation de l'image. Les dérivées horizontales et verticales de l'image deviendront dérivées circulaire et radiale dans le plan catadioptrique. Ces dérivées sont définies par :

$$\begin{cases} \frac{\partial I}{\partial \rho} = \frac{\partial I}{\partial u} \frac{\partial u}{\partial \rho} + \frac{\partial I}{\partial v} \frac{\partial v}{\partial \rho} \\ \frac{\partial I}{\partial \phi} = \frac{\partial I}{\partial u} \frac{\partial u}{\partial \phi} + \frac{\partial I}{\partial v} \frac{\partial v}{\partial \phi} \end{cases} \quad (2.29)$$

où la dérivée partielle $\frac{\partial I}{\partial u}$ représente la différence entre cercles voisins (tous les points du cercle partagent le même degré de polarisation) et $\frac{\partial I}{\partial v}$ représente la différence entre radiales voisines (tous les points sur les radiales partagent le même angle de polarisation).

Détection de point d'intérêt. Nous avons ré-écrit le détecteur de HARRIS [Harris 1988] puisqu'il est basé sur la dérivation de l'image et la convolution avec un noyau Gaussien. Dans le cas perspectif traditionnel, la matrice M , représentant la convolution des dérivées de l'image et du noyau Gaussien G , doit être calculée pour chaque pixel :

$$M = \begin{bmatrix} \frac{\partial I^2}{\partial x} & \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} \\ \frac{\partial I}{\partial x} \frac{\partial I}{\partial y} & \frac{\partial I^2}{\partial y} \end{bmatrix} \otimes G \quad (2.30)$$

$\frac{\partial I}{\partial x}$ et $\frac{\partial I}{\partial y}$ sont les dérivées horizontales et verticales. Un point d'intérêt est détecté si $R = \det(M) - k \cdot \text{trace}(M)$ est positif, k étant fixé empiriquement. Dans le plan catadioptrique, on obtient :

$$M = \begin{bmatrix} \frac{\partial I^2}{\partial \rho} & \frac{\partial I}{\partial \rho} \frac{\partial I}{\partial \phi} \\ \frac{\partial I}{\partial \rho} \frac{\partial I}{\partial \phi} & \frac{\partial I^2}{\partial \phi} \end{bmatrix} \otimes G \quad (2.31)$$

Appariement de points. L'appariement de points par corrélation requiert la définition d'un voisinage. Pour l'exemple, nous considérerons la NSSD (*Normalized Squared of Normal Differences*), définie par :

$$NSSD(p_i, p_j) = \frac{\|neigh(p_i) - neigh(p_j)\|^2}{\|neigh(p_i)\| \|neigh(p_j)\|} \quad (2.32)$$

où p_i et p_j sont les pixels de l'image i et j respectivement ; $neigh(p)$ est l'ensemble des pixels voisins de p . L'adaptation est directe : au lieu d'utiliser le voisinage orthogonal, on utilisera un voisinage adapté sur lequel effectué les calculs.

2.3.4 Résultats Expérimentaux

Les opérateurs dérivés dans la sous-section précédentes ont été implémentés et testés. La figure 2.21 présente les résultats de la détection de contours sur un voisinage de 5×5 . Nous proposons une comparaison visuelle avec la méthode perspective. On constate clairement que les contours sont mieux localisés quand le voisinage est adapté. On remarque également des artefacts dans l'image de gauche (sur la partie gauche).

Nous avons effectué le même type de comparaison pour le détecteur de points d'intérêt (figure 2.22). On constate que quand le voisinage est adapté, on parvient à détecter plus de points et qu'ils sont mieux répartis dans l'image (on rencontre moins d'amas). On peut conjecturer que ces artefacts sont provoqués par le voisinage carré en lui-même, créant artificiellement des coins dans l'image catadioptrique.

Enfin, nous présentons un exemple de mise en correspondance (figure 2.23). Dans l'exemple, représentant un cas très certainement extrême, le voisinage classique est mis en défaut presque

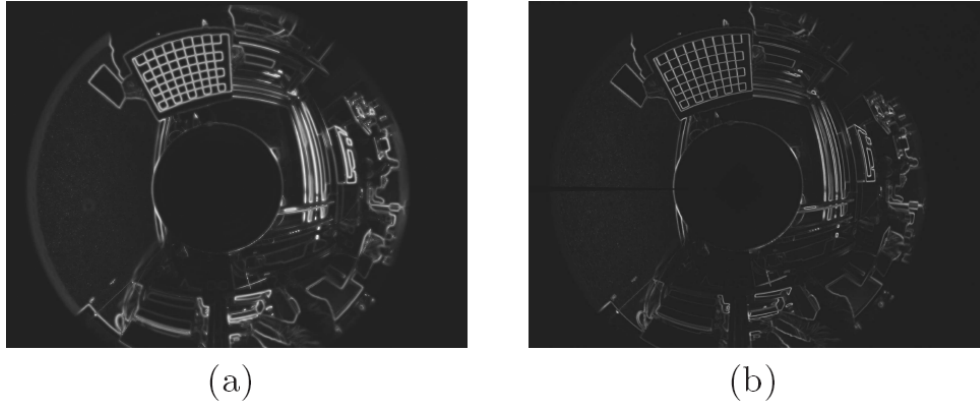


FIGURE 2.21: Détection de contours. (a) Par la méthode perspective. (b) Par la méthode adaptée

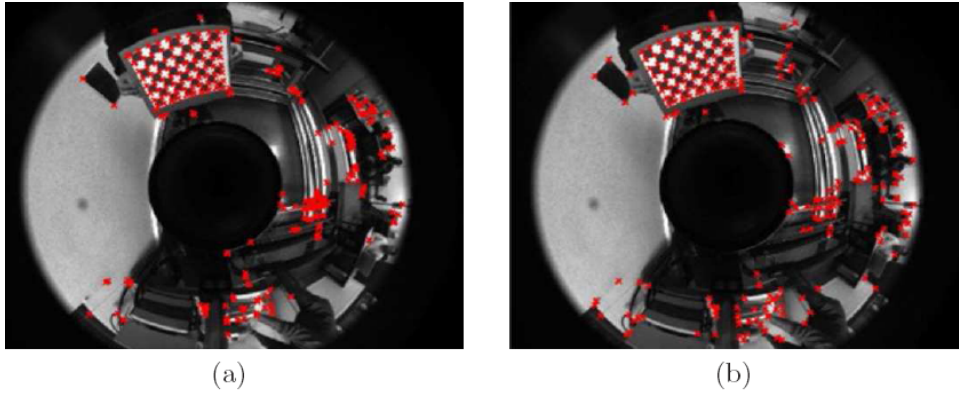


FIGURE 2.22: Détection de points d'intérêt par la méthode de Harris. (a) Sur un voisinage classique. (b) Sur un voisinage adapté

systématiquement alors que le voisinage adapté permet une mise en correspondance correcte de tous les points.

Nous manquons encore de recul sur la méthode d'adaptation du voisinage pour fournir une comparaison quantitative et pertinente avec la méthode classique. Nous devons également entreprendre une évaluation comparative avec les méthodes concurrentes. Aussi, nous éprouvons les opérateurs adaptés sur des capteurs catadioptriques non-centraux : capteurs désalignés, dos de cuillère, etc. Nous cherchons à démontrer les apports, mais aussi les limites de la méthode développée et à justifier l'équivalence entre les paramètres de polarisation et le voisinage spatial quand le point de vue unique n'est pas respecté.

2.4 Mise en Correspondance de Droites Catadioptriques

La mise en correspondance de droites consiste à trouver, au sein d'une séquence d'images (au moins deux), les portions de droites correspondant à la même droite physique 3D. C'est une étape fondamentale dans l'utilisation d'un capteur de vision pour la navigation d'un robot mobile. En effet, elle permet tout à la fois d'apparier des primitives homologues, d'estimer le déplacement du capteur d'une prise de vue à l'autre, d'alimenter un algorithme d'auto-calibrage et de reconstruire en 3D la scène observée. Nous avons mené ces travaux préliminaires en collaboration avec le MIS de l'Université de Picardie Jules Verne via le co-encadrement d'un stagiaire de master. Nous les pour-

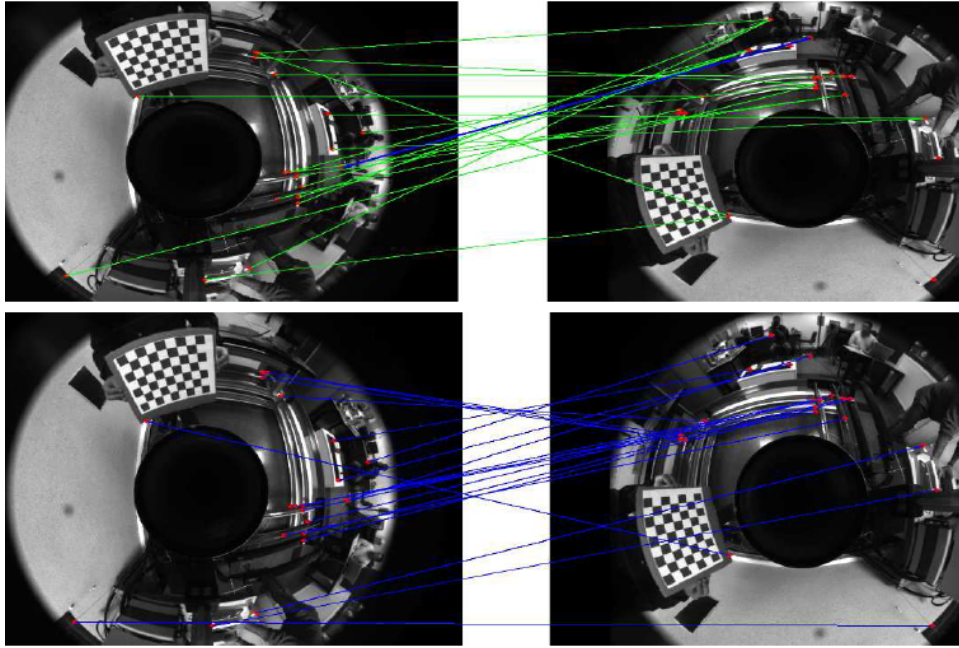


FIGURE 2.23: Mise en correspondance. (En Haut) Sur un voisinage classique. (En Bas) Sur un voisinage adapté

suivons dans le cadre d'une thèse de doctorat, débutée en octobre 2008, qui impliquera également les deux laboratoires (MIS et Le2i).

La projection d'une droite de l'espace sur le plan image d'un capteur catadioptrique central donne une portion de conique. Si on considère la projection sur la sphère équivalente, une droite donne un grand cercle. L'objectif sera d'apparier ces grands cercles d'une image à l'autre et d'estimer simultanément le déplacement. Il est entendu que, sans information complémentaire, il est difficile de calculer la translation séparant les deux prises de vue; la connaissance de points physiques (extrémités de segment par exemple) permet de contraindre un peu plus le système et d'envisager ce calcul. En première approximation, nous proposons une méthode permettant d'effectuer l'appariement dans le cas où la translation est courte (*short baseline motion*) ou nulle.

Une droite sera représentée par le vecteur normal du plan passant par le centre de la sphère équivalente et la droite 3D dans l'espace. Nous parlerons d'*intersection* pour désigner le vecteur unitaire dirigé du centre de la sphère à point d'intersection de deux droites 3D projeté sur la sphère; de *point d'intersection* pour désigner le point où deux lignes viennent se couper dans le plan image; d'*angle d'intersection* pour désigner l'angle entre deux grands cercles; de *lignes* quand nous désignerons la projection des droites 3D sur le plan image ou sur la sphère équivalente.

2.4.1 Relation entre les Droites

Supposons que la droite \mathbf{n}_1 est l'homologue de la droite \mathbf{n}_2 (figure 2.24). Connaissant un ensemble de correspondances est-il possible d'extraire la matrice de transformation reliant deux positions du capteur? On démontre que cette relation dépend de la rotation \mathbf{R} , de la translation \mathbf{t} , mais également de l'orientation de la droite dans l'espace :

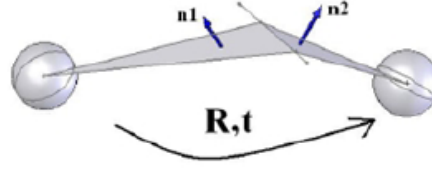


FIGURE 2.24: Droites 3D dans l'espace et projection sur la sphère équivalente. \mathbf{n}_1 et \mathbf{n}_2 sont les vecteurs normaux

$$\begin{aligned}
 \mathbf{n}_2 &= (\mathbf{R}\mathbf{X}_1 + \mathbf{t}) \times (\mathbf{R}\mathbf{X}_2 + \mathbf{t}) = \dots \\
 \det(\mathbf{R}) \mathbf{R}^{-t} (\mathbf{X}_1 \times \mathbf{X}_2) + [\mathbf{t}]_{\times} \mathbf{R} (\mathbf{X}_1 - \mathbf{X}_2) &= \dots \\
 \det(\mathbf{R}) \mathbf{R}^{-t} \mathbf{n}_1 + [\mathbf{t}]_{\times} \mathbf{R} (\mathbf{X}_1 - \mathbf{X}_2) &= \mathbf{R}^* \mathbf{n}_1 + [\mathbf{t}]_{\times} \mathbf{R} (\mathbf{X}_1 - \mathbf{X}_2) \quad (2.33)
 \end{aligned}$$

Où \mathbf{X}_1 et \mathbf{X}_2 sont des points reposant sur la droite d'orientation \mathbf{n}_1 et \mathbf{n}_2 respectivement et \mathbf{R}^* est la matrice de co-facteur de \mathbf{R} .

En supposant que le déplacement est une rotation pure ou que la translation est faible par rapport à la distance séparant le capteur de la scène (en robotique aérienne, par exemple), le second terme peut être négligé :

$$\mathbf{n}_2 = \mathbf{R}^* \mathbf{n}_1 \quad (2.34)$$

La connaissance de \mathbf{R}^* permet donc d'établir la correspondance de toutes les droites de la scène à partir de la mise en correspondance initiale de trois droites uniquement.

2.4.2 Algorithme de Mise en Correspondance

L'algorithme de mise en correspondance se sub-divise en cinq grandes étapes, que nous détaillons dans la suite de la section :

1. l'extraction des lignes sur le plan image et le calcul des intersections ;
2. l'estimation d'une mise en correspondance initiale en utilisant la contrainte de consistance des angles ;
3. l'affinage des données en utilisant la corrélation et la contrainte de consistance antipodale ;
4. le calcul de \mathbf{R} par RANSAC à partir de la mise en correspondance initiale ;
5. la mise en correspondance finale par \mathbf{R}^* .

Nous avons identifié deux contraintes, appelées *contrainte de consistance des angles* et *contrainte de consistance antipodale*, que nous définissons ainsi :

Contrainte de consistance des angles : l'angle à l'intersection de deux lignes sur la sphère équivalente est invariant si le déplacement est une rotation pure. En effet, l'angle entre les images de deux droites 3D sur la sphère équivalente est l'angle entre les deux plans qui passent par le centre de la sphère et ces droites. Puisque, dans le cas d'une rotation pure, le centre de la sphère est immobile et que les droites sont supposées fixes dans la scène, l'angle restera invariant après rotation du capteur.

Contrainte de consistance antipodale : si deux points d'intersection dans deux images différentes sont en correspondance, alors leurs points antipodaux sont également en correspondance (la figure 2.25 donne un exemple de grands cercles, d'intersections et de points antipodaux).

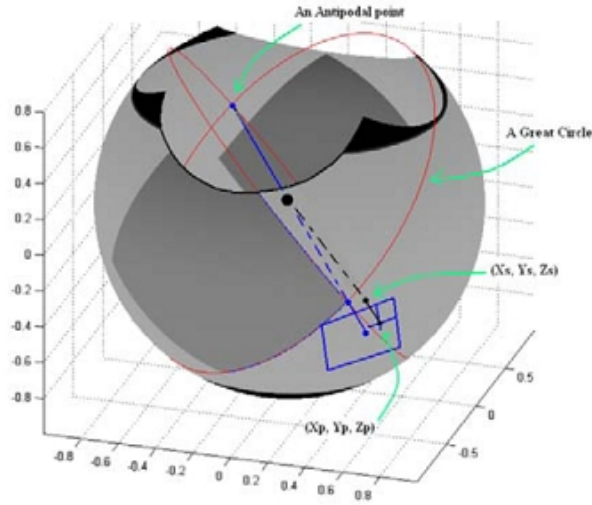


FIGURE 2.25: Projection d'un pixel sur la sphère équivalente, grands cercles et points antipodaux

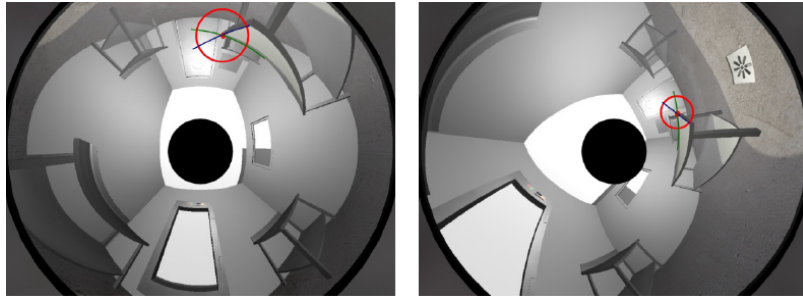


FIGURE 2.26: Consistance des angles sous une rotation pure

2.4.2.1 Extraction des Lignes, Intersections et Angles

Nous utilisons l'algorithme proposé en [Bazin 2007] pour segmenter les lignes de l'image. Ces lignes sont des coniques qui se coupent en deux points dits antipodaux quand ils sont projetés sur la sphère équivalente. A partir des lignes segmentées, nous calculons toutes les intersections possibles dans le plan image. La contrainte de consistance des angles nous permet d'obtenir une mise en correspondance initiale en appariant les intersections partageant le même angle (figure 2.26).

2.4.2.2 Corrélation au Voisinage des Intersections

Un point d'intersection divise son voisinage en quatre quadrants. En supposant la scène rigide et une rotation pure du capteur (ou, pour le moins, une translation négligeable par rapport à la rotation), la portion de scène "derrière" les quadrants est invariante. Nous pouvons donc utiliser ce voisinage pour estimer une corrélation entre points d'intersection. En fait, il est suffisant de considérer deux fenêtres (*patches*) centrées sur le point d'intersection et orientées le long des lignes comme sur la figure 2.27. Ces *patches* sont adaptés à la géométrie de la sphère équivalente.

Nous utilisons une simple corrélation-croisée normalisée (CC) pour appairer les *patches* – cette étape a pour but d'affiner la correspondance obtenue à l'étape précédente et à rejeter les éventuelles faux-positifs ; les scores de corrélation inférieurs à un seuil donné sont automatiquement rejetés :

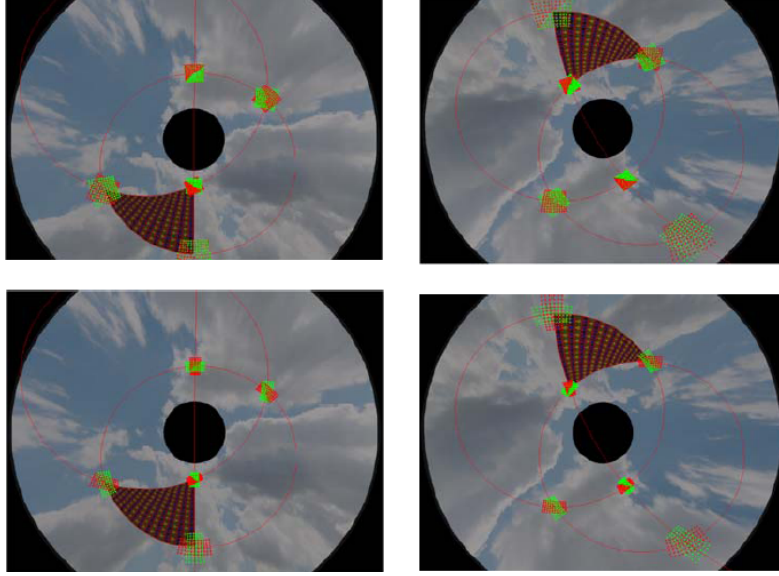
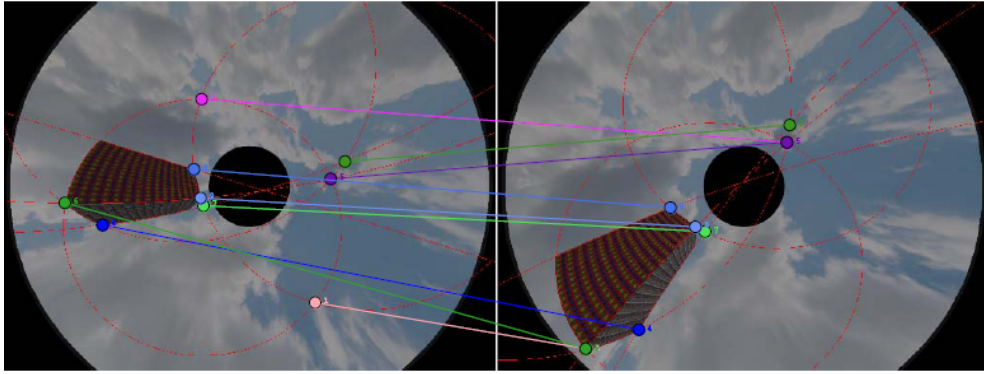
FIGURE 2.27: Extraction des *patches*

FIGURE 2.28: Exemple de mise en correspondance des intersections

$$CC = \frac{\sum_{j=1}^R \sum_{i=1}^C (p_t(i, j) - \bar{p}_t) (p_s(i, j) - \bar{p}_s)}{\sqrt{\sum_{j=1}^R \sum_{i=1}^C (p_t(i, j) - \bar{p}_t)^2 \sum_{j=1}^R \sum_{i=1}^C (p_s(i, j) - \bar{p}_s)^2}} \quad (2.35)$$

où p_t et p_s sont les niveaux de gris dans les deux *patches* à tester ; \bar{p}_t et \bar{p}_s , sont les moyennes des niveaux de gris ; R et C correspondent au nombre de lignes et de colonnes du *patch*. En complément, nous utilisons la contrainte de consistance antipodale pour rejeter les points mal appariés : si un point d'intersection i est apparié au point j dans l'autre image, alors son point antipodal doit être apparié au point antipodal de j . La figure 2.28 montre le résultat d'une mise en correspondance. Il faut noter qu'à ce stade, il peut encore subsister des appariements incorrects.

2.4.2.3 Calcul de la Rotation

La dernière étape consiste à estimer la rotation par la méthode RANSAC, tout en rejetant les derniers faux appariements [Fischler 1981]. L'équation 2.34, liant deux grands cercles mis en correspondance, servira de fonction d'ajustement (*fitting function*) pour l'algorithme. RANSAC

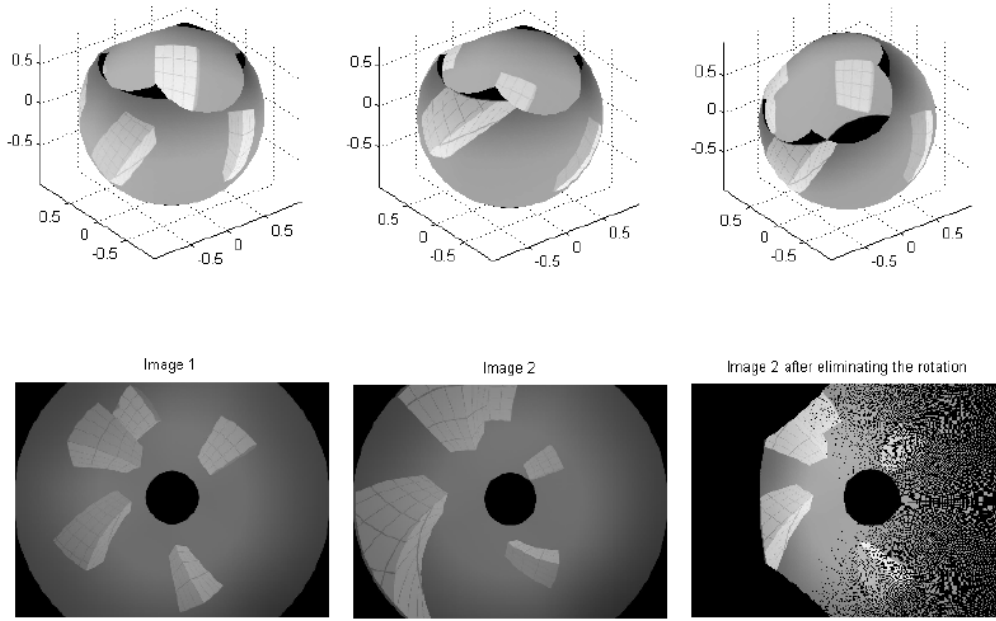


FIGURE 2.29: Translation longue. La première et la deuxième colonne sont des images prises à différentes positions. La troisième colonne est l'image dont on a compensé la rotation

peut être appliqué sur les lignes (ou grands cercles), mais également directement sur les points d'intersection :

$$\mathbf{I}_2 = \mathbf{n}'_1 \times \mathbf{n}'_2 = (\mathbf{R}^* \mathbf{n}_1) \times (\mathbf{R}^* \mathbf{n}_2) = \mathbf{R}^{**} (\mathbf{n}_1 \times \mathbf{n}_2) = \mathbf{R} \mathbf{I}_1 \quad (2.36)$$

où les \mathbf{I}_i représentent des intersections, les \mathbf{n}'_i sont les vecteurs normaux aux grands cercles qui donnent l'intersection. Une fois \mathbf{R} (ou \mathbf{R}^*) estimé, il est possible d'obtenir la mise en correspondance des lignes (et des points d'intersection) via l'équation 2.34. Il faut, au préalable, s'assurer que la matrice \mathbf{R} est bien une matrice de rotation et l'orthonormalisée.

2.4.3 Rotation et Translation Longue

Si \mathbf{R} est connue, il est possible de compenser l'effet de la rotation dans l'image en y appliquant une rotation de même norme mais de signe opposé sur la sphère équivalente (figure 2.29). En re-projetant la sphère obtenue sur le plan image, on constate que les lignes en correspondance sont parallèles et que la distance qui les séparent dépend de la valeur de la translation (figure 2.30).

2.4.4 Résultats Expérimentaux

L'algorithme décrit précédemment a été éprouvé sur des données synthétiques et réelles. La méthode est globalement validée. Il reste cependant à évaluer le comportement de l'algorithme vis-à-vis de la translation : quelle est l'ampleur de la translation autorisée – qu'est-ce qu'une *short baseline* ? Jusqu'où peut-on tenir l'hypothèse ? Nous devons également multiplier les résultats sur des données réelles contenant un nombre plus conséquent de lignes et quantifier plus justement la robustesse au bruit et aux points aberrants. Enfin, la méthode peut être considérablement améliorée dans ses détails : corrélation, utilisation de points physiques pour prendre en compte des translations importantes, etc.

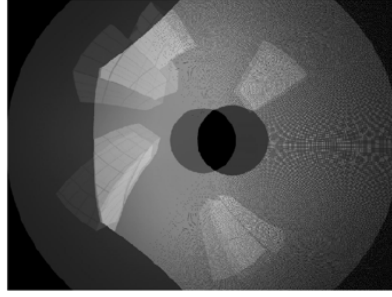


FIGURE 2.30: Surimpression de l'image initiale et de l'image compensée : les lignes sont parallèles

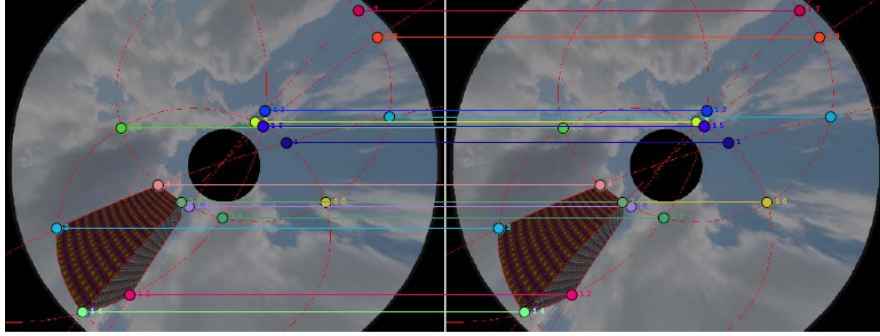


FIGURE 2.31: Mise en correspondance sous un déplacement nul

2.4.4.1 Données Synthétiques

La figure 2.31 montre les résultats de mise en correspondance sous un déplacement nul (ni translation, ni rotation). La matrice des co-facteurs de la rotation obtenue est :

$$\mathbf{R}^* = \begin{bmatrix} 0.5774 & 0.0000 & -0.0000 \\ 0.0000 & 0.5774 & 0.0000 \\ 0.0000 & 0.0000 & 0.5774 \end{bmatrix} \longrightarrow \mathbf{R}_{ortho}^* = \begin{bmatrix} 1.0000 & -0.0000 & 0.0000 \\ 0.0000 & 1.0000 & -0.0000 \\ -0.0000 & 0.0000 & 1.0000 \end{bmatrix}$$

où \mathbf{R}_{ortho}^* est la matrice orthonormalisée (la matrice des co-facteurs de la matrice identité est la matrice identité, donc $\mathbf{R}_{ortho} = \mathbf{I}_{3 \times 3}$).

Nous proposons ensuite l'estimation de la mise en correspondance sous une faible rotation (figure 2.32) – 40° autour de l'axe z et 15° autour de l'axe y . Nous ne donnons pas les résultats intermédiaires de la mise en correspondance initiale. On peut néanmoins souligner le fait que la fonction de corrélation utilisée n'est sans doute pas la plus efficace et qu'il faudrait assurément revoir la méthode de mise en correspondance des intersections (en utilisant les traitements adaptés que nous avons décrits plus haut). L'estimation de la rotation donne :

$$\mathbf{R}^* = \begin{bmatrix} 0.6631 & -0.5502 & -0.5076 \\ -0.7080 & -0.6810 & -0.1868 \\ 0.2429 & -0.4832 & 0.8411 \end{bmatrix} \longrightarrow \mathbf{R} = \begin{bmatrix} 0.7567 & 0.6071 & -0.2428 \\ -0.6114 & 0.7886 & 0.0664 \\ 0.2318 & 0.0982 & 0.9678 \end{bmatrix}$$

Nous avons calculé l'expression suivante afin de vérifier si la rotation donnée par l'algorithme correspondait à la rotation opérée initialement sur l'image :

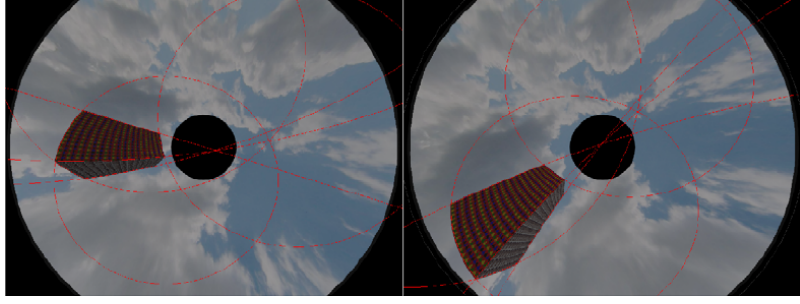


FIGURE 2.32: Images synthétiques séparées par une faible rotation

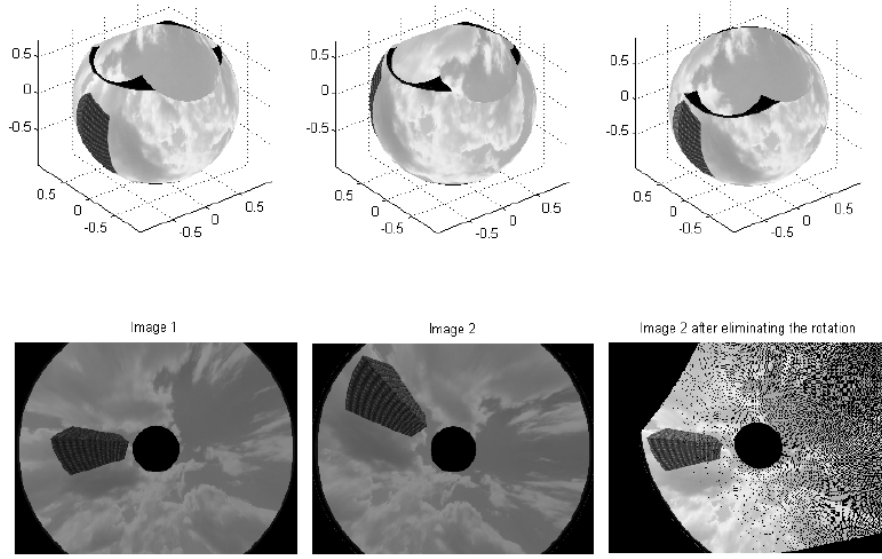


FIGURE 2.33: Compensation de la rotation sur l'image 2

$$\begin{aligned}
 & \mathbf{R}_{-40}^z \mathbf{R}_{-15}^y \mathbf{R}^{-1} = \dots \\
 \dots = & \begin{bmatrix} 0.7660 & 0.6428 & 0 \\ -0.6428 & 0.7660 & 0 \\ 0 & 0 & 1.000 \end{bmatrix} \begin{bmatrix} 0.9659 & 0 & -0.2588 \\ 0 & 1.0000 & 0 \\ 0.2588 & 0 & 0.9659 \end{bmatrix} \begin{bmatrix} 0.7603 & -0.6119 & 0.2182 \\ 0.6048 & 0.7893 & 0.1061 \\ -0.2371 & 0.0513 & 0.9701 \end{bmatrix} = \dots \\
 \dots = & \begin{bmatrix} 0.9964 & 0.0195 & 0.0828 \\ -0.0315 & 0.9889 & 0.1453 \\ -0.0790 & -0.1474 & 0.9859 \end{bmatrix} \simeq \mathbf{I}_{3 \times 3}
 \end{aligned} \tag{2.37}$$

Nous pouvons également vérifier visuellement la correcte estimation de la rotation (figures 2.33 et 2.34). L'erreur est visuellement minime et l'estimation de la rotation est suffisante pour garantir la mise en correspondance de toutes les lignes.

2.4.4.2 Données Réelles

Nous avons testé l'algorithme sur des données réelles (figure 2.35). La rotation estimée est :

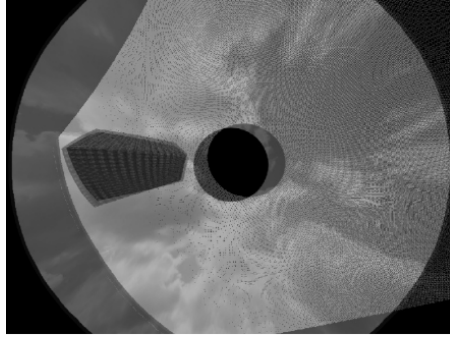


FIGURE 2.34: Surimpression de l'image initiale et de l'image compensée

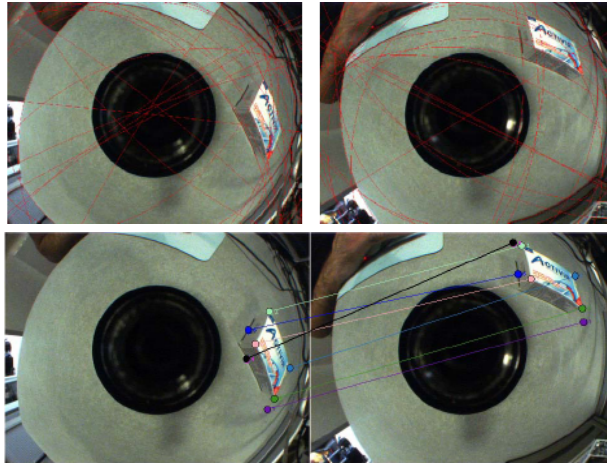


FIGURE 2.35: Couple d'images et mise en correspondance

$$\mathbf{R} = \begin{bmatrix} 0.7620 & 0.6476 & -0.0008 \\ -0.6476 & 0.7620 & 0.0095 \\ 0.0068 & -0.0067 & 1.0000 \end{bmatrix} \quad R_x = -2.3752^\circ, R_y = -0.2256^\circ, R_z = -41.9013^\circ$$

Elle semble cohérente avec l'analyse visuelle des images. Nous travaillons actuellement à la mise en place d'un dispositif expérimental permettant de connaître *a priori* le déplacement et ainsi d'évaluer quantitativement nos résultats et de tester l'hypothèse de *short baseline*.

2.5 Conclusions

Le projet CaViAR prendra fin en décembre 2008. Nous aurons abordé les thématiques du traitement adapté des images, du calibrage et de la reconstruction, de la mise en correspondance de lignes catadioptriques. Nous poursuivons ces travaux dans le cadre de la thèse de doctorat de Saleh MOSADDEGH. A très court terme, nous envisageons d'établir un protocole expérimental visant à évaluer finement les algorithmes et méthodes proposés :

- quantification des erreurs de reconstruction et/ou de déplacement pour le calibrage ;
- évaluation des performances des traitements adaptés et comparaison avec les méthodes usuelles ;

- estimation des limites de l’hypothèse de *short baseline*.

Notre ambition a été jusque-là de rester le plus général possible, aussi bien pour le calibrage (la contrainte du point de vue unique peut être relâchée) que pour les traitements (nous n’utilisons pas la projection sur la sphère équivalente). Nous voudrions étendre ce principe à la mise en correspondance de lignes (peut-on se défaire de la contrainte de point de vue unique?). Nous étudierons également la mise en correspondance de lignes dans le cas d’une translation longue ou *wide baseline*. L’objectif est d’aborder peu à peu les problèmes de *structure-from-motion* (estimation de la pose relative, auto-calibrage) et d’aboutir à une chaîne intégrée de traitements suffisamment générale ou générique.

En trois ans, la cellule Vision 3D du Le2i a pu intégrer la vision catadioptrique comme l’une de ses thématiques de recherche : une thèse a débuté en janvier 2008 sur des aspects combinant l’imagerie polarimétrique et catadioptrique, une autre en octobre 2008 sur des aspects liés à la vision par ordinateur et la reconstruction 3D. Nous avons fait un abondant travail bibliographique en amont (notamment, un état de l’art sur les méthodes de calibrage, non publié) et profité de la confrontation interne au projet avec des équipes plus expérimentées pour nous “mettre à niveau”. Nous sommes aussi impliqués, avec le KAIST (Séoul, Corée du Sud) et le MIS de l’Université de Picardie Jules Verne, dans un projet de collaboration financé dans le cadre du programme STAR Hubert Curien de l’Egide.

Chapitre 3

Une Application à la Chirurgie Virtuelle

3.1 Contexte

L'objectif de ces travaux est la réalisation d'un dispositif à base de réalité virtuelle afin d'assister le chirurgien orthopédiste lors de la substitution de l'articulation gléno-humérale par une prothèse. Ce dispositif a pour but l'optimisation des débattements de l'articulation prothésée de l'épaule à partir d'un choix de différents scénarios de pose. Les travaux se basent sur de nombreuses thématiques telles que la chirurgie orthopédique, le traitement d'image, l'informatique graphique et la réalité virtuelle. La démarche choisie comprend une (i) phase de modélisation de la partie osseuse de l'articulation par des formes géométriques simples, ce modèle est paramétrable à partir de radios ou de coupes scanners; (ii) une phase de simulation des débattements de l'articulation gléno-humérale saine; ainsi que (iii) une phase de simulation des opérations de coupes et de perçages avec choix des implants prothétiques pour une optimisation des débattements de l'articulation en préopératoire (cette optimisation se fait grâce une comparaison de la cinématique de l'épaule saine avec les différents scénarios issus de la simulation préopératoire).

Cette thèse a été financée par la SGM¹, le Conseil Régional de Bourgogne et le Fonds Social Européen. Les travaux ont été menés en étroite collaboration avec le Professeur Trouilloud, chirurgien orthopédiste au CHU de Dijon. L'objectif général était de développer un dispositif d'aide à la chirurgie, en pré-opératoire et en per-opératoire, pour le placement d'une prothèse d'épaule, à l'image de ce qui existe déjà pour la hanche. Une simple transposition de la méthode s'est avérée inopérante pour plusieurs raisons : essentiellement par la différence notable de l'anatomie et de la biomécanique des deux articulations et par les contraintes techniques et ergonomiques imposées par le chirurgien. Les principales questions auxquelles nous avons tenté de répondre sont : comment modéliser tri-dimensionnellement la jointure scapulo-humérale à partir de données médicales minimales (quelques radiographies) ? quelles mesures pertinentes extraire des images médicales pour parvenir à reconstruire un modèle adapté au patient ?

Ces travaux sont très appliquées. Il s'agissait de répondre à un besoin très précis de chirurgiens orthopédistes du CHU de Dijon et de la société partenaire commercialisant les prothèses. La contribution est essentiellement méthodologique. Au terme du projet, le livrable était composé de deux logiciels, le premier pour l'extraction des paramètres et le second pour la simulation en pré-opératoire. Il est utile de préciser que l'intégration du système a fait l'objet d'un dialogue avancé avec une société spécialisée dans la conception de plateformes de chirurgie en réalité augmentée.

3.2 De l'Epaule au Modèle

Cette section vise principalement à introduire la terminologie médicale. Elle décrit succinctement l'anatomie de l'épaule, les prothèses utilisées et la technique opératoire dans le cadre d'un placement de prothèse, autrement dit, ce que nous devons modéliser et simuler.

1. Société de Génie Médical - Aston Medical

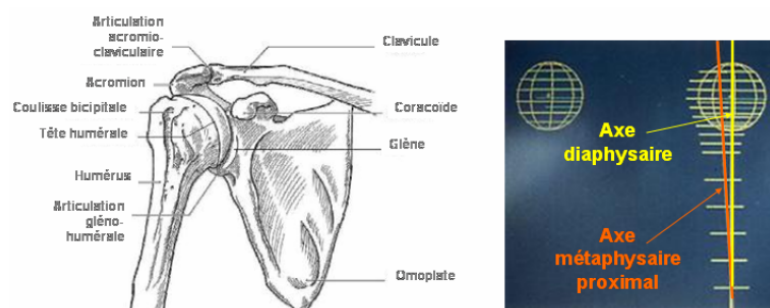


FIGURE 3.1: (A gauche) Partie osseuse de l'épaule. (A droite) Les deux axes principaux de l'humérus

3.2.1 Anatomie et Biomécanique de l'Epaule

L'épaule est un ensemble d'articulations qui permet à la main de s'orienter dans un grand volume d'espace, volume supérieur à une demi-sphère. Il s'agit de l'articulation la plus mobile de l'organisme mais cette grande mobilité a pour corollaire une grande instabilité pourvoyeuse de la plupart des pathologies mécaniques. La partie osseuse de l'épaule est constituée de la tête de l'humérus et de la cavité articulaire de l'omoplate ou glène, recouvertes de cartilage articulaire. L'omoplate s'étend pour former des excroissances osseuses : l'acromion entoure l'épaule d'arrière en avant pour former son toit, et la coracoïde en avant (voir la figure 3.1, à gauche).

L'extrémité supérieure de l'humérus peut être modélisée par une sphère épiphysaire et un cylindre métaphysaire [Roberts 1991]. Sur la figure 3.1 (à droite), nous définissons les deux axes de l'humérus : l'axe diaphysaire (en jaune), axe de rotation de l'humérus ; l'axe métaphysaire proximal (en orange), le long duquel une tige prothétique peut être descendue.

Les cinq articulations de l'épaule sont : le scapulo-huméral ou omo-huméral (omoplate-humérus) ; l'acromio-claviculaire ; le sterno-claviculaire ; l'omo-thoracique ; le sous-acromial. Seule l'articulation scapulo-humérale est concernée lors du placement d'une prothèse de l'épaule. On dénombre cinq mouvements participant au débattement de l'épaule, illustrés en figure 3.2 :

- l'abduction ;
- l'adduction ;
- l'antépulsion ;
- la rétropulsion ;
- la rotation externe ;
- la rotation interne.

Au final, l'objectif était de proposer un outil informatique permettant d'assurer les mêmes débattements pour une épaule prothésée que pour une épaule saine ; c'est-à-dire, estimé le débattement attendu en fonction du type, de la taille, de la position de la prothèse et de la morphologie du patient avant l'intervention, en plus d'offrir une assistance pendant l'intervention.

3.2.2 La Prothèse de l'Epaule

Nous avons exclusivement travaillé sur la prothèse Duocentric ; elle se décline en deux versions : la prothèse anatomique et la prothèse inversée (figure 3.3). Elle existe en différentes tailles afin de s'adapter au mieux à la morphologie du patient.

La première est composée d'une tige humérale anatomique, d'une calotte humérale et d'un support glène en acier inox ainsi que d'une glène anatomique en polyéthylène. Les vis à os spongieux sont en titane. Cette prothèse permet de respecter l'architecture biomécanique du patient grâce à l'intégration des paramètres de positionnement de la tête humérale et du col de glène et ce afin d'éviter les luxations et les usures précoces de la glène en polyéthylène. La tige humérale anatomique est conçue pour une coupe humérale à 140° .

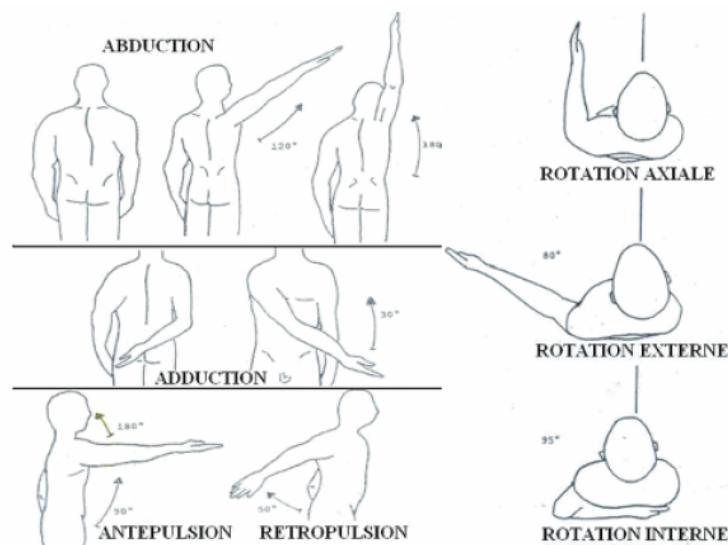


FIGURE 3.2: Les mouvements de l'épaule



FIGURE 3.3: (En haut) Prothèse anatomique. (En bas) Prothèse inversée

La prothèse inversée est composée d'une tige humérale inversée, d'une platine humérale, d'une duoglène et d'un support glène en acier inox ainsi que d'un insert huméral en polyéthylène. Elle a pour principe une inversion des géométries des surfaces articulaires en plaçant la partie concave du côté huméral et la partie convexe du côté glénoïdien. L'humérus est abaissé et le bras de levier du deltoïde est donc augmenté. On obtient ainsi une plus grande amplitude et une meilleure motricité du membre supérieur. La tige humérale inversée est conçue pour une coupe humérale à 125°.

3.2.3 Technique Opératoire (Prothèse Inversée)

Nous décrivons la technique opératoire pour le placement d'une prothèse inversée. Afin de ne pas surcharger inutilement le mémoire, nous omettrons dans la suite du chapitre tout ce qui a trait à la prothèse anatomique ; le principe et la méthodologie restent les mêmes.

Préparation de l'opération. Afin de préparer l'intervention, des radiographies du patient sont réalisées selon différentes incidences (figure 3.4). A l'aide d'un jeu de calques mis à disposition, le chirurgien peut alors planifier les tailles des différents implants. Lors de l'opération, le patient est installé en position semi-assise, le bras complètement libre reposant sur un appui le long du corps, la tête étant très bien maintenue. La voie d'abord est choisie à l'appréciation du chirurgien, la plus utilisée étant la voie supéro-externe ou antéro-supérieure. L'incision est de 8 à 10 cm, verticale,



FIGURE 3.4: (A gauche) Position du patient pendant l'intervention. (A droite) Emplacement en bleu de l'incision

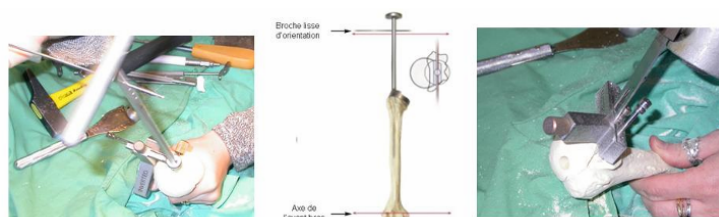


FIGURE 3.5: Préparation de l'humérus (gauche) tige centro-médullaire, guide de coupe et broche lisse de la duocentric (centre) coupe de la tête humérale (à droite)



FIGURE 3.6: Râpe, embout porte râpe et broche lisse

partant de l'articulation acromio-claviculaire et suivant le bord antérieur de l'acromion dans l'axe du deltoïde. Cette voie permet d'avoir une excellente exposition glénoïdienne telle que le travail se fasse perpendiculairement à la glène.

Première étape : préparation de l'humérus. La tête humérale est percée suivant l'axe centro-médullaire à l'aide d'un foret de diamètre 8mm. Le chirurgien monte ensuite le guide de coupe et son support sur la tige centro-médullaire avant d'installer cette tige dans l'humérus (figure 3.5). A l'aide d'une broche lisse, on simule l'angle de rétroversion par rapport à l'axe de l'avant-bras et avec la jauge de simulation, on visualise la future coupe. Une fois le guide de coupe positionné, il est fixé à l'humérus. On peut alors retirer le support du guide et la tige centro-médullaire puis couper la tête humérale. Le plan de coupe doit être minutieusement choisi car la mobilité de l'épaule prothésée en dépend grandement.

On monte ensuite la première râpe sur l'embout porte râpe correspondant à la version de l'implant et on insère la broche lisse pour régler l'angle de rétroversion (figure 3.6). On impacte le tout dans l'humérus et on réitère jusqu'à ce que la taille de tige adaptée soit atteinte.

Deuxième étape : travail de la glène. Un écarteur spécial dégage la glène (figure 3.7). Le pourtour de la glène est alors nettoyé afin d'assurer un bon contact osseux avec le support de glène.



FIGURE 3.7: Ecarteur de glène



FIGURE 3.8: Fantôme 1



FIGURE 3.9: Fraisage

On utilise ensuite un fantôme qui déterminera le centre de fonctionnement de la sphère glénoïdienne (figure 3.8). Le fantôme est positionné de manière à obtenir la meilleure couverture latérale et à avoir la patte supérieure en direction du pied de la coracoïde. Il matérialise ainsi le bord inférieur de la glène que l'on va découper. Le choix du centre de fonctionnement de la sphère glénoïdienne reste un paramètre essentiel pour optimiser la mobilité de l'épaule prothésée.

On perce, à l'aide d'un foret de diamètre 3.5mm, les deux trous à travers le fantôme et on place le *pin's* n°1 dans le trou inférieur avant d'enlever le fantôme. Le *pin's* n°1 servant de guidage, on peut alors surfacer la glène à l'aide d'une fraise puis retirer le *pin's* et mettre en place le *pin's* n°2 dans le deuxième trou percé avec le fantôme (figure 3.9).

On installe ensuite un deuxième fantôme sur ce *pin's* et on perce le trou diamètre 7.5mm du plot d'expansion ainsi qu'une série de trous diamètre 3.5mm sur la partie inférieure de la glène (figure 3.10). Le fantôme et le *pin's* sont alors retirés afin de terminer la découpe de la glène suivant le bord supérieur des trous diamètre 3.5mm.

Il ne reste plus qu'à mettre en place le support de glène sur le patient. Un perçage sera nécessaire dans le pied de la coracoïde. Enfin, la duoglène est mise en place sur le support de glène (figure 3.11).

Troisième étape : finalisation de l'humérus. Une tige d'essai est mise en place dans l'humérus. On teste alors des platines inserts d'essai de différentes tailles et suivant différents offsets (figure 3.12). Une fois la mobilité souhaitée obtenue, la platine définitive est mise en place sur la tige finale. L'insert est ajouté à la platine et l'ensemble est enfin impacté dans l'os. L'opération sera suivie d'une période de rééducation d'une durée de 3 à 6 mois.

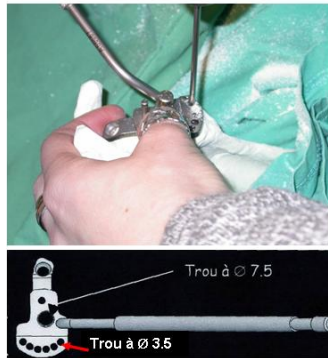


FIGURE 3.10: Fantôme 2

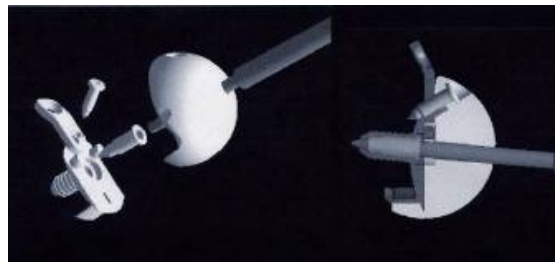


FIGURE 3.11: Montage de la duogène et du support de glène

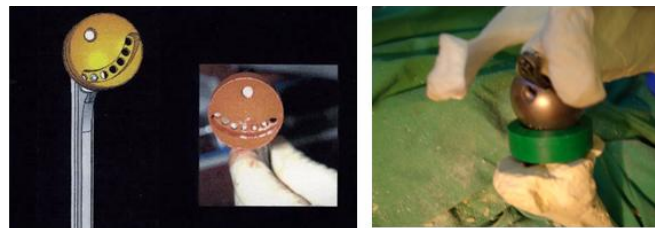


FIGURE 3.12: Platine-insert d'essai

La finalité de cette thèse est de réaliser un dispositif d'aide à la chirurgie de l'épaule. Pour cela, on doit modéliser la maquette numérique de l'épaule qui servira par la suite aux différentes simulations virtuelles de découpe et de perçage de l'épaule virtuelle, permettant d'envisager une assistance per-opératoire. Dans ce qui suit, nous allons voir les différentes familles de modélisation existantes ainsi que la méthode retenue pour la suite de nos travaux.

3.3 Quel Modèle ?

La reconstruction tri-dimensionnelle de structures anatomiques (os ou organes) à partir de coupes, ou tranches, scanner est une procédure bien maîtrisée aujourd'hui. Elle permet une reconstruction de la structure par des techniques de rendu surfacique (*surface rendering*) ou volumique (*volume rendering*). Des logiciels tels que *RapidForm*² proposent un module de reconstruction à partir d'une séquence d'images tomographiques. En général, les techniques de reconstruction tridimensionnelle par coupes séquentielles sont basées sur une segmentation automatique et/ou manuelle

2. INUS Technology

des structures à reconstruire dans les images 2D (il faut détourner les structures que l'on souhaite reconstruire). Cette segmentation est un traitement d'image permettant de rechercher les pixels de même intensité lumineuse et représentant le même objet pour les relier. Il suffit ensuite d'empiler les coupes ainsi obtenues, connaissant le pas de déplacement entre chaque coupe et le rapport de conversion pixel/mètre. C'est sans doute la technique qui permet la reconstruction la plus précise et la plus dense en imagerie médicale. En regard des contraintes que nous nous sommes imposées, le seul reproche qu'on pourrait lui faire est qu'elle nécessite justement une séquence complète de coupes scanners. Puisque nous voulons une solution générique, puisque nous posons comme principe que la méthode doit pouvoir être fonctionnelle à partir de seules images radiographiques, cette voie fut rapidement abandonnée.

Le principe de la reconstruction tridimensionnelle à partir de radiographies bi-planaires (ou multi-planaires) est basé sur l'acquisition de deux radiographies (ou plusieurs) capturées sous différents angles d'incidence [Le Bras 2004]. À partir des projections 2D obtenues, il est alors possible de définir la position tridimensionnelle de repères anatomiques dans l'environnement radiologique. Néanmoins, avant d'appliquer une méthode de reconstruction 3D, il est nécessaire de définir spatialement l'environnement radiologique, c'est-à-dire, le positionnement relatif des plans images et des sources de rayons X. Cette étape est appelée phase de *calibrage*.

Le principe du calibrage d'un appareil radiographique est d'estimer, au moyen d'une mire de calibrage, la transformation géométrique entre les coordonnées 3D d'un point de l'espace et ses coordonnées 2D projetées sur les différents plans radiologiques. Il existe de nombreux algorithmes permettant d'estimer cette transformation : l'algorithme DLT (*Direct Linear Transformation*), bien connu en vision par ordinateur ; les algorithmes de calibrage explicite, basés sur la connaissance de l'angle entre deux vues ; etc.

Le point crucial de la reconstruction multi-planaire est la mise en correspondance des points homologues, c'est-à-dire l'identification des points stéréo-correspondants [Selvik]. Pour simplifier cette opération, délicate sur des images radiographiques, l'implantation in vivo de marqueurs biocompatibles opaques est fréquente : elle n'est évidemment pas adaptée à notre problème. Une approche de type *Shape from Contour* est également envisageable [Laporte 2003] - *NSCC, Non-Stereo Corresponding Contour*. Cette méthode nécessite un objet générique représentant la structure à reconstruire.

Ces deux méthodes ne permettent que de reconstruire quelques points à la surface de l'os, elles souffrent en outre des problèmes d'occultation (fréquents sur une articulation, où les os sont emboîtés) et de segmentation des images et elles nécessitent le calibrage du système d'acquisition : ce qui est difficilement conciliable avec les critères de généricité et d'ergonomie demandés par les chirurgiens. D'autre part, les méthodes de modélisation (*marching cubes*, *Radial Basis Function*, *snakes*, etc.) fournissent une description tri-dimensionnelle précise, mais régie par des paramètres mathématiques dont le lien avec les paramètres physiques de l'épaule du patient n'est pas simple à estimer. Pour pallier ce premier problème, une technique de *morphing* peut être utilisée. Celle-ci requiert cependant suffisamment de points de contrôle pour converger. Comment mesurer ces points de contrôle – le palpation étant à exclure puisqu'il empêcherait d'obtenir le modèle adapté au patient avant l'opération et rendrait impossible l'outil de simulation pré-opératoire ?

Deux solutions s'offraient à nous :

1. Combiner l'approche reconstruction avec l'approche modélisation (modéliser une articulation générique et utiliser les points reconstruits comme points de contrôle pour le *morphing*). C'est une approche lourde et dont rien ne garantit qu'elle donne des résultats satisfaisants - ou même simplement des résultats : la distribution des points reconstruits peut ne pas permettre à l'algorithme de *morphing* de converger, par exemple.
2. Utiliser une méthode de reconstruction par contraintes, c'est-à-dire décrire la jointure scapulo-humérale par des formes géométriques simples connues *a priori* en adaptant les dimensions des primitives aux mesures morphométriques du patient.

Nous nous sommes orientés vers cette seconde solution. Nous détaillerons dans la section qui suit, la modélisation définie et la méthode d'extraction des paramètres morphologiques du patient à

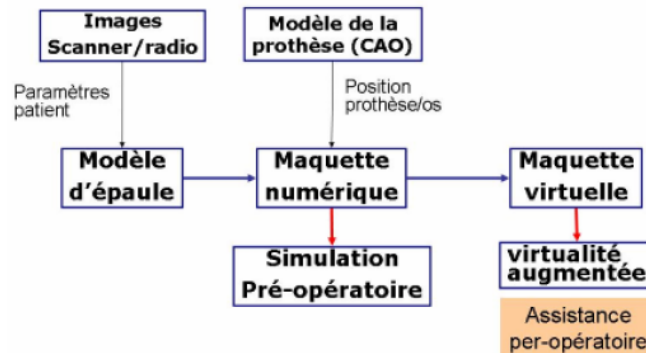


FIGURE 3.13: Démarche suivie pour la réalisation du dispositif d'aide à la chirurgie

partir d'images radiographiques et de coupes scanners.

3.4 Maquette Numérique et Chirurgie Virtuelle

Nous nous sommes donc orientés vers un modèle à base de formes simples liés entre elles par un arbre CSG (*Constructive Solid Geometry*). Ce modèle sera utilisé pour simuler la mécanique de l'épaule en préopératoire et per-opératoire en y introduisant le modèle CAO d'une prothèse. Il servira aussi pour simuler une opération de chirurgie virtuelle, mais également comme aide à l'opération dans un dispositif de virtualité augmentée. Pour atteindre ce but, nous avons identifié les étapes suivantes : modélisation de la partie osseuse de l'articulation à partir des données médicales disponibles du patient radios et/ou scanners ; simulation de l'opération à partir de ce modèle (chirurgie préopératoire virtuelle) ; conception d'un système de virtualité augmentée temps réel pour le travail peropératoire. La figure 3.13 décrit les étapes à suivre pour réaliser ce dispositif d'aide à la chirurgie de l'épaule.

3.4.1 Modélisation Osseuse

Par modèle à formes simples, nous entendons un modèle construit par un arbre CSG dont les primitives sont des quadriques, plans, etc. La modélisation obtenue au final est simplissime. Le rendu est peu réaliste, mais suffisant pour permettre au chirurgien d'opérer virtuellement. En outre, la précision du modèle permet un calcul des débitements correct, au regard des attendus des utilisateurs et des pratiques usuelles des chirurgiens.

Partie humérale. Le tableau 3.1 synthétise les principales formes géométriques utilisées dans notre application . Ces formes ont été choisies pour leur simplicité et parce que les principaux paramètres qui règlent leurs dimensions et, quelques fois, leur forme ont un sens *physique* direct : le diamètre de la sphère est directement donné par le diamètre mesuré de la tête humérale, par exemple. Deux primitives particulières ont été ajoutées pour effectuer la jointure des différentes parties du modèle : une cyclide de Dupin et une courbe de Bézier. Ces primitives s'ajustent automatiquement aux paramètres des primitives avec lesquelles elles sont liées.

Partie scapulaire. Nous illustrons par les figures qui suivent (3.15, 3.16 et 3.17) la modélisation de l'omoplate. Nous indiquons également quelques paramètres anatomiques utiles à prendre en compte lors de la modélisation.

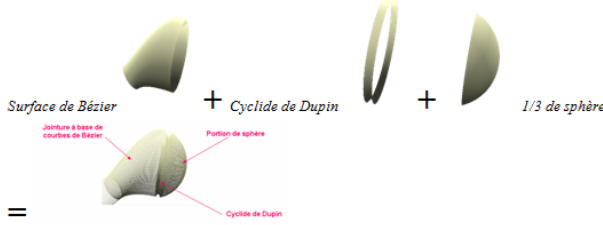



Parties représentées	Modèles utilisés
Tête humérale	 <p>Surface de Bézier + Cyclide de Dupin + 1/3 de sphère</p> <p>=</p> <p>Jointure à base de courbes de Bézier</p> <p>Parton de sphère</p> <p>Cyclide de Dupin</p>
Axe de l'épicondyle	 <p>Cylindre horizontal</p>
Gouttière bicipitale	 <p>Cylindre transparent</p>
Cylindre métaphysaire	 <p>de longueur 40mm et rayon 8mm (valeur par défaut).</p>

TABLE 3.1: Formes géométriques utilisées pour modéliser l'humérus

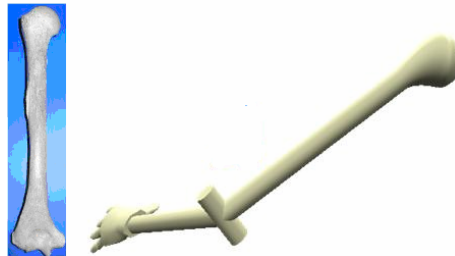


FIGURE 3.14: Humérus réel (à gauche) et humérus reconstruit avec la méthode des formes simples (à droite)

3.4.2 Extraction des Paramètres

Nous prendrons quelques exemples d'extraction des paramètres permettant d'ajuster le modèle à la morphologie du patient. Les paramètres peuvent être indifféremment extraits d'images scanners ou radiographiques; automatiquement pour certains et en mode supervisé pour d'autres. Nous illustrerons la méthodologie employée avec ces dernières. Tous ces paramètres ont été définis en accord avec les chirurgiens-anatomistes. Un logiciel spécialement développé à cet effet a été fourni aux chirurgiens pour validation.

La mesure du rayon de la base de la calotte humérale est illustrée en figure 3.18. Trois points sont

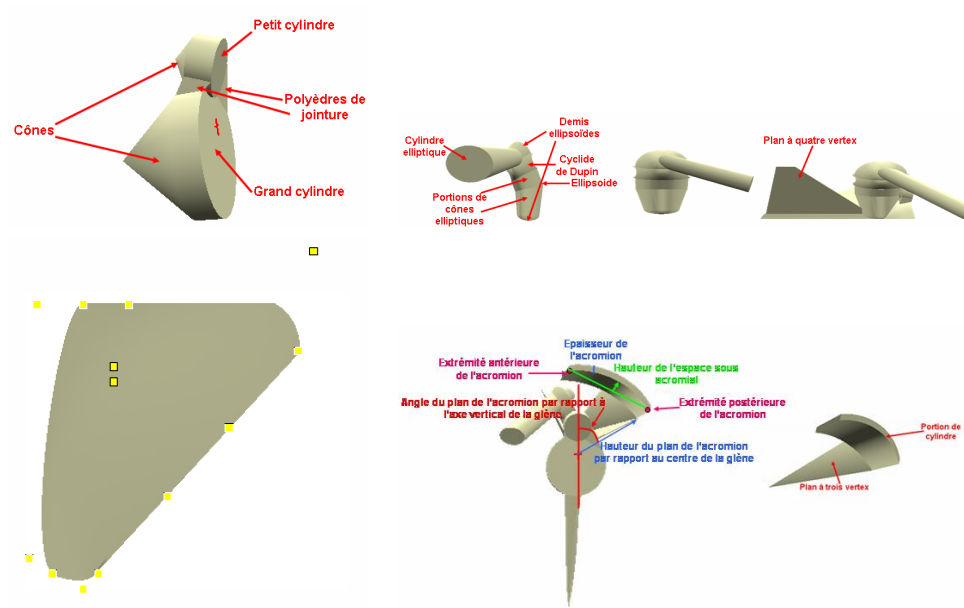


FIGURE 3.15: Principales formes utilisées pour modéliser l'omoplate et les paramètres correspondants

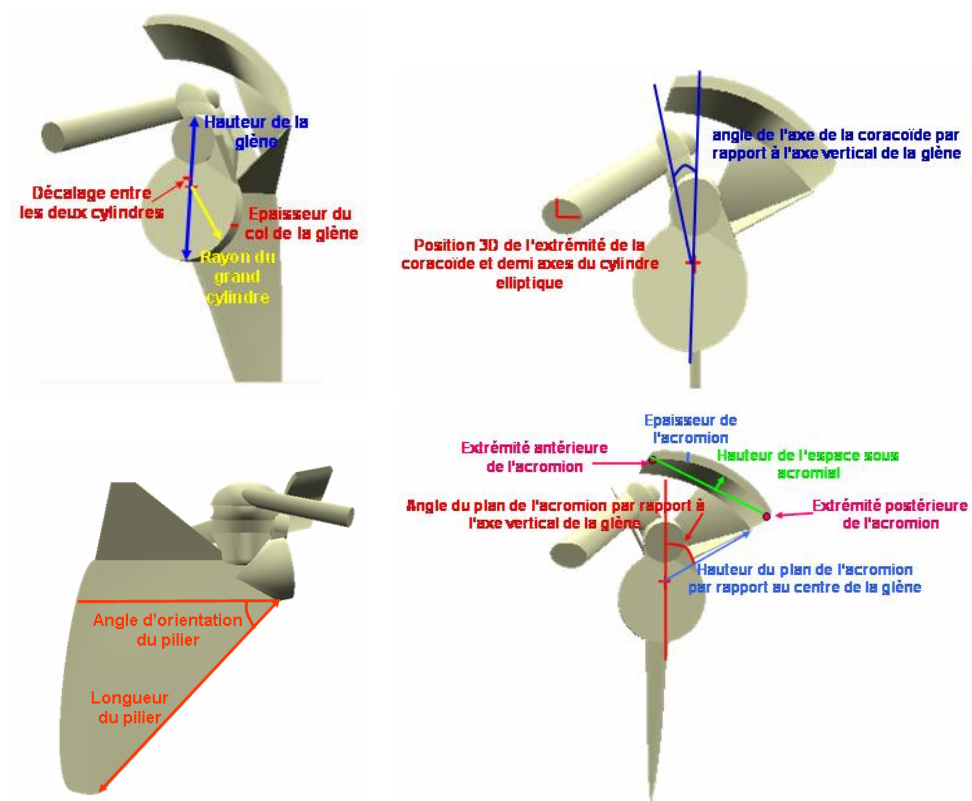


FIGURE 3.16: D'autres paramètres principaux et secondaires de l'omoplate

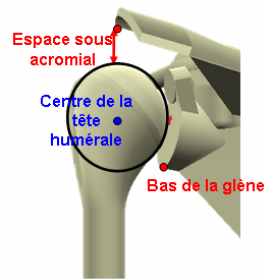


FIGURE 3.17: Espace sous acromial (flèche rouge)



FIGURE 3.18: Extraction du rayon de la tête humérale à partir de radios



FIGURE 3.19: Mesure du demi-rayon du cylindre métaphysaire

choisis par le chirurgien sur le détour de la calotte humérale. Un cercle est ajusté automatiquement à ces trois points et le rayon est extrait pour ajuster la dimension de la tête de l'humérus. La mesure du rayon du cylindre métaphysaire (figure 3.19) repose sur le même principe. La taille du segment (en rouge sur la figure) est plus précisément ajustée en détectant la différence de luminosité aux bords. Nous n'allons pas égrener les mesures de tous les paramètres. Certains, comme le décalage entre l'axe de la tête humérale et l'axe de la diaphyse humérale, nécessite une procédure un peu plus complexe (figure 3.20). Notons aussi que certains paramètres nécessitent des images radiographiques prises sous différents angles de vue (figure 3.21, 3.22). Au total, trois vues sont nécessaire : de face, de dessus et de côté.



FIGURE 3.20: Extraction du décalage entre l'axe de la tête humérale et l'axe de la diaphyse humérale

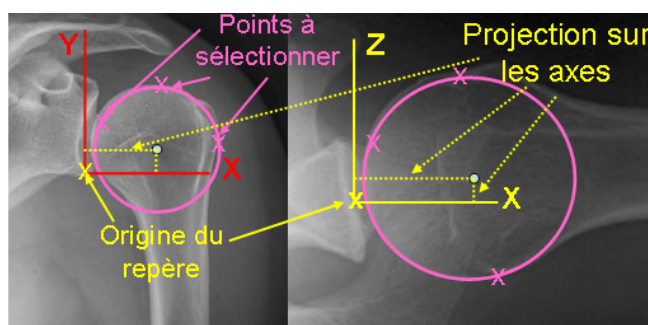


FIGURE 3.21: Mesure de l'espace sous acromial



FIGURE 3.22: Position 3D de l'extrémité de la coracoïde

3.4.3 Simulation Chirurgicale Pré-Opératoire

La maquette numérique de la prothèse (son modèle CAO, figure 3.23) et le modèle à formes simples de la jointure scapulo-humérale doivent être assemblés. De cet assemblage dépendra l'amplitude des débats possibles. La technique opératoire détaillée en 3.1.1.3 doit donc être reproduite virtuellement pour effectuer la simulation pré-opératoire qui permettra au chirurgien de tester et comparer différents scénarios et de choisir celui qui lui semblera le meilleur au regard des performances en termes de débats.

Un second logiciel (figure 3.24), permettant la visualisation du modèle et l'intégration de la prothèse et puisant ses paramètres dans la phase d'extraction, a été livré aux chirurgiens. Il permet :

- **La préparation de l'humérus** : réglage de l'axe de perçage et perçage (figure 3.25), réglage du plan de coupe de la calotte humérale (figure 3.26), choix du modèle de tige inversée de la prothèse.
- **Le travail de la glène** : placement des trous pour les pins (figure 3.27), positionnement et utilisation des fantômes (figure 3.28), travail du bas de la glène (figure 3.29) et réglage de



FIGURE 3.23: Modèle CAO assemblé de la prothèse

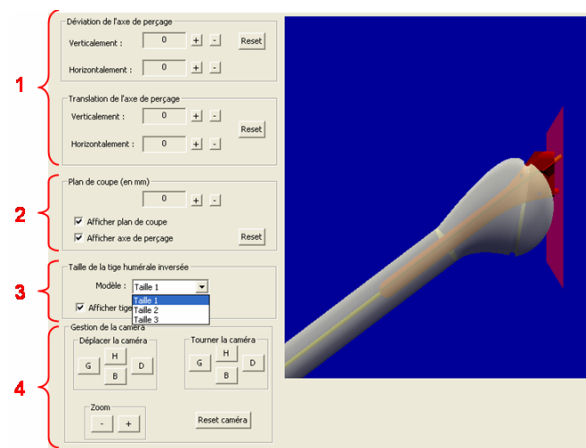


FIGURE 3.24: Interface utilisateur pour l'opération de découpe de la tête humérale

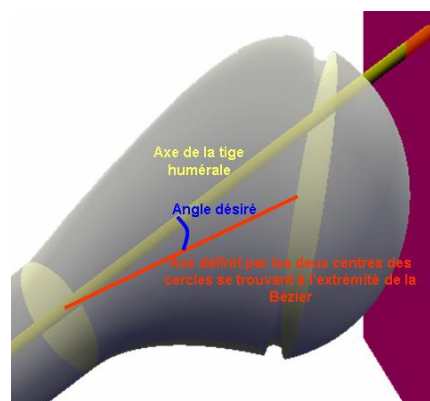


FIGURE 3.25: Angle entre l'axe de la tige humérale et l'axe du cône métaphysaire

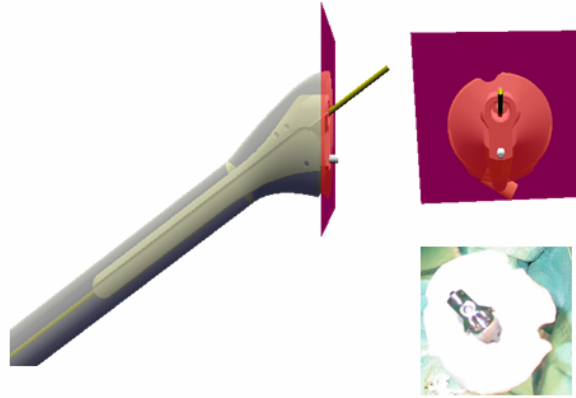


FIGURE 3.26: Résultat obtenu lors d'une découpe de la tête humérale suivant le plan en rouge et l'axe de perçage en jaune

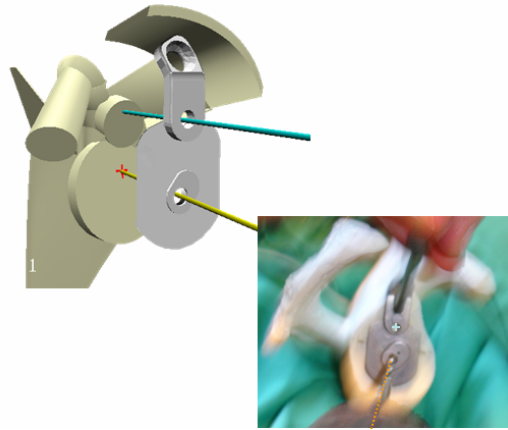


FIGURE 3.27: Placement des trous pour le pin's 1 et pin's 2 suivant le fantôme 1

l'offset de l'insert huméral (figure 3.30).

3.5 Simulation de Débattements

Il est maintenant possible, par un calcul de collisions, d'estimer les débattements pour chaque scénario mémorisé. On peut tout aussi bien estimer les débattements théoriques (pour un patient donné, de corpulence connue), estimer les débattements du modèle du patient (figure 3.31) et estimer les débattements du modèle prothésé du patient. La mesure des débattements est basée sur la détection des collisions entre les parties fixes liées à l'omoplate (glène, acromion, coracoïde, duogène et support glène) et les parties mobiles liées à l'humérus (humérus, insert huméral et platine humérale). Par exemple, l'abduction de l'épaule prothésée est stoppée par la collision entre l'insert huméral et le support de glène (figure 3.32) ; pour l'antépulsion, le mouvement de l'humérus est stoppé par une détection de collision entre l'humérus et la coracoïde (figure 3.33) ; pour le mouvement de rétropulsion, l'humérus est bloqué par l'acromion (figure 3.34) ; lors du mouvement de rotation interne, la rotation de l'humérus prothésée est stoppée par une collision entre l'omoplate et l'insert huméral (figure 3.35) ; enfin, lors du mouvement de rotation externe la détection entre l'humérus et l'omoplate intervient au niveau du pilier (figure 3.36).

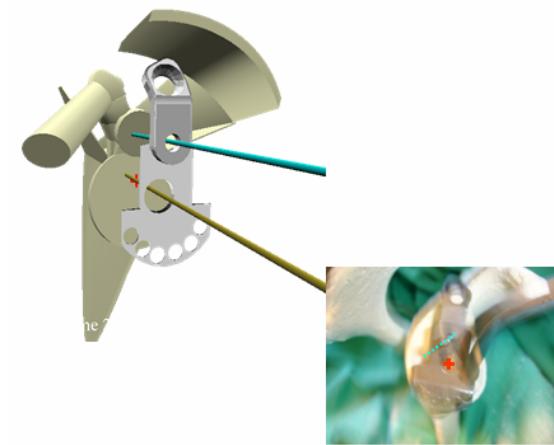


FIGURE 3.28: Utilisation du fantôme 2 pour le placement du support glène après avoir travaillé le bas de la glène



FIGURE 3.29: Travail du bas de glène et placement du support glène avec un choix entre les tailles S et M

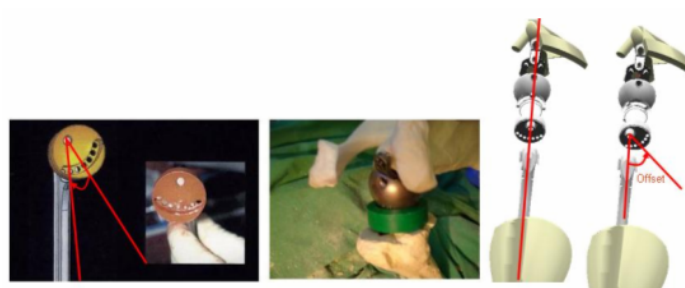


FIGURE 3.30: Simulation de l'offset de l'insert huméral

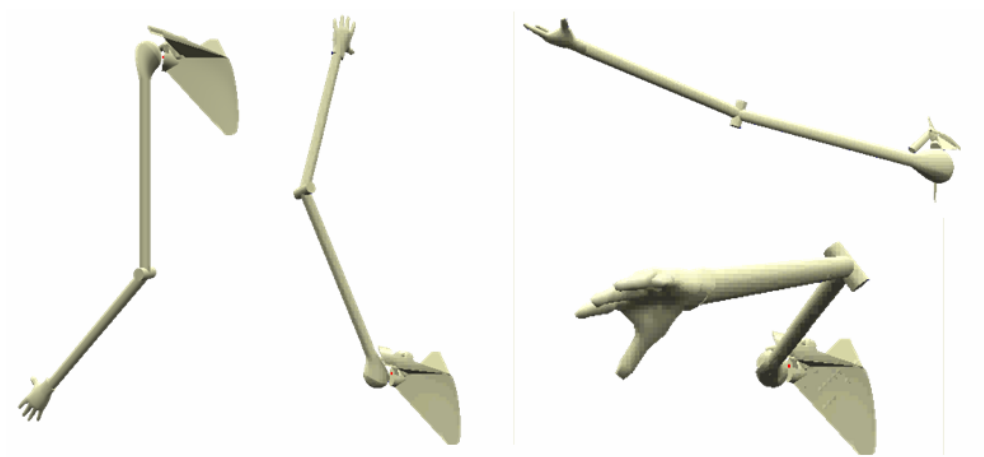


FIGURE 3.31: Différents débattements d'une épaule saine

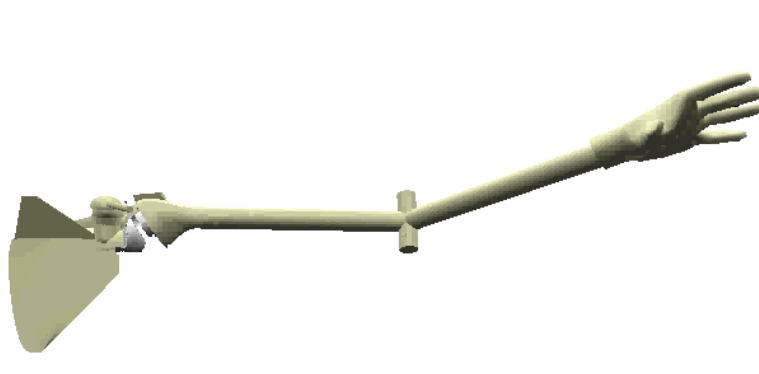


FIGURE 3.32: Mouvement d'abduction avec une épaule prothésée

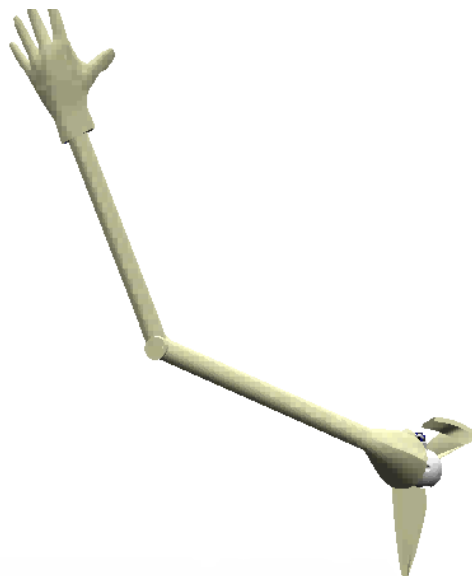


FIGURE 3.33: Mouvement d'ante pulsion d'une épaule prothésée

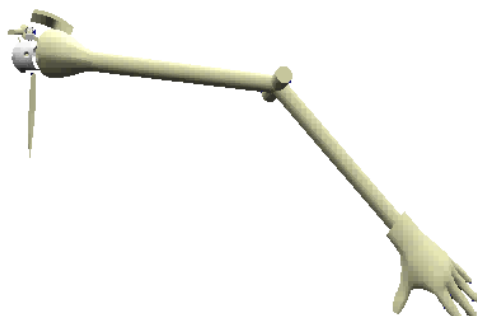


FIGURE 3.34: Mouvement de rétropulsion d'une épaule prothésée

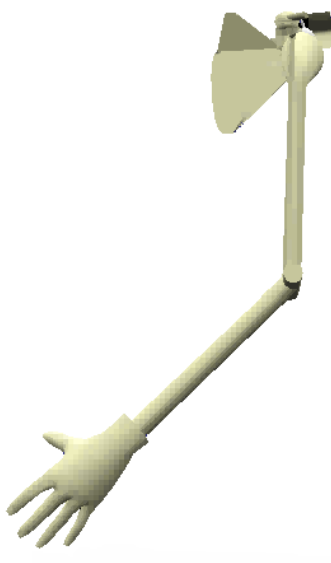


FIGURE 3.35: Mouvement de rotation interne d'une épaule prothésée

Une aide à la décision est proposée par comparaison des débattements (figure 3.37).

3.6 Conclusions

La partie extraction des paramètres du modèle peut donner lieu à des améliorations, comme une extraction complètement automatique à partir des radios ou des coupes scanners. Ces améliorations nécessiteraient un long travail de développement d'algorithmes de traitement adaptés à ces types d'images et à la forme des structures à segmenter. Il est également envisageable d'automatiser la modélisation de l'épaule en segmentant tri-dimensionnellement un modèle générique (os préalablement scanné, par exemple) et en ajustant des primitives aux différentes régions. L'utilisation de primitives plus complexes, telles que les *Supershapes* [Fougerolle 2006], et de transformations géométriques (torsion, cisaillement) applicables sur ces primitives, pourrait accroître la précision du modèle et le rendu graphique.

Nous avons expérimenté nos algorithmes, la précision du modèle et des paramètres extraits, mais uniquement de manière "technicienne", du point de vue du traiteur d'image. C'est évidemment très insuffisant pour un dispositif médical, censé être utilisé pour préparer une opération.



FIGURE 3.36: Mouvement de rotation externe d'une épaule prothésée

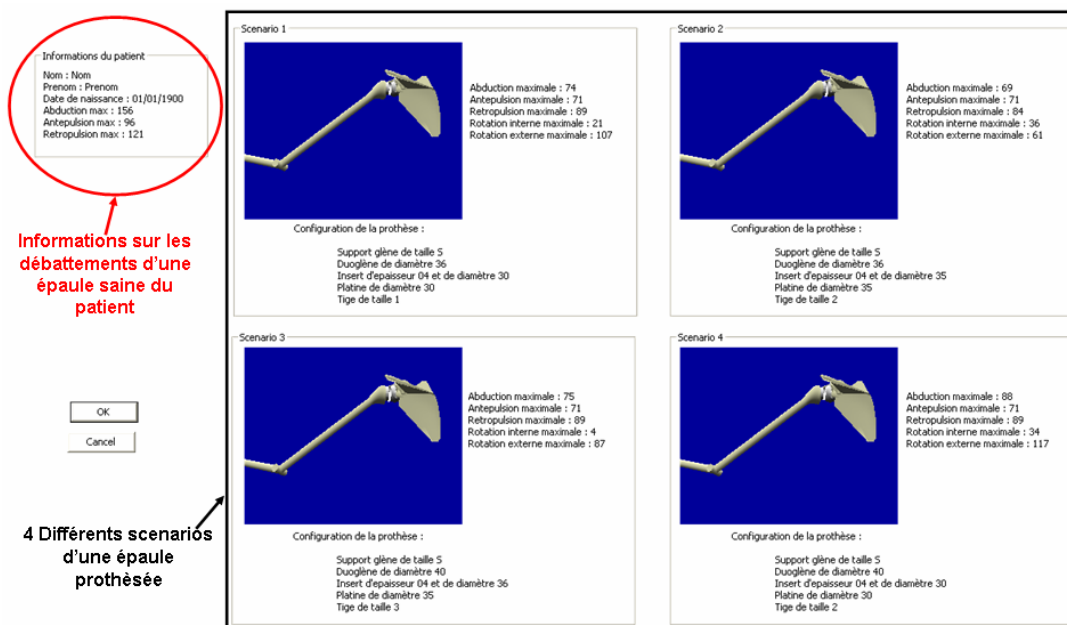


FIGURE 3.37: Comparaison des différents mouvements de l'épaule avec ou sans prothèse

Une validation médicale des algorithmes et méthodes proposés dans ce projet nous semble donc indispensable. Nous avons évoqué avec les chirurgiens le fait de pouvoir travailler sur cadavres (c'est-à-dire, radiographier un cadavre, estimer le modèle à formes simples et comparer avec son modèle morphométrique, et entreprendre un travail semblable pour la mesure des débattements) – sans en avoir eu la possibilité.

Chapitre 4

Conclusions et Perspectives

Nous présentons dans ce chapitre quelques pistes à suivre, à court, moyen ou long terme, liées à la vision en lumière structurée, à la vision catadioptrique ou, plus généralement, à la vision tridimensionnelle. Nous présenterons de manière synthétique nos principales contributions à la vision par ordinateur. Nous décrirons ensuite les développements thématiques sur lesquels nous souhaitons travailler. Nous préciserons finalement comment nous entendons structurer ces recherches au sein de la cellule, du laboratoire ou de collaborations nationales ou internationales.

4.1 Contributions

J’ai présenté dans ce mémoire l’ensemble de mes travaux de recherche depuis mon doctorat. Il est possible de classer les contributions auxquelles ils ont donné lieu en deux groupes. Le premier est constitué des contributions théoriques ou méthodologiques :

- une méthode d’auto-calibrage pour les systèmes projecteur-caméra basée sur la génération de contraintes géométriques directement issue des relations internes au motif structuré ;
- un état de l’art complet et comparé des méthodes de recalage 3D rigide ;
- une méthode de minimisation de l’erreur de recalage pour une séquence d’images ;
- une méthode de calibrage par imagerie polarimétrique dédiée aux capteurs catadioptriques ;
- une méthode de traitement adapté des images catadioptriques pouvant également profiter des paramètres de polarisation du miroir ;
- une méthodologie générale pour la modélisation de la jointure scapulo-humérale.

Le second groupe est constitué d’applications ou de contributions appliquées :

- le développement d’un prototype de scanner à main ;
- le développement d’une *toolbox* sur les méthodes de recalage ;
- l’intégration d’outils de vision pour la texturation virtuelle ;
- le développement de logiciels permettant d’extraire des paramètres morphologiques et biomécaniques de l’épaule dans le cadre d’une opération de placement d’une prothèse.

Dans la suite du chapitre, nous détaillerons les travaux et développements que nous entendons mener sur ces thématiques.

4.2 Développement des Thématiques

4.2.1 Vision en Lumière Structurée

En lumière structurée, le choix de la codification est fait en fonction de l’application ou des objets à numériser. On préférera ainsi projeter une lumière structurée binaire sur des objets co-

lorés ou texturés et une lumière structurée colorée sur des objets neutres colorimétriquement ; on préférera une codification directe à un multiplexage temporel quand la scène est dynamique et inversement si l'on souhaite privilégier la précision ou la résolution. L'idéal serait de pouvoir analyser la scène en amont et d'adapter la structure du motif aux couleurs et à la géométrie de celle-ci. Des solutions ont été publiées [Caspi 1998, Koninckx 2005], mais elles envisagent la scène comme un tout homogène. En d'autres termes, la structure du motif reste uniforme. Nous pensons qu'il est possible de généraliser encore ce processus. Imaginons qu'une image de la scène soit capturée au préalable (sous lumière structurée ou non) et qu'à partir de celles-ci, des informations géométriques (courbure) et colorimétriques peuvent en être tirées. On peut envisager d'adapter un sous-motif à chaque zone de l'image, adapté en structure et en couleur à la portion de scène sur laquelle il sera projeté. Le motif ainsi obtenu formera un *patchwork* de lumière structurée et codée. Ce *patchwork* pourrait en outre être adapté dynamiquement en fonction des changements dans la scène et intégré dans une stratégie de lumière structurée imperceptible [Cotting 2004]. Nous aurions un véritable codage local adaptatif et dynamique et pourrions mettre en oeuvre un véritable système générique en lumière structurée. C'est une voie que nous souhaitons vivement explorer à court ou moyen terme. De manière plus ciblée, nous envisageons aussi d'entreprendre une étude visant à combiner les travaux sur la codification de la lumière structurée et ceux sur les automates cellulaires [Wolfram 2002]. Peut-on définir des règles simples (de construction d'un automate cellulaire) pour concevoir un motif ou une classe de motifs donnés ? Une séquence de De Bruijn peut-elle être reproduite par un automate cellulaire ? etc.

Nous souhaitons également revenir sur le problème de l'autocalibrage d'un système projecteur-caméra, en commençant par l'adaptation directe de la méthode flexible de POLLEFEYS et en la complétant par les contraintes propres à la projection d'une lumière structurée [Pollefeys 1999] (connaissance a priori de l'image, organisation et structure du motif, etc.) Idéalement, cela consisterait à implémenter la méthode d'auto-calibrage pour deux vues et à se servir des contraintes liées à la projection pour lever les ambiguïtés et rendre la résolution plus robuste. En guise d'étape préliminaire, nous avons commencé à étudier et implémenter la méthode de calibrage proposée par RASKAR [Raskar 2001], qui utilise des outils géométriques et de résolution similaires. Pour automatiser le calibrage, nous cherchons à simuler le déplacement de la mire plane en transformant directement l'image-projecteur et en la projetant sur un plan d'orientation et de position invariantes.

4.2.2 Vision Catadioptrique

A long terme, notre objectif est de proposer une méthode générique de *structure-from-motion* allant du traitement des images au modèle 3D, adaptée à tout type d'anamorphose catadioptrique (complètement indépendant de la forme et de l'alignement du miroir) et donc, en configuration limite, adaptée également à la vision perspective. La première étape consistera à imaginer une méthode d'auto-calibrage d'un capteur catadioptrique quelconque (central ou non-central). La mise en correspondance de lignes catadioptriques peut être vue comme une toute première proposition allant dans ce sens (elle permet d'apparier des primitives et donc les paramètres intrinsèques du capteur [Barreto 2005], et d'estimer le déplacement d'une prise de vue à l'autre, et donc les paramètres extrinsèques). Il faut bien sûr la généraliser au cas de la *wide baseline* et la rendre indépendante du point de vue unique.

La seconde étape consistera à proposer une méthode unifiée de traitement des images (anamorphosées ou non) permettant d'adapter l'opérateur à la morphologie de celles-ci. L'idée est d'obtenir des opérateurs (auto-)adaptatifs utilisables indépendamment sur des images perspectives, *fish-eye*, catadioptriques, etc. Nous pourrions ainsi combiner modèle géométrique générique et méthode unifiée de traitements pour offrir une boîte à outils complète comprenant la détection de points d'intérêt, la mise en correspondance, l'auto-calibrage, la triangulation, etc. et intégrer le tout dans une méthode générique de *structure-from-motion*, adaptable au capteur ou aux éléments du capteur (capteur mixte composé d'une caméra perspective et catadioptrique, par exemple). Nous nous inspirerons évidemment des travaux sur le modèle générique de caméra [Sturm 2004, Sturm 2005, Ramalingam 2006] en essayant, soit de proposer une méthode d'ajus-

tement de faisceaux adaptée, soit en proposant une voie intermédiaire permettant d’avoir une paramétrisation du modèle (et appliquer l’ajustement de faisceau classique). *In fine*, nous souhaitons pouvoir donner des images quelconques en entrée de l’algorithme et en obtenir un nuage de points 3D en sortie, quel que soit le type et le nombre d’images ; le tout, en y appliquant le même modèle et les mêmes équations pour chacune.

En parallèle, nous souhaitons poursuivre les travaux menés autour de l’imagerie polarimétrique (pour la vision catadioptrique). L’extension de la méthode aux caméras perspectives est en cours d’étude. Sa généralisation à des miroirs plans par exemple, ou même à des systèmes dioptriques pourraient également parachever ces travaux sur le calibrage.

4.2.3 Reconstruction d’Objets Transparents

Ces travaux sont tout récents et la problématique abordée ne sera évoquée ici qu’en termes généraux ; le but étant d’ouvrir sur une application nouvelle, comprenant l’intégration d’outils et de méthodes divers (optiques, géométriques, etc.). Depuis quelques années déjà, le Le2i travaille sur la reconstruction 3D de surfaces spéculaires par imagerie polarimétrique ; des solutions concrètes ont été proposées [Morel 2006a, Morel 2006b]. L’extension de ces méthodes aux surfaces transparentes vient logiquement. Des travaux sont menés au sein de l’équipe sur la polarisation ou la triangulation pour apporter une solution à ce problème. Notre démarche vise quant à elle à intégrer ses différents outils, venant de différentes cultures scientifiques, pour parvenir à un prototype de reconstruction de surfaces transparentes suffisamment précis et modérément contraint. Quelques pistes peuvent être données :

- utilisation de la vision infrarouge : en effet, un objet n’est pas transparent dans toutes les longueurs d’onde ;
- utilisation de l’imagerie polarimétrique, qui a fait ses preuves pour la reconstruction de surfaces spéculaires et qui est déjà étudié au laboratoire pour la reconstruction de surfaces transparentes ;
- utilisation de la stéréovision ou de la triangulation : l’ajout de différents points de vue peut permettre de lever certaines ambiguïtés inhérentes à l’observation de surfaces transparentes.

On a vu plus haut que le *shape from polarization* prenait souvent l’hypothèse d’une caméra orthographique. Nous souhaitons dans un premier temps relâcher cette contrainte (un peu comme l’ont fait PRADOS et FAUGERAS pour le *Shape from Shading* [Prados 2003]). Ceci nous amène à étudier le comportement optique et géométrique du couple polariseur-caméra pour établir une correspondance univoque entre les pixels et les paramètres de polarisation. Les surfaces transparentes ont la propriété de transmettre et réfléchir la lumière. Les informations qui arrivent à la caméra sont donc une combinaison de réflexions multiples aux interfaces et aux surfaces. Nous envisageons d’étudier l’apport de la vision binoculaire (et de ses outils géométriques, contrainte épipolaire notamment) à l’imagerie polarimétrique. Une fois ces aspects optiques et géométriques étudiés, nous pourrions évaluer l’apport de la vision infrarouge et proposer un prototype pour la reconstruction de surfaces transparentes.

4.3 Collaborations Présentes et Futures

La recherche menée au laboratoire autour de la vision catadioptrique s’est structurée autour du projet CaViAR. Nous envisageons de poursuivre ces travaux dans un projet d’envergure européenne, sur des thématiques proches de celles que nous avons évoquées plus haut (et regroupant une partie des partenaires du projet CaViAR). Par ailleurs, la collaboration avec le MIS de l’Université de Picardie Jules Verne sur la vision catadioptrique perdure via le co-encadrement d’une thèse de doctorat et l’implication commune dans le projet STAR Hubert Curien avec la Corée du Sud. Enfin, un premier contact a été pris avec le laboratoire VIPS (*Vision, Image Processing and Sound*) de l’Université de Vérone avec lequel nous envisageons de travailler sur les méthodes de *structure-*

from-motion. En interne, cette thématique est complétée par l’axe “imagerie non-conventionnelle” et notamment les travaux sur le *shape from polarization*.

Au sein de la cellule Vision 3D, les travaux sur la lumière structurée peuvent trouver un appui et un complément sur les travaux menés au laboratoire sur l’optimisation du *scanning* d’objet (stratégie de *best-next-view*). On peut très aisément imaginer de combiner l’auto-calibrage et le choix d’une position nouvelle et optimale du capteur afin de numériser au mieux un objet donné. Un bras robotisé a été commandé par la Plateform3D ; il sera mis à la disposition des chercheurs du laboratoire pour servir de plateforme expérimentale aux travaux sur la numérisation et la reconstruction tridimensionnelle. En termes d’échanges et de collaborations extérieures, les quelques pistes évoquées ici doivent encore mûrir et se concrétiser par des résultats locaux avant de solliciter des partenaires.

Plus généralement, et en sortant un peu des activités de recherche que je mène ou dirige, une réflexion est conduite au laboratoire pour définir des axes de recherche principaux, au recouvrement suffisamment important pour favoriser l’échange et la dynamique de groupe. La vision catadioptrique, par les projets et les collaborations qu’elle génère, mais aussi par son positionnement thématique, qui la place en interlocuteur privilégié de l’axe “imagerie non-conventionnelle” et des aspects plus fondamentaux de la géométrie des vues multiples, réunit ces critères. Nos travaux sur la vision en lumière structurée peuvent proposer des outils et méthodes à un axe “numérisation 3D” visant à fournir des modèles denses et précis, et agglomérant les recherches sur le *best-next-view*.

Les outils théoriques que partagent vision en lumière structurée et vision catadioptrique et les méthodes que nous souhaitons mettre en oeuvre et développer sont basiquement les outils de la stéréovision et des vues multiples. En ce sens, ils sont d’une portée très transversale et ajustable à un nombre conséquent d’applications de robotique et de numérisation.

4.4 De la Recherche à l’Enseignement

Le métier d’enseignant-chercheur est à la croisée de deux activités complémentaires. Il ne s’agit pas de juxtaposer recherche et enseignement, de manière décorrelée et tout à fait hermétique, mais bien de redistribuer les acquis de la recherche dans l’enseignement. Faire bénéficier aux étudiants d’un enseignement de pointe, sur des sujets et des problématiques récents. La recherche garantit la fraîcheur, l’actualité, la vivacité de l’enseignement. A l’inverse, l’enseignement parfois, ou les activités pédagogiques plus généralement, peuvent aider, soutenir, alimenter nos travaux de recherche. L’expérience que nous avons avec nos étudiants du master VIBOT est à ce titre édifiante. Nous les formons à la vision par ordinateur, nous leur proposons des projets à mi-chemin entre nos activités de recherche et d’enseignement et nous bénéficions des expérimentations, résultats, idées, propositions qu’ils peuvent nous faire en retour. Nous pouvons ainsi, et grâce à eux, planifier et tester des thématiques de recherche sur quelques années, explorer des idées nouvelles ou finaliser des projets en cours ; faire de la recherche au-delà du laboratoire.

Chaque année, je propose des sujets de projets et des sujets de stage en lien avec mes activités de recherche. Certains donnent lieu à une suite, sous forme de thèse de doctorat par exemple, d’autres sont abandonnés. A titre d’exemple, j’encadre cette année un projet sur le *structure-from-motion*, un autre sur le calibrage d’un système projecteur-caméra et un dernier sur la vision catadioptrique. J’ai également proposé un stage sur la mise en correspondance dense d’images catadioptriques. Notre objectif est d’ajuster l’offre de projets et de sujets de stage à la politique scientifique et aux thématiques défendues par le laboratoire.

Deuxième partie

Curriculum Vitae

Chapitre 5

Encadrement de Thèses et de Stage

Globalement, j'ai pu, depuis ma prise de fonction en septembre 2002, encadrer 2 séjours post-doctoraux, 5 thèses de doctorat (2 soutenues, 1 dont la soutenance est prévue fin 2008 et 2 qui ont débuté cette année), 6 stages de master et 1 stage d'ingénieur dans le cadre d'un échange Erasmus avec l'Université de Gérone (Espagne).

5.1 Post-doctorat

J'ai encadré deux séjours post-doctoraux de 12 mois dans le cadre du projet CaViAR, financé par l'ANR. Olivier MOREL a travaillé sur le calibrage d'un capteur catadioptrique par imagerie polarimétrique et Samia AINOUIZ, sur le traitement adapté des images catadioptriques.

Nom, Prénom	Période	Titre	Co-encadrant	Financement
Ainouz, Samia	2007/2008 (12 mois)	Reconstruction 3D Dense par Vision Catadioptrique	Morel, Olivier (50%)	ANR
Morel, Olivier	2006/2007 (12 mois)	Calibrage d'un Capteur Catadioptrique par Imagerie Polarimétrique	-	ANR

TABLE 5.1: Encadrements post-doctoraux

5.2 Doctorat

Le tableau 5.2 résume les co-encadrements doctoraux que j'ai ou ai eu depuis 2004. Il faut noter que Carles MATABOSCH était inscrit à l'Université de Gérone en Espagne et que le co-encadrement s'est matérialisé sous la forme de deux stages de recherche d'une durée de 5 mois chacun, effectué par le doctorant dans l'équipe Vision 3D du Le2i en 2004 et 2005. Une convention de collaboration a été signée par les deux universités à cet effet.

5.3 Master

Les étudiants en stage de master (ou DEA) que j'ai encadré proviennent essentiellement du master VIBOT que nous coordonnons et du master 3i (Image, Informatique, Ingénierie) porté par nos collègues dijonnais. J'ai également encadré le stage de deux étudiants de l'Université de Gérone : le premier dans le cadre du programme doctoral de l'Université (équivalent au DEA) et la seconde dans le cadre de son projet de fin d'études d'ingénieur.

Nom, Prénom	Période	Titre	Directeur de thèse	Financement
Mosaddegh, Saleh	2008/...	Vision Omnidirectionnelle appliquée à la Robotique Aérienne	Gorria, Patrick, Le2i (30%)	Région Bourgogne
Rantson, Rindra	2008/...	Numérisation Tridimensionnelle d'Objets Transparents basée sur la Stéréoscopie dans le Domaine de l'Infrarouge Thermique	Mériaudeau, Fabrice, Le2i (33%); Stolz, Christophe, Le2i (33%)	CIFRE
Molinier, Thierry	2004/...	Approche Coopérative pour l'Acquisition et l'Observation de Formes Tridimensionnelles	Gorria, Patrick, Le2i (50%)	MESR
Atmani, Hakim	2004/2008 (39 mois)	Dispositif Virtuel d'Aide à la Chirurgie de l'Epaule	Mérienne, Frédéric, Le2i / ENSAM (30%)	Région / Industrie
Matabosch, Carles	2004/2007 (33 mois)	Hand-Held 3D Scanner for Large Surface Registration	Salvi, Joaquim, VICOROB / UdG (60%)	Espagne

TABLE 5.2: Encadrements doctoraux

Nom, Prénom	Titre	Mention	Jury
Atmani, Hakim	Dispositif Virtuel d'Aide à la Chirurgie de l'Epaule	Très honorable	Fabrice Mériaudeau (uB, Président) Edwige Pissaloux (Univ. Paris 6, Rapporteur) Hervé Tanneguy Redarce (INSA Lyon, Rapporteur) Pierre Dubois (CHU de Lille, Examineur) Pierre Trouilloud (CHU de Dijon, Examineur) Jean-Luc Martinez (ENSAM, Invité) David Fofi (uB, Co-encadrant)
Matabosch, Carles	Hand-Held 3D Scanner for Large Surface Registration	Félicitations du jury	Joan Batlle (UdG, Président) Peter Sturm (INRIA, Rapporteur) Yvan Pétillot (HWU, Rapporteur) José Maria Martinez Montiel (Univ. de Saragosse, Examineur) El Mustapha Mouaddib (UPJV, Examineur)

TABLE 5.3: Jury et mention des thèses soutenues

Nom, Prénom	Année	Titre	Co-encadrant	Diplôme
Mosaddegh, Saleh	2008	Line Matching in Catadioptric Images	Vasseur, Pascal, MIS, UPJV (50%)	Erasmus Mundus VIBOT Master
Muhammad, Naveed	2008	Visual SLAM for a 6 dof Robot	-	Erasmus Mundus VIBOT Master
Lertrudachakul, Intuon	2008	Photorealistic Reconstruction by Voxel Coloring	-	Erasmus Mundus VIBOT Master
Prat Altimira, Gemma	2005	Reconstruction 3D de Scènes Urbaines à partir d'une Séquence d'Images Photographiques	-	Ingénieur, Universitat de Girona, Espagne
Molinier, Thierry	2004	Auto-Calibrage d'un Capteur de Vision en Lumière Structurée Imperceptible	-	Master 3i, Université de Bourgogne
Atmani, Hakim	2004	Conception d'une Maquette Numérique pour l'Aide à la Chirurgie de l'Epaule	Mérienne Frédéric, Le2i, ENSAM (70%)	Master 3i, Université de Bourgogne
Matabosch, Carles	2004	A State-of-Art of Range Image Registration	Salvi, Joaquim, VICOROB, UdG (50%)	Universitat de Girona, Espagne

TABLE 5.4: Encadrements de masters

Chapitre 6

Responsabilités, Enseignement, Animation et Valorisation

6.1 Responsabilités Scientifiques

Responsable de la cellule Vision 3D du Le2i (depuis novembre 2007). La cellule compte 17 permanents (6 Professeurs des Universités, 9 Maître de Conférences, 1 Ingénieur de Recherche CNRS, 1 Secrétaire-Gestionnaire CNRS), 1 Ingénieur d'Etudes et une quinzaine de doctorants. La responsabilité consiste essentiellement en l'animation scientifique de l'équipe et en sa gestion financière et administrative. La cellule Vision 3D travaille en partenariat avec la Plateform3D (labellisée Plateforme Technologique par le Ministère de l'Enseignement Supérieur et de la Recherche) pour les prestations industrielles et le transfert de technologie. Elle propose des formations en vision industrielle via l'INSACAST et CNRS Formation Entreprises. Les principales collaborations académiques de la cellule sont montées avec les laboratoires et universités suivantes : Oak Ridge National Laboratory (USA), IRIS (USA), le VICOROB (Espagne), l'Université de Heriot-Watt (Ecosse), l'Université de Sao Paulo (Brésil), le MIS (Amiens).

	Revues	Brevets	Editions et livres	Conférences internationales	Conférences nationales
Depuis 2001	124	10	21	300	82
Depuis 2006	45	3	12	91	14

TABLE 6.1: Bilan des publications de la cellule

Coordinateur local du projet ANR CaViAR (décembre 2005 - décembre 2008). Le projet CaViaR (Catadioptric Vision for Aerial Robots) a été sélectionné par l'ANR pour la période 2005-2008 dans le cadre de ses programmes blancs. Il compte 5 partenaires institutionnels : le MIS de l'Université de Picardie Jules Verne (coordinateur), le LAAS CNRS (Toulouse), AROBAS INRIA Sophia-Antipolis, PERCEPTION INRIA Rhône-Alpes et le Le2i. La dotation globale est de 450 000€ pour les 3 ans, soit 90 000€ environ par équipe.

Partenaire du projet NOVA (2008 et 2009) Le projet NOVA a été sélectionné dans le cadre des projets STAR de l'Egide (Programme Hubert Curien). Il regroupe, autour de la vision omnidirectionnelle, le MIS (Amiens, Université de Picardie Jules Verne), le KAIST (Séoul, Corée du Sud) et le Le2i - Vision 3D. 4 mois de mobilité par an sont financés (3 dans le sens France-Corée, 1 dans le sens Corée-France) pour les années 2008 et 2009.

Responsable scientifique de la collaboration entre le VICOROB et le Le2i (depuis 2002). Une convention liant le VICOROB (Université de Gérone) et le Le2i a été signée quelques mois après mon arrivée en Bourgogne. Elle prévoit l'échange de chercheurs (doctorants et permanents) et de ressources. Elle a débouché sur le co-encadrement de Carles MATABOSCH, le séjour de 4 mois d'Alexandra LATHUILIÈRE (doctorante au Le2i) et les deux séjours (4 mois chacun) de Thierry MOLINIER à Gérone.

Contributeur pour la SPIE Newsroom (depuis 2008). La *SPIE Newsroom* est un bulletin en ligne¹ publiant de courts articles de science, à mi-chemin entre l'article de vulgarisation et l'article d'actualité. Il a pour but de faire connaître rapidement des avancées technologiques ou scientifiques, des produits innovants ou des brevets récents, des laboratoires travaillant sur une thématique particulière et originale, etc. J'en ai été désigné contributeur par SPIE en 2008 pour la rubrique *Electronic Imaging & Signal Processing*. Le tâche consiste à proposer des idées d'article et à solliciter des chercheurs en traitement du signal et des images ou en vision par ordinateur pour publication.

6.2 Autres Responsabilités

Directeur des Etudes du Master Erasmus Mundus VIBOT (depuis 2006). Le Master VIBOT (*European Masters in Vision and roBOTics*) a été labellisé par la commission Européenne en 2006, dans le cadre du programme d'excellence Erasmus Mundus, dont l'ambition est d'attirer en Europe les meilleurs étudiants du monde entier. Je suis, avec MM. Fabrice MÉRIAUDEAU et Bernard LAMALLE, à l'origine de la création du master et du montage de dossier pour la labellisation par la Commission Européenne. Le consortium regroupe l'Université d'Heriot-Watt à Edimbourg (où les étudiants passent le premier semestre), l'Université de Gérone en Espagne (où les étudiants passent le deuxième semestre) et le Centre Universitaire Condorcet au Creusot (où les étudiants passent le troisième semestre). Nous percevons, chaque année, 20 bourses d'étude (42000€/ 2 ans) pour une vingtaine d'étudiants extra-européens versées par l'Europe et une petite dizaine de bourses ou d'aides pour les étudiants européens versées par des partenaires privés ou institutionnels. Nous avons droit également à 9 mois de professeurs invités, financés également par l'Europe. Les trois premières promotions du VIBOT comptent respectivement 26, 35 et 27 étudiants pour environ 300 candidatures chaque année. Je suis plus particulièrement en charge des emplois du temps, de l'organisation de l'attribution des stages de master, des projets de recherche et de l'encadrement pédagogique général. Trois réunions sont prévues chaque année avec nos partenaires espagnols et écossais : une en septembre pour la rentrée des étudiants au Creusot, une en janvier/février pour la sélection des nouveaux étudiants, une en juin pour les soutenances. Elles sont toutes trois complétées par une réunion pédagogique, un bilan sur les enquêtes qualité et la préparation des *progress reports* à soumettre à la Commission Européenne.

Responsable des Relations Internationales de l'IUT du Creusot (depuis 2005). L'IUT du Creusot envoie chaque année une soixante d'étudiants en stage à l'étranger et une vingtaine d'étudiants en DUETI (Diplôme Universitaire d'Etudes Technologiques Internationales). Il accueille en outre une trentaine d'étudiants chinois dans ses différents DU (Diplôme Universitaire). Enfin, l'IUT du Creusot est centre agréé de préparation et de passage du TOEIC. La responsabilité du service Relations Internationales de l'IUT consiste à superviser et articuler l'ensemble des programmes liés à l'international et à être le garant de la politique menée par l'IUT dans ce domaine. Le service coordonne le travail d'une quinzaine d'enseignants et enseignants-chercheurs impliqués.

Membre du Conseil Scientifique de l'Université de Bourgogne (depuis 2005). Je suis membre élu du Conseil Scientifique de l'Université de Bourgogne depuis décembre 2005 et membre de son bureau jusqu'au printemps 2008.

1. <http://www.spie.org>

Partenaire du projet Erasmus Mundus EACOVIREO (2008-2011). Le projet EACOVIREO - *Enhance the Attractiveness of Computer Vision and Robotics in Europe* est porté par l'antenne creusotaine de l'Université de Bourgogne et a pour objectif de renforcer l'attractivité de la vision et de la robotique en Europe. Il rassemble 18 partenaires institutionnels européens et extra-européens. Je suis personnellement responsable d'un des deux *Work Packages* du projet consistant à identifier et recenser les formations européennes dans le domaine de la vision par ordinateur et la robotique et de consigner l'information dans un portail web.

Responsabilités passées.

- Responsable pédagogique de la Licence Professionnelle C.I.V.A. (Contrôle Industriel par Vision Artificielle) en 2006/2007.
- Directeur des études de la Licence Professionnelle C.I.V.A. en 2004/2005 et 2005/2006.
- Responsable du recrutement et Directeur des Etudes du Diplôme Universitaire Sciences pour l'Ingénieur réservé aux étudiants chinois en 2004/2005 et 2005/2006.
- Membre de la Commission de Spécialistes 61/63 de l'Université de Bourgogne (2003-2008)
- Membre extérieur de la Commission de Spécialistes 61/63 de l'Université Picardie Jules Verne (2004-2008).

6.3 Relectures et Evaluations

Relecteur pour des revues scientifiques. J'ai été sollicité pour la relecture d'une vingtaine d'articles soumis à des revues dans le domaine de la vision par ordinateur et du traitement des images depuis 5 ou 6 ans :

- *International Journal of Computer Vision*, Springer
- *Image and Vision Computing*, Elsevier
- *Journal of Electronic Imaging*, IS&T / SPIE
- *Machine Vision and Applications*, Springer
- *Optical Engineering*, SPIE
- *Pattern Recognition*, Elsevier
- *International Journal of Image and Graphics*, World Scientific
- *Traitement du Signal*, GRETSI

Jury de thèse. J'ai été récemment sollicité pour rapporter sur la thèse de Mlle Elisabet BATLLE de l'Université de Gérone (il n'y a pas d'équivalent à l'HDR en Espagne et ce diplôme n'est pas requis pour être rapporteur de thèse). La soutenance est prévue dans les toutes prochaines semaines. La soutenance de thèse de M. Vincent PAQUIT est prévue le 12 décembre 2008.

- Elisabet Batlle, "Large-Scale Surface Registration", Université de Gérone, Espagne, 2008. **Rapporteur.**

La thèse traite du recalage rigide de vues 3D sur une grande échelle. La contribution majeure réside dans le développement d'une stratégie de minimisation de l'erreur basée sur la détection de cycles et la théorie des graphes. Un état de l'art complet, regroupant méthodes d'estimation de la pose relative, de recalage, d'auto-calibrage et de SLAM est proposé.

- Vincent Paquit, "Analyse multispectrale des tissus biologiques – Application à la localisation des structures subcutanées", Université de Bourgogne, France, 2008. **Examineur.**

La thèse décrit le développement d'un capteur permettant la détection subcutanée du système vasculaire par imagerie infrarouge et sa localisation dans le volume du bras par vision en lumière structurée. L'application visée est l'insertion automatique de cathéter.

Responsabilités dans des conférences. Le tableau 6.2 synthétise les diverses responsabilités que j'ai tenues dans l'organisation ou l'animation de conférences. Il faut souligner ceci : je suis, pour la deuxième année consécutive (2008 et 2009), *General Co-Chair* de la conférence *Image Processing : Machine Vision Applications* organisée dans le cadre du symposium *IS&T/SPIE Electronic Imaging* à San José (USA). Cette responsabilité comprend la promotion de la conférence, la composition du comité de programme et l'attribution des articles pour relecture, la décision, la constitution du programme et l'édition des *proceedings*. 34 articles ont été acceptés et présentés en 2008. 41 articles sont acceptés pour l'édition 2009.

En 2007, j'ai pris la présidence du comité d'organisation du QCAV (conférence sur les applications de la vision au contrôle qualité), qui a eu lieu au Creusot et a regroupé environ 90 participants. 48 articles ont été sélectionnés (pour un peu moins de 80 soumissions). La conférence a été parrainée par *IEEE* et *SPIE Europe*. Elle a donné lieu à l'édition de *proceedings*, d'un numéro spécial de *Electronic Letters on Computer Vision and Image Analysis* (CVC Press) et d'une section spécial du *Journal of Electronic Imaging* (IS&T / SPIE) parue en juillet 2008. Le comité scientifique du QCAV est présidé par Philippe BOLON (LISTIC, Université de Savoie) et Kenneth TOBIN (Oak Ridge National Laboratory).

Sigle	Année(s)	Nom de la conférence	Rôle
QCAV	2005, 2007	Quality Control on Artificial Vision	Président et co-président du comité d'organisation
PSIVT	2007, 2009	Pacific-Rim Symposium on Image and Video Technology	Membre du comité de programme
VISAPP	2009	International Conference on Computer Vision, Theory and Applications	Membre du comité de programme
EI	2008, 2009	Electronic Imaging	Président du comité de programme
ICIP	2006, 2007, 2008	International Conference on Image Processing	Relecteur
ICRA	2008, 2009	International Conference on Robotics and Automation	Relecteur
ITSC	2007, 2008	Intelligent Transportation Systems	Relecteur
ACCV	2007	Asian Conference on Computer Vision	Relecteur
IbPRIA	2007	Iberian Conference on Pattern Recognition and Image Analysis	Relecteur
RFIA	2006	Reconnaissance des Formes et Intelligence Artificielle	Relecteur

TABLE 6.2: Responsabilités dans des conférences

6.4 Activités Pédagogiques

Depuis 2002, j'ai effectué l'essentiel de mon service dans le département GEII de l'IUT du Creusot, où j'enseigne l'information industrielle, la programmation et la C.A.O., et en Licence Professionnelle CIVA (Contrôle Industriel par Vision Artificielle) où j'enseigne le traitement des images et les réseaux de neurones. J'effectue également quelques heures de reconnaissance des formes dans le Master STIC de Dijon (3i : Instrumentation, Image et Ingénierie; VIIA : Vision Industrielle et Intelligence Artificielle). Plus marginalement, je suis intervenu dans le master Ingénierie Numérique co-habité par l'ENSAM et l'Université de Bourgogne pour un cours sur le recalage 3D rigide.

Mes enseignements en Master VIBOT traitent de la vision par ordinateur (géométrie des vues multiples, auto-calibrage) et proposent également une introduction générale à la perception visuelle. Je suis également en charge de l'organisation et du suivi des projets de recherche confiés à ces étudiants. En quelques dates :

- **Depuis septembre 2002** : Maître de Conférences au département GEII de l'IUT du Creusot : voir tableau ci-dessous.
- **2001-2002** : Professeur Contractuel au département Informatique de l'IUT d'Amiens (Université de Picardie Jules Verne) : mathématiques du signal, traitement des images, graphes et algorithmes, langage C.
- **1999-2001** : A.T.E.R. au département Informatique de l'IUT d'Amiens (Université de Picardie Jules Verne) : simulation numérique, algorithmique / langage C, bureautique.
- **1997-1999** : Chargé d'Enseignement Vacataire (IUT Amiens, IUP GEII Amiens, ESIEE Amiens, INSSET Saint-Quentin) : traitement des images, base de données, simulation numérique, bureautique.

Le tableau ci-dessous regroupe les matières et le volume horaire précis des enseignements effectués pour l'année scolaire 2005/2006 (depuis l'année 2006/2007 pour les enseignements au master VIBOT) . Il est représentatif des enseignements que j'ai dispensés depuis mon arrivée à l'Université de Bourgogne.

Matière enseignée	Formation	CM	TD	TP
Reconnaissance des formes	Master Recherche 3i / Master Pro. VIIA	10	0	0
Traitement des images	Licence Pro. CIVA	2	14	24
Réseaux de neurones	Licence Pro. CIVA	3	0	8
Projets tutorés	Licence Pro. CIVA	0	40	0
Mesures 3D sans contact	IUT Mesures Physiques, 2e année	6	0	0
Informatique industrielle	IUT GEII, 2e année	10	0	0
C.A.O.	IUT GEII, 1e année	0	0	45
Informatique	DU SPI Chine	0	0	25
Human Vision Psychophysics	Master VIBOT	5	0	0
3D Digitization	Master VIBOT	5	0	0
Research Project	Master VIBOT	Responsable		

TABLE 6.3: Charge d'enseignement

J'ai obtenu une année complète de délégation CNRS en 2007/2008, renouvelée pour 6 mois en 2008/2009.

6.5 Valorisation et Transfert

Le tableau 6.4 recense les prestations, expertises, activités de valorisation et de transfert effectuées pendant et après ma thèse de doctorat. C'est la plateforme technologique du Creusot, Platform3D, qui structure les activités de transfert et de valorisation de la cellule Vision 3D.

Les opérations sont, pour la plupart, collectives et mobilisent une équipe de 2 ou 3 enseignants-chercheurs, ingénieurs ou techniciens autour d'un projet.

Société	Année	Intitulé	Type d'action	Lieu
CDI Saint Gobain Vitrages	1999	Développement d'un logiciel de photo-émaillage d'images en couleurs à l'aide d'un laser	Prestation industrielle	Thourotte (60)
High-Tech Design	2002	Validation d'un outil de conception de pièces décorées et dispositif de contrôle par VAO de films imprimés par Héliographie	Prestation de Conseil Technologique	Montchanin (71)
COM'AIR	2005	Cartographie à partir de Prises de Vues Aériennes	Prestation de Conseil Technologique	Montcoy (71)
Systel Electronique	2005	Détection de Défauts par Acquisition Vidéo Sous-Marine d'Assemblage de Combustibles Nucléaires	Prestation de Conseil Technologique	Cersot (71)
NT2i	2007	Etude et Développement d'un Profilomètre Laser	Prestation, transfert	Saint-Etienne (42)
Qualup	2007	Rapport d'expertise sur le logiciel de photogrammétrie PixDim	Expertise OSEO-ANVAR	Montbellet (71)
InterControle / AREVA	2007	Etat de l'Art sur les Méthodes de Métrologie 3D pour la Caractérisation de Surfaces Métalliques	Prestation, Expertise	Chalon-sur-Saône (71)

TABLE 6.4: Prestations, Valorisation, Transfert

Chapitre 7

Bibliographie de l'Auteur

Depuis 1998, j'ai publié dans une dizaine de revues scientifiques et participé à une quarantaine de conférences internationales et nationales. J'ai également été éditeur de trois *proceedings* de conférences internationales et éditeur invité de deux numéros spéciaux dans les revues *Journal of Electronic Imaging* (SPIE/IS&T) et *Electronic Letters on Computer Vision and Image Analysis* (CVC Press).

Le logiciel *Harzing's Publish or Perish*, dont les calculs se fondent sur la base de données de *Google Scholar*, m'attribue un facteur H de 4, c'est-à-dire au moins 4 articles cités au moins 4 fois. Ce sont mes travaux sur la vision en lumière structurée qui sont les plus cités (18 fois pour l'article publié dans *Pattern Recognition* en 2003 et 15 fois pour la communication faite à ICRA 2001) ; ce sont aussi les plus anciens... L'article publié en mai 2007 dans la revue *Image and Vision Computing*, proposant un état de l'art des méthodes récentes de recalage rigide 3D, recueille déjà 17 citations ; il est parmi les articles les plus téléchargés de la revue en 2007 selon le site *ScienceDirect*. Au total, j'obtiens 93 citations pour un ratio de 7,75 citations par an.

7.1 Articles de Revue

1. O. Morel, R. Seulin, D. Fofi, "Measurement of the Three-Dimensional Mirror Parameters by Polarization Imaging applied to Catadioptric Camera Calibration", *Journal of Electronic Imaging*, IS&T/SPIE, July 2008.
2. C. Matabosch, D. Fofi, J. Salvi, E. Batlle, "Registration of Surfaces Minimizing Error Propagation for a One-Shot Multi-Slit Hand-Held Scanner", *Pattern Recognition*, Elsevier, 41(6), pp. 2055-2067, June 2008.
3. J. Salvi, C. Matabosch, D. Fofi, J. Forest, "A Review of Recent Range Image Registration Methods with Accuracy Evaluation", *Image and Vision Computing*, Elsevier, 25(5), pp. 578-596, May 2007.
4. T. Molinier, D. Fofi, F. Meriaudeau, R. Seulin, "Trifocal Tensor as a Tool for Modelling an Imperceptible Structured Light Sensor", *Journal of Optical Technology*, Optical Society of America, 74(4), April 2007.
5. H. Atmani, F. Mérienne, D. Fofi, P. Trouilloud, "Computer-Aided Surgery System for Shoulder Prosthesis Placement", *Computer Aided Surgery*, Taylor & Francis, 12(1), pp. 63-73, January 2007.
6. R. Seulin, C. Stolz, D. Fofi, G. Millon, F. Nicolier, "3D Tools for Ancient Wooden Stamps Analysis and Conservation", *The Imaging Science Journal*, Maney Publishing, 54 (2), pp. 111-121, June 2006.
7. P. Trouilloud, N. Cheynel, M. Benkhadra, O. Trost, H. Atmani, D. Fofi, F. Mérienne, E. Baulot, "Modélisation de l'Articulation de l'Epaule et Navigation", *Morphologie*, Elsevier, 89(287), December 2005.

8. C. Matabosch, D. Fofi, J. Salvi, J. Forest, "Registration of Moving Surfaces by means of One-Shot Laser Projection", *Lecture Notes in Computer Science (Proc. IbPRIA'2005)*, 3522, pp. 145-152, Springer-Verlag, June 2005.
9. T. Molinier, D. Fofi, "Auto-Calibrage à partir de Deux Vues et de la Projection d'Un Motif Lumineux", *J3eA - Journal sur l'enseignement des sciences et technologies de l'information et des systèmes*, volume 4, Hors-série 3, (2005) 002.
10. D. Fofi, E. Mouaddib, J. Salvi, "Estimation du Mouvement d'un Capteur de Vision en Lumière Structurée", *Moroccan Journal of Control, Computer Science and Signal Processing*, Special Issue ICISP'03, 2005.
11. T. Sliwa, D. Fofi, Y. Voisin, A. Diou, "Utilisation d'une Caméra à Haute Vitesse d'Acquisition pour la Détection de Lignes lors d'une Illumination non Contrôlée", *Moroccan Journal of Control, Computer Science and Signal Processing*, Special Issue ICISP'03, 2005.
12. D. Fofi, J. Salvi, E. Mouaddib, "Uncalibrated Reconstruction : An Adaptation to Structured Light Vision", *Pattern Recognition*, 36(7), pp. 1631-1644, July 2003.
13. B. Marhic, E. Mouaddib, D. Fofi, E. Brassart, "Localisation Absolue par le Capteur Omnidirectionnel SYCLOP", *Traitement du Signal*, numéro spécial « Fusion de données pour véhicules intelligents », vol. 17, n°3, pp. 195-206, 2000.

7.2 Communications Internationales

1. N. Muhammad, D. Fofi, S. Ainouz, "Current State-of-the-Art of Vision-based SLAM", *IS&T / SPIE Electronic Imaging - Image Processing : Machine Vision Applications II (IP-MVA'2009)*, San José, USA, January 2009.
2. S. Mosaddegh, D. Fofi, P. Vasseur, S. Ainouz, "Line Matching Across Catadioptric Images under Short Baseline Motion", *8th Workshop on Omnidirectional Vision (OMNIVIS'2008)* - in conjunction with ECCV'2008, Marseille, France, November 2008.
3. S. Ainouz, O. Morel, D. Fofi, "Mirror-Adapted Matching of Catadioptric Images", *IEEE International Conference on Image Processing (ICIP'2008)*, San Diego, California, USA, October 2008.
4. S. Ainouz, O. Morel, D. Fofi, "Image Processing in Catadioptric Plane", *The International Topical Meeting on Optical Sensing and Artificial Vision (OSAV'2008)*, Saint Petersburg (Russia), May 2008.
5. T. Molinier, D. Fofi, P. Gorria, J. Salvi, "2D Virtual Texture on 3D Real Object with Coded Structured Light", *SPIE Electronic Imaging - Image Processing : Machine Vision Applications*, San José, USA, January 2008.
6. T. Molinier, D. Fofi, P. Gorria, J. Salvi, "Automatic Texture Mapping on Real 3D Model", *IEEE Workshop on Projector Camera Systems (PROCAMS'2007)* - in conjunction with CVPR'2007, Minneapolis (Minnesota), June 2007.
7. O. Morel, R. Seulin, D. Fofi, "Catadioptric Camera Calibration by Polarization Imaging", *IAPR 3rd Iberian Conference on Pattern Recognition and Image Analysis (IbPRIA'2007)*, Girona (Spain), June 2007.
8. O. Morel, R. Seulin, D. Fofi, "Measurement of the Three-Dimensional Mirror Parameters by Polarization Imaging applied to Catadioptric Camera Calibration", *IEEE/SPIE 8th International Conference on Quality Control by Artificial Vision (QCAV'2007)*, Le Creusot (France), May 2007.
9. T. Molinier, D. Fofi, P. Gorria, J. Salvi, "Detection of Surfaces for Projection of Texture", *IEEE/SPIE 8th International Conference on Quality Control by Artificial Vision (QCAV'2007)*, Le Creusot (France), May 2007.
10. H. Atmani, F. Mérienne, D. Fofi, P. Trouilloud, "From Medical Data to Simple Virtual Mock-Up of Scapulo-Humeral Articulation", *IEEE/SPIE 8th International Conference on Quality Control by Artificial Vision (QCAV'2007)*, Le Creusot (France), May 2007.

11. O. Morel, D. Fofi, "Calibration of Catadioptric Sensors by Polarization Imaging", *IEEE International Conference on Robotics and Automation (ICRA'2007)*, Roma (Italia), April 2007.
12. C. Matabosch, E. Batlle, D. Fofi, J. Salvi, "A Variant of Point-to-Plane Registration Including Cycle Minimization", *ISPRS Photogrammetric Computer Vision (PVC'2006)*, Bonn (Germany), September 2006.
13. H. Atmani, D. Fofi, P. Trouilloud, F. Mérienne, "Virtual Mock-Up of Shoulder Articulation for an Aided Surgery System", *2nd Workshop on Computer Assisted Diagnosis and Surgery*, Santiago de Chile (Chile), March 2006.
14. H. Atmani, F. Mérienne, D. Fofi, P. Trouilloud, "Towards a Computer-Aided Surgery System for Shoulder Prosthesis Placement", *Medecine Meets Virtual Reality 14 (MMVR'14)*, Long Beach (California), January 2006.
15. C. Matabosch, J. Salvi, D. Fofi, F. Meriaudeau, "A Refined Range Image Registration Technique for Multi-Stripe Laser Scanner", *SPIE Electronic Imaging – Machine Vision Applications in Industrial Inspection XIV (MVAII'2006)*, San José (USA), January 2006.
16. T. Molinier, D. Fofi, P. Gorria, "Novel View Synthesis for Projective Texture Mapping on Real 3D Objects", *SPIE Electronic Imaging – Machine Vision Applications in Industrial Inspection XIV (MVAII'2006)*, San José (USA), January 2006.
17. H. Atmani, D. Fofi, F. Mérienne, P. Trouilloud, "Constructing a Simple Parametric Model of Shoulder from Medical Images", *SPIE Electronic Imaging – Machine Vision Applications in Industrial Inspection XIV (MVAII'2006)*, San José (USA), January 2006.
18. T. Molinier, D. Fofi, F. Meriaudeau, "Self-Calibration of a Trinocular Sensor with Imperceptible Structured Light and Varying Intrinsic Parameters", *7th International Conference on Quality Control by Artificial Vision (QCAV'2005)*, Nagoya (Japan), May 2005.
19. C. Matabosch, J. Salvi, D. Fofi, "A New Proposal to Register Range Images", *7th International Conference on Quality Control by Artificial Vision (QCAV'2005)*, Nagoya (Japan), May 2005.
20. E. Thomas, D. Fofi, F. Nicolier, G. Millon, R. Seulin, "Low Cost System for Ancient Stamps Range Image Acquisition", *SPIE Electronic Imaging – Machine Vision Applications in Industrial Inspection XIII (MVAII'2005)*, San José (USA), January 2005.
21. C. Matabosch, J. Salvi, D. Fofi, F. Meriaudeau, "Range Image Registration for Industrial Inspection", *SPIE Electronic Imaging – Machine Vision Applications in Industrial Inspection XIII (MVAII'2005)*, San José (USA), January 2005.
22. T. Molinier, D. Fofi, R. Seulin, F. Meriaudeau, "Trifocal Tensor Estimation Applied to Imperceptible Structured Light", *The International Topical Meeting on Optical Sensing and Artificial Vision (OSAV'2004)*, Saint Petersburg (Russia), October 2004.
23. R. Seulin, F. Nicolier, D. Fofi, G. Millon, C. Stolz, "3D Imaging Applications for Ancient Wooden Stamps Analysis", *The International Topical Meeting on Optical Sensing and Artificial Vision (OSAV'2004)*, Saint Petersburg (Russia), October 2004.
24. D. Fofi, T. Sliwa, Y. Voisin, "A Comparative Survey on Invisible Structured Light", *SPIE Electronic Imaging – Machine Vision Applications in Industrial Inspection XII (MVAII'2004)*, pp. 90-97, San José (USA), January 2004.
25. R. Seulin, Y. Voisin, D. Fofi, F. Meriaudeau, "Complete Machine Vision Solution for Tube Inspection in Nuclear Industry", *SPIE Electronic Imaging – Machine Vision Applications in Industrial Inspection XII (MVAII'2004)*, pp. 128- 136, San José (USA), January 2004.
26. T. Sliwa, D. Fofi, Y. Voisin, A. Diou, "Utilisation d'une Caméra à Haute Vitesse d'Acquisition pour la Détection de Lignes lors d'une Illumination non Contrôlée", *International Conference on Image and Signal Processing (ICISP'2003)*, pp. 497-508, Agadir (Maroc), June 2003.
27. D. Fofi, E. Mouaddib, J. Salvi, "Estimation du Mouvement d'un Capteur de Vision en Lumière Structurée", *International Conference on Image and Signal Processing (ICISP'2003)*, pp. 468-475, Agadir (Maroc), June 2003.

28. D. Fofi, E. Mouaddib, J. Salvi, "Segmentation et Décodage d'un Patron de Lumière Structurée", *Troisièmes Journées d'Optique et de Traitement de l'Information (OPTIQUE'2002)*, pp. 162-169, Kénitra (Maroc), Octobre 2002.
29. D. Fofi, E. Mouaddib, J. Salvi, "How to Self-Calibrate a Structured Light Sensor?", *Proceedings of the 9th International Symposium on Intelligent Robotic System (SIRS'2001)*, Toulouse (France), July 2001.
30. D. Fofi, J. Salvi, E. Mouaddib, "Uncalibrated Vision based on Structured Light", *IEEE International Conference on Robotics and Automation (ICRA'2001)*, Séoul (Corée), May 2001.
31. D. Fofi, J. Salvi, E. Mouaddib, "Euclidean Reconstruction by means of an Uncalibrated Structured Light Sensor", *Vth Ibero-American Symposium on Pattern Recognition (SIARP'2000)*, pp. 159-170, Lisboa (Portugal), September 2000.

7.3 Communications Nationales

1. T. Molinier, D. Fofi, P. Gorria, J. Salvi, "Projection de Texture sur des Objets 3D Réels", *Colloque du Groupe de Recherche sur le Traitement du Signal et des Images (GRETSI'2007)*, Troyes (France), Septembre 2007.
2. O. Morel, D. Fofi, "Calibrage de Capteurs Catadioptriques par Imagerie Polarimétrique", *Journée Vision Omnidirectionnelle*, GDR ISIS / Club EEA, Amiens (France), 27 mars 2007.
3. H. Atmani, F. Mérienne, D. Fofi, P. Trouilloud, "Conception d'un Maquette Numérique de l'Epaule : Exploitation d'Images Médicales", *18es Journées de l'Association Française d'Informatique Graphique (AFIG'2005)*, Strasbourg (France), Novembre 2005.
4. C. Matabosch, D. Fofi, J. Salvi, "Reconstruction et Recalage de Surfaces en Mouvement par Projection Laser Multi-Lignes", *9es Journées des Jeunes Chercheurs en Vision par Ordinateur (ORASIS'2005)*, Clermont-Ferrand, Mai 2005.
5. D. Fofi, P. Gorria, "Analyse du Temps d'Oxydation de Tubes Métalliques par Polytopes de Contraintes dans l'Espace Couleur", *Ecole de Printemps Images Numériques Couleurs (EDPC'2003)*, Dijon (France), Mars 2003.
6. D. Fofi, E. Mouaddib, J. Salvi, "Décodage d'un Motif Structurant Codé par la Couleur", *Colloque du Groupe de Recherche et d'Etude en Traitement du Signal et des Images (GRETSI'2001)*, Toulouse (France), Septembre 2001.
7. D. Fofi, E. Mouaddib, "La Vision en Lumière Structurée Appliquée à la Navigation d'un Robot Mobile", *Réunion du groupe de travail GT 3.2 Couleur*, GDR-PRC ISIS, 15 Juin 1999.
8. D. Fofi, E. Mouaddib, "Vision en Lumière Structurée : Application à la Détection d'Obstacles", *Actes des 9es Journées des Jeunes Chercheurs en Robotique (JJCR '9)*, pp. 95-99, Clermont-Ferrand (France), Juin 1998.

7.4 Editions

1. *Image Processing : Machine Vision Applications II*, edited by K. Niel, D. Fofi, Proceedings of SPIE, vol. 6813, January 2009.
2. *Special Section on Quality Control by Artificial Vision*, Journal of Electronic Imaging, co-edited by H. Sari-Sarraf, N. Yung, D. Fofi, SPIE and IS&T, 2008.
3. *Special Issue on Quality Control by Artificial Vision*, Electronic Letters on Computer Vision and Image Analysis, co-edited by D. Fofi, R. Seulin, CVC Press, 2008.
4. *Image Processing : Machine Vision Applications*, edited by K. Niel, D. Fofi, Proceedings of SPIE, vol. 6813, January 2008.
5. *Eighth International Conference on Quality Control by Artificial Vision*, edited by D. Fofi, F. Mériaudeau, Proceedings of SPIE, vol. 6356, May 2007.

Bibliographie

- [Armangué 2003] X. Armangué, J. Salvi, “Overall View Regarding Fundamental Matrix Estimation”, *Image and Vision Computing*, 21(2), pp 205-220, February 2003.
- [Baker 1999] S. Baker, S.K. Nayar, “A Theory of Single-Viewpoint Catadioptric Image Formation”, *International Journal of Computer Vision*, 35(2), pp. 175-196, 1999.
- [Barreto 2005] J.P. Barreto, H. Araujo, “Geometric Properties of Central Catadioptric Line Images and their Application in Calibration”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(8), pp. 1327-1333, 2005.
- [Batlle 1998] J. Batlle, J. Salvi, E. Mouaddib, “Recent Progress in Coded Structured Light as a Technique to Solve the Correspondence Problem : A Survey”, *Pattern Recognition*, 31(7), pp. 963-982, July 1998.
- [Bazin 2007] J.C. Bazin, C. Démonceaux, P. Vasseur, “Fast Central Catadioptric Line Extraction”, *Iberian Conference on Pattern Recognition and Image Analysis*, pp. 25-32, 2007.
- [Besl 1992] J.P. Besl, N.D. McKay, “A method for registration of 3-D shapes”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2), pp. 239-256, 1992.
- [Caspi 1998] D. Caspi, N. Kiryati, J. Shamir, “Range Imaging With Adaptive Color Structured Light”, *IEEE Transactions Pattern Analysis and Machine Intelligence*, 20(5), pp. 470-480, 1998.
- [Chen 1987] C.H. Chen and A.C. Kak, “Modeling and Calibration of a Structured Light Scanner for 3D Robot Vision”, *Proceedings of the IEEE Conference on Robotics and Automation (ICRA '1987)*, pages 807-815, 1987.
- [Chen 1991] Y. Chen, G. Medioni, “Object modeling by registration of multiple range images”, *IEEE International Conference on Robotics and Automation*, pp. 2724-2729, 1991.
- [Chen 2002] S.Y. Chen, Y.F. Li, “Self Recalibration of a Structured Light Vision System from a Single View”, *IEEE International Conference on Robotics and Automation (ICRA '2002)*, vol. 3, pp. 2539-2544, May 2002.
- [Chung 1998] D. Chung, Y.D.S. Lee, “Registration of multiple-range views using the reverse-calibration technique”, *Pattern Recognition*, 31(4), pp. 457-464, 1998.
- [Cotting 2004] D. Cotting, M. Naef, M. Gross, H. Fuchs, “Embedding imperceptible patterns into projected images for simultaneous acquisition and display”, *3rd IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 100-109, Washington, DC, USA, 2004.
- [Daniilidis 2002] K. Daniilidis, A. Makadia, T. Blow, “Image Processing in Catadioptric Planes : Spatiotemporal Derivatives and Optical Flow Computation”, *IEEE 3rd Workshop on Omnidirectional Vision (OMNIVIS)*, pp. 3-10, 2002.
- [Devernay 2002] F. Devernay, O. Bantiche, E. Coste-Manière, “Structured light on dynamic scenes using standard stereoscopy algorithms”, *Rapport de Recherche INRIA*, 4477, 2002.

- [Faugeras 1986] O. Faugeras, G. Toscani, "The calibration problem for stereo", *Computer Vision and Pattern Recognition (CVPR)*, pp. 15-20, June 1986.
- [Fischler 1981] M.A. Fischler, R.C. Bolles, "Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography", *Communications of the ACM*, 24(6), pp. 381-395, 1981.
- [Fofi 2001] D. Fofi, "Navigation d'un Véhicule Intelligent à l'aide d'un Capteur de Vision en Lumière Structurée et Codée", *Thèse de Doctorat*, Université de Picardie Jules Verne, Septembre 2001.
- [Fougerolle 2006] Y. Fougerolle, A. Gribok, S. Foufou, F. Truchetet, M. Abidi, "Radial supershapes for solid modeling", *Journal of Computer Science and Technology*, 21(2), pp. 238-243, 2006.
- [Forest 2004] J. Forest, "New Methods for Triangulation-Based Shape Acquisition using Laser Scanners", *PhD Thesis*, Universitat de Girona, December 2004.
- [Frankot 1988] R. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms", vol. 10, pp. 439-451, July 1988.
- [Geyer 2000] C. Geyer, K. Daniilidis, "A Unifying Theory for Central Panoramic Systems and Practical Implications", *Lecture Notes in Computer Science*, 1843/2000, Springer-Verlag, 2000.
- [Geyer 2001] C. Geyer, K. Daniilidis, "Catadioptric Projective Geometry", *International Journal of Computer Vision*, Springer-Verlag, 2001.
- [Griffin 1992] P. Griffin, L. Narasimhan, S. Yee, "Generation of uniquely encoded light patterns for range data acquisition", *Pattern Recognition*, 25(6), pp. 609- 616, 1992.
- [Grossberg 2001] M. D. Grossberg and S. K. Nayar, "A General Imaging Model and a Method for Finding its Parameters", *IEEE International Conference on Computer Vision*, Vol. 2, pp. 108-115, 2001.
- [Harris 1988] C. Harris, M. Stephens, "A Combined Corner and Edge Detector", *Fourth Alvey Conference*, pp. 147-151, 1988.
- [Hartley 2000] R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge Press University, 2000.
- [Huynh 1999] D.Q. Huynh, R.A. Owens, P.E. Hartmann, "Calibrating a Structured Light Stripe System : A Novel Approach", *International Journal of Computer Vision*, 33(1), pp. 73-86, 1999.
- [Koninckx 2005] T. P. Koninckx, P. Peers, P. Dutré, L. Van Gool, "Scene-adapted Structured Light", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp 611-619, 2005.
- [Laporte 2003] S. Laporte, W. Skalli, J.A. De Guise, F. Lavaste, D. Mitton, "A Biplanar Reconstruction Method Based on 2D and 3D Contours : Application to the Distal Femur", *Computer Methods in Biomechanics and Biomedical Engineering*, 6, pp. 1-6, 2003.
- [Le Bras 2004] A. Le Bras, "Exploration des Potentialités du Système EOS pour la Caractérisation Mécanique de Structure Osseuse : Application à l'Extrémité Supérieure du Fémur", *Thèse de Doctorat*, ENSAM Paris, 2004.
- [Li 2003] Y.F. Li, S.Y. Chen, "Automatic Recalibration of an Active Structured Light Vision System", *IEEE Transactions on Robotics and Automation*, 19(2), pp. 259-268, 2003.
- [Marquardt 1963] D.W. Marquardt, "An Algorithm for the Estimation of Nonlinear Parameters", *Soc. Ind. Appl. Math. J.*, 11, pp. 431-441, 1963.

- [Matsuda 1996] T. Masuda, K. Sakaue, N. Yoyoka, "Registration and integration of multiple range images for 3D model construction", *International Conference on Pattern Recognition (ICPR)*, vol.1, pp. 879-883, 1996.
- [Matsuda 2001] T. Masuda, "Generation of geometric model by registration and integration of multiple range images", *3rd International Conference on 3-D Digital Imaging and Modeling*, pp. 254-261, 2001.
- [Matabosch 2007] C. Matabosch, "Hand-Held 3D Scanner for Large Surface Registration", *PhD Thesis*, Universitat de Girona, Spain, July 2007.
- [Matabosch 2008] C. Matabosch, D. Fofi, J. Salvi, E. Batlle, "Registration of Surfaces Minimizing Error Propagation for a One-Shot Multi-Slit Hand-Held Scanner", *Pattern Recognition*, Elsevier, 41(6), pp. 2055-2067, June 2008.
- [Mathieu 1995] H. Mathieu, F. Devernay, "Système de Miroir pour la Stéréoscopie", *Rapport Technique INRIA*, 172, Juin 1995.
- [Mei 2007] C. Mei, P. Rives, "Single View Point Omnidirectional Camera Calibration from Planar Grids", *IEEE International Conference on Robotics and Automation*, 2007.
- [Miyazaki 2004] D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent Surface Modeling from a Pair of Polarization Images", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(X), pp. 73-82, January 2004.
- [Morel 2006a] O. Morel, C. Stolz, F. Meriaudeau, and P. Gorria, "Active lighting applied to 3D reconstruction of specular metallic surfaces by polarization imaging", *Applied Optics*, vol. 45, pp. 4062-4068, June 2006.
- [Morel 2006b] O. Morel, C. Stolz, P. Gorria, "Polarization Imaging for 3D Inspection of Highly Reflective Metallic Objects", *Optics and Spectroscopy*, 101 (1), pp. 15-21, July 2006.
- [Morel 2007] O. Morel, D. Fofi, "Calibration of Catadioptric Sensors by Polarization Imaging", *IEEE International Conference on Robotics and Automation (ICRA'2007)*, Roma (Italia), April 2007.
- [Morel 2008] O. Morel, R. Seulin, D. Fofi, "Measurement of the Three-Dimensional Mirror Parameters by Polarization Imaging applied to Catadioptric Camera Calibration", *Journal of Electronic Imaging*, IS&T/SPIE, July 2008.
- [Mouaddib 2003] E. Mouaddib, "La Vision Omnidirectionnelle", *Journées Nationales de la Recherche en Robotique*, Clermont-Ferrand, France, Octobre 2003.
- [Park 2003] S.-Y. Park, M. Subbarao, "A fast point-to-tangent plane technique for multi-view registration", *4th International Conference on 3D Digital Imaging and Modeling*, pp. 276-284, 2003.
- [Pless 2003] R. Pless, "Using Many Cameras as One", *IEEE Computer Vision and Pattern Recognition*, 2003.
- [Pollefeys 1999] M. Pollefeys, R. Koch and L. Van Gool, "Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Internal Camera Parameters", *International Journal of Computer Vision*, 32(1), 7-25, 1999.
- [Prados 2003] E. Prados, O. Faugeras, "'Perspective Shape from Shading" and Viscosity Solutions", *IEEE International Conference on Computer Vision*, 2003.
- [Ramalingam 2006] S. Ramalingam, S. Lodha, P. Sturm, "A Generic Structure-from-Motion Framework", *Computer Vision and Image Understanding*, Elsevier, 103(3), pp. 218-228, 2006.

- [Raskar 1998] R. Raskar, G. Welch, M. Cutts, A. Lake, L. Stesin, H. Fuchs, "The office of the future : a unified approach to image-based modeling and spatially immersive displays", *International Conference on Computer Graphics and Interactive Techniques*, pp. 179-188, 1998.
- [Raskar 2001] R. Raskar, G. Welch, K.-L. Low, D. Bandyopadhyay, "Shader Lamps : Animating Real Objects With Image-Based Illumination", *12th Eurographics Workshop on Rendering Techniques*, pp. 89-102, 2001.
- [Raskar 2001] R. Raskar, P.A. Beardsley, "A Self-Correcting Projector", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'2001)*, vol. 2, pp. 504-508, December 2001.
- [Ressl 2003] C. Ressel, "Geometry, Constraints and Computation of the Trifocal Tensor", *PhD Thesis*, Institut für Photogrammetrie und Fernerkundung, Germany, 2003.
- [Roberts 1991] S.N.J. Roberts, A.P.J. Foley, H.M. Swallow, W.A. Wallace, D.P. Coughlan, "The geometry of the humeral head and the design of the protheses", *Journal of Bone and Joint Surgery*, 73-B, p. 647-650, 1991.
- [Robinson 2004] A. Robinson, L. Alboul, M. Rodrigues, "Methods for indexing stripes in uncoded structured light scanning systems", *Journal of Winter School of Computer Graphics*, 2004.
- [Rusinkiewicz 2001] S. Rusinkiewicz, M. Levoy, "Efficient variant of the ICP algorithm", *3rd International Conference on 3-D Digital Imaging and Modeling*, pp. 145-152, 2001.
- [Salvi 2002] J. Salvi, X. Armangué, J. Batlle, "A Comparative Review of Camera Calibrating Methods with Accuracy Evaluation", *Pattern Recognition*, 35(7), pp. 1617-1635, July 2002.
- [Salvi 2004] J. Salvi, J. Pagès, J. Batlle, "Pattern Codification Strategies in Structured Light Systems", *Pattern Recognition*, 37(4), pp. 827-849, April 2004.
- [Selvik] G. Selvik, P. Alberius, A.S. Aronson, "A roentgen stereophotogrammetric system. Construction, calibration and technical accuracy", *Acta Radiologica Diagnosis*, 24, pp. 343-52, 1983.
- [Sharp 2004] C.G. Sharp, S.W. Lee, D.K. Wehe, "Multiview registration of 3D scenes by minimizing error between coordinate frames", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(8), pp. 1037-1050, 2004.
- [Strauss 2005] O. Strauss, F. Comby, "Fuzzy Morphology for Omnidirectional Images", *International Conference on Image Processing (ICIP)*, Genova, Italia, 2005.
- [Sturm 1999] P. Sturm, S. J. Maybank, "On Plane-Based Camera Calibration : A General Algorithm, Singularities, Applications", *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 432-437, Fort Collins, Colorado, June 1999.
- [Sturm 2004] P. Sturm, S. Ramalingam, "A Generic Concept for Camera Calibration", *European Conference on Computer Vision*, Prague, Czech Republic, Vol. 2, pp. 1-13, May 2004.
- [Sturm 2005] P. Sturm, "Multi-View Geometry for Generic Camera Model", *IEEE Computer Vision and Pattern Recognition*, Vol. 1, pp. 206-212, 2005.
- [Swaminathan 2001] R. Swaminathan, M.D. Grossberg, S.K. Nayar, "Caustics of Catadioptric Cameras", *Proceedings of the 8th International Conference on Computer Vision (ICCV)*, 2, pp. 2-9, 2001.
- [Turk 1996] G. Turk, M. Levoy, "Zippered polygon meshes from range images", *21st annual Conference on Computer graphics and Interactive Techniques (SIGGRAPH)*, Orlando, Florida, pp. 311-318, 1996.

- [Will 1971] P.M. Will, K.S. Pennington, “Grid Coding : A Preprocessing Technique for Robot and Machine Vision”, *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 66-70, 1971.
- [Wolff 1990] L. B. Wolff, “Polarization-based Material Classification from Specular Reflection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(11), pp. 1059–1071, November 1990.
- [Wolff 1991] L. B. Wolff and T. E. Boulton, “Constraining Object Features using a Polarization Reflectance Model”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(X), pp. 635–657, July 1991.
- [Wolfram 2002] S. Wolfram, *A New Kind of Science*, Wolfram Media, 2002.
- [Zhang 1998] Z. Zhang, “Determining the Epipolar Geometry and its Uncertainty : A Review”, *International Journal of Computer Vision*, 27(2), pp. 61-198, 1998.
- [Zhang 1999] Z. Zhang, “Flexible Camera Calibration by Viewing a Plane from Unknown Orientations”, *International Conference on Computer Vision (ICCV)*, pp. 666-673, Kerkyra, Greece, 1999.
- [Zinsser 2003] T. Zinsser, H. Schmidt, J. Niermann, “A refined ICP algorithm for robust 3-d correspondences estimation”, *International Conference on Image Processing*, pp. 695–698, 2003.

Measurement of three-dimensional mirror parameters by polarization imaging applied to catadioptric camera calibration

Olivier Morel
Ralph Seulin
David Fofi

Institut Universitaire de Technologie Le Creusot
Laboratory of Electronics, Computer Science and Image
Unités Mixtes de Recherche Centre National de la Recherche Scientifique 5158
Le Creusot, Burgundy 71200, France
E-mail: olivier.morel@u-bourgogne.fr

Abstract. We present a new efficient method for calibration of catadioptric sensors. The method is based on an accurate measurement of the three-dimensional parameters of the mirror through polarization imaging. While inserting a rotating polarizer between the camera and the mirror, the system is automatically calibrated without any calibration patterns. Moreover, this method permits most of the constraints related to the calibration of catadioptric systems to be relaxed. We show that, contrary to our system, the traditional methods of calibration are very sensitive to misalignment of the camera axis and the symmetry axis of the mirror. From the measurement of three-dimensional parameters, we apply the generic calibration concept to calibrate the catadioptric sensor. We also show the influence of the disturbed measurement of the parameters on the reconstruction of a synthetic scene. Finally, experiments prove the validity of the method with some preliminary results on three-dimensional reconstruction. © 2008 SPIE and IS&T. [DOI: 10.1117/1.2958290]

1 Introduction

Conventional perspective cameras have limited fields of view that make them restrictive in some applications such as robotics, video surveillance, and so on. One way to enhance the field of view is to place a mirror with a surface of revolution in front of the camera so that the scene reflects on the mirror omnidirectionally. Such a system, comprised of both lenses (dioptric) and mirrors (catoptric) for image formation, is called *catadioptric*. Several configurations exist, those satisfying the single-viewpoint constraint are described in Ref. 1.

Catadioptric vision systems available on the market have been extensively studied. Commercial devices are not adapted to our requirements, because optical components need to be placed between the camera and the mirror. Mirrors have therefore been produced in our own facilities

thanks to the Plateform3D Department.³² A high-speed machining center has been used to produce a very high-quality surface, which is polished after production.

We developed a new approach for calibrating catadioptric sensors by polarization imaging. This method enables the calibration of any mirror shape, since it is based on the measurement of three-dimensional parameters such as height and normal orientations of the surface. The only constraint is that an orthographic camera has to be used. To calibrate the system, we apply the generic calibration concept developed by Ramalingam and Sturm.^{2,3}

The paper is structured as follows. Section 2 recalls previous work on paracatadioptric calibration, since the measurement of the surface normals by polarization imaging induces orthographic projection, and most of the calibration methods developed for catadioptric systems rely on the single-viewpoint constraint. We show the misalignment sensitivity of these methods for the reconstruction of a synthetic scene. Then, after presenting some basic knowledge about polarization imaging, we detail in Section 3 how to calibrate the sensor with the generic calibration concept. In Section 4, we present simulations to illustrate the influence of the parameter measurement on the quality of the reconstruction. Preliminary results on a calibrated spherical mirror are also described. The paper ends with a conclusion and a few words about future work.

2 Calibration of Catadioptric Cameras

2.1 Previous Work

The most obvious calibration method that can be used is an approach based on the image of the mirror's bounding circle.^{4,5} It has the main advantage of being easily automated, but the drawbacks are that the surface mirror has to be very accurate and the mirror boundary has to accurately encode the intrinsic parameters. Other self-calibration methods can be found.^{6,7} Another approach of calibrating catadioptric sensors is to use geometric invariants on the image such as lines or circles.⁸⁻¹¹ Finally, the sensor can also be calibrated by using some calibration patterns with control points whose 3D world coordinates are known.¹²⁻¹⁶

Paper 07123SSRR received Jun. 22, 2007; revised manuscript received Feb. 8, 2008; accepted for publication Feb. 8, 2008; published online Aug. 18, 2008. This paper is a revision of a paper presented at the SPIE Conference on Quality Control by Artificial Vision, May 2007, Le Creusot, France. The paper presented there appears (unrefereed) in SPIE proceedings Vol. 6356.

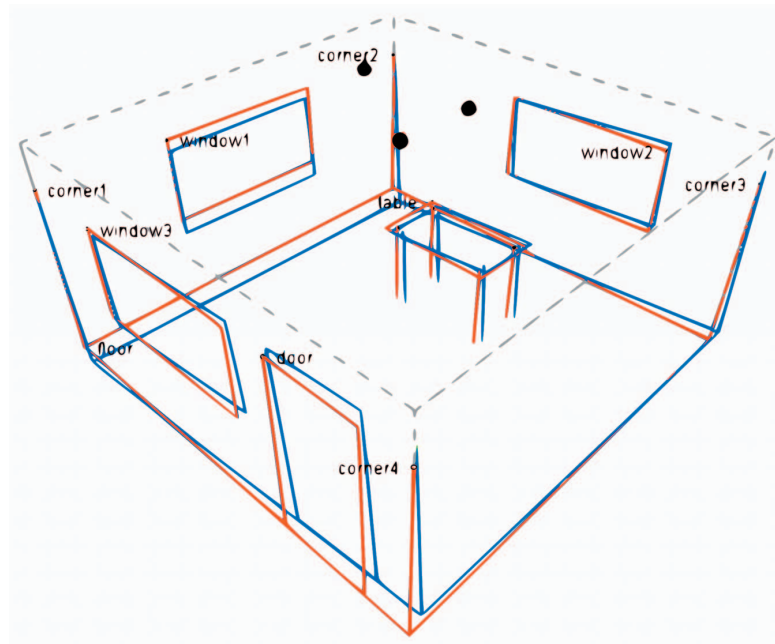


Fig. 1 Simulation of the three-dimensional reconstruction. The theoretical scene, in blue, represents a room with elements such as windows, door, and table; black dots depict the sensor's three locations. The reconstructed scene with the Vanderportael calibration method (a 2° misalignment) is drawn in red.

Since our approach requires a telecentric lens, we recall some methods devoted to paracatadioptric camera calibration (the single-viewpoint constraint involves a telecentric lens and a paraboloidal mirror). In this case, the more robust methods are based on the fitting of lines projected onto the mirror.^{8,17,18} This approach also has some shortcomings: Lines have to be precisely detected and the optical axis of the camera is assumed to be aligned with the symmetry axis of the paraboloid. To illustrate the misalignment effect, the three-dimensional reconstructions of a synthetic scene based on calibrations using the three preceding methods are simulated (Fig. 1).

The scene represents a room of size $500 \times 500 \times 250$ cm with elements such as windows, doors, and a table. Three images of the catadioptric sensor are used to triangulate the points of the scene. For the calibration process, 20 lines are computed and perfectly detected on the mirror. Then the calibration parameters are used to reconstruct the scene, according to the linear-eigen method.¹⁹ As presented in Fig. 2, the misalignment of the paracatadioptric system leads to the introduction of an important error on the reconstruction even if the calibration is performed with perfect line fittings.

To deal with the alignment errors between the mirror and the lens, more sophisticated algorithms were introduced.^{20–22} The calibration process in Refs. 20 and 21, which requires a large number of parameters to estimate, involves a three-step algorithm in order to compute all the extrinsic and intrinsic parameters. Ieng and Benosman proposed a more flexible solution based on caustic curves.²² The main advantage of this method is that it is applicable for every surface of revolution. In order to relax constraints

on the mirror shape, the method presented here is based on the generic calibration concept: It leads to a flexible, easy-to-perform, and shape-independent method of calibration.

2.2 The Generic Calibration Concept

The previous calibration methods for omnidirectional catadioptric sensors assume that (1) the mirror shape is perfectly known, (2) the alignment of the sensor is perfect, so that the single-viewpoint constraint is satisfied, and (3) the projection model can easily be parameterized. Some methods relax the second constraint and a few relax the first one, but prior to recent work,^{3,12,23} calibrating methods always underlied an explicit parametric model of projection. This

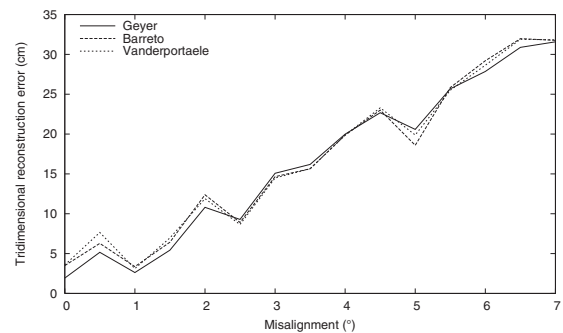


Fig. 2 Reconstruction error induced by the misalignment between camera's optical axis and the paraboloid's symmetry axis. Calibration methods used are from Geyer, Barreto, and Vanderportael. (The size of the room is $500 \times 500 \times 250$ cm.)

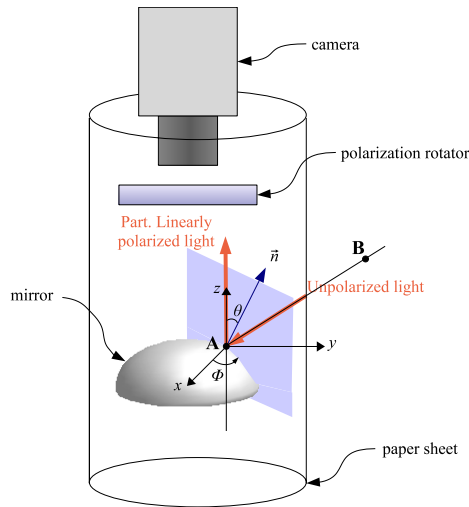


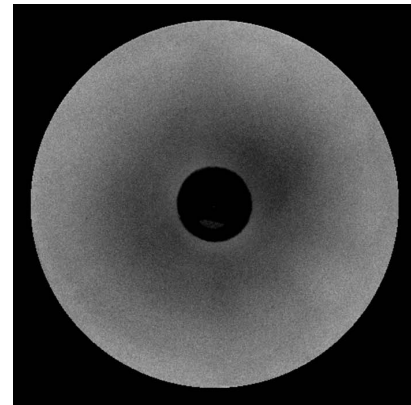
Fig. 3 Polarization imaging: After being reflected by the mirror, the light becomes partially linearly polarized.

new model has the advantage of working for any type of camera (catadioptric systems, central cameras with or without distortion, axial cameras, etc.) and of handling heterogeneous systems² (for instance, a sensor composed of an omnidirectional camera and a perspective camera). However, developing an efficient and easy-to-use calibration method based on this model is not trivial. Our proposed new method enables catadioptric sensor calibration by polarization imaging. It relaxes the three constraints listed above, and the calibration can even be performed by a non-specialist, as it only requires an optical apparatus and no image processing.

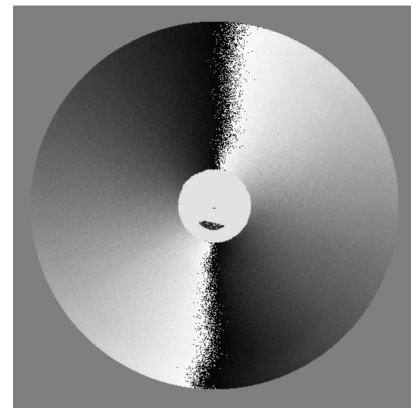
3 Polarization Imaging

Polarization imaging enables the study of the polarization state of a light wave. The most common applications in artificial vision involve segmenting dielectric and metallic objects²⁴ and detecting transparent surfaces. Polarization imaging likewise enables three-dimensional information of specular objects to be detected (the “shape from polarization” method^{25,26}). The physical principle behind “shape from polarization” is the following: After being reflected, an unpolarized light wave becomes partially linearly polarized, depending on the surface normal and the refractive index of the media on which it impinges. Partially linearly polarized light has three parameters: the light magnitude I , the degree of polarization ρ , and the angle of polarization φ .

To calibrate the mirror used in our catadioptric sensor, the polarization state of the reflected light has to be measured. A rotating polarizer placed between the camera and the mirror is used. The complete sensor (mirror and camera) and the polarizer are placed into a diffuse light environment composed of a backlit cylindrical diffuser (Fig. 3). The light intensity of each pixel is linked to the angle of the polarizer and to the polarization parameters by the following equation:



(a)



(b)

Fig. 4 Images of the polarization parameters that are needed to reconstruct the mirror shape. (a) Degree of polarization ($\rho \in [0, 1]$); (b) angle of polarization ($\varphi \in [0, \pi]$).

$$I_p(\alpha) = \frac{I}{2}(\rho \cos(2\alpha - 2\varphi) + 1), \quad (1)$$

where α is the polarizer angle. The purpose of polarization imaging is to compute the three parameters I , φ , and ρ by interpolating this formula. Because three parameters need to be determined, at least three images are required. Images are taken with different orientations of the polarizer. To get an automatic calibration of the catadioptric system, a liquid crystal polarization rotator is used instead of the polarizer. It acts as a rotating polarizer, which has the ability to be electrically controlled. Figure 4 shows an image of the degree and the angle of polarization of a spherical mirror.

3.1 The Relationship Between the Polarization Parameters and the Normals

Wolff and Boulton demonstrated how to determine constraints on surface normals using the Fresnel reflectance model.²⁷ The surface of the mirror is assumed to be continuous and described by a Cartesian expression: $z=f(x,y)$. Therefore, each surface normal is given by the following nonnormalized expression:

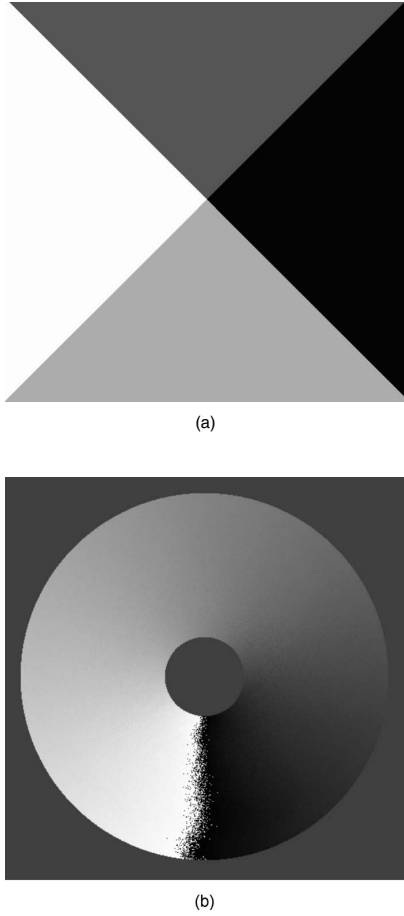


Fig. 5 Disambiguation of the azimuth angle. (a) Segmented image ($I_{quad} \in \{0, 1, 2, 3\}$); (b) image of the resulting azimuth angle ϕ ($\phi \in [-\pi/2, 3\pi/2]$).

$$\vec{n} = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ -\frac{\partial f(x,y)}{\partial y} \\ 1 \end{bmatrix} = \begin{bmatrix} p = \tan \theta \cos \phi \\ q = \tan \theta \sin \phi \\ 1 \end{bmatrix}. \quad (2)$$

The aim of the “shape from polarization” technique is to compute the normals from the angles θ and ϕ . By combining the Fresnel formulas and the Snell–Descartes law, one can find a relationship between the degree of polarization ρ

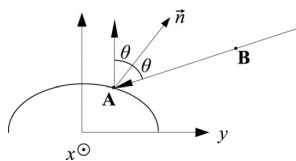


Fig. 6 Description of the 3D ray given by points **A** and **B**.

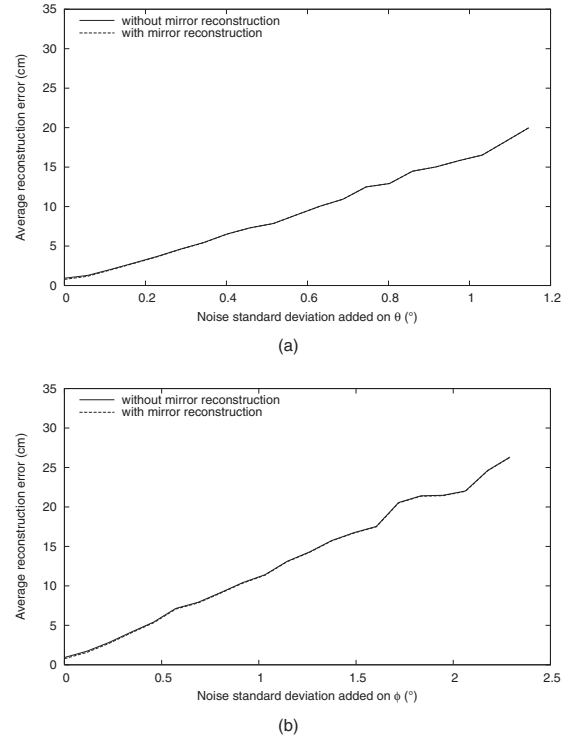


Fig. 7 Reconstruction error induced by noisy measurement of the normals' parameters. (a) θ angle; (b) ϕ angle.

and the zenith angle θ .²⁵ For specular metallic surfaces, the following formula can be applied:²⁸

$$\rho(\theta) = \frac{2n \tan \theta \sin \theta}{\tan^2 \theta \sin^2 \theta + |\hat{n}|^2}, \quad (3)$$

where $\hat{n} = n(1 + i\kappa)$ is the complex refractive index of the mirror. In the case of a perfect specular mirror, Eq. (3) can be directly used if the complex refractive index of the material is known. However, if the refractive index is unknown or if the mirror is not perfectly specular, a pseudo-refractive index²⁹ can be applied. This consists of estimating a pseudo-refractive index that best fits the relation between the angle θ and the degree of polarization ρ measured on a known shape of the same material. Finally, this pseudo-refractive index, which has no physical meaning, enables the degree of polarization image of the new object to be computed using the same relation [Eq. (3)].

The azimuth angle ϕ is linked to the angle of polarization φ since the reflected light becomes partially linearly polarized according to the normal of the plane of incidence. Because our imaging system uses a telecentric lens, an orthographic projection onto the sensor is assumed and the azimuth angle ϕ can be inferred from the angle of polarization φ according to the following equation:

$$\phi = \varphi \pm \frac{\pi}{2}. \quad (4)$$

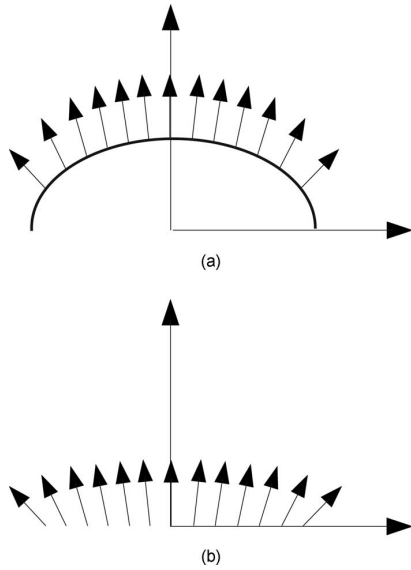


Fig. 8 Three-dimensional parameters used. (a) Normals θ , ϕ , and surface height z ; (b) only normals θ , ϕ (without integration process).

3.2 Disambiguation of the Normals

From Eqs. (3) and (4), the surface normals are determined with an ambiguity. Since mirrors used in catadioptric vision are convex and of a revolution shape, a segmented image I_{quad} can be directly computed from the near center of the mirror [Fig. 5(a)]. This segmented image is an image with four gray levels that represent the four quadrants oriented with an angle in $]0, \pi/2[$. The algorithm of the disambiguation process described in Ref. 29 is applied with the segmented image I_{quad} and with the angle of polarization image φ . The azimuth angle ϕ is computed as follows:

1. $\phi = \varphi - \pi/2$,
2. $\phi = \phi + \pi$ if $[(I_{\text{quad}}=0) \wedge (\phi \leq 0)] \vee [I_{\text{quad}}=1] \vee [(I_{\text{quad}}=3) \wedge (\phi \geq 0)]$,

where \wedge and \vee represent, respectively, the logical operators AND and OR. The result of the disambiguation is presented in Fig. 5(b).

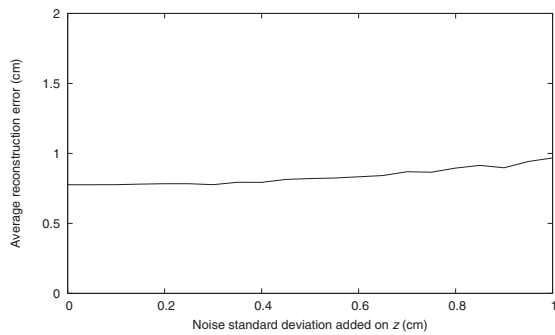
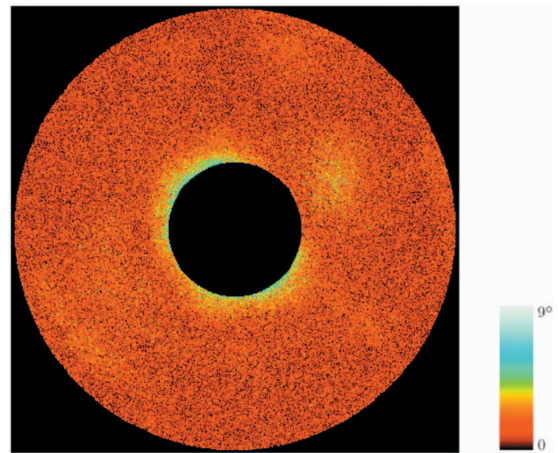
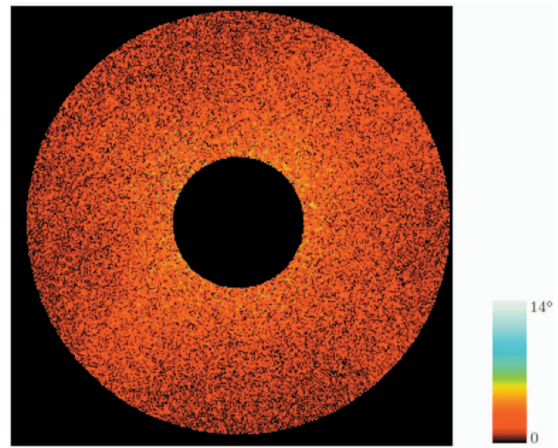


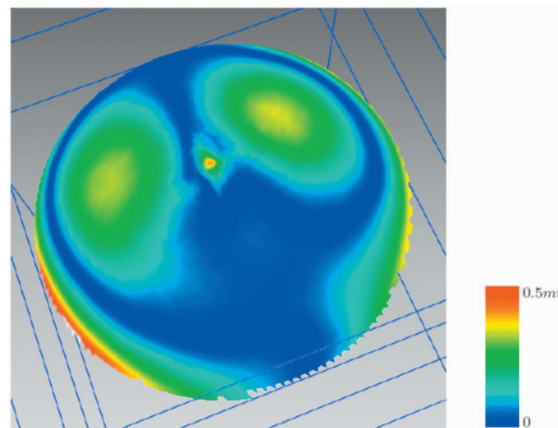
Fig. 9 Reconstruction error induced by noisy calculation of the mirror height z .



(a)



(b)



(c)

Fig. 10 Measurement errors of the three-dimensional parameters. (a) Angle θ ; (b) angle ϕ ; and (c) deviation map of the mirror z .

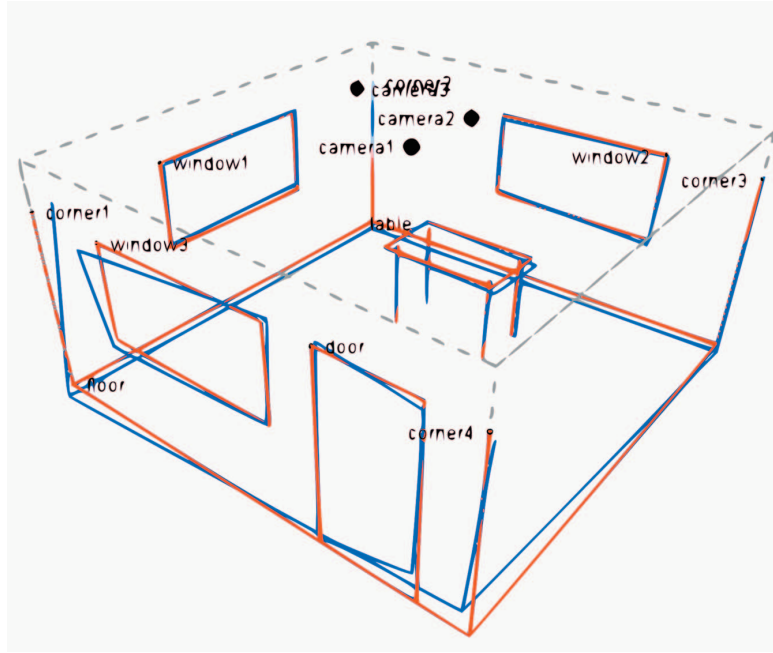


Fig. 11 Simulation of the three-dimensional reconstruction by using the calibration from the polarization imaging.

3.3 Calibration

To calibrate our imaging system, we use the generic calibration concept introduced by Sturm and Ramalingam.³⁰ The concept considers an image as a collection of pixels, and each pixel measures the light along a particular 3D ray. Thus, calibration can be seen as the determination of all projection rays and their correspondence with pixels. A 3D ray is represented here by two points (**A**, **B**) that belong to the ray. The points' coordinates are defined as described in Eq. (5):

$$\mathbf{A} = \begin{bmatrix} x_a \\ y_a \\ z_a \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} x_b \\ y_b \\ z_b \end{bmatrix}. \quad (5)$$

To calibrate the sensor, let us take point **A**, which belongs to both the mirror surface and the 3D ray, to be the first point of the ray (Fig. 6). Since an orthographic projection is assumed, the x_A - and y_A -coordinates can be directly deduced from the (u, v) -coordinates of each pixel (up to a scale factor given by the lens). Finally, the computation of the z_a -coordinate requires the determination of the 3D mirror surface. The 3D surface of the mirror can be computed from the normals (given by polarization imaging) thanks to the Frankot–Chellappa algorithm.³¹ Denoting the Fourier transforms of, respectively, the surface height and the x , y gradients as \tilde{f} , \tilde{p} , and \tilde{q} , we have

$$\forall (u, v) \neq (0, 0), \quad \tilde{f}(u, v) = \frac{-ju\tilde{p} - jv\tilde{q}}{u^2 + v^2}. \quad (6)$$

The three-dimensional surface is obtained by taking the inverse Fourier transform of the former equation. This integration process gives us the surface height of the mirror with a constant of integration. Nevertheless, this constant is not required, because of the orthographic projection assumption.

As shown in Fig. 6, the second point **B** of the ray can be written as

$$\mathbf{B} = \mathbf{A} + k \begin{bmatrix} \tan 2\theta \cos \phi \\ \tan 2\theta \sin \phi \\ 1 \end{bmatrix}, \quad (7)$$

where k is a nonnull constant.

4 Experiments

In the previous section, we showed that the three-dimensional parameters of the mirror, θ , ϕ , and z , are required to calibrate the catadioptric system, according to the generic calibration concept introduced by Sturm. The azimuth and zenith angles, ϕ and θ , are directly given by the measurement of the polarization parameters (ρ, φ) of the light reflected by the mirror. The surface height z , which represents the 3D shape of the mirror, is obtained by an integration process.

In this section, simulations are presented to illustrate the influence of the parameters on reconstruction quality. Then,

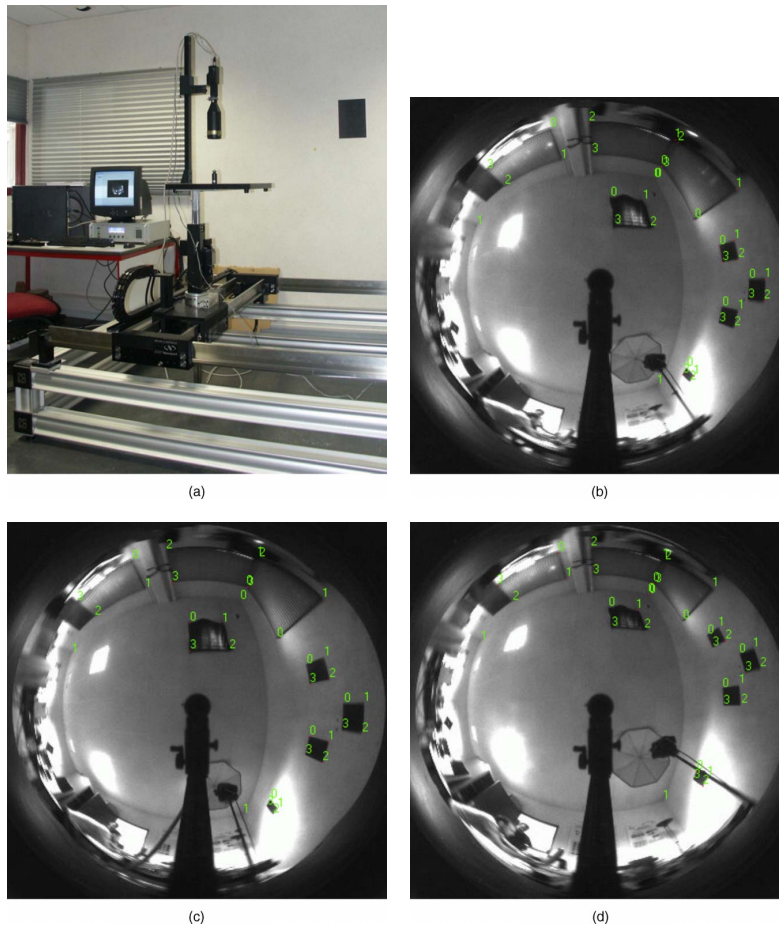


Fig. 12 Three-dimensional reconstruction of a real scene with the catadioptric system. (a) Experimental setup; (b)–(d) points of interest.

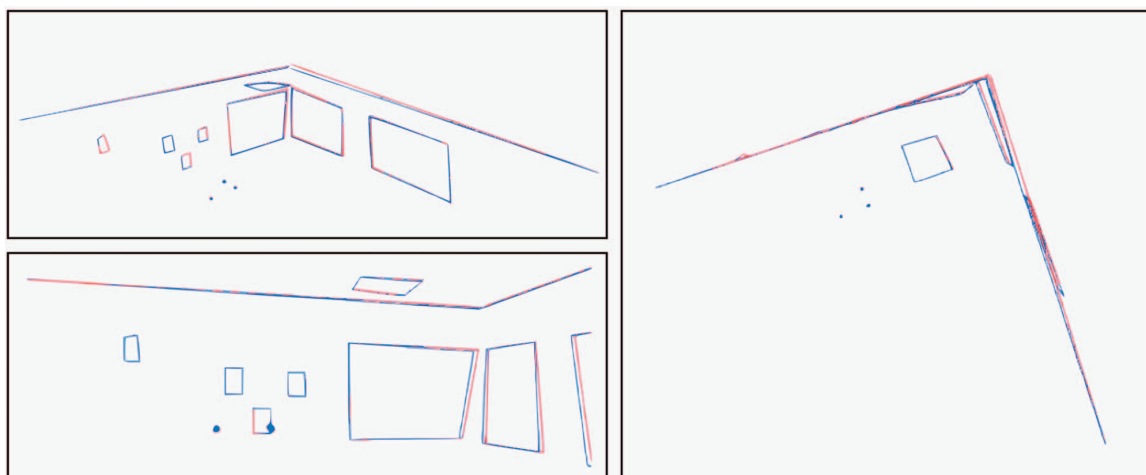


Fig. 13 Results of the three-dimensional reconstruction. Reconstruction according to the linear-eigen and midpoint methods are respectively drawn in red and blue, while the gray points represent the sensor's 3 positions.

preliminary results on a calibrated spherical mirror show the good accuracy of the three-dimensional parameters measurement by polarization imaging. Experiments on the reconstruction of a real scene are also presented.

4.1 Simulations

To simulate the three-dimensional reconstruction error, the synthetic scene introduced in Section 2 (Fig. 1) is reconstructed thanks to the generic calibration concept. The normals' angles θ and ϕ were computed from a perfect parabolic mirror with a 7° misalignment between the optical axis and the symmetry axis of the mirror. The parameters are then disturbed by adding various levels of Gaussian noise. At each level of noise, the experiment is regenerated 50 times, and the average error reconstruction is computed. Figures 7(a) and 7(b), respectively, show the reconstruction error of the scene induced by noisy measurement of θ and ϕ . The synthetic scene is reconstructed with or without mirror reconstruction, depending on whether or not the integration process from Eq. (6) is carried out (Fig. 8).

Figure 7 shows that scene reconstruction is quite sensitive to the measurement of the parameters θ and ϕ ; on the other hand, the integration process is not required, and we can assume that the z parameter is negligible. In addition, Fig. 9 shows the reconstruction error of the scene by only adding Gaussian noise to the mirror height z . In this case, the normals are assumed to be perfectly measured and noise is only added on the surface height of the mirror. The reconstruction quality remains good even if the mirror height is very noisy (the mirror height is 1 cm and the radius is 2 cm) or badly reconstructed.

4.2 Preliminary Results

4.2.1 Simulated scene

Preliminary results were carried out with a catadioptric sensor made of a camera with a telecentric lens and a calibrated spherical mirror (radius=1 cm). Let us notice that our system does not satisfy the single-viewpoint constraint. Nevertheless, this property is not required here for the three-dimensional reconstruction of the scene. As described in Section 3, our catadioptric sensor is calibrated by measuring the three-dimensional parameters of the mirror with a liquid crystal polarization rotator placed between the camera and the mirror. To evaluate the accuracy of our system, we compare the parameters (θ , ϕ , and z) obtained with our system to the theoretical parameters of the mirror (Fig. 10).

The mean quadratic errors of the angles θ and ϕ are, respectively, 0.49° and 1.02° . Figure 11 shows the reconstruction of the synthetic scene by using the calibration made by polarization imaging. Since the mirror is spherical, three-dimensional reconstruction errors greatly increase. Nevertheless, the synthetic scene is well reconstructed, with an average error of 9.68 cm.

4.2.2 Real scene

As presented in Fig. 12, the three-dimensional reconstruction of a real scene was also carried out. The imaging system is made of a CCD camera with a telecentric lens and a spherical mirror. In order to precisely control its displacement, it has been placed on a precise three-axes stage. The

$x \times y \times z$ operating area is $1.5 \text{ m} \times 1.6 \text{ m} \times 0.4 \text{ m}$, and the $x \times y \times z$ room size to reconstruct is about $6 \text{ m} \times 5 \text{ m} \times 2.5 \text{ m}$. After calibrating our catadioptric sensor by polarization imaging, 3D points of interest have been triangulated by moving the system.

Once the imaging system is automatically calibrated by polarization imaging, three images with three different known positions have been acquired. Several points of interest were picked on each image: four rectangular targets, two edges of the ceiling, three windows, and one fluorescent light [Figs. 12(b)–12(d)]. The result of the triangulation process presented in Fig. 13 shows that the reconstruction is qualitatively satisfactory. The global topology of the scene and the relative distances are respected; right angles appear as (near-) right angles. In fact, the overall reconstruction is as expected, considering

- the small displacement of the sensor in comparison with the size of the scene,
- the fact that we only performed linear triangulation without refinement, and
- the reconstruction also depends on the accuracy of point picking (the correspondence has been established manually and a slight error in the location of 2D points led to erroneous 3D reconstruction).

5 Conclusion

We presented, a new efficient calibration method for catadioptric sensors. This method is based on the measurement of three-dimensional parameters of the mirror thanks to polarization imaging. The calibration can be performed “in one click” even by a nonspecialist, because it only requires an optical apparatus—no image processing and no calibration pattern. Contrary to traditional methods, it deals with misalignment of the sensor and works for any shape mirror (regular or not). Experimental results prove that the sensor is properly calibrated and a satisfactory three-dimensional reconstruction of the scene can be obtained. We have also shown that the 3D shape of the mirror can be neglected in comparison with the normals' orientations. The Platform3D Department is currently manufacturing a parabolic mirror, and future work will consist of creating a paracatadioptric sensor in order to compare our method to other methods known in the literature for real scenes.

References

1. S. Baker and S. K. Nayar, “A theory of catadioptric image formation,” in *Proc. Int. Conf. Comp. Vis.*, pp. 35–42 (1998).
2. S. Ramalingam, S. K. Lodha, and P. Sturm, “A generic structure-from-motion framework,” *Comput. Vis. Image Underst.* **103**, 218–228 (2006).
3. P. Sturm, “Multi-view geometry for general camera models,” *IEEE Comp. Vis. Patt. Recog.* **1**, 206–212 (2005).
4. S. Kang, “Catadioptric self-calibration,” *IEEE Comp. Vis. Patt. Recog.* 1201–1207 (2000).
5. Y. Yagi, S. Kawato, and S. Tsuji, “Real-time omnidirectional image sensor (copis) for vision-guided navigation,” *IEEE Trans. Rob. Autom.* **10**, 11–22 (Feb. 1994).
6. J. Fabrizio, J.-P. Tarel, and R. Benosman, “Calibration of panoramic catadioptric sensors made easier,” *Omnivis* **00**, 45 (2002).
7. C. Geyer and K. Daniilidis, “Structure and motion from uncalibrated catadioptric views,” *IEEE Comp. Vis. Patt. Recog.* **1**, 279–293 (2001).
8. C. Geyer and K. Daniilidis, “Paracatadioptric camera calibration,” *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(5), 687–695 (2002).
9. C. Geyer and K. Daniilidis, “Catadioptric camera calibration,” in *Proc. Int. Conf. Comp. Vis.*, p. 398 (1999).

10. J. Barreto and H. Araujo, "Geometric properties of central catadioptric line images," in *Proc. Eu. Conf. Comp. Vis.* (2002).
11. X. Ying and Z. Hu, "Catadioptric camera calibration using geometric invariants," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 1260–1271 (2004).
12. M. D. Grossberg and S. K. Nayar, "A general imaging model and a method for finding its parameters," in *Proc. Int. Conf. Comp. Vis.* **2**, 108–115 (2001).
13. D. Aliaga, "Accurate catadioptric calibration for real-time pose estimation of room-size environments," in *Proc. Int. Conf. Comp. Vis.*, pp. 127–134 (2001).
14. R. Swaminathan, M. D. Grossberg, and S. K. Nayar, "Caustics of catadioptric cameras," in *Proc. Int. Conf. Comp. Vis.*, p. 2 (2001).
15. C. Mei and P. Rives, "Single view point omnidirectional camera calibration from planar grids," in *Proc. IEEE Int. Conf. Robo. Automat.* (2007).
16. S. Ramalingam, P. Sturm, and S. Lodha, "Theory and calibration algorithms for axial cameras," in *Proc. Asian Conf. Comp. Vis.* **1**, 704–713 (2006).
17. J. P. Barreto and H. Araujo, "Paracatadioptric camera calibration using lines," in *Proc. Int. Conf. Comp. Vis.*, pp. 1359–1365 (2003).
18. B. Vanderportaele, M. Cattoen, P. Marthon, and P. Gurdjos, "A new linear calibration method for paracatadioptric cameras," in *Proc. Int. Conf. Patt. Recog.* **4**, 647–651 (2006).
19. R. Hartley and P. Sturm, "Triangulation," *Comput. Vis. Image Underst.* **68**(2), 146–157 (1997).
20. J.-J. Gonzalez-Barbosa, "Vision panoramique pour la robotique mobile: stéréovision et localisation par indexation d'images," Ph.D. thesis, Université Toulouse III (2004).
21. J.-J. Gonzalez-Barbosa and S. Lacroix, "Fast dense panoramic stereo-vision," in *Proc. IEEE Int. Conf. Robo. Automat.*, pp. 1210–1215, 18–22 April 2005.
22. S. Ieng and R. Benosman, "Geometric construction of the caustic curves for catadioptric sensors," in *Proc. IEEE Int. Conf. Image Proc.*, pp. 3387–3390 (2004).
23. R. Pless, "Using many cameras as one," *IEEE Comp. Vis. Patt. Recog.* **2**, 587–593 (2003).
24. L. B. Wolff, "Polarization-based material classification from specular reflection," *IEEE Trans. Pattern Anal. Mach. Intell.* **12**, 1059–1071 (1990).
25. D. Miyazaki, M. Kagesawa, and K. Ikeuchi, "Transparent surface modeling from a pair of polarization images," *IEEE Trans. Pattern Anal. Mach. Intell.* **26**, 73–82 (2004).
26. S. Rahmann, "Reconstruction of quadrics from two polarization views," in *Proc. Iberian Conf. Patt. Recog. Image Analysis (Ib-PRIA03)*, Springer, LNCS 2652, pp. 810–820 (2003).
27. L. B. Wolff and T. E. Boulton, "Constraining object features using a polarization reflectance model," *IEEE Trans. Pattern Anal. Mach. Intell.* **13**, 635–657 (1991).
28. O. Morel, C. Stolz, F. Meriaudeau, and P. Gorria, "Three-dimensional inspection of highly-reflective metallic objects by polarization imaging," *Electron. Imag. News.* **15**(2), 4 (2005).
29. O. Morel, C. Stolz, F. Meriaudeau, and P. Gorria, "Active lighting applied to 3D reconstruction of specular metallic surfaces by polarization imaging," *Appl. Opt.* **45**, 4062–4068 (June 2006).
30. P. Sturm and S. Ramalingam, "A generic concept for camera calibration," in *Proc. Eur. Conf. Comp. Vis.*, **2**, 1–13 May (2004).
31. R. Frankot and R. Chellappa, "A method for enforcing integrability in shape from shading algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.* **10**, 439–451 (July 1988).
32. Platform3D Department, "3-dimensional solid modeling and prototyping," <http://www.platform3d.com>.



Olivier Morel received his MSc degree in computer vision and image processing from the University of Burgundy in 2002. In November 2005, he received his PhD degree in computer vision from the University of Burgundy. Since September 2007, he has worked as a lecturer in the 3D vision team of the Le2i Lab. His main research interests are polarization imaging, 3D vision and catadioptric vision.



Ralph Seulin received his PhD degree in computer vision from the University of Burgundy in 2002. He is a research engineer at the National Center for Scientific Research (CNRS) in the Le2i Laboratory. His main research interests include machine vision for industrial parts inspection and 3D digitization automation. He also manages industrial projects for machine vision applications in the nuclear industry.

David Fofi received his MSc degree in computer science (computer vision, signal and image processing) in the University of Cergy-Pontoise, ENSEA, in 1997. In September 2001, he received his PhD degree in robotics and computer vision from the University of Picardie Jules Verne. Since September 2002, he works as a lecturer in the Le2i Lab where he became head of the 3D vision team in 2007. His main research interests are 3D vision, projector-camera systems, structured light and catadioptric vision.

Registration of surfaces minimizing error propagation for a one-shot multi-slit hand-held scanner[☆]

C. Matabosch^{a,*}, D. Fofi^b, J. Salvi^a, E. Batlle^a

^a*Institute of Informatics and Applications of the University of Girona, Spain*

^b*University of Burgundy, France*

Received 17 April 2007; received in revised form 20 September 2007; accepted 13 October 2007

Abstract

We propose an algorithm for the on-line automatic registration of multiple 3D surfaces acquired in a sequence by a new hand-held laser scanner. The laser emitter is coupled with an optical lens that spreads the light forming 19 parallel slits that are projected to the scene and acquired with subpixel accuracy by a camera. Splines are used to interpolate the acquired profiles to increase the sample of points and Delaunay triangulation is used to obtain the normal vectors at every point. A point-to-plane pair-wise registration method is proposed to align the surfaces in pairs while they are acquired, conforming paths and eventually cycles that are minimized once detected. The algorithm is specially designed for on-line applications and can be classified as a closing-the-loop technique, where there are not that many competing methods, though it has been compared to the literature. Experiments providing qualitative and quantitative evaluation are shown by means of synthetic and real data and we demonstrated the reliability of our technique.

© 2007 Elsevier Ltd. All rights reserved.

Keywords: Computer vision; 3D registration; Multi-view; Closing-the-loop

1. Introduction

The complete 3D acquisition of a given object, surface or even scene has many research challenges and it is definitely in the research interests of the computer vision community. Besides, there are several applications which may benefit by such complete acquisition such as reverse engineering, modeling, metrology, visual inspection and even robot navigation.

Overall, there are diverse 3D acquisition systems which are basically based on laser triangulation [1,2] and pattern projection [3], especially when dense images are required. These systems gather range images from which 3D information can be extracted. Besides, other acquisition systems are based on processing one or many 2D images captured by cameras. One of

these techniques is *Shape from Silhouettes*, in which several images of the measuring object are acquired and the complete 3D model of the object is computed by means of its silhouette at every image. Although some authors tend to use this technique without knowing the camera poses from where images were acquired [4], such positions are usually required and only a rough accuracy is obtained. Other techniques are based on directly processing the set of images by means of solving the matching problem [5]. In this case, features such as points/lines/regions of interest must be determined within the images reducing drastically the resolution and hence obtaining a sparse acquisition. In conclusion, laser triangulation is considered one of the most reliable techniques of acquiring 3D data, so that data is dense and accurate and the correspondence problem alleviated.

In general, laser triangulation techniques are based on the use of a laser emitter coupled to a cylindrical lens that spreads the light forming a plane that is projected to the measured surface. The projection of a laser plane only lets us acquire a profile of the measuring surface. Hence, in most cases a mechanical system is added to permit the scanning, so that: (a) the laser plane is projected onto a rotating mirror and reflected towards the

[☆] This research has been partly supported by Spanish Project MCYT DPI2007-66796-C03-02.

* Corresponding author. Tel.: +34 972 419812; fax: +34 972 418976.

E-mail addresses: cmatabos@eia.udg.es (C. Matabosch), d.fofi@iutlecreusot.u-bourgogne.fr (D. Fofi), qsalvi@eia.udg.es (J. Salvi), bbatlle@eia.udg.es (E. Batlle).

surface; (b) the laser beam is attached to a moving worm gear; (c) the laser beam keeps motionless while the object is placed on a rotating table, obtaining dense acquisitions. However, the accuracy of the 3D acquisition depends on the mechanical system as potential vibrations are likely to produce misalignments. Furthermore, the sequence of images that are captured in the scanning process forces the object to be motion controlled reducing the number of potential applications. Finally, an uncompleted acquisition of the object is usually obtained due to object occlusions and the limited field of view of the sensor.

Summarizing, laser triangulation is a reliable technique to acquire dense and accurate 3D data. However, existing commercial sensors usually only acquire a partial view of the object, others are constraint to mechanical structures such as rotating tables and moving gears. There are few sensors that can compute their ego-pose in a free space. For instance, the ZScanner700 of ZCorporation [6] can compute such ego-pose, but a set of reflective targets must be stuck on the measuring surface. The set of views are directly aligned once the ego-pose is computed. Registration is a technique that can permit to perform such alignment without adding markers or any other reference to the object or to the measuring scenario.

Pair-wise registration is a well-studied problem in the literature. It is known that the published techniques can be classified in coarse and fine registration techniques depending on the way of solving the registration: (a) in a closed-form solution from a reduced set of points of interest obtained from surface features and solving a global matching; (b) in an iterative minimization solution using a sample of points and solving the matching locally. Multi-view registration is a more difficult problem and there are two strategies to solve the alignment: local (sequential) and global (multi-view). The sequential registration of views does not give an optimal solution due to the accumulation and propagation of errors. Besides, global registration tries to distribute the registration errors evenly among all the views but it is not suitable for on-line registration since all the views are first required to initiate the registration. In this paper, we propose a closing-the-loop technique for the on-line registration of sequence of views in which cycles are minimized once detected preserving an optimal solution. The pair-wise registration is solved by using a variant of the point-to-plane technique and cycles are minimized considering only the views involved constricting the computing time.

The remaining of the paper is structured as follows. First, a brief overview of registration techniques is presented in Section 2, discussing the pros and cons of the existing methods with the aim of justifying our proposal. Then, Section 3 presents our proposal including pair-wise registration, cycle detection and cycle minimization. Experimental results provided by both synthetic and real data are presented in Section 4. The article ends with conclusions.

2. Overview of range image registration techniques

According to a recent study of the state-of-art concerning surface registration techniques [7], all the existing methods are classified in two main groups of techniques: (a) Coarse

Registration and (b) Fine Registration. In the following part, pros and cons of these techniques are described and every technique is summarized.

Coarse registration techniques obtain a rough alignment by means of finding correspondences between two surfaces. Most coarse registration techniques are based on searching for points (curves) in the second surface that are similar to points (curves) in the first surface. Overall, there are only two different methods to choose the interest points: (a) feature-to-points and (b) point-to-features. The first selects points in the first surface that are similar to a predefined feature [8,9]. In the second, some points in the first surface are arbitrarily selected and characterized considering the position of its neighbors and searched in the second surface [10–12]. In both techniques, all points in the second surface must be compared with the selected points to establish correspondences. Once the correspondence problem is solved, the Euclidean motion that aligns both surfaces is computed in a closed-form solution.

Coarse registration techniques have two main drawbacks. The long time required to solve the matching among points and the poor quality of the registration. Hence, a fine registration technique is usually applied *a posteriori* to improve the results by minimization.

Fine registration techniques search for an accurate alignment of two acquired surfaces by minimizing the distance between both surfaces iteratively. Such distance is computed in each iteration by means of temporal matching among points from both surfaces. Overall, the existing techniques solve the temporal matching using one of the following three approaches: (1) point-to-point, in which correspondences are established by searching for the points in the second surface that are closest to a set of points in the first [13]; (2) point-to-plane, in which the points in the second surface are computed at every step by the intersection of a plane and a line. The line is defined by a point and a vector, so that the point is a given point in the first surface; and the vector is orthogonal to the surface defined by a neighborhood around the given point. The plane is located where that line intersects with the second surface and it is oriented in such a way that the plane is tangent to that surface [14]; and finally (3) point-to-projection, in which correspondences are established between points in the first surface and the points obtained by projecting the first points onto the second surface using the point of view of the second surface [15].

Although point-to-projection is the fastest technique because searching is avoided, results obtained are not very satisfactory. Besides, point-to-plane provides the best results because it is not influenced by local minima [12,16]. Despite the difficulty in determining the intersection of a line with a cloud of points in 3D, several authors have presented proposals to facilitate this computation [17,18].

One-to-one alignment of views in a sequence causes a drift that is propagated throughout the sequence. Hence, some techniques have been proposed to reduce the propagating error benefiting from the existence of cycles and re-visited regions and considering the uncertainty in the alignment.

In order to minimize the propagating error, some authors have improved their algorithms by adding a final step that aligns all

the acquired views at the same time. This approach spreads one-to-one pair-wise registration errors throughout the sequence of views, known as multi-view registration [19].

Early approaches proposed the aggregation of subsequent views in a single metaview which is progressively enlarged each time another view is registered [14]. Here, the main constraint is the lack of flexibility to re-register views already merged in the metaview due to the greedy approach of the technique. In 1999, Pulli proposed an ICP¹ relaxation method based on the previous metaview approach but considering all the potential alignments between views before proceeding with the multi-view registration. In addition this method takes into account the information of all the overlapping areas and the already registered regions can be analyzed again for further transformations [20]. Later on, Nüchter proposed a global relaxation method based on Pulli's approach with the main difference that no iterative pair-wise alignment is required. However the success of this method drastically depends on the disposal of an accurate initial estimation of the pose [21].

A different approach was proposed by Bergevin [22], who presented a multi-view registration technique based on the graph theory: views are associated to nodes and transformations to edges. Authors consider all views as a whole and align all of them simultaneously. The same idea was proposed later on by Silva [23], Huber [24] and Krishnan [25]. Besides, Masuda presented a multi-view registration algorithm based on the Matching Signed Distance Fields in which outliers are automatically removed obtaining a more robust method [26]. Lu's technique is based on cycle minimization, though the relationship (edges) among views (nodes) are established prior to minimization [27]. Overall, multi-view techniques suffer two main drawbacks: (a) the whole set of 3D views has to be acquired before the algorithm starts; (b) an accurate estimation of the motion between views is needed as initial guesses to ensure convergence. Thus, multi-view techniques are not considered for on-line applications.

Few authors have faced the challenge of registering 3D views in a sequence while they are acquired avoiding or at least controlling error propagation. For instance, Sharp [28] proposed the registration of pairs of consecutive views until a cycle is found. Since only pair-wise registration is required, the method becomes very fast. Here, the interest is in the way of distributing the motion (and hence the propagation error) among the different views. The author proposed to use weights directly related to the residue obtained in the pair-wise registration. Actually, this is not very accurate especially in the presence of misalignments between end views in the cycle as a matter of noise and object occlusions. In this case, the whole motion of such a cycle is also distributed to all the views increasing the error in the registration.

Finally, in the last few years, a photogrammetric technique called Bundle Adjustment has increased popularity in the computer vision community and it is growing in interest in robotics. Bundle adjustment is the problem of refining a visual reconstruction to produce jointly optimal 3D structures and

viewing parameter (camera pose and/or calibration) estimates [29]. Therefore, bundle adjustment techniques can be used in both robot/camera localization and 3D mapping in many fields such as camera calibration, robot navigation, and scene reconstruction. Since bundle adjustment is a non-linear minimization problem, it is solved by means of iterative non-linear least squares or total squares methods such as Levenberg–Marquardt or M-estimator techniques [7,30]. Although bundle adjustment is commonly classified as a multi-view technique, some authors have used it in consecutive pair-wise alignment as a technique to reduce error propagation [31].

In summary, we conclude that methods based on the metaview approaches present good results when initial guesses are accurate and the surface to be registered does not have a large scale. Otherwise, the method suffers a large propagation error producing drift and misalignments and its greedy approach usually falls in local minima. The use of methods based on graphs has the advantage of minimizing the error in all the views simultaneously but these techniques usually require a previous pair-wise registration step, whose accuracy can be determinant in the global minimization process. Besides, closing the loop strategies provide trustworthy constraints for error minimization but require a huge amount of memory and usually involve a high computational cost. Bundle adjustment techniques provide good results in the presence of outliers, but need a good enough initial guess and it is hardly used in large robot missions or large scale objects.

All these pros and cons of the existing methods have been considered to present a new surface registration technique which is presented and discussed in the rest of the paper.

3. Registering a sequence of surfaces

This section describes the proposing method for continuously registering a sequence of 3D views while they are acquired. The method first aligns the consecutive views by means of point-to-plane pair-wise registration. When a cycle is detected, a multi-view technique is applied only in the views conforming to the cycle leading to fast and accurate results and preserving the on-line registration for many and varied applications (see Fig. 1).

3.1. Pair-wise registration

Pair-wise registration is divided into a first coarse registration to estimate an initial alignment, followed by a fine registration computed by means of minimization techniques. In our case, views are acquired consecutively and a slight movement between views is assumed. Slight movement is defined like the movement that guarantees at least a 60% of overlapping with consecutive views. We initialize fine registration considering motionless views, avoiding the expensive computation required to compute initial guesses and preserving a high accuracy, as demonstrated in the following paragraphs and shown in the experimental results.

Point-to-plane has been chosen as the most suitable fine registration technique as discussed in the previous section. The technique we propose is based on the fast variant proposed

¹ Iterative closest point.

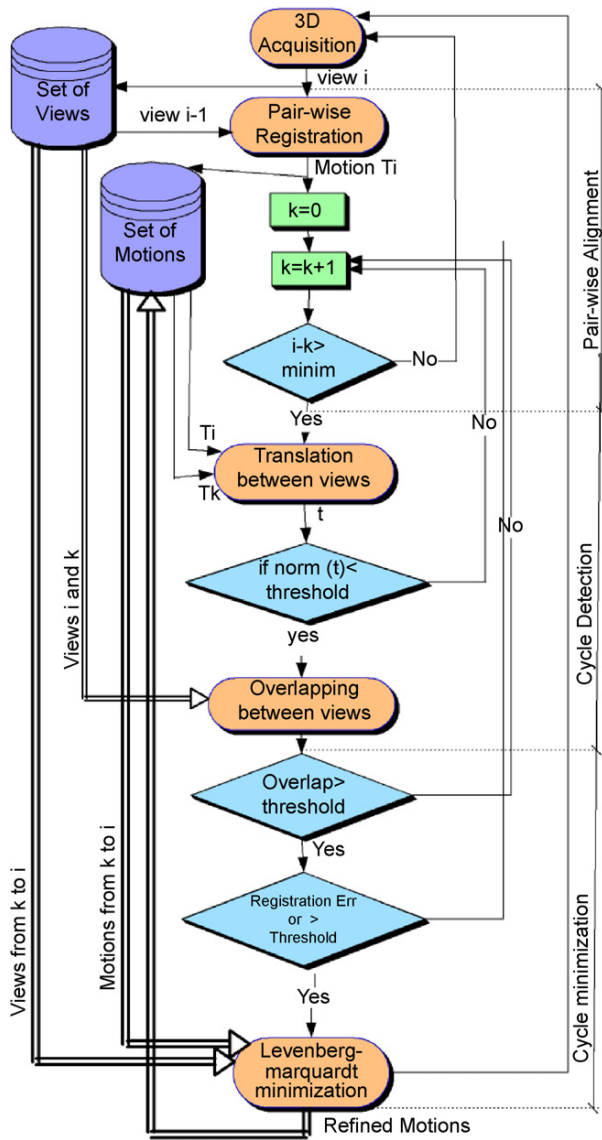


Fig. 1. Flow diagram of the proposed method.

by Park [18] from the original point-to-plane registration proposed by Chen [14], although some modifications have been implemented to increase accuracy, which are explained in the following paragraph.

First, we remove the non-overlapping area of the present view before this view is registered with the former. In theory, this area is unknown because the movement is also unknown. However, as the views are taken in a sequence with slight movements between them, we can assume that points located in the center of the view are good candidates for the matching. Besides, most of the points located in the boundary of the surface might be hardly matched. In consequence, the boundary area of the present view is not considered in the fine registration step. In fact, the bounding area coincides with the boundary in the image formed by projecting the present view to the XY plane of the camera (orthogonal to the focal axis), so the selection of

points to remove becomes very easy. In the image plane, the bounding box is computed. A rectangle whose dimensions are 80% of the bounding box is centered to the image projection and all points out of this rectangle are not taken into account in the registration step.

Second, only a sample of the remaining points of the present view is preserved for the fine registration. There are several types of sampling: uniform sampling [26,32], random sampling [33], and normal sampling [16], among others. Although sampling is normally used to speed up the algorithm by selecting a reduced set of points, sampling can be also used to increase accuracy by selecting also the most appropriate points. Note that, in smooth surfaces with even shape registration becomes difficult. In this situation, only a small percentage of points give useful shape information. For instance, consider a flat surface with two perpendicular cuts. If all the points are considered in the registration, results are not accurate because of the low influence of points in cuts with respect to the rest of the points. However, if the registration is done with a high percentage of points on the uneven area, accuracy increases (see Fig. 2).

The goal of normal sampling is to select the most representative points to increase the quality of the registration. Hence, all points are first transformed to a 2D normal space defined by α and β as follows:

$$\alpha = a \tan 2 \left(n_x, \sqrt{n_z^2 + n_y^2} \right),$$

$$\beta = a \tan 2(n_y, n_z), \quad (1)$$

where α and β are the coordinates in the normal space, and n_x , n_y and n_z are the three components of the normal vector of each point. Then, every point is placed in a 2D grid. Finally only one point from every grid cell is randomly selected, so that a single point is chosen among all points with similar normal vectors. These selected points actually conform to the reduced set of points used to register the present surface. The percentage of points that remains after the sampling step depends a lot on the unevenness of such surface but tends to be between a 20% and a 30% of the total points.

As stated before, the fine registration technique we propose is based on the fast variant proposed by Park [18] from the original point-to-plane registration proposed by Chen [14]. Here we use a recursive method to compute the intersection between lines and surfaces which is actually the main difficulty of the method. Hence, initially the selected points of the previous view are projected orthographically onto the XY plane of the camera. A grid composed of 50×50 square cells is scaled so that it contains the projection of all points. Second, a point p_0 of the current view is projected to such a grid, in whose cell we search for the closest point obtaining the point q_{p_0} in the previous surface. The projection of point q_{p_0} to the normal vector of p_0 defines a new point p_1 , which is actually an approximation of the intersection. This approximation is refined recursively by projecting new points p_i until $\text{norm}(p_i - q_{p_i})$ is smaller than a threshold (see Fig. 3). Finally, the process is repeated for all the points conforming to the current view and a set of correspondences is obtained.

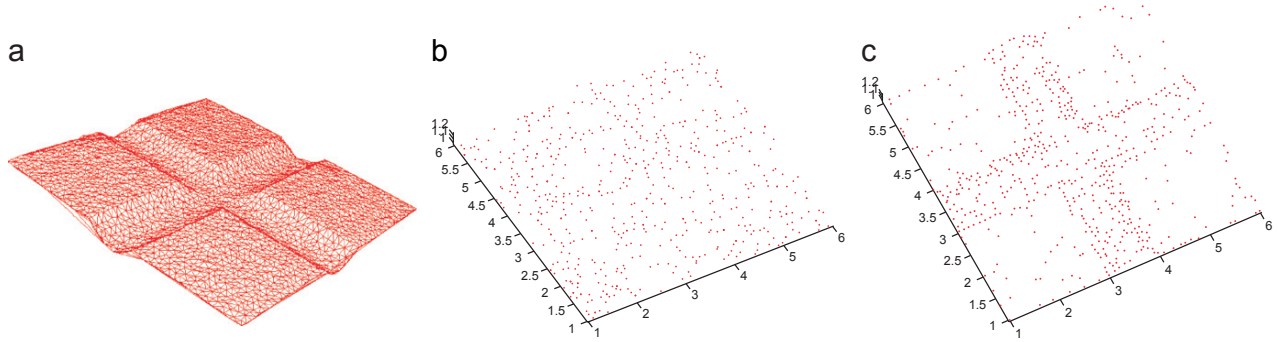


Fig. 2. Effects of sampling: (a) Original surface; (b) Random sampling; (c) Normal sampling.

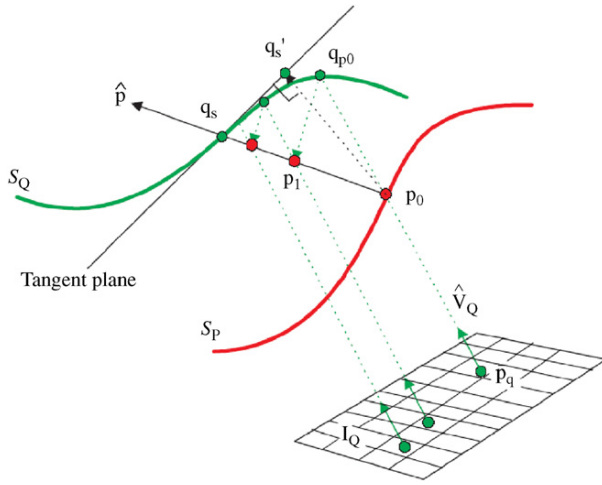


Fig. 3. Strategy used to compute the intersection between the tangent plane and the surface S_q along the orthogonal vector \hat{p} . See Park [18] for an extended review.

Once correspondences are established, minimization is applied to compute the motion between both surfaces (the previous and the current) as defined by Eq. (2).

$$f = \frac{1}{N_p} \sum_{i=1}^{N_p} \|m_i - R p_i - t\|^2, \quad (2)$$

where N_p is the number of correspondences; m_i is the set of points selected in the former view that have a correspondence in the present view; p_i are the correspondences of m_i in the present view; and R and t are the rotation matrix and the translation vector that aligns both views, respectively.

Eq. (2) is minimized by means of quaternions [13] so that R and t are refined iteratively. In each iteration, the correspondences must be recomputed because initial correspondences are not usually correct. The algorithm stops when: the mean of the square errors (distances between correspondences) is smaller than a given threshold; or the mean of the square errors does not decrease.

Note that the views are registered consecutively, so that every registered view is referenced with respect to the first by

means of the product of all the consecutive Euclidean motions defined by the sequence of views. Hence, registration inaccuracies are propagated through the sequence. In the following sections, we aim to minimize the propagation error by detecting cycles and minimizing the views conforming the cycle all together.

3.2. Cycle detection

Now the interest is to detect every time the scanner re-visits the same object surface obtaining cycles of views that are used to reduce the propagation error significantly.

Cycle detection complexity varies depending on whether views are unorganized or views are acquired sequentially. In the former, the relationship among views is unknown and each view has to be registered to all the others to detect potential links [34]. The problem is simplified when such relationship is previously known [27]. However, in the latter, error propagation requires a robust cycle detection method. Nüchter proposed the registration of the last view to some of the already acquired views to detect such cycles [35], in which such views are selected based on some hypotheses concerning the laser range and the sensor pose.

Note that once any two views are registered, the Euclidean transformation between them is known and a link established. These links form paths through the views in which the motion of the scanner can be estimated from the product of the consecutive Euclidean transformations. Hence, the translation vector of such movement is considered, so that if this vector is smaller than a threshold and the views are not neighbors, a potential cycle is considered. The given threshold is computed dynamically considering the object scale and the number of views forming the potential cycle, so that the threshold increases proportional to the propagation error.

However, a sequence of views with an overall slight translation does not always conform a cycle, especially when rotation is relevant. Hence, the total rotation within the path is computed, so that such rotation has to be close to identity to consider the path like a cycle.

Finally, a cycle is detected if both end views also share a common surface, that is a significant overlapping area.

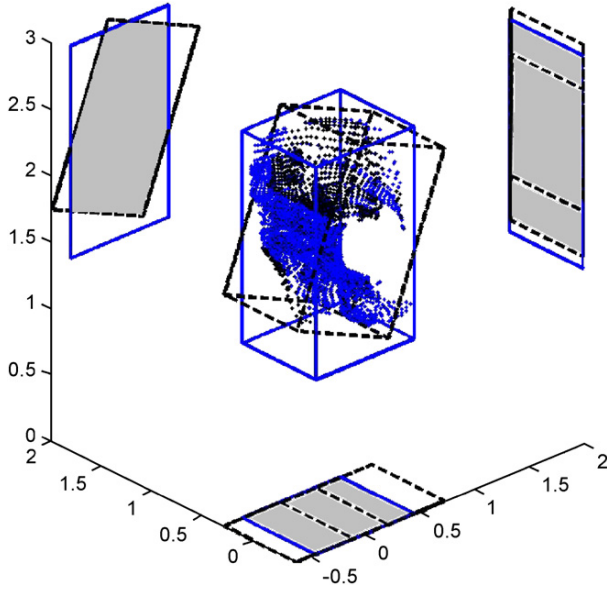


Fig. 4. Example of the projection of the bounding boxes of two different views in the XY, XZ and YZ planes. The grey area represents the overlapping.

The accurate computation of the percentage of overlapping would imply the fine registration between both end views and the computation of corresponding points. In order to avoid this expensive step, a fast technique is proposed based on the overlapping of bounding boxes, which is just an approximation of the convex hull of both surfaces, but accurate enough to detect cycles.

The bounding box of a given surface is defined as the minimum parallelepiped that contains all the points of the surface. The intersection of 3D bounding boxes is complex so that it is alleviated by projecting such boxes to the planes XY, XZ and YZ (see Fig. 4), defining two 2D bounding boxes in every plane and thus computing three overlapping areas. If the maximum of the three overlapping areas exceeds a given threshold of the total area and the distance between both bounding box centers is small enough, a cycle is considered.

The reason to choose the maximum overlapping value among the three planes instead of the product of overlapping values is in virtue of preserving the detection of potential cycles in the presence of almost flat surfaces. In this case, the bounding boxes in some of the three planes are usually not relevant.

3.3. Cycle minimization

Cycle minimization consists of a simultaneous minimization of all the correspondences between points of all the views that conform the cycle. In cycle minimization we assume that the overall motion in the cycle is null and hence the position of both end views coincides. This is actually impossible and that is the reason why a virtual view is added between both end views. This virtual view is nothing other than the first view of the cycle registered to the last one. We can assume that the overall motion in the cycle is null which means that the motion between both end views must be zero.

The significant points for every view are used to search for correspondences among all the other views in the cycle by using again our variant of the point-to-plane registration technique. This technique is based on the iterative minimization of the distances between temporal correspondences. However, at last iteration, temporal correspondences can be assumed to be accurate correspondences.

A threshold in the relative motion between views is used to ensure a significant overlapping area between views and hence many point correspondences. Obviously, this decision leads to a quite fast method without losing robustness. Otherwise, the algorithm wasted a lot of time searching for correspondences where it was known they are either not available or not significant.

Finally, a Levenberg–Marquardt minimization is applied to determine a more accurate registration among views in the cycle. The minimizing parameters are the rotation matrices (represented as quaternion vectors) and translation vectors of the Euclidean transformations between consecutive views. The minimizing function is the sum of distances between point correspondences because the distance between point correspondences should be minimized, and hopefully may reach zero, as shown in the following equation:

$$\min \left\{ \sum_{i=1}^{N-1} \sum_{j=i+1}^N \sum_{k=1}^{N_p} \|P_i(k) - T_j^i \times P_j(k)\| + \|T_i^j \times P_i(k) - P_j(k)\| \right\}, \quad (3)$$

where $P_i(k)$ and $P_j(k)$ are the points that configure the k correspondence between views i and j ; N_p is the number of points correspondences; N is the number of views; and T_i^j and T_j^i are the Euclidean motions that transform points from i to j and from j to i , respectively, computed as follows:

$$T_j^i = \prod_{k=i+1}^j T_k^{k-1}, \quad (4)$$

and

$$T_i^j = \left(\prod_{k=j}^{N-1} T_{k+1}^k \right) T_1^N \left(\prod_{k=2}^i T_k^{k-1} \right), \quad (5)$$

where $j > i$.

The closing-the-loop constraint ε_{cr} is added to the optimization function in Eq. (3), where

$$\varepsilon_{cr} = \varepsilon_R + s_f \varepsilon_T, \quad (6)$$

where ε_R is the rotation constraint; ε_T is the translation constraint; and s_f is the scale factor that weights the translation constraint to be adequately compared to the rotation constraint.

The rotation constraint ε_R is

$$\varepsilon_R = \text{sum}(\text{abs}(R_{\text{accum}} - I_{3 \times 3})), \quad (7)$$

where R_{accum} is the product of all the rotation matrices conforming the cycle, and $I_{3 \times 3}$ is the identity matrix.

The translation constraint ε_T is

$$\varepsilon_T = \text{norm}(t), \quad (8)$$

where t is the translation vector between both initial and end views of the cycle, computed as follows:

$$\begin{bmatrix} R & t \\ 0 & 1 \end{bmatrix} = T_{\text{cycle}} = \left(\prod_{i=2}^n T_i^{i-1} \right) \cdot T_n^1. \quad (9)$$

The whole process leads to quite accurate results, but if they are not good enough, they can be repeatedly refined by selecting new significant point correspondences at the end of every refinement.

4. Experimental results

The proposed method has been implemented and compared to one of the most similar methods present in the literature, the method proposed by Sharp et al. in 2004 [28]. Both methods have been tested under the same conditions in order to evaluate their advantages and constraints. So, we have used the same point-to-plane method to determine the motion between consecutive views.

A low-cost multi-slit laser acquisition system has been used. The set-up is composed of an off-the-self camera, a 635 nm laser emitter and an optical lens that spreads the laser beam conforming 19 parallel planes. The depth field of the system is in the range of 100 to 300 mm in the Z-axis of the camera, which is limited due to system baseline, laser power, and camera focus. The camera and the laser emitter conform the one-shot hand-held 3D acquisition system especially developed to test our proposal. The 19 parallel planes are projected onto the measuring surface obtaining 19 profiles with subpixel accuracy. 3D profiles are obtained by triangulation and finally splines are used to interpolate a surface, increasing the sample of points used in the registration process. Once the 3D surface is acquired, the Delaunay triangulation is applied to obtain triangles and hence estimate the normal vector at every surface point which is used further on in the pair-wise alignment. Fig. 5 shows an acquired image and the process of spline interpolation. The reader is pointed to Matabosch [36] for more details about the acquisition sensor.

The performance of our method has been compared to the method of Sharp [28] both quantitatively and qualitatively. Our cycle detection method has been used in both methods, so they are tested in the presence of the same number of cycles. Note that in the method of Sharp cycles are detected manually. Experiments and results are presented in the following paragraphs.

4.1. Quantitative evaluation

Quantitative evaluation is analyzed from both synthetic and real data. Synthetic data is obtained from the 3D synthetic models courtesy of INRIA.² A synthetic scanner has been programmed to simulate the acquisition of a set of consecutive views acquired by the one-shot hand-held scanner (see Fig. 6). Here, the pose of the scanner for every acquisition is obviously given by the simulator. So, the accuracy of the registration can be precisely evaluated. The experiment is repeated adding some Gaussian noise to the 3D points. In addition, a real object has been placed on a motion-controlled table in which our one-shot hand-held scanner has been attached (see Fig. 7). The object is moved in three degrees of freedom (X , Y and Z) so that 29 consecutive views are acquired. In this case, the position of the sensor is given by the mechanics of the table.

Parameters and thresholds used during this experiment are reported in Table 1. Their value differs depending on the size of the measuring object. NSS_{grids} specifies the number of cells used in the normal space sampling grid. $Intersection_{\text{error}}$ is the threshold used in the point-to-plane registration and it corresponds to the maximum value of $\text{norm}(p_i - q_i)$, so that a small value increases registration accuracy but decreases the number of correspondences, specially in case of important misalignments. Concerning cycle detection, the first parameter fixes the minimum number of views to consider a potential cycle. Translation error specifies in millimeters the maximum distance between both end-views in a cycle to be considered a potential cycle. Overlapping area specifies the minimum overlapping required between both end-views of a potential cycle. Finally, a scale factor is introduced in the cycle minimization step to normalize translation and rotation errors.

In order to provide quantitative evaluation, the motion estimated by registration has to be compared to the real motion provided by the simulator or the 3D table in terms of both translation and rotation. Rotation is represented as a directional vector which can be easily extracted from any rotation matrix. Then, the estimated rotation is compared to the known one. So, the error is determined as the norm of the difference between both axes of rotation. The translation error is defined as the distance between the origin of the coordinate system estimated by registration with respect to the known origin, which is the norm of the difference between both translation vectors.

Additionally, the MSE (mean squared error) is computed. For each point of the registered acquisition, the nearest point in the set composed by the rest of acquisitions is found, determining a correspondence. The mean of all distances give us the estimation of the discrepancy between registered views.

Finally, our proposed method and Sharp's method are both compared to the precise alignment in terms of accuracy so that quantitative results are reported in Table 2 and Fig. 8. In Table 2 we have also included a fast variant of our approach. This fast variant differs only in the cycle minimization step, so that the minimization uses the correspondences previously obtained by the pair-wise registration avoiding the search for new

² <http://www-c.inria.fr/gamma/download/download.php>

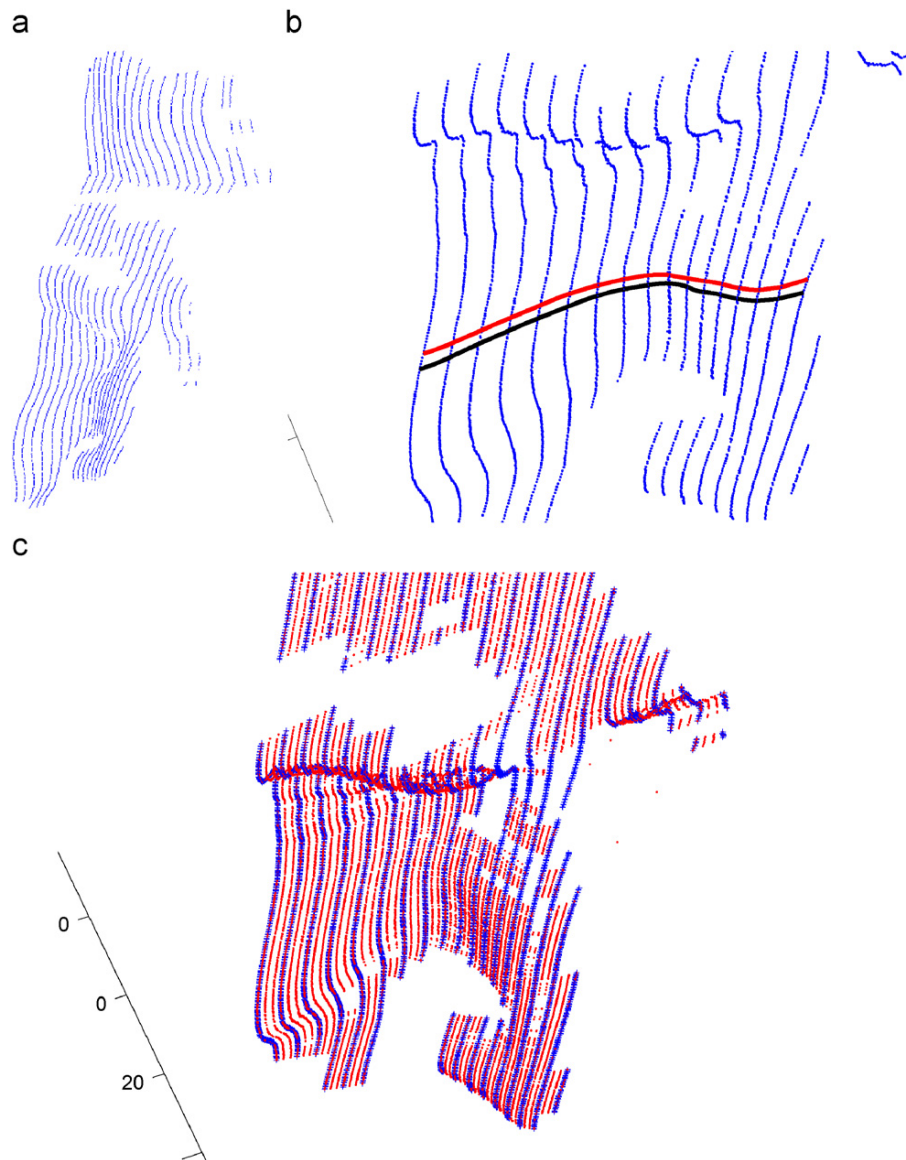


Fig. 5. Acquisition examples: (a) Acquired cloud of points from the 19 slits; (b) Spline curve computation (in blue the acquired profiles, in red and black two samples of splines); (c) Cloud of points obtained after spline sampling (in blue the original points, in red the new points computed).

matching among all the views in the cycle. Actually, matching is one of the most expensive steps. So, the fast variant approach consumes a computing time similar to Sharp's approach but preserves a good accuracy, as shown in Table 2.

Table 2 shows for every experiment the mean and the standard deviation computed from the set of rotation and translation errors. Note that both methods obtain similar results when acquisition noise is unimportant. When the acquisition noise becomes significant, pair-wise registration is not accurate enough and hence Sharp's approach distributes a large error in the cycle. Besides, it is shown that our proposal obtains better results in both synthetic and real data. A special attention requires the results obtained when noise was 1.25% and 3.75%. In both cases, pair wise registrations between both end-views of cycles

were not accurate enough. Such inaccuracies produce significant errors in the method of Sharp that are minimized by our method thanks to the multi-view registration performed inside every cycle.

Fig. 8 presents the MSEs after register all views with our robust method and Sharp's method. Cycle detection algorithm determines a cycle between views 1 and 23. Our robust method obtained better results on these views. Sharp's method obtains good results results in the last view of the cycle, due to the closing-the-loop constrain. However, inside the cycle, propagation error is not always correctly distributed through the views. After view 23, all methods obtain the same results, because there is no cycle, and results are directly obtained from the pair-wise registration.

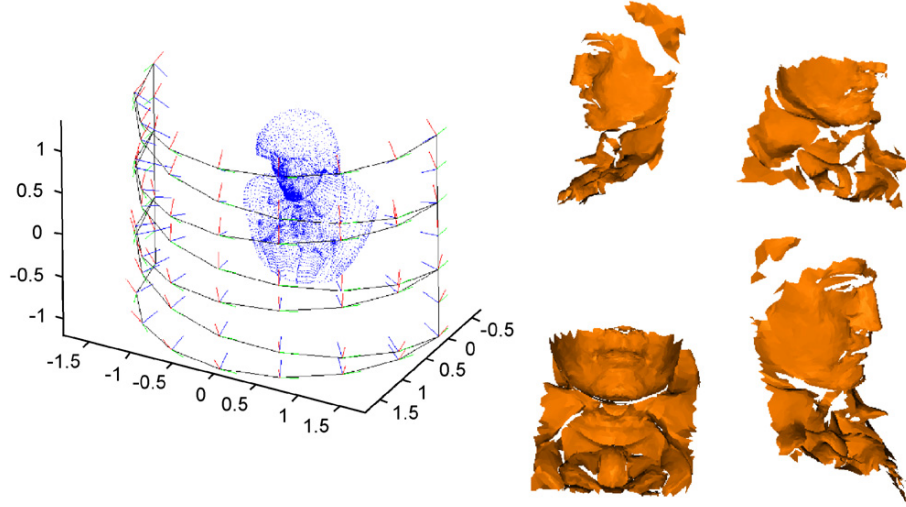


Fig. 6. Left: Path described by the simulator to scan a synthetic object (Beethoven). Right: Some of the acquired synthetic images.

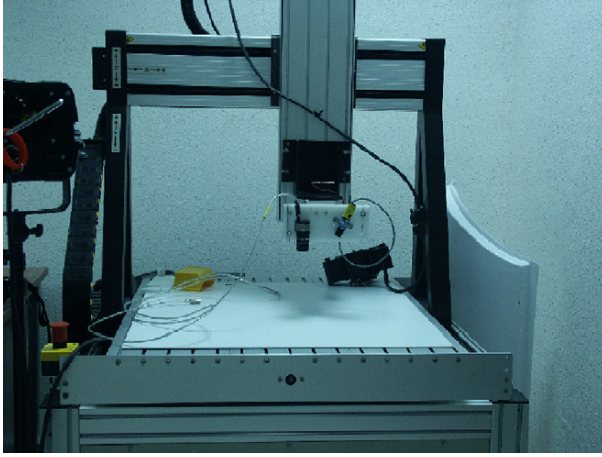


Fig. 7. Accurate motion-controlled table used in the quantitative evaluation.

Table 1
Quantitative experiment settings

Steps	Parameter	Beethoven	Sun/moon
Pair-wise	NSS_{grids}	2500	2500
	$Intersection_{error}$	0.3	6.0
Cycle detection	Minimum number views	8.0	8.0
	Translation error	7.0	15.0
	Overlapping area	50%	50%
Cycle minimization	Sf	1.0	0.1

4.2. Qualitative evaluation

In order to evaluate the performance of the methods, it is also useful to observe the registration of a real object and analyze it from a qualitative point of view. In this experiment,

the one-shot hand-held scanner is coupled to a FANUC industrial manipulator. The manipulator describes a trajectory so that a given object is scanned obtaining a sequence of views. As the kinematics of the manipulator is known, the views can be aligned without applying any registration and hence such raw alignment is provided for comparison.

Note that the kinematics of the manipulator provides the position of the robot hand H with respect to the coordinate frame of the robot base R (see Fig. 9). Besides, registration is referenced with respect to the frame S of the camera of the one-shot hand-held scanner. The rigid transformation between H and S is unknown and hence has to be first estimated.


The computation of ${}^H T_S$ is known as the eye-to-hand problem in the robotics community and it is based on solving equation $AX = XB$, where X is the matrix we are looking for. So, X transforms points from the coordinate frame of the scanner S to the coordinate frame of the hand H , A is the motion of the hand between two different positions of the robot given by the robot control system, and B is the motion computed by means of triangulating the movement in the image of the one-shot hand-held scanner.

There are several papers addressing the computation of $AX = XB$ [37,38]. In our case, we have acquired 10 views of a calibrating pattern and the X matrix is estimated by using the algorithm of Shiu [38]. First, the algorithm determines a set of A and B matrices from every view. Then, a system of equations with the form $AX - XB = 0$ is defined and solved. Theoretically X can be computed with only three views, though it is more accurate to solve the equation of an over-determined system by using singular value decomposition.

Once X is known, all views can be represented in the same reference using the following equation:

$${}^W T_S = {}^W T_R \times {}^R T_H \times X, \quad (10)$$

Table 2
Quantitative results

Scene	Our method			Fast variant			Sharp's method		
	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE	$error_R$	$error_T$	MSE
$\sigma = 0$	0.516	0.008	0.003	0.339	0.079	0.001	0.511	0.074	0.002
	1.120	0.004	0.003	0.867	0.191	0.006	1.006	0.034	0.002
$\sigma = 1.25\%$	0.675	0.154	0.004	1.177	0.459	0.006	2.225	4.403	0.026
	1.115	0.305	0.005	1.388	0.265	0.006	1.385	1.196	0.051
$\sigma = 2.5\%$	1.1286	0.4698	0.005	1.202	0.316	0.002	1.472	1.367	0.001
	1.1905	0.2149	0.005	1.410	0.217	0.006	1.202	0.704	0.002
$\sigma = 3.75\%$	0.246	0.056	0.003	1.552	0.875	0.007	2.601	3.485	0.026
	0.732	0.024	0.002	1.169	0.425	0.006	1.134	2.106	0.046
$\sigma = 5.0\%$	1.570	0.890	0.005	1.533	0.828	0.007	2.753	3.126	0.017
	1.284	0.682	0.005	1.144	0.497	0.006	1.212	2.225	0.020
	1.2804	5.063	0.334	1.3485	5.4103	0.389	1.3863	4.640	0.432
	0.303	2.459	0.335	0.285	2.425	0.529	0.291	2.308	0.380

Both our original method and its fast variant are compared to the method of Sharp: $error_R$ is the norm of the difference between both axes of rotation; $error_T$ is the norm of the difference between both translation vectors (distance between the points of origin of both coordinate systems); MSE is the mean squared error. Every table cell indicates the mean (up) and standard deviation (down) of the error for a set of synthetic experiments varying the Gaussian noise (σ) and one experiment with real data. The synthetic object (Beethoven) consists in 48 views composed of about 5000 points per view and with the presence of 6 detected cycles. The real object (sun/moon) consists in 27 views composed of about 8000 points per view and with the presence of 1 detected cycle between views 1 and 23.

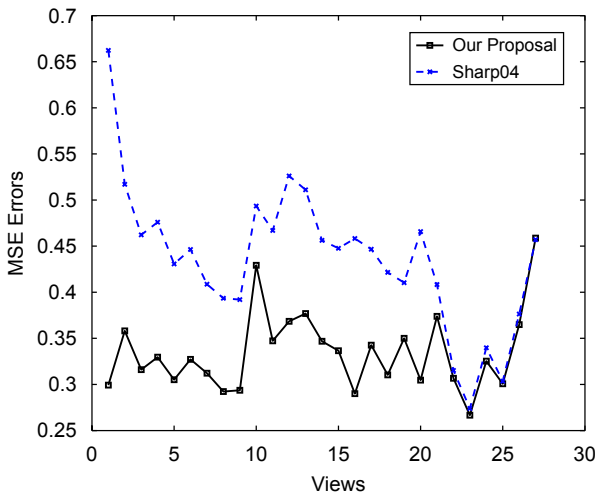


Fig. 8. Evolution of the MSE registration errors in the registration of the real object (sun/moon). Scale of the measured object: 180 mm (width) \times 200 mm (height) \times 56 mm (depth).

where ${}^W T_S$ is the Euclidean motion that transforms points in S to the world coordinate system W (used by the one-shot hand-held scanner to refer 3D points), ${}^W T_R$ is the Euclidean motion that relates the world coordinate system W to the robot base R , ${}^R T_H$ is the motion given by the kinematics of the robot arm, and X is the Euclidean transformation between the camera of the one-shot hand-held scanner and the robot hand.

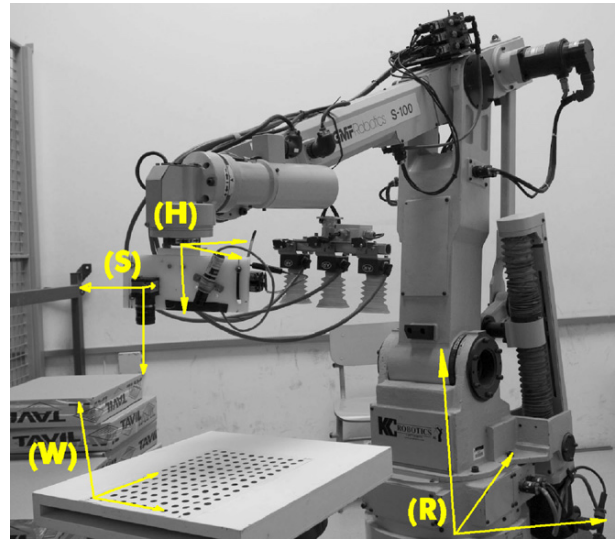


Fig. 9. Industrial manipulator used in experiments. The four coordinate frames are represented: W (world), R (robot), H (Hand) and S (Scanner).

Now we can proceed with the experiment. The parameters of that experiment are presented in Table 3. The manipulator has been programmed so that an 8-shape trajectory is done over a ceramic object acquiring up to 41 images and hence 41 3D partial views of the object. Note that the trajectory ensures cycles which will be used in the registration. First, all the views are referenced with respect to the same frame by means of the X

Table 3
Qualitative experiment settings

Steps	Parameter	Value
Pair-wise	NSS_{grids}	2500
	$Intersection_{error}$	6.0
Cycle detection	Minimum number views	8.0
	Translation error	15.0
	Overlapping area	50%
Cycle minimization	Sf	0.1

matrix. Second, a volumetric integration algorithm is applied to get a continuous surface [39]. Third, the sequence of views are aligned according to: (a) the registration algorithm proposed in this article; (b) the multi-view algorithm proposed by Sharp [28]; and (c) the kinematics of the robot. Finally, any surface smooth technique is applied to enhance the visualization. Qualitative results are shown in Fig. 10. Registration really improves the alignment provided by the kinematics of the robot. Note that the alignment directly obtained from the kinematics of the robot suffers not only from inaccuracies given by the mechanics but especially inaccuracies in the computation of X . Besides, the experiment also shows that our approach provides a surface with more details and less artefacts compared to the method proposed by Sharp. Registration time is presented in Table 4. These experiments are performed using Matlab 6.5 in a pentium IV 2.6 GHz.

The acquisition system obtains the structure of the surface with a single-shot and hence vibrations produced by a human operator do not affect the acquisition. Actually, there is no difference in the obtained results performing the experiments by a manipulator or by a human operator.

Table 4
Registration time

Method	Time (s)
Robust	1192.47
Fast	900.69
Sharp	642.66

5. Conclusions

There are several techniques to register a set of views, though most of them are based on the multi-view approach. In general, multi-view techniques are constrained by the following drawbacks: (a) all the views must be first acquired before the aligning algorithm starts leading to off-line applications; (b) guesses to roughly align the views are needed to initialize the algorithm so that an expensive coarse registration technique is needed; and (c) matching is searched among all the views without considering neighborhood which is inefficient and computing intensive, especially in large data sets. Besides, multi-view techniques are not suitable for registering views that form sequences and loops because of the error propagation problem.

This paper presents a new multi-view registration technique which includes cycle minimization and it is updated in the measure that new views are acquired. Although the technique can be applied in short sequences of views, it is designed to deal with large data sets and with the presence of multiple cycles. First, a fast point-to-plane with normal space sampling and non-overlapping area removal is applied between consecutive views to obtain an accurate alignment. Second, in the measure that new views are acquired, the method searches for cycles considering neighborhood and overlapping percentage. Finally, once a cycle is detected it is minimized by means of a Levenberg–Marquardt approach, so that the system always ensures the most accurate global registration.

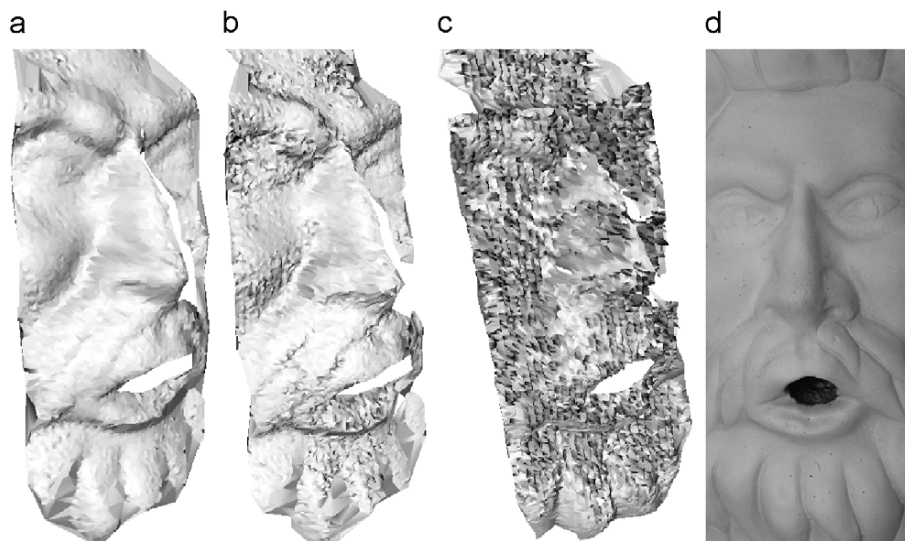


Fig. 10. Results of the registration: (a) Our method; (b) Sharp's method; (c) Mechanical alignment; (d) Real object.

Experiments with both synthetic and real data have been overcome. Synthetic data has been downloaded from well-known databases. A one-shot hand-held scanner composed of a camera and a multi-slit laser emitter has been developed to acquire real data. The scanner has been coupled to a commercial manipulator to acquire sequences of views. Our approach has been compared to: (a) the mechanical alignment provided by the kinematics of the manipulator; and (b) the multi-view alignment method proposed by Sharp [28], which from our point of view is one of the most similar. Results show from both a quantitative and a qualitative point of view that our approach provides a more accurate alignment.

References

- [1] J. Forest, J. Salvi, An overview of laser slit 3D digitizers, in: Proceedings of the International Conference on Intelligent Robots and Systems, Lausanne, October 2002, pp. 73–78.
- [2] J. Forest, J. Salvi, E. Cabruja, C. Pous, Laser stripe peak detector for 3D scanners. A FIR filter approach, in: International Conference on Pattern Recognition, Cambridge, August 2004, pp. 646–649.
- [3] J. Salvi, J. Pagès, J. Batlle, Pattern codification strategies in structured light systems, *Pattern Recognition* 37 (4) (2004) 827–849.
- [4] A. Bottino, A. Laurentini, Shape-from-silhouettes with unknown relative position of the viewpoints: investigating a new problem, in: Signal Processing, Pattern Recognition, and Applications (SPPRA 2002), June 2002.
- [5] Z. Zhang, Q.T. Luong, O. Faugeras, Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction, Technical Report RR-2079, 1993.
- [6] Zcorporation. zscanner700, in: (<http://www.zcorp.com/products/zscanner700.asp?ID=1>), 2007.
- [7] J. Salvi, C. Matabosch, D. Fofi, J. Forest, A review of recent range image registration methods with accuracy evaluation, *Image Vision Comput.* 25 (5) (2007) 578–596.
- [8] I. Stamos, M. Leordeanu, Automated feature-based range registration of urban scenes of large scale, in: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, vol. 2, June 2003, pp. 555–561.
- [9] J. Vanden Wyngaerd, Automatic crude patch registration: toward automatic 3d model building, *Comput. Vision Image Understand.* 87 (2002) 8–26.
- [10] C.S. Jarvis, R. Chua, Point signatures: a new representation for 3d object recognition, *Int. J. Comput. Vision* 25 (1) (1997) 63–85.
- [11] A. Johnson, M. Hebert, Surface registration by matching oriented points, in: International Conference on Recent Advances in 3-D Digital Imaging and Modeling, May 1997, pp. 121–128.
- [12] N. Gelfand, N.J. Mitra, L.J. Guibas, H. Pottmann, Robust global registration, in: Symposium on Geometry Processing, 2005, pp. 197–206.
- [13] P.J. Besl, N.D. McKay, A method for registration of 3-d shapes, *IEEE Trans. Pattern Anal. Mach. Intell.* 14 (2) (1992) 239–256.
- [14] G. Chen, Y. ad Medioni, Object modeling by registration of multiple range images, in: IEEE International Conference on Robotics and Automation, April 1991, pp. 2724–2729.
- [15] P. Neugebauer, Geometrical cloning 3d objects via simultaneous registration of multiview range images, in: Shape Modeling and Applications, 1997, pp. 130–139.
- [16] S. Rusinkiewicz, M. Levoy, Efficient variant of the icp algorithm, in: Third International Conference on 3-D Digital Imaging and Modeling, 2001, pp. 145–152.
- [17] H. Gagnon, M. Soucy, R. Bergevin, D. Laurendeau, Registration of multiple range views for automatic 3-d model building, in: Computer Vision Pattern Recognition, June 1994, pp. 581–586.
- [18] S.-Y. Park, M. Subbarao, A fast point-to-tangent plane technique for multi-view registration, in: 3DIM, Fourth International Conference on 3D Digital Imaging and Modeling, 2003, pp. 276–284.
- [19] E. Batlle, C. Matabosch, J. Salvi, Summarizing image/surface registration for 6dof robot/camera pose estimation, in: Third Iberian Conference on Pattern Recognition and Image Analysis, Girona, Spain, June 2007, pp. 105–112.
- [20] K. Pulli, Multiview registration for large data sets, in: 3-D Digital Imaging and Modeling, Ottawa, Ont., Canada, 1999, pp. 160–168.
- [21] A. Nüchter, H. Surmann, K. Lingemann, J. Hertzberg, S. Thrun, 6d slam with an application in autonomous mine mapping, in: IEEE International Conference on Robotics and Automation, ICRA '04, vol. 2, May 2004, pp. 1998–2003.
- [22] R. Bergevin, M. Soucy, H. Gagnon, D. Laurendeau, Towards a general multiview registration technique, *Pattern Anal. Mach. Intell.* 18 (5) (1996) 540–547.
- [23] L. Silva, O.R.P. Bellon, K.L. Boyer, Multiview range image registration using the surface interpenetration measure, *Image Vision Comput.* 25 (1) (2006) 114–125.
- [24] D. Huber, M. Hebert, Fully automatic registration of multiple 3d data sets, *Image Vision Comput.* 21 (7) (2003) 637–650.
- [25] S. Krishnan, P.Y. Lee, J.B. Moore, S. Venkatasubramanian, Global registration of multiple 3d point sets via optimization-on-a-manifold, in: Symposium on Geometry Processing, 2005, pp. 187–196.
- [26] T. Masuda, Generation of geometric model by registration and integration of multiple range images, in: Third International Conference on 3-D Digital Imaging and Modeling, May 2001, pp. 254–261.
- [27] F. Lu, E.E. Milios, Globally consistent range scan alignment for environment mapping, *Auton. Robots* 4 (4) (1997) 333–349.
- [28] G.C. Sharp, S.W. Lee, D.K. Wehe, Multiview registration of 3d scenes by minimizing error between coordinate frames, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (8) (2004) 1037–1050.
- [29] B. Triggs, P. McLauchlan, R. Hartley, A. Fitzgibbon, Bundle adjustment—a modern synthesis, in: International Workshop on Vision Algorithms, London, UK, 2000, pp. 298–372.
- [30] A. Fitzgibbon, Robust registration of 2d and 3d point sets, in: The British Machine Vision Conference, 2001, pp. 662–670.
- [31] M. Pollefeys, M.R. Koch, M. Vergauwen, L. Van Gool, Automated reconstruction of 3d scenes from sequences of images, *Photogrammetry and Remote Sensing* 55 (2000) 251–267.
- [32] G. Turk, M. Levoy, Zippered polygon meshes from range images, in: SIGGRAPH '94: Proceedings of the 21st Annual Conference on Computer Graphics and Interactive Techniques, Orlando, FL, July 1996, pp. 311–318.
- [33] T. Masuda, K. Sakaue, N. Yoyoka, Registration and intergration of multiple range images for 3-d model construction, in: ICPR '96: Proceedings of the 1996 International Conference on Pattern Recognition, vol. 1, Washington, DC, USA, August 1996, pp. 879–883.
- [34] D.F. Huber, Automatic 3d modeling using range images obtained from unknown viewpoints, in: International Conference on 3D Digital Imaging and Modeling, May 2001, pp. 153–160.
- [35] A. Nüchter, K. Lingemann, J. Hertzberg, H. Surmann, Heuristic-based laser scan matching for outdoor 6d slam, in: 28th German Conference on Artificial Intelligence, 2005, pp. 304–319.
- [36] C. Matabosch, J. Salvi, D. Fofi, F. Meriaudeau, A refined range image registration technique for multi-stripe laser scanner, in: Machine Vision Applications in Industrial Inspection XIV, vol. 6070, SPIE, January 2006, pp. 1–8.
- [37] I. Fassi, G. Legnani, Hand to sensor calibration: a geometrical interpretation of the matrix equation $ax = xb$, *J. Robot. Systems* 22 (9) (2005) 497–506.
- [38] Y.C. Shiu, S. Ahmad, Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form $ax = xb$, *IEEE Trans. Robot. Autom.* 5 (1) (1989) 16–29.
- [39] B. Curless, M. Levoy, A volumetric method for building complex models from range images, in: SIGGRAPH-96, 1996, pp. 303–312.

About the Author—CARLES MATABOSCH received the degree in industrial engineering from the University of Girona, Spain in 2002, and received the D.E.A. degree in computer science in October 2004. He is currently studying toward the Ph.D. degree in computer vision, also at the University of Girona, under the supervision of Dr. J. Salvi. His research interests include 3D reconstruction, laser projection and multi-view registration.

About the Author—DAVID FOFI studied electric and industrial informatics engineering as an undergraduate at the University of Picardie Jules Verne, Amiens and received the M.Sc. degree in Image & Signal Processing from the University of Cergy-Pontoise/ENSEA in 1997. He completed the Ph.D. degree in the area of coded structured light in 2001 at the University of Picardie Jules Verne. He is currently an Associate Professor at the University of Burgundy. His main research interests are in 3D vision, active vision, camera-projector systems, coded and structured light.

About the Author—JOAQUIM SALVI was graduated in computer science from the Technical University of Catalonia, Catalonia, Spain in 1993, and received the D.E.A. degree in computer science in July 1996, and the Ph.D. degree in industrial engineering in 1998, from the Computer Vision and Robotics Group, University of Girona, Girona, Spain. He is currently an associate professor with the Electronics, Computer Engineering and Automation Department, University of Girona. He is involved in some governmental projects and technology transfer to industrial environments. His current interests are in the field of computer vision and mobile robotics, focused on structured light, stereovision, and camera calibration. He is the leader of the 3-D Perception Lab. Dr. Salvi received the Best Thesis Award in engineering for his Ph.D. dissertation.

About the Author—ELISABET BATLLE received her B.S. degree in computer science from the University Autònoma of Barcelona in 2003 before joining the Computer Vision and Robotics Group at the University of Girona. She received her D.E.A. degree in computer science in June 2002 and she is currently studying toward her Ph.D. degree in computer vision, at the University of Girona. She is also an Assistant Professor with the Electronics, Computer Engineering and Automation Department, University of Girona. Her research interests include computer vision, 3D reconstruction, robotics, SLAM and multi-view registration.



PERGAMON

Pattern Recognition 36 (2003) 1631–1644

PATTERN RECOGNITION

THE JOURNAL OF THE PATTERN RECOGNITION SOCIETY

www.elsevier.com/locate/patcog

Uncalibrated reconstruction: an adaptation to structured light vision

David Fofi^a, Joaquim Salvi^b, El Mustapha Mouaddib^{a,*}

^aCREA, Université de Picardie Jules Verne, 7 rue du Moulin Neuf, 80000 Amiens, France

^bComputer Vision and Robotics Group, IliA, Universitat de Girona, Avda. Lluís Santalo, s/n, 17071 Girona, Spain

Received 15 May 2002; accepted 23 September 2002

Abstract

Euclidean reconstruction from two uncalibrated stereoscopic views is achievable from the knowledge of geometrical constraints about the environment. Unfortunately, these constraints may be quite difficult to obtain. In this paper, we propose an approach based on structured lighting, which has the advantage of providing geometrical constraints independent of the scene geometry. Moreover, the use of structured light provides a unique solution to the tricky correspondence problem present in stereovision. The projection matrices are first computed by using a canonical representation, and a projective reconstruction is performed. Then, several constraints are generated from the image analysis and the projective reconstruction is upgraded into an Euclidean one—as we will demonstrate, it is assumed that the sensor behaviour is affine without loss of generality so that the constraints generation is simplified. The method provides our sensor with adaptive capabilities and permits to be used in the measurement of moving scenes such as dynamic visual inspection or mobile robot navigation. Experimental results obtained from both simulated and real data are presented.

© 2003 Pattern Recognition Society. Published by Elsevier Science Ltd. All rights reserved.

Keywords: Uncalibrated system; Projective reconstruction; Euclidean constraints; Structured light; Computer vision

1. Introduction

The perception of the three-dimensional structure of the environment is an important task in computer vision. In mobile robotics, it forms the basis of obstacle detection, map building, scene analysis, etc. The challenge is to infer 3D information of a scene by starting from at least two images of it [1]. From two images, the reconstruction is thus possible but we need to calibrate the cameras, i.e. determining its optical parameters and internal geometry (focal distance, principal point, pixel adjustment) and its geometrical parameters (position and orientation with respect to a reference frame). The correspondence among the 3D object points and their projections has to be established; the obtained matrix makes

possible to relate each point to its line of sight [2]. This technique is named *hard-calibration* and is carried out off-line by using a calibrating pattern whose 3D points co-ordinates of interest are precisely known. This process has to be repeated each time that one of the parameters of the camera is modified. Hard-calibration is exclusively adapted to applications that keep the sensor unchanged during the measuring process. Nevertheless, a visual adaptation to the environment is essential in the measurement of moving scenes such as dynamic visual inspection or mobile robot navigation. Then, the visual adaptation permits to use a camera with auto-focus (to increase the quality of the image), zoom (to concentrate on relevant regions of the image) and aperture (in case of illumination changes), which is the first step to develop strategies of observation and/or exploration.

It is well known that the major drawback of stereoscopy is the correspondence problem, i.e. the matching of homologue points among the images. With the aim of reducing

* Corresponding author. Tel./fax: +33-3-22-82-76-68.

E-mail address: mouaddib@u-picardie.fr (E.M. Mouaddib).

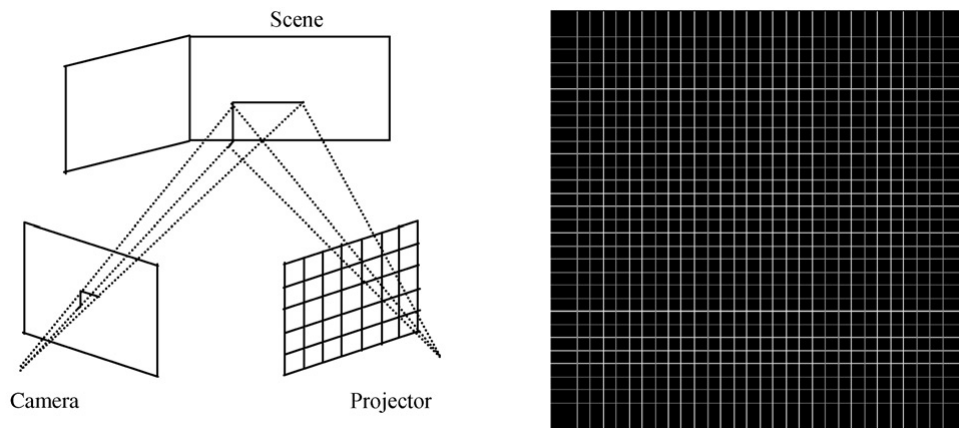


Fig. 1. The structured light system: (a) the geometrical principle; (b) the colour-encoded pattern.

this problem, coded structured light techniques have been developed [3]. In a structured light system, the second camera is replaced by a light source that projects a known pattern of light onto the scene, as shown in Fig. 1a. Since a projector can be seen as a camera acting in reverse, it can be modeled in the same way a camera is.

Our pattern is composed by a set of vertical and horizontal slits, uniquely colour-encoded in a single pattern projection (Fig. 1b). The reader is pointed to Salvi et al. [4] to get deeper into the pattern design. The coloured codification permits to solve the correspondence problem finding out for each imaged point, its corresponding point in the projecting plane.

The main goal of this paper is, firstly, to provide adaptive capabilities to our structured light vision sensor and, secondly, to adapt the techniques of uncalibrated reconstruction to structured light. Our contribution is to demonstrate how the projection of a grid of light and the analysis of the coded images permit to generate Euclidean constraints for a three-dimensional reconstruction of the scene and, more generally, to show how to self-calibrate a structured light sensor. It is assumed throughout the paper that the camera and the projector behaviour can be approximated by an affine model. However, no assumption on the scene geometry is imposed, although a planar piece-wise environment ensures a major number of constraints.

The next points summarize our approach:

- Extraction of the image points by a specific image processing (see the section experimental results).
- Projective reconstruction from one view and one pattern projection using the canonical representation.
- Automatic generation of constraints in order to reach an Euclidean reconstruction of the scene.

This article is organized as follows. Section two presents the related work about reconstruction methods that could be adapted to a structured light system. Then, section three de-

tails the Euclidean reconstruction through structured lighting, which contains the major contribution of this paper. Furthermore, section four deals with some experimental results considering both simulated and real scenes. The article ends with conclusions.

2. Reconstruction and structured light system

This section goes deeper into what has been proposed on structured light, taking advantages of the specificities of projection and pattern structure and considering that the principle of calibration is known. For instance, Salvi et al. [4] proposed to model the projector like a camera acting in reverse. First, they calibrate the camera by using a calibrating plane and then an image of the projected pattern on the calibrating plane is grabbed and used to get the 3D points to calibrate the projector. Proesmans et al. [5] proved that the reconstruction could be performed whether the angle between directions of projection and capture is known, assuming an orthographic model. This particular way to calibrate consists in observing a blank calibration pattern whose angle, made up by the two planes that compose it, is precisely known. Sotoca et al. [6] proposed a calibration method adapted to large surface. Beforehand, the pattern is projected onto a *base plane* and onto a *reference plane* and an image is grabbed for each of these planes. By positioning the object to analyze between these planes, the authors show that it is possible to obtain a depth image through some simple calculations based on the similarity of triangles. Finally, let us conclude by the method developed by Huynh et al. [7] which has been proposed for light plane projections but it can be generalized to pattern projections. Four sets of three coplanar points, whose coordinates are precisely known, are positioned on the two planes of a calibration pattern, depicting four lines on it. While the light plane intersects these lines, a fourth point is obtained on each of them: the cross-ratio of these points is equal to the cross-ratio of the images of these points. It

provides a 3D measurement of the lighted traces fairly precise (up to the stability of cross-ratios). A classical calibration process is performed from these measurements.

Uncalibrated vision has generated an increasing number of publications since the end of the 1980s. Aware of the drawbacks of hard-calibration for some applications in which the sensor has to adapt its behaviour to the variations of the environment and to the strategy of observation, many authors looked into the problem that consists of inferring 3D structure of the scene from the pixel coordinates only. It is proved that an Euclidean reconstruction cannot be obtained without calibrating but, at best, a projective one [2]; this reconstruction has to be constrained by means of additional information or assumptions (about the intrinsic parameters, the movement of the sensor, the scene geometry) to achieve the self-calibration of the sensor, i.e. the Euclidean reconstruction of the scene.

Koenderink and Van Doorn [8] can be regarded as pioneers. In 1989, they proposed a method which allows to recover the affine scene structure from at least two images of it, by using a shape invariant computed from a reference plane. Later, Faugeras [2] and, independently, Hartley et al. [9] proved that from a weakly calibrated sensor (i.e. which epipolar geometry is known) a projective reconstruction is possible. Mohr et al. [10] reached similar results through a global estimation of the unknowns by minimizing the residual errors between image points and their back-projections. However, these methods do not allow to compute the Euclidean structure of the environment. With this aim, Faugeras et al. [11] proposed a method that takes advantage of the invariance under similarities of the absolute conic (in other words, the image of the conic only depends on intrinsic parameters of the camera). By such performing, the authors have rediscovered the Kruppa equations. Besides, Hartley [12] demonstrated that these equations could be explicitly obtained through the singular value decomposition of the fundamental matrix \mathbf{F} . In order to solve them, three fundamental matrices have to be computed, which are given from a single displacement of the stereo head.

A second class of methods assumes that intrinsic parameters remain constant during the measuring process. A projective reconstruction is first performed. Then, the constancy assumption leads to an equation solvable from three views of the scene; the reader can refer to the work of Hartley [13] or Heyden and Aström [14]. At last, if the constancy of intrinsic parameters cannot be assumed, it is possible to upgrade a projection reconstruction into an Euclidean one by generating Euclidean constraints grabbed from the scene geometry. Boufama et al. [15] pioneered this method; Zhang et al. [16] later on proposed a similar method.

Some considerations have to be taken in mind when structured light is used. Any movement of the sensor, and particularly of the projector, produces a sliding of the projected points on the observed surfaces. That is to say, the points illuminated before the movement are different than the ones illuminated after the movement. As a consequence, stere-

ovision algorithms using more than two views cannot be adapted to structured light vision. Besides, due to the heterogeneity of the sensor, composed by a camera and a projector, the constancy of intrinsic parameters cannot be assumed either. Hence, methods based on Kruppa's equations and methods based on constant intrinsic parameters are unsuited to structured light vision.

There is only one choice left: performing a projective reconstruction first and rectifying it by using Euclidean constraints grabbed from the scene geometry. It is shown in the next section what kind of projective reconstruction method may be used and how to generate constraints by using the geometry of light patterns.

3. Uncalibrated reconstruction and structured light system

This section details a method that permits to locate a point in the three-dimensional space from a pair of uncalibrated perspective views (which is equivalent to one view from a camera and a known projected pattern). First, the method performs a reconstruction in a projective frame. Then, the reconstruction of the scene is transformed to an Euclidean frame by using some a priori knowledge between the view of the scene and the projected pattern, less restrictive than point co-ordinates, such as parallelism, orthogonality, angles, length ratio, and so on.

The following section presents the theoretical basis and methods about the uncalibrated reconstruction adapted to structured light. Then, Section 3.5 details the proposed algorithm that describes the whole process.

3.1. Projective reconstruction

It is known, since the work of Luong and Vieville about the canonical representation of the geometry of multiple views [17], that it is possible to estimate the camera projection matrices from the knowledge of epipolar geometry. Then, considering two 2D images (an image of the scene and the projected pattern), we have:

$$\mathbf{P}_{proj} = [\mathbf{M} \ \mathbf{e}], \quad \mathbf{P}'_{proj} = [\mathbf{I} \ \mathbf{0}]$$

$$\text{with } \mathbf{M} = -\frac{1}{\|\mathbf{e}\|^2} [\mathbf{e}]_{\times} \mathbf{F}, \quad (1)$$

where \mathbf{F} denotes the fundamental matrix and \mathbf{e} the epipole of the first image, \mathbf{P} is the camera matrix and \mathbf{P}' the projector matrix. Subscript *proj* denotes matrices, vectors or scalars expressed in a projective frame, in contrast with *eucl* which will denote matrices, vectors or scalars expressed in an Euclidean frame.

Whereas $\mathbf{F}^T \mathbf{e} = 0$, so that the coordinates of the epipole are given by the eigenvector of the matrix $\mathbf{F}\mathbf{F}^T$ associated with the smallest eigenvalue. Numerically, better results are obtained by normalizing the epipole in the way that $\|\mathbf{e}\| = 1$.

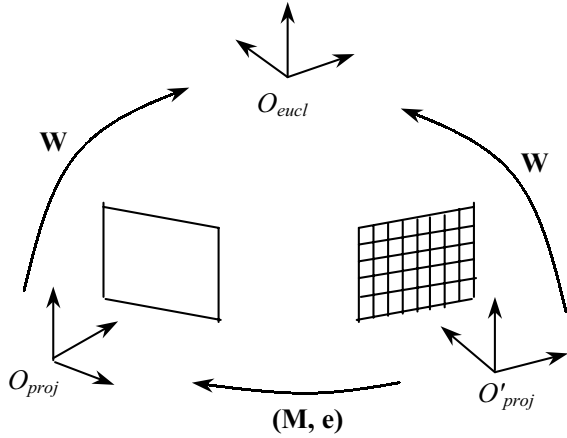


Fig. 2. Reference frames and projection matrices for the projective reconstruction.

This formulation satisfies the epipolar constraint and fixes the projective basis on the camera frame; where the projection matrices are computed so that the 3D observed points could be reconstructed with respect to *that* frame. We will perform it in two steps: first, by a linear method, fast but not accurate enough; then, by a non-linear method initialized with the results of the linear method. We describe the two steps in the following notations and reference frames are depicted in Fig. 2.

3.1.1. Linear method

Given a pair of points in correspondence, $\mathbf{m} = [u \ v \ 1]^T$ and $\mathbf{m}' = [u' \ v' \ 1]^T$, and their corresponding 3D point in space $\mathbf{M}_{proj} = [x \ y \ z \ t]^T$ expressed in the projective frame, it is obtained:

$$\lambda [u \ v \ 1]^T = \mathbf{P}_{proj} [x \ y \ z \ t]^T \quad (2)$$

$$\lambda' [u' \ v' \ 1]^T = \mathbf{P}'_{proj} [x' \ y' \ z' \ t']^T \quad (3)$$

where λ and λ' are two non-zero scale factors. Eliminating the scale factors and re-arranging equations (2) and (3) yields to

$$\mathbf{Q} \mathbf{M}_{proj} = 0. \quad (4)$$

\mathbf{Q} is a 4×4 matrix given by

$$\mathbf{Q} = [\mathbf{p}_1 - u\mathbf{p}_3 \quad \mathbf{p}_2 - v\mathbf{p}_3 \quad \mathbf{p}'_1 - u'\mathbf{p}'_3 \quad \mathbf{p}'_2 - v'\mathbf{p}'_3], \quad (5)$$

where \mathbf{p}_i and \mathbf{p}'_i are the vectors corresponding to the i th row of \mathbf{P} and \mathbf{P}' , respectively. As \mathbf{M}_{proj} is defined up to a scale factor, we can impose $\|\mathbf{M}\| = 1$. The solution is given by the eigenvector of the matrix $\mathbf{Q}^T \mathbf{Q}$ associated to the smallest eigenvalue.

3.1.2. Non-linear method

It seems difficult, in the previous approach, to give a good physical interpretation to the criterion that is minimized. Besides, the accuracy of the results can be significantly improved. A way to alleviate these drawbacks is to

use a non-linear iterative method of minimization. The error to minimize is the difference between the observation and the back-projection of the reconstructed points or *residual error*; in other words:

$$\begin{aligned} & \left(u - \frac{\mathbf{p}_1^T \mathbf{M}_{proj}}{\mathbf{p}_3^T \mathbf{M}_{proj}} \right)^2 + \left(v - \frac{\mathbf{p}_2^T \mathbf{M}_{proj}}{\mathbf{p}_3^T \mathbf{M}_{proj}} \right)^2 + \left(u' - \frac{\mathbf{p}'_1^T \mathbf{M}_{proj}}{\mathbf{p}'_3^T \mathbf{M}_{proj}} \right)^2 \\ & + \left(v' - \frac{\mathbf{p}'_2^T \mathbf{M}_{proj}}{\mathbf{p}'_3^T \mathbf{M}_{proj}} \right)^2. \end{aligned} \quad (6)$$

In practice, a traditional algorithm of minimization like Levenberg–Marquardt is used [18]. The results provided by the direct method are used to initialize the algorithm.

3.2. Towards an Euclidean reconstruction

It is known that the Euclidean geometry is a particular case of the projective geometry. In other words, a collineation exists which brings the solution to an Euclidean one. Finding this collineation, it is thus possible to recover the Euclidean structure of the scene.

Let us consider \mathbf{M}_{proj} , a point with projective co-ordinates, and \mathbf{M}_{eucl} the same point with Euclidean co-ordinates.

$$\mathbf{M}_{proj} = \begin{pmatrix} x_{proj} \\ y_{proj} \\ z_{proj} \\ t_{proj} \end{pmatrix}, \quad \mathbf{M}_{eucl} = \begin{pmatrix} x_{eucl} \\ y_{eucl} \\ z_{eucl} \\ 1 \end{pmatrix}. \quad (7)$$

The problem is to determine \mathbf{W} such as

$$\mathbf{M}_{eucl} = \mathbf{W} \cdot \mathbf{M}_{proj}. \quad (8)$$

With the aim of computing the collineation, geometrical knowledge about the scene is translated into constraints on the entries of \mathbf{W} .

We have to fix affine or Euclidean constraints, which reports geometrical properties of the scene extracted from the images. \mathbf{W} has 15 degrees of freedom; therefore 15 independent and coherent constraints have to be found. Hereafter, a non-exhaustive list of constraints with their mathematical formulations is given and it is shown how the use of structured light leads to generate such constraints.

3.3. Euclidean constraints from grid coding

As described in Section 1, grid coding is the way a grid of light is projected onto the scene to be analyzed. It is the sub-class of structured light techniques that used a grid pattern. It is shown in this section how geometrical knowledge about the scene can be obtained analyzing the grid-coded images. This knowledge can be used as constraints to bring a projective reconstruction to an Euclidean one.

3.3.1. Plane detection in space

The elementary cell of a grid is a square. Each cell projected onto a planar surface is captured by the camera as a

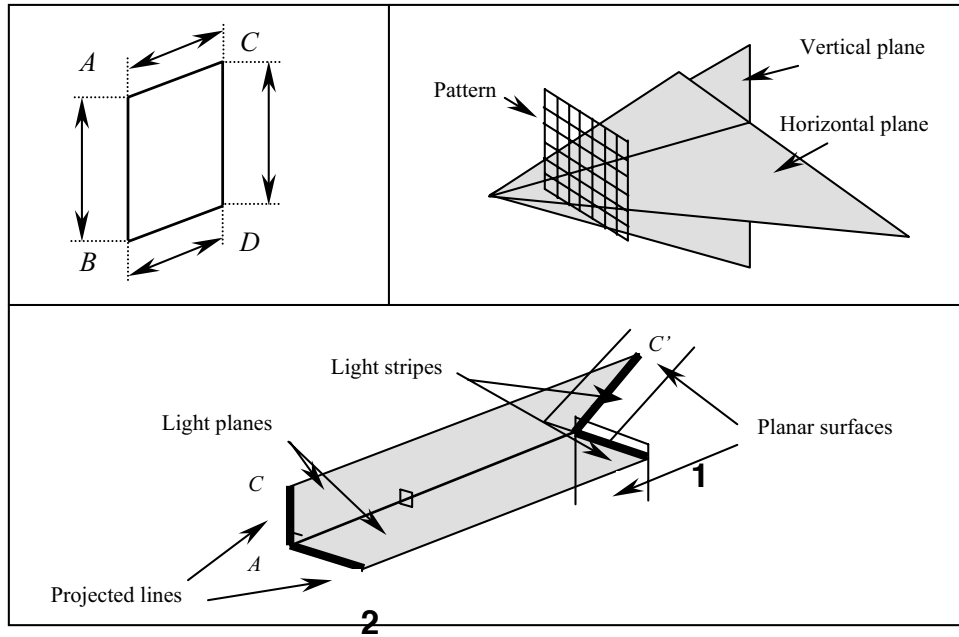


Fig. 3. Some of the geometrical constraints used in Euclidean reconstruction: (a) Parallelogram constraint; (b) Horizontal and vertical plane; (c) Orthogonality.

quadrilateral. Quadrilaterals can theoretically either be captured as squares, rectangles, rhombuses, parallelograms or trapezoids, depending on the position and orientation of the camera and projector, and the planar surface observed. Although quadrilateral detection within the image is not equivalent to plane detection in 3D space (there are configurations of curved surfaces that can yield quadrilaterals in the image), it is quite probable that a quadrilateral within the image corresponds to a plane in space: the equivalence is assumed throughout the paper and, moreover, this can be verified, thanks to the test of coplanarity we propose in Section 4.1.

3.3.2. Parallelogram constraints

Assuming that the projector have an approximately affine behaviour, we obtain that if a square is projected onto a planar surface, the more generic quadrilateral formed on the surface is a parallelogram. Furthermore, a parallelogram captured by an affine camera forms a parallelogram onto the retina. Hence, a parallelogram within the image corresponds to the image of a parallelogram on a 3D plane. Geometrical knowledge about the scene can thus be deduced.

Relative positioning of the four points A , B , C and D of the parallelogram (see Fig. 3a) in space is such as

$$\overline{AB} = \overline{CD}, \quad \overline{AC} = \overline{BD}, \quad (9)$$

$$(AB)/(CD), \quad (AC)/(BD). \quad (10)$$

It leads to a *redundant* set of constraints on W . Besides, knowing Eq. (11), parallelism constraints can be simplified

as shown in Eqs. (11) and (12):

$$(AB)/(CD) \Leftrightarrow \frac{\overrightarrow{AB}}{\|\overrightarrow{AB}\|} = \frac{\overrightarrow{CD}}{\|\overrightarrow{CD}\|}, \quad (11)$$

$$\begin{aligned} (x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2 \\ = (x_D - x_C)^2 + (y_D - y_C)^2 + (z_D - z_C)^2, \\ (x_C - x_A)^2 + (y_C - y_A)^2 + (z_C - z_A)^2 \\ = (x_D - x_B)^2 + (y_D - y_B)^2 + (z_D - z_B)^2, \end{aligned} \quad (12)$$

$$\begin{aligned} (x_B - x_A) = (x_D - x_C), (y_B - y_A) = (y_D - y_C), (z_B - z_A) \\ = (z_D - z_C), \end{aligned}$$

$$\begin{aligned} (x_C - x_A) = (x_D - x_B), (y_C - y_A) = (y_D - y_B), (z_C - z_A) \\ = (z_D - z_B). \end{aligned} \quad (13)$$

Since projective geometry keeps unchanged the alignment and the coplanarity, Eqs. (12) and (13) determine the same configuration of points (redundant constraints). Note that a parallelogram completely determines a 3D plane. Therefore, for each plane composing the scene, a unique set of parallelogram constraints is sufficient.

Now, let us consider the two configurations of points shown in Fig. 4. Whether the points o_i , p_i , q_i , r_i and s_i ($i = 1$ or 2) are projected onto a plane, the cross-ratio within the pattern is equal to the cross-ratio of the five points formed onto this plane; moreover, the cross-ratio of the homologue points within the image is equal to both. The change from projected points to imaged points is obtained by two successive homographies. It can be deduced that if

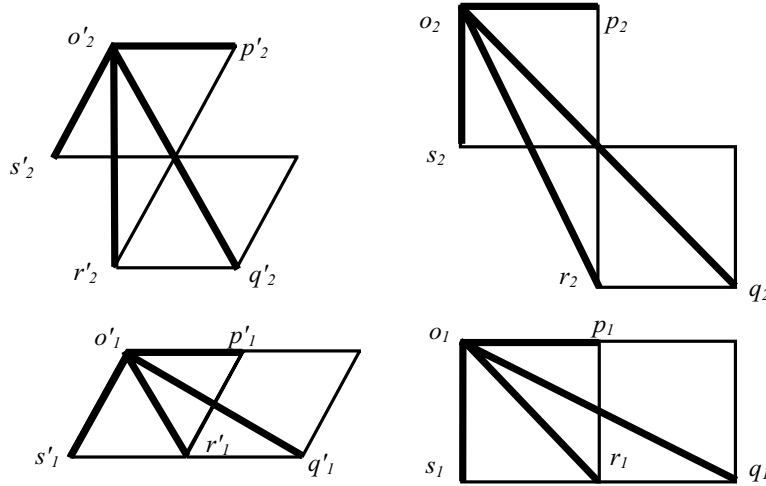


Fig. 4. Test of coplanarity. Points configuration: image plane (left) and projector plane (right).

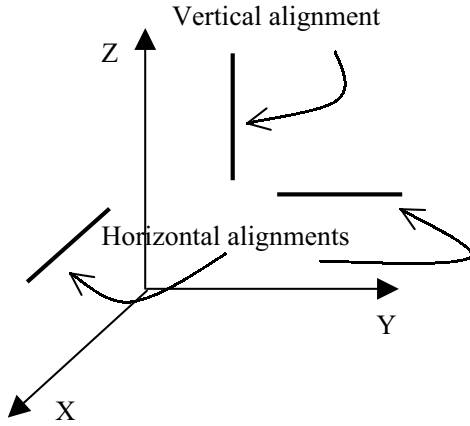


Fig. 5. Euclidean constraint: alignment.

Eq. (14) is verified, the corresponding object points O, P, Q, R and S are coplanar.

$$\{o_i; p_i, q_i, r_i, s_i\} = \{o'_i; p'_i, q'_i, r'_i, s'_i\} \quad \text{with } i = 1 \text{ or } 2. \quad (14)$$

3.3.3. Horizontal and vertical plane

If a point P belongs to the horizontal plane passing through the origin, then $z_P = 0$, which permits to obtain the following linear constraint (Fig. 5):

$$w_{31}x_{P'} + w_{32}y_{P'} + w_{33}z_{P'} + w_{34}t_{P'} = 0. \quad (15)$$

Replacing w_{3i} by w_{2i} or w_{1i} , the homologue constraints for $y_P = 0$ or $x_P = 0$, respectively, are expressed. Noting that each projected horizontal line of the pattern generates a light plane in space, which can be considered as a 3D horizontal plane in the projector co-ordinate system (see Fig. 3b). And each projected vertical line of the pattern generates a light

plane in space, which can be considered as a vertical 3D plane in the projector co-ordinate system (considering either an affine or projective camera model). Their corresponding lines captured by the camera can be used to generate such kind of constraints. Indeed, what it is imaged by the camera are the intersections of the projecting planes of light with the scene surfaces, therefore points belong to horizontal or vertical planes.

Furthermore, an arbitrary distance can be set between two successive horizontal or vertical planes. If the distance between two points A and B is assumed to be d . Then, $(x_A - x_B)^2 + (y_A - y_B)^2 + (z_A - z_B)^2 = d^2$ and, as a consequence, the following non-linear constraint is obtained:

$$\begin{aligned} & \left(\frac{w_{11}x_{A'} + w_{12}y_{A'} + w_{13}z_{A'} + w_{14}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} \right. \\ & \quad \left. - \frac{w_{11}x_{B'} + w_{12}y_{B'} + w_{13}z_{B'} + w_{14}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}} \right)^2 \\ & + \left(\frac{w_{21}x_{A'} + w_{22}y_{A'} + w_{23}z_{A'} + w_{24}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} \right. \\ & \quad \left. - \frac{w_{21}x_{B'} + w_{22}y_{B'} + w_{23}z_{B'} + w_{24}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}} \right)^2 \\ & + \left(\frac{w_{31}x_{A'} + w_{32}y_{A'} + w_{33}z_{A'} + w_{34}t_{A'}}{w_{41}x_{A'} + w_{42}y_{A'} + w_{43}z_{A'} + w_{44}t_{A'}} \right. \\ & \quad \left. - \frac{w_{31}x_{B'} + w_{32}y_{B'} + w_{33}z_{B'} + w_{34}t_{B'}}{w_{41}x_{B'} + w_{42}y_{B'} + w_{43}z_{B'} + w_{44}t_{B'}} \right)^2 = d^2. \quad (16) \end{aligned}$$

This constraint permits to assign a metric to the 3D space. It is possible to give an arbitrary value to d but the reconstruction will be achieved up to a scale factor.

Without particular knowledge, a plane can be arbitrarily chosen as a horizontal or vertical plane; in this case, the

reconstruction will be performed up to a rotation and a translation along the z -axis.

3.3.4. Fixing the origin

If the Euclidean co-ordinates of a point p in space are known, it is obtained that

$$\begin{aligned} x_p &= \frac{w_{11}x_{p'} + w_{12}y_{p'} + w_{13}z_{p'} + w_{14}t_{p'}}{w_{41}x_{p'} + w_{42}y_{p'} + w_{43}z_{p'} + w_{44}t_{p'}}, \\ y_p &= \frac{w_{21}x_{p'} + w_{22}y_{p'} + w_{23}z_{p'} + w_{24}t_{p'}}{w_{41}x_{p'} + w_{42}y_{p'} + w_{43}z_{p'} + w_{44}t_{p'}}, \\ z_p &= \frac{w_{31}x_{p'} + w_{32}y_{p'} + w_{33}z_{p'} + w_{34}t_{p'}}{w_{41}x_{p'} + w_{42}y_{p'} + w_{43}z_{p'} + w_{44}t_{p'}}. \end{aligned} \quad (17)$$

Then, these equations give three linear constraints. As we said before, this knowledge is barely available; nevertheless, we can fix the origin of the Euclidean co-ordinates frame by equaling these equations to zero. The cross-point (which appears in the image as the intersection of two light stripes) of the planes $y = 0$ and $x = 0$ is considered to be the origin of the Euclidean co-ordinate frame.

3.3.5. Orthogonality constraint

Orthogonality is an important feature in Euclidean reconstruction. The detection of orthogonal planes permits to define, at least partially, a 3D Euclidean frame of the scene. Let us consider again an affine model for the projector. The projection of a line produces a light plane in space. The projection of two orthogonal lines (AB) and (AC) produces two orthogonal light planes (Fig. 3c). When the projecting light planes intersect planar surfaces, they produce light stripes on them which will be imaged by the camera. We have thus two lines ($A'B'$) and ($A'C'$) in space, which belong to orthogonal planes. Since A' and B' belong to the same horizontal plane and A' and C' belong to the same vertical plane, considering the world co-ordinate system is fixed at the projector, it is obtained

$$x_{A'} = x_{B'}, \quad y_{A'} = y_{C'}, \quad (18)$$

$$\begin{aligned} \overrightarrow{A'B'} \cdot \overrightarrow{A'C'} &= (x_{A'} - x_{B'})(x_{A'} - x_{C'}) \\ &\quad + (y_{A'} - y_{B'})(y_{A'} - y_{C'}) \\ &\quad + (z_{A'} - z_{B'})(z_{A'} - z_{C'}) \\ &= (z_{A'} - z_{B'})(z_{A'} - z_{C'}). \end{aligned} \quad (19)$$

So

$$(A'B') \perp (A'C') \Leftrightarrow z_{A'} = z_{B'} \text{ or } z_{A'} = z_{C'}. \quad (20)$$

If the conditions imposed by (21) are satisfied, we obtain an *orthogonality constraint*, otherwise we obtain a *reduced orthogonality constraint*:

$$(x_{A'} - x_{B'})(x_{A'} - x_{C'}) + (y_{A'} - y_{B'})(y_{A'} - y_{C'}) = 0. \quad (21)$$

3.4. Resolution of the system

The projective reconstruction is first performed by solving the set of equations (1) previously described. In a least-squares optimization, it leads to minimize the following error:

$$(\hat{A}, \hat{P}) = \arg \min_{A, P} (p - AP)^T C^{-1} (p - AP), \quad (22)$$

where \hat{A} and \hat{P} are the estimated values of A and P , respectively; and C is the covariance matrix. As the location of points within the images is the major factor of noise, all the other factors are neglected. Then, C is a diagonal matrix and its elements are all equal to the variance since imprecise location induces decorrelated noise.

The Levenberg–Marquardt algorithm [18] is used to solve this set of non-linear equations. As 3D points and projective matrices can only be known up to a scale factor, a scale constraint must be added for each point and each matrix in order to lead to a unique solution. Eq. (23) is the constraint for points and Eq. (24) for matrices:

$$x_i^2 + y_i^2 + z_i^2 + t_i^2 - 1 = 0, \quad (23)$$

$$m_{34}^{(j)} = 1, \quad j = 1, \dots, s. \quad (24)$$

Once the projective reconstruction is performed, the matrix W has to be estimated to obtain the Euclidean reconstruction. The Levenberg–Marquardt algorithm is also used. Equations which should be minimized are the ones that provide Euclidean constraints (Eqs. (14)–(18) and (22)). The scale constraint that has to be added is given by Eq. (25).

$$\sum_{i,j} (w_{ij})^2 = 1. \quad (25)$$

3.5. Algorithm

Let us now summarize the steps that are necessary to perform an Euclidean reconstruction without any a priori knowledge about the observed scene, but a single image.

1. Image processing

Input: camera image and projected pattern

Output: the two sets of matched points (U_{ij} , V_{ij} , the co-ordinates of the i th point in the j th image)

- Segmentation and decoding.
- Solve the correspondence problem decoding the pattern.

2. Projective reconstruction (Section 3.1):

Input: matching points

Output: 3D reconstructed points in a projective frame

- Estimation of the fundamental matrix.
- Estimation of the projection matrices.
- 3D reconstruction by linear method.

- 3D reconstruction by iterative method (using the previous method as initialization).

3. Euclidean reconstruction:

Input: 3D projective points and projection matrices, segmented and decoded lines and cross-points.

Output: 3D reconstructed points in a Euclidean frame.

- Fix an arbitrary point as the origin of the world co-ordinate system (Section 3.3.4).
- Fix the horizontal (in the pattern) line to which it belongs as the $y = 0$ plane in space (Section 3.3.3).
- Fix the vertical (in the pattern) line to which it belongs as the $x = 0$ plane in space (Section 3.3.3).
- Extract parallelograms within the image and generate parallelogram constraints, if the test of coplanarity is ok (Section 3.3.2).
- Extract crossing lines within the image and generate (reduced) orthogonality constraints (Section 3.3.5).
- Fix an arbitrary distance d between two points in space (Section 3.3.5).
- Compute the collineation in order to upgrade the projective reconstruction into Euclidean one (Section 3.2).

4. Experimental results

First of all, the stability of cross-ratio is discussed in order to evaluate the efficiency of our test of coplanarity and some results are presented. Then, experiments on reconstruction have been performed with simulated data. A set of 3D points is fixed with respect to a world co-ordinate system and these points are observed by two virtual cameras. Five of these points are chosen as the reference basis; all the other ones are reconstructed using the method previously described, i.e. by using only their pixel coordinates and the matching in both images. Euclidean reconstruction by adding geometrical constraints is obtained and results validated. Furthermore, we have performed the reconstruction method with real images using our structured light sensor. In the following, we detailed the implementation of the algorithm and we summarize the experimental results obtained. Quantitative results are given for the experiments performed with simulated data; since we do not have reliable measurements of the real scenes, only qualitative results are given for the experiments with real data.

All the experiments have been performed by using Matlab, so time consuming is not very significant. The test of coplanarity and the linear method for projective reconstruction, based on matrix algebra, are achieved in less than one second. In contrast, the Euclidean reconstruction (i.e. the Levenberg–Marquardt algorithm) is performed in a few iterations (from 5 to 20, depending on the Euclidean constraints), that is, in a few minutes. Of course, by programming the algorithm in C/C++ code, it is possible to considerably improve time consuming (a set of non-linear

equations can be solved in less than one second in C/C++).

4.1. Test of coplanarity

We have tested the stability of the cross-ratio for the configurations of points required by the test of coplanarity. We took five points separated by the distance d (on Fig. 4, d is the distance $o_1 p_1$, $o_2 p_2$, etc.) A gaussian noise, varying from 0 to $0.5 \times d$, is added on the points co-ordinates. The results are depicted by Fig. 6. The left part shows the stability of cross-ratio with a noise of $\pm 5\%$ for 100 computed values (theoretically, cross-ratio is 2 in this example). The right part shows the evolution of the error against the noise level (which depends on d).

To be able to compare the theoretical cross-ratios with the cross-ratios computed from the images (i.e. to compute the error) we used a *projective distance* based on the method of the random cross-ratios, detailed in Ref. [19]. The tolerance error is empirically fixed to 5×10^{-3} . Under these conditions, a noise up to 15% is allowed to well discriminate configurations of coplanar points. Obviously, as it can be deduced from the results of Fig. 6: the larger the distance d is, the more robust the measure of cross-ratio will be. The left part of Fig. 6 shows that, with a moderate noise ($\pm 5\%$), the measured cross-ratio is very near to the theoretical one. Hence, the stability of cross-ratio is good enough for applications of uncalibrated reconstruction.

We have tried out the test of coplanarity by performing three experiments. In the first one, a planar configuration of points is detected (theoretical cross-ratio = 2, measured cross-ratio = 1.96, projective error = 2.2×10^{-3}). In the second one, the pattern is projected onto an irregular surface and the test classifies these points as non-coplanar (theoretical cross-ratio = 2, measured cross-ratio = 2.186, projective error = 5.9×10^{-3}). Finally, in the last experiment, the points are projected onto a cube corner (clearly not coplanar) and the points are well-classified (theoretical cross-ratio = 2, measured cross-ratio = 2.2055, projective error = 9.9×10^{-3}). As this test is only based on cross-ratio computing, its time computing is near-instantaneous.

4.2. Simulated data

4.2.1. Five known points

It is assumed here that five points of the scene are taken as landmarks whose Euclidean co-ordinates are known. Let us assume that the camera is set at the origin of the world co-ordinate system. Only four independent parameters have to be estimated in order to obtain the projection matrix of the camera. The co-ordinates of the principal point are initialized with the co-ordinates of the geometrical image centre. The 3D point co-ordinates are initialized as the co-ordinates of the barycentre of the points to be reconstructed. Obviously, with simulated data and no noise, the discrepancy between

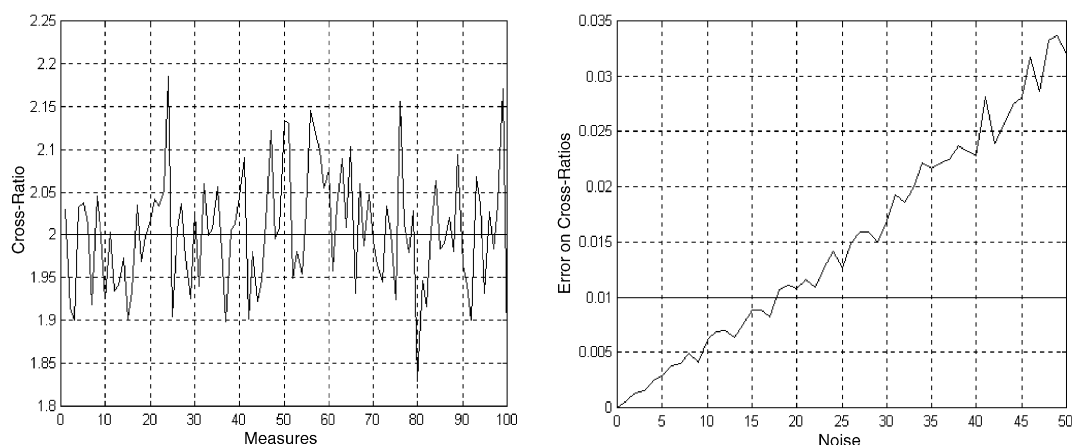


Fig. 6. Cross-ratio stability. 100 computations with a 5% noise level (left) and measuring error with a noise varying from 0% to 50% (right).

the real 3D points and the reconstructed 3D points is nearly zero. The error is due to round off in digital computation. So, in order to analyze the robustness of the method, noise is added on pixel co-ordinates. In a structured light system, the projected image is perfectly known so that error is not present when the point co-ordinates are measured. Then, noise is only added on the pixel co-ordinates of the camera.

The reconstruction was performed on 40 points. Table 1 presents the result for 10 of these 40 points with a uniform noise of $\pm 1\%$.

It has to be noticed that using five known 3D points results degrade quickly. This method appears to be very sensitive to the location of the five points used as landmarks: better results are obtained if no noise is added on landmarks co-ordinates even if noise is added on the co-ordinates of the rest of points.

4.2.2. Euclidean constraints

The reconstruction is performed in two steps: a projective reconstruction assigning the reference points to an arbitrary projective basis and then an Euclidean reconstruction performed from the previous projective reconstruction to which geometrical knowledge about the scene is added. Projective reconstruction provides projection matrices and 3D co-ordinates with respect to a projective frame. In order to validate this reconstruction, the 3D co-ordinates are back-projected onto the image planes through the projection matrices and the residual error is evaluated (see Fig. 7, where projection parameters are given by \mathbf{P}_{proj} and \mathbf{P}'_{proj}). Our conclusion is that projective reconstruction performed well in most cases. However, in order to ensure convergence of the algorithm, the relative positioning of the 3D points must correspond more or less to the configuration of the chosen basis i.e. the Euclidean reference points must be in adequacy with the projective co-ordinates given to them.

We have used different Euclidean constraints as fixing the origin, parallelism, distance, etc. Re-scaling and

re-positioning the computed reconstruction, it is possible to validate Euclidean reconstruction. In a representative example of our experimental results, we found that mean absolute error is less than 8 mm and max absolute error is about 45 mm; the standard deviation is 7.42, 4.76, and 27.08 mm for, respectively, the x -, y - and z -component. The range of each component is $[100; 1000 \text{ mm}]$ for x , $[-400; 1000 \text{ mm}]$ for y and $[500; 4000 \text{ mm}]$ for z .

4.3. Real data

Hereafter, we present some results achieved from real images. The image processing method is described in Ref. [20]. Let us recall the key points. The original coloured image is first converted into the CIE-Lab space. Within the L-image, a self-adaptive thresholding is performed, followed by a morphological squeletization, a Hough transform and the recovering of intersecting points. Within the ab-image, a process to determine the projected colours from the apparent ones is performed which permits to decode the pattern.

The structured light system is composed by an RGB camera, a computer and an electronic slide projector. The coloured pattern is shaped in a 512×512 RGB image which is projected on the measuring scene using the projector, and the scene is then captured by the camera into the computer memory. The reader is pointed to Refs. [4,20] to focus on pattern segmentation and decoding.

In the following, we go on to the reconstruction results, giving some qualitative results.

4.3.1. Scene I

The scene is composed by three geometric and achromatic objects illuminated by the coloured pattern, as shown in Fig. 8. We proceeded in two steps: first a projective reconstruction using the canonical representation, and then a Euclidean reconstruction by adding constraints obtained

Table 1

Errors on reconstruction with uniform noise ± 1

Real co-ordinates			Errors on estimate co-ordinates		
X	Y	Z	ΔX	ΔY	ΔZ
100	−50	4000	0.518	−0.267	3.95
300	−50	2000	−0.65	−0.242	−1.5
700	−50	4000	0.614	−0.33	6.43
500	−400	4020	−1.132	−1.768	−4.332
300	50	4000	0.091	0.397	2.597
500	50	2000	0.076	−0.119	0.449
900	50	4000	0.13	0.171	2.007
300	−430	3000	0.505	−0.911	5.079
450	75	2500	0.76	−1.154	4.016
705	−120	1000	0.603	−0.827	0.829
Mean relative error (%)			0.518	1.539	0.169
Standard deviation			0.610	0.655	3.222

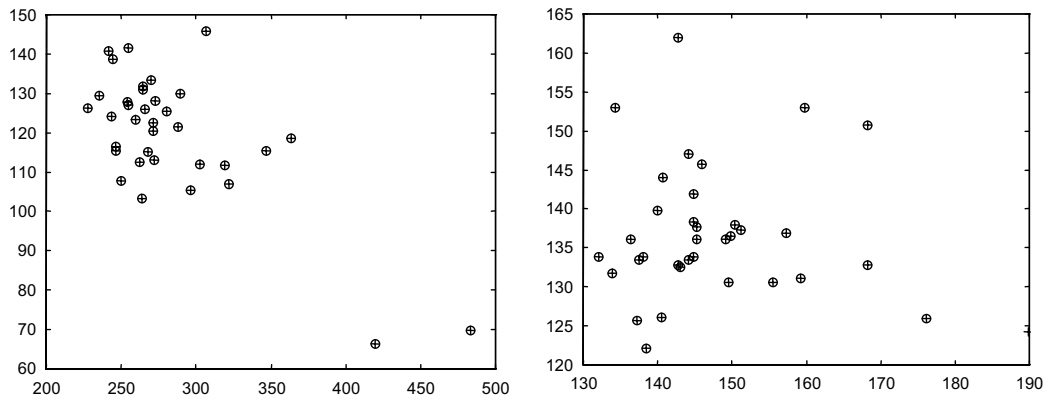


Fig. 7. Validation of the projective reconstruction, left and right image planes. Circles represent real image points and crosses represent projectively reconstructed ones.

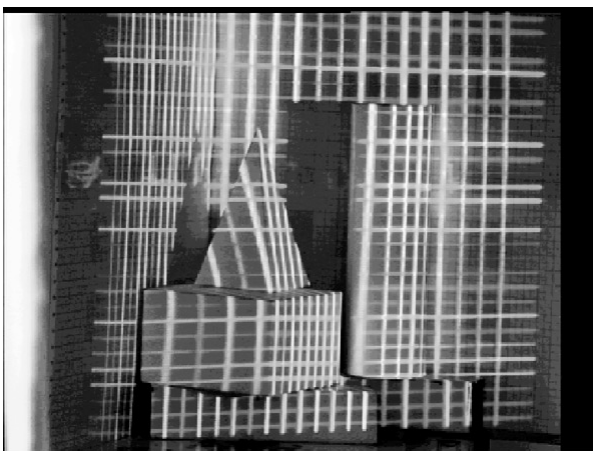


Fig. 8. Structured lighted image of the first scene.

from image analysis (note that, at the time, the constraints are generated manually).

In Fig. 9, we present the back-projection of the projective reconstruction onto the image plane and the projector plane obtained with the linear method (through \mathbf{P}_{proj} and \mathbf{P}'_{proj}). Circles represent the real 2D points and crosses the back-projected ones. The projection matrices and the 3D projective points computed with the linear method are used as initializations for the iterative method: the results are clearly improved as shown in Fig. 10 and quantified in Table 2 (this time \mathbf{P}_{proj} and \mathbf{P}'_{proj} are given by the non-linear method). The maximum absolute 2D error is 3.069 pixels and the mean absolute 2D error is 0.204 pixels in the projector plane and 2.715 pixels and 0.169 pixels, respectively, in the image plane.

At this point, the projective reconstruction is validated (the point 2 in Section 3.5 is performed). We now have to

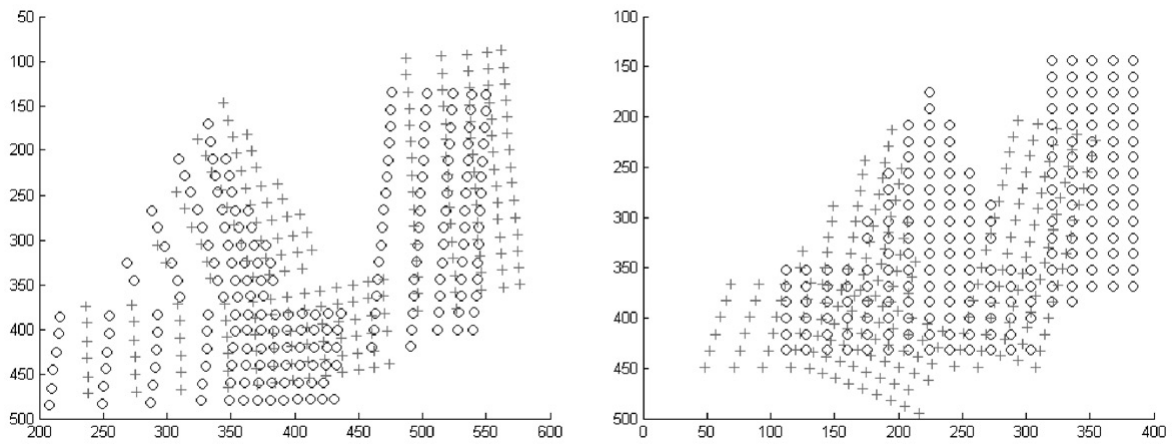


Fig. 9. Back-projection of the linear reconstruction method. Image plane (left) and projector plane (right).

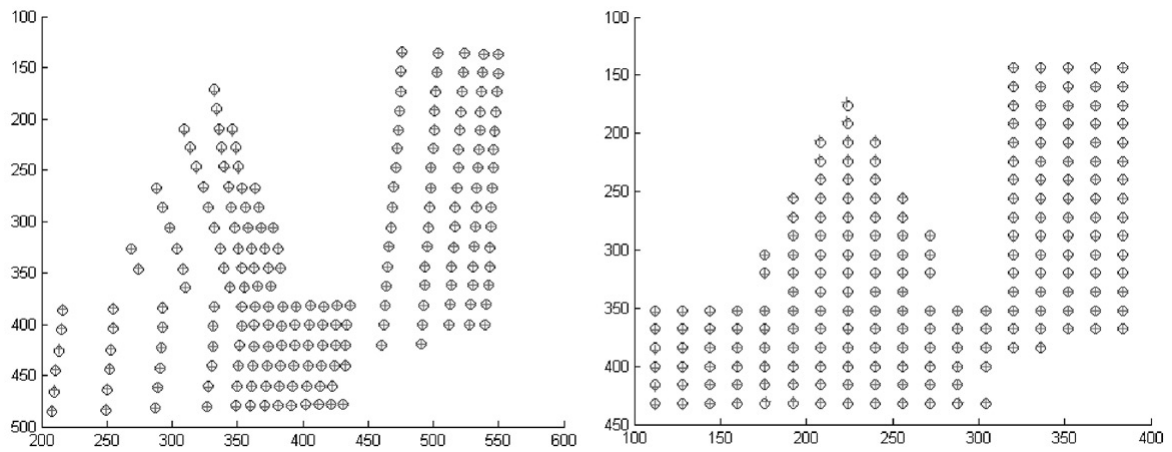


Fig. 10. Back-projection of the non-linear reconstruction method. Image plane (left) and projector plane (right).

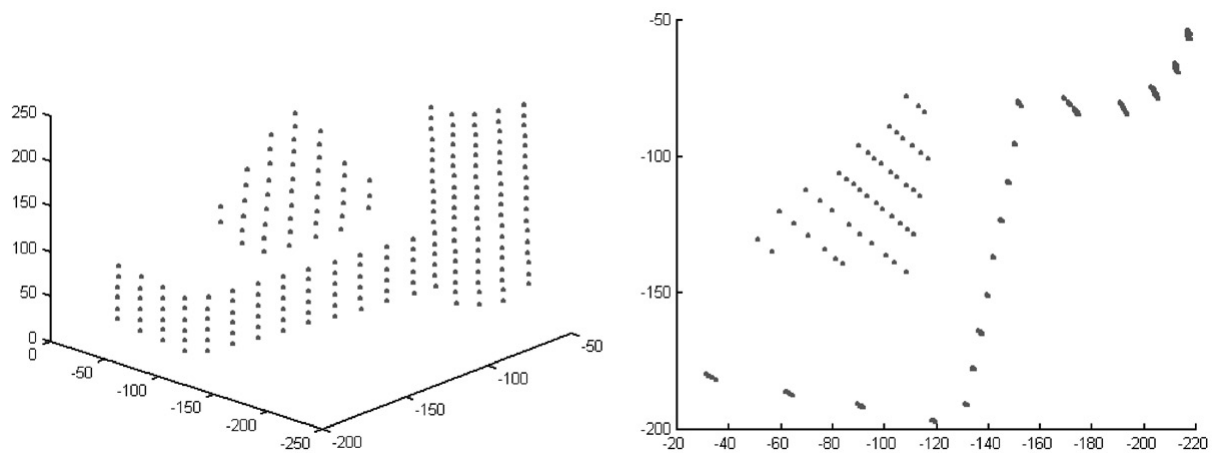


Fig. 11. Two views of the reconstructed scene (I).

Table 2
Residual 2D error

		Max. absolute error (pixels)	Mean absolute error (pixels)
Linear method	Camera	50.084	18.428
	Projector	88.143	32.751
Iterative method	Camera	3.069	0.204
	Projector	2.715	0.169

rectify it into the Euclidean space by following the steps of point 3 in Section 3.5: fix the origin, fix horizontal and vertical planes, generate parallelogram and orthogonality constraint, fix an arbitrary scale factor, etc. The results obtained are shown in Fig. 11. It can be seen that the three objects are globally well-reconstructed. Parallelism and orthogonality are recovered with a sufficient precision and proportions seem to be preserved.

4.3.2. Scene II

The goal of this experimentation is to validate the reconstruction method on a more realistic scene. The one shown in Fig. 12 is grabbed under real conditions of illumination; its size is about $1 \text{ m} \times 1 \text{ m}$. It is in an office environment, the image has been shot under the desk.

The highlighted lines of the figure have been reconstructed; the results are presented in Fig. 13. An arbitrary metric has been assigned, parallelogram and orthogonality

constraints have been generated. Vertices of highlighted polygons are the reconstructed points and the lines which compose it show the geometrical constraints (parallel lines give parallelism constraints, orthogonal lines give orthogonality constraints, etc.)

Similarly to the previous scene, it can be noticed that parallelism and orthogonality are satisfactorily reconstructed, as well as the image proportion.

5. Conclusions

This article presents a method to perform Euclidean reconstruction from an uncalibrated structured light sensor independently of the scene geometry, by assuming that the sensor behaviour is affine or that it can be approximated by an affine camera model. Through pixel correspondences and without knowing neither extrinsic nor intrinsic parameters of the sensor, a projective reconstruction is first computed by choosing five arbitrary points of the scene as a reference frame. Such a reconstruction is only possible up to a projective transformation, which depends on the world reference frame that it has been chosen. Since Euclidean geometry is a particular case of projective geometry, a collineation exists which brings projective reconstruction to Euclidean reconstruction. This collineation can be assessed by translating geometrical information about the scene into constraints on the elements of the collineation matrix. Besides, we show that projecting a known grid pattern of light onto the scene permits to retrieve intrinsic geometrical knowledge about this scene as parallelism and orthogonality. The major

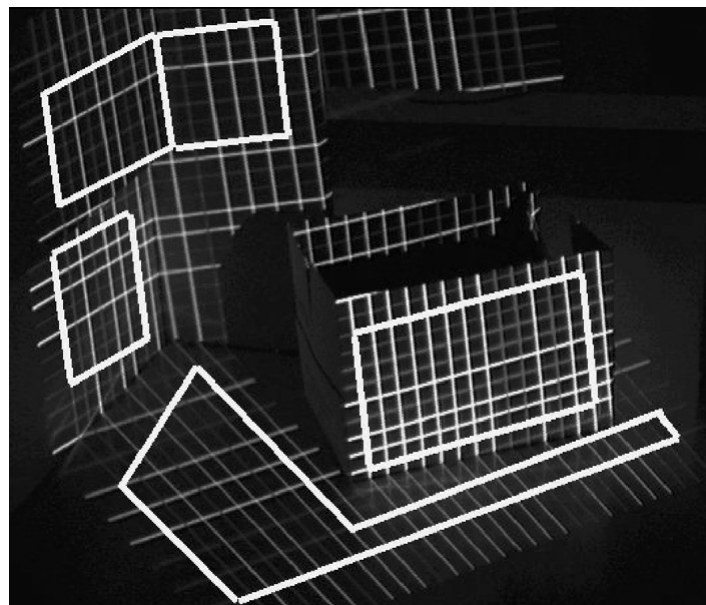


Fig. 12. Structured lighted image of the second scene.

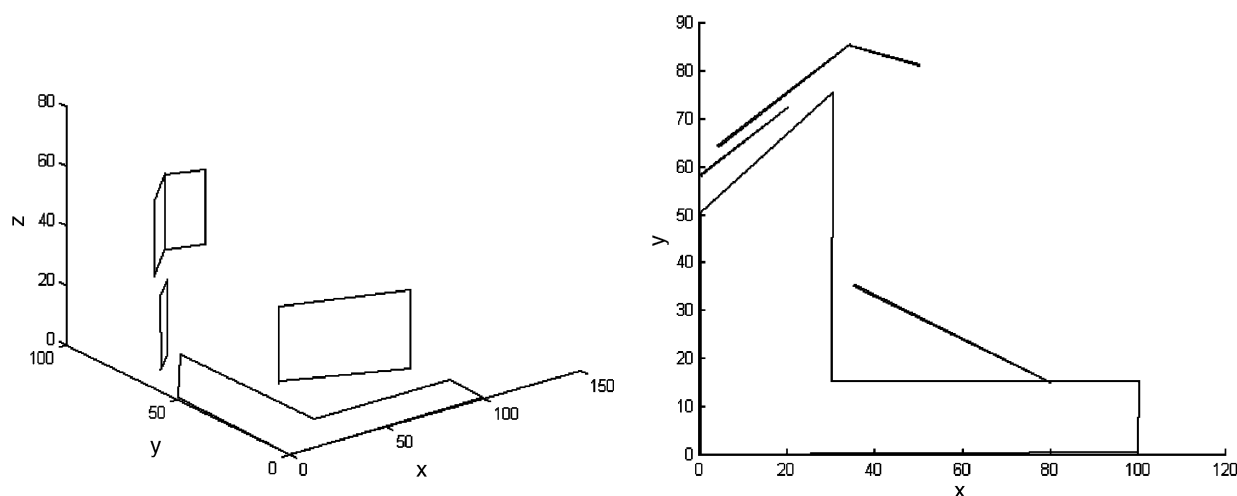


Fig. 13. Two views of the reconstructed scene (II).

contribution of the paper is to show that structured light can be used to deduce geometrical constraints of the scene, which are used to reconstruct the scene without any previous calibration. As no constraint is required on the projection matrices, this approach allows us to reconstruct without considering potential change of the focus, the aperture and the zoom of both the camera and the projector. Structured lighting permits to ensure there is known scene structure which can be used to upgrade the reconstruction to Euclidean and provides numerous constraints which are useful for the convergence of non-linear optimisation methods as Levenberg–Marquardt algorithm.

Experimental results validate the method. However, the automation of the whole process is necessary. A particular care has to be taken in image segmentation, e.g. straight lines, parallels, crossing lines and parallelograms must be accurately extracted from the image. As a further work, we intend to automate the constraints generation, that is, to formulate them mathematically and to solve them from the image segmentation and decoding.

6. Summary

This paper deals with uncalibrated reconstruction through structured lighting. In a structured light system, unlike classical stereovision, the second camera is replaced by a light source that projects a known pattern of light onto the scene. The main goal of this work is to provide adaptive capabilities to this kind of sensor which allows to use it in mobile robotics or dynamic visual inspection. First, we present a survey of the most relevant techniques of uncalibrated reconstruction. It is shown that, due to the fact that any movement of the light source produces a movement of the pattern (i.e. of the 3D points), the reconstruction has to be performed from a single camera shot and a single pattern projection.

Thus, we first focus on a projective reconstruction method based on the canonical representation of views, which requires only pixel correspondences, one view and one pattern projection. The reconstruction is performed in a projective frame, up to a projective transformation.

An Euclidean reconstruction can be recovered from a projective one since Euclidean transformations are a sub-group of projective transformations. In other words, there exists a collineation matrix which permits to pass from projective to Euclidean. This matrix can be assessed by constraining its entries with geometrical knowledge grabbed from the scene. We describe how the pattern projection is used to acquire geometrical knowledge as parallelism, orthogonality, horizontality and verticality. Moreover, structured lighting permits to ensure there is known scene structure which can be used to upgrade the reconstruction to Euclidean and provides numerous constraints which are useful for the convergence of non-linear optimisation methods as Levenberg–Marquardt algorithm. Experimental results, performed both on simulated and real data, are presented and discussed.

References

- [1] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*, MIT Press, Cambridge, MA, 1993.
- [2] O. Faugeras, What can be seen in three dimensions with an uncalibrated stereo rig? *Proceedings of the Second European Conference on Computer Vision*, Santa Maria Ligure, Italia, 1992, pp. 563–578.
- [3] J. Batlle, E. Mouaddib, J. Salvi, Recent progress in coded structured light as a technique to solve the correspondence problem. A survey, *Pattern Recognition* 31 (7) (1998) 963–982.

- [4] J. Salvi, J. Batlle, E. Mouaddib, A robust-coded pattern projection for dynamic measurement of moving scenes, *Pattern Recognition Lett.* 19 (1998) 1055–1065.
- [5] M. Proesmans, L. Van Gool, A. Oosterlinck, One-shot active range acquisition, *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria, 1996, pp. 336–340.
- [6] J.M. Sotoca, M. Buendia, J.M. Iñesta, A new structured light calibration method for large surface topography, *Pattern Recognition Appl. Frontiers Artif. Intell.* 56 (2000) 261–270.
- [7] D.Q. Huynh, R.A. Owens, P.E. Hartman, Calibrating a structured light stripe system: a novel approach, *Int. J. Comput. Vision* 33 (1) (1999) 73–86.
- [8] J.J. Koenderink, A.J. Van Doorn, Affine structure from motion, Technical Report, Utrecht University, Utrecht, Netherlands, October 1989.
- [9] R. Hartley, R. Gupta, T. Chang, Stereo from uncalibrated cameras, *Proceedings of the Conference on Computer Vision and Pattern Recognition*, Urbana-Champaign, Illinois, USA, 1992, pp. 761–764.
- [10] R. Mohr, B. Boufama, P. Brand, Accurate projective reconstruction, *Proceedings of the Second ESPRIT-ARPA-NSF Workshop on Invariance*, Azores, 1993, pp. 257–276.
- [11] O. Faugeras, Q.-T. Luong, S.J. Maybank, Camera self-calibration: theory and experiments, *Proceedings of the Second European Conference on Computer Vision*, Santa Maria Ligure, Italia, 1992, pp. 321–334.
- [12] R.I. Hartley, Kruppa's equations derived from the fundamental matrix, *IEEE Trans. Pattern Anal. Mach. Intell.* 19 (2) (1997) 133–135.
- [13] R. Hartley, Euclidean reconstruction from uncalibrated views, *Lecture Notes in Computer Science*, Vol. 825, 1994, pp. 237–256.
- [14] A. Heyden, K. Åström, Euclidean reconstruction from constant intrinsic parameters, *Proceedings of the International Conference on Pattern Recognition*, Vienna, Austria, 1996, pp. 339–343.
- [15] B. Boufama, R. Mohr, F. Veillon, Euclidean constraints for uncalibrated reconstruction, *Proceedings of the Fourth International Conference on Computer Vision*, Berlin, Germany, 1993, pp. 466–470.
- [16] Z. Zhang, K. Isono, S. Akamatsu, Euclidean structure from uncalibrated images using fuzzy domain knowledge: application to facial images synthesis, *Proceedings of the International Conference on Computer Vision*, Bombay, India, 1998, pp. 784–789.
- [17] Q.-T. Luong, T. Viéville, Canonic representations for the geometries of multiple projective views, *Proceedings of the Third European Conference on Computer Vision*, Stockholm, Sweden, 1994, pp. 589–599.
- [18] D.W. Marquardt, An algorithm for the estimation of nonlinear parameters, *Soc. Ind. Appl. Math. J.* 11 (1963) 431–441.
- [19] K.E. Åström, L. Morin, Random cross ratios, *Proceedings of the Ninth Scandinavian Conference on Image Analysis*, Uppsala, Sweden, 1995, pp. 1053–1061.
- [20] D. Fofi, Navigation d'un véhicule intelligent à l'aide d'un capteur de vision en lumière structurée et codée, Ph.D. Thesis, CREA, Université de Picardie Jules Verne, 2001.

About the Author—DAVID FOFI received the M.Sc. degree in Computer Science (computer vision, signal and image processing) in the University of Cergy-Pontoise, ENSEA, in 1997. Since 1998, he is involved in the study of structured light vision for mobile robotics. In September 2001, he received the Ph.D. degree in Robotics and Computer Vision from the University of Picardie Jules Verne. Since September 2002, he works as an associate professor in the Le2i—IUT Le Creusot.

About the Author—JOAQUIM SALVI received the B.Sc. degree in Computer Science in the Polytechnic University of Catalunya in 1993. He received the M.Sc. degree in Computer Science in 1996 and the Ph.D. degree in Electrical Engineering in 1998 both from the University of Girona. At present he is an associate professor in the Electronics, Computer Science and Automation Department of University of Girona where he is involved in research projects on mobile robotics and computer vision. His current interests are in the field of stereovision, structured light and scene reconstruction.

About the Author—EL MUSTAPHA MOUADDIB received the Ph.D. degree in robotics at the University of Picardie Jules Verne, Amiens, France, in 1991 and HDR in December 1999. He is professor at the same University where he teaches Robotics and Computer Vision. His main interests are computer vision, omnidirectional vision, mobile robotics and perception. Since 2001, he is the head of the “Centre de Robotique, d'Electrotechnique et d'Automatique”, where he is involved in research projects on robot navigation with a conic sensor and in particular localisation and obstacle detection.