



Modélisation 3D d'objets par un capteur visuel déplacé par un opérateur

Benjamin Coudrin

► To cite this version:

Benjamin Coudrin. Modélisation 3D d'objets par un capteur visuel déplacé par un opérateur. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université Toulouse 3 Paul Sabatier, 2011. Français. NNT: . tel-01201505

HAL Id: tel-01201505

<https://theses.hal.science/tel-01201505>

Submitted on 18 Sep 2015

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.



Université
de Toulouse

THÈSE

En vue de l'obtention du DOCTORAT DE L'UNIVERSITÉ DE TOULOUSE

Délivré par :

Université Toulouse 3 Paul Sabatier (UT3 Paul Sabatier)

Discipline ou spécialité :

Systèmes Embarqués

Présentée et soutenue par :

Benjamin Coudrin

le : mercredi 2 mars 2011

Titre :

Modélisation 3D d'objets par un capteur visuel déplacé par un opérateur

Ecole doctorale :

Systèmes (EDSYS)

Unité de recherche :

LAAS-CNRS / ICA

Directeur(s) de Thèse :

Michel Devy

Jean-José Orteu

Rapporteurs :

Françoise Prêteux

Yvon Voisin

Autre(s) membre(s) du jury

Michel Dhome

Jean-Pierre Jessel

Ludovic Brèthes

à Jean-Michel

Remerciements

Les travaux présentés ici sont issus d'une démarche personnelle et de plusieurs années de travail. Si mon nom est sur la couverture, ce travail est, néanmoins, le fruit de l'accompagnement, du soutien moral, technique ou financier, ainsi que, parfois, de l'acharnement d'un réseau de personnes que je tiens à associer à la réussite de ce projet.

Tout d'abord, je tiens à remercier Michel Devy, Directeur de Recherche au CNRS, et Pr. Jean-José Orteu, mes directeurs de thèse, pour notre collaboration et pour avoir cru en mes idées et en ma capacité à les réaliser. J'espère que nous aurons l'occasion de retravailler ensemble.

Je souhaite également remercier chaleureusement Dr. Ludovic Brèthes, Directeur Scientifique de la société Noomeo, pour avoir mené ce projet avec moi et m'avoir accompagné tout au long de sa réalisation.

Pour avoir accepté de rapporter sur ces travaux ainsi que pour leurs conseils, je tiens à exprimer ma reconnaissance aux Prs. Françoise Prêteux et Yvon Voisin. De même, je salue l'aide et le soutien de Michel Dhome, Directeur de Recherche au CNRS, et Pr. Jean-Pierre Jessel, pour leurs avis, les idées qu'ils m'ont données et pour avoir pris part à mon jury de thèse.

Bien entendu, je tiens à remercier les appuis structurels et financiers de ces travaux, en commençant par Dr. Vincent Lemonde, Président de la société Noomeo, pour avoir monté ce projet, financé cette thèse mais aussi pour son implication personnelle et scientifique. Je remercie, de plus, Raja Chatila, Directeur de Recherche au CNRS, pour m'avoir accueilli au sein du LAAS-CNRS.

Je tiens à exprimer toute ma gratitude au Dr. Benoît Bocquillon pour son aide et sa collaboration directe dans mes travaux. J'adresse également toute ma reconnaissance aux collaborateurs de la société Noomeo, qui se sont tous impliqués dans ma réussite : MM. Pierrick Le Quément, Matthieu André, Romain Rioux, Mlles Sophie Neveu, Hélène Mis, MM. Sébastien Bach, Benjamin Amanrich, Olivier Miazek, Dr. Frédéric Courteille, M. Rémi Parlouar, Dr. Jacques Harvent et Mlle Ludivine Vergonnier.

De même j'adresse toute ma reconnaissance aux confrères et consœurs du LAAS-CNRS. Particulièrement, je remercie MM. Nizar Sallem, Benoît Ducarouge, Julien Bonnal, Adrien Durand-Petiteville ainsi que les Drs. Mathias Fontmarty, Thierry Germa, Xavier Broquère et Mokhtar Gharbi.

Enfin, je n'oublie pas ma famille, mes amis et tous mes relecteurs. Je leur dois un support sans faille et une patience inhumaine.

Merci au Dr. Jorge Cham et à PhDComics.

Et merci à tous.

Table des matières

Introduction	15
1 La numérisation 3D et ses applications	15
2 Principales techniques de numérisation	16
3 Quelques systèmes existants	18
4 Spécifications de notre projet	21
5 Principales contributions	23
6 Organisation du manuscrit	23

1	
Description du capteur et méthodologie d'utilisation	
1.1 Conception mécanique et électronique du système	26
1.1.1 Description de la conception et des composants	26
1.1.2 Mécanismes d'acquisition	29
1.2 Mode d'acquisition nominal	32
1.2.1 Description de la séquence	32
1.2.2 Revue des données	33
1.3 Cas d'utilisation	34
1.3.1 Interactions avec l'appareil	34
1.3.2 Analyse opérationnelle	37
1.3.3 Problématiques	39

2	
Stéréovision	
2.1 Bases pour la stéréovision	42
2.1.1 Modèle de caméra	42
2.1.2 Modèle d'une paire stéréoscopique	45
2.1.3 Géométrie épipolaire	46
2.1.4 Rectification	47
2.2 Mise en correspondance dans les images stéréoscopiques	47
2.2.1 Mise en correspondance dense	47
2.2.2 Corrélation précise	49
2.2.3 Mise en correspondance de primitives	51
2.3 Reconstruction 3D	52
2.3.1 Chaîne de reconstruction	52
2.3.2 Triangulation	53

3	
Recalage paire à paire	
3.1 Recalage grossier	56
3.1.1 État de l'art	56
3.1.2 Estimation de pose par appareillage	57
3.1.3 Estimation de pose par mise en correspondance de descripteurs locaux	58

3.1.4	Estimation de pose par approche globale	64
3.1.5	Méthode hybride mise en œuvre pour le recalage rapide	67
3.2	Recalage précis	69
3.2.1	État de l'art	69
3.2.2	Iterative Closest Points	71
3.2.3	Point-to-plane	74
3.2.4	Point-to-surface	75
3.2.5	Interpolation dans les images, approche pyramidale	75
3.3	Évaluations	78
3.3.1	Pièce « Bunny »	78
3.3.2	Pièce « Étalon »	81
3.4	Conclusions	82

4

Approche globale pour le recalage

4.1	État de l'art	90
4.2	Présentation liminaire	92
4.2.1	Problématique	92
4.2.2	Données et pré-traitements	93
4.3	Champs de rotations	93
4.3.1	Sélection de vue	94
4.3.2	Score de reprojection	96
4.3.3	Erreur inertielle et correction	97
4.3.4	Algorithme complet	98
4.4	Propagation par graphe	99
4.4.1	Détection de fermeture de boucle	99
4.4.2	Optimisation à la fermeture	101
4.5	Optimisation hors-ligne	102
4.5.1	Stratégie multi-vues	102
4.5.2	Critère	102
4.5.3	Sélection de points et pondération	103
4.6	Évaluations	104
4.7	Conclusion	105

5

Évaluations quantitatives

5.1	Protocole expérimental	114
5.2	Cylindre étalon	115
5.3	Statue	116
5.4	Pièce emboutie	117
5.4.1	Comparaisons au modèle CAO	119
5.4.2	Comparaison avec le modèle B	119
5.5	Lien avec la qualité de reconstruction	120
5.6	Conclusion	122

6	
Vers l'analyse de qualité pour la sélection de point de vue	
6.1	Évolution de la qualité d'un modèle 126
6.2	Sélection de point de vue 128
6.2.1	Méthodes notables 128
6.2.2	Fonction d'utilité d'un point de vue 129
6.3	Conclusions 133
Conclusions et perspectives	135
Bibliographie	139

Table des figures

1	Proposition de classification des méthodes de numérisation 3D	17
2	Polhemus FastSCAN TM	19
3	Creaform HandyScan 3D TM REVscan TM	19
4	Leica [®] T-Scan [®] TS50	20
5	3D-Modeller (3DMo)	20
6	GOM ATOS III	21
7	Konica Minolta VIVID 9i	21
1.1	Le capteur OptiNum TM	26
1.2	Plan transversal d'OptiNum TM	27
1.3	Le motif aléatoire projeté sur la scène est basé sur un bruit de Perlin [Orteu 06]	28
1.4	Construction d'une séquence contenant des scénarios différents. (a) scénario 1 : il dure 10 <i>ms</i> ; la première est utilisée par le séquenceur, puis on déclenche une mesure inertielle synchronisée avec une exposition des caméras et l'illumination secondaire durant 6 <i>ms</i> . (b) scénario 2 : il dure 10 <i>ms</i> ; la première est toujours utilisée par le séquenceur, puis on déclenche une mesure inertielle. (c) Composition des scénarios dans la séquence : le scénario 1 est joué une fois puis le scénario 2 est répété 9 fois ; on obtient ainsi une mesure inertielle toutes les 10 <i>ms</i> (100 <i>Hz</i>) et une capture des caméras toutes les 100 <i>ms</i> (10 <i>Hz</i>) avec un temps d'exposition de 6 <i>ms</i> par capture. Ce type de scénario permet, par exemple, le suivi du mouvement par intégration de la mesure inertielle dans un programme exécuté sur l'ordinateur.	30
1.5	États et transition du trigger	31
1.6	Description de la séquence du mode « normal » de fonctionnement. Ici, t_e correspond au temps d'exposition des caméras, et T est la période d'acquisition. Le temps final est égal à la période d'acquisition tronquée du temps d'exposition pour chacune des deux acquisitions sur les caméras ainsi que les trois millisecondes nécessaires aux changements de contexte du séquenceur. Dans ce scénario, la centrale inertielle réalise matériellement l'intégration de ses mesures pour le suivi de mouvement. On récupère donc directement le résultat du suivi.	33
1.7	Mode de prévisualisation temps-réel	34
1.8	Mode de modélisation incrémentale	35
1.9	Mode de relocalisation du capteur	36
1.10	Mode d'aide à l'opérateur pour compléter le modèle	36
1.11	Mode de finalisation	37
2.1	Représentation du modèle sténopé et de la projection d'un point dans l'image	43
2.2	Organisation des repères d'un capteur stéréoscopique	45
2.3	(a) Les projections d'un point du monde dans les deux images sont coplanaires avec ce point et les centres optiques. (b) Le lieu géométrique du stéréo-correspondant d'un point image est une ligne épipolaire.	46

2.4	Le point $\mathbf{X}$ n'est pas observable par la caméra $\mathcal{C}_1$ bien qu'il admette $\mathbf{q}_n$ comme projeté. Ainsi la paire $(\mathbf{p}_i, \mathbf{q}_n)$ est valide pour la triangulation mais aura une mesure de similitude $\delta(\mathbf{p}_i, \mathbf{q}_n)$ très faible puisque les caméras observent des parties différentes de la scène.	48
2.5	Autocorrélation par le score ZNCC (Zero-mean Normalized Cross-Correlation) [Solà Ortega 07]. On corrèle une imagerie de 15×15 pixels avec son voisinage horizontal. <i>Gauche</i> : une texture pauvre induit une autocorrélation forte dans tout le voisinage et de ce fait une mauvaise discrimination. <i>Centre</i> : dans une zone d'information riche la discrimination est facile et offre un pic de corrélation large permettant une mise en correspondance correcte lors de variations raisonnables d'apparence. <i>Droite</i> : une texture de type <i>bruit blanc</i> ou approchant fait tendre la fonction d'autocorrélation vers un pic de Dirac, rendant difficile sa détection dès lors que des légers changements d'apparence se produisent.	49
2.6	Carte de disparité illustrant la mise en correspondance de points dans une paire d'images	50
2.7	Approximations de la déformation de la région de référence au premier et au second ordre dans le cas rectifié.	51
2.8	Variations de luminance sur une ligne d'une paire d'images. Les pics de luminance sont extraits et mis en correspondance.	52
2.9	Chaine de reconstruction par stéréovision.	52
3.1	Point signature [Chua 97]. (a) Définition du contour de surface (b) Expression des points du contour dans la nouvelle référence (c) Signature du point. . . .	59
3.2	Spin-images de quelques points d'un modèle [Johnson 97]	60
3.3	Descripteur SURF [Bay 06], Le voisinage est décomposé en 4×4 régions pour le calcul des réponses d_x et d_y , chaque région est décrite par les sommes des réponses et les sommes des valeurs absolues des réponses.	63
3.4	Chaque acquisition d'images – avec ou sans illumination – donne lieu à l'estimation d'une attitude par la centrale inertielle. En composant les informations on détermine la pose de recalage recherchée	67
3.5	Mise en correspondance de points d'intérêt entre les images coaxiales aux temps t et $t + \tau$ après rectification perspective de l'image au temps $t + \tau$. (a) Avant filtrage. (b) Après filtrage.	70
3.6	La distribution des directions d'appariements permet d'éliminer les paires les plus incohérentes	71
3.7	Point-to-plane	74
3.8	Mise en correspondance par calibrage inverse. La surface $\mathbf{S}_t$ peut être reconstruite depuis la paire à l'instant t et la surface $\mathbf{S}_{t+\tau}$ peut être reconstruite depuis la paire à l'instant $t + \tau$. Le correspondant $\mathbf{q}$ du point $\mathbf{p}$ est la projection de ce dernier sur la surface $\mathbf{S}_t$ selon le point de vue du capteur à l'instant t . La fonction de pondération appliquée à cette paire est liée à la distance du point $\mathbf{q}$ à la normale ponctuelle au point $\mathbf{p}$	77
3.9	Lors de la recherche de correspondant on reconstruit une partie du voisinage du point image déterminé. Le point 3D que l'on apparie finalement est déterminé selon une distance point à point.	78
3.10	Erreur RMS de recalage pour la pièce « Bunny »	79

3.11	Évolution de la pose estimée. La ligne rouge représente la vérité terrain. . . .	80
3.12	Nuages de points de la scène test « Bunny ». (a) Estimée initiale (b) ICP (c) Méthode de Chen (d) Calibrage inverse (e) Notre méthode.	84
3.13	Paires d'images sources de la pièce de test « Étalon » (a) Vue n (b) Vue $n + 1$	85
3.14	Erreur RMS de recalage pour la pièce « Étalon »	85
3.15	Nuages de points de la scène test « Étalon ». (a) Estimée initiale (b) ICP (c) Méthode de Chen (d) Calibrage inverse (e) Notre méthode.	86
3.16	Comparaison des nuages recalés au modèle CAO. (a) ICP (b) Méthode de Chen (c) Calibrage inverse (d) Notre méthode.	87
3.17	Distributions des erreurs par comparaison au modèle théorique	88
4.1	Données et pré-traitements pour l'acquisition à l'instant n . Les informations stockées pouvant être utilisées sont identifiées en rouge.	94
4.2	Application de la déformation sur les coins de l'image et le centre. On applique une translation pour ramener le centre transformé au centre réel de l'image. .	96
4.3	Algorithme de sélection de vue et recalage	98
4.4	Principe de la fermeture de boucle. (a) Lors de l'acquisition des neuf premières vues, on construit le graphe de relations en ligne. (b) Lorsque la possibilité de cycle est détectée, on fait l'ajout de la relation $\mathbf{T}^{91}$ pour effectivement créer ce cycle.	99
4.5	Après recalages successifs, la position estimée de la dixième vue est proche de celle de la première. On va chercher la présence d'une fermeture de boucle sur ces vues.	100
4.6	Consistance par n -uplet. La composition de toutes les transformations d'un cycle dans le sens de l'orientation du graphe est une identité. De plus $\mathbf{T}^{ac} = \mathbf{T}^{ab}\mathbf{T}^{bc}$	101
4.7	Pour une position de caméra, les points intervenant dans l'optimisation de sa pose sont choisis en fonction du maillage global et des normales locales. <i>En rouge</i> : les points dont l'angle entre leur normale et la direction de visée est trop importante ne sont pas observables. <i>En bleu</i> : le rayon liant le centre optique et le point e est occulté par une maille du modèle, il n'est pas observable non plus.	103
4.8	Séquence du bras de fauteuil. (a) Sans fermeture de boucle, l'erreur se propage (b) Après optimisation, l'erreur a été corrigée sur l'ensemble du modèle . . .	105
4.9	Variations des paramètres de pose estimés à chaque acquisition.	106
4.10	Correction apportée par l'optimisation finale.	107
4.11	Trajectoires du capteur projetées sur les plans (X, Y) , (Y, Z) , (X, Z)	108
4.12	Résultat sur la séquence complète (nuage de point en niveaux de gris)	109
4.13	Résultat sur la séquence complète (maillage en niveaux de gris)	110
4.14	Détail des plis du vêtement (faces et arêtes du maillage)	111
4.15	Détail de l'épaule et du dossier du fauteil (points, arêtes, faces et faces en niveaux de gris)	112
5.1	Statue utilisée pour notre évaluation	116
5.2	Résultat de numérisation après la modélisation incrémentale et avant la finalisation.	117
5.3	Modélisation finalisée et texturée de la statue à l'aide de notre système	118

5.4	Detail de l'objet. Dans les zones gauches, certaines faces peuvent être difficiles à observer du fait des occultations. (a) L'observation de surfaces gauches avec notre système est facilitée par son mode de manipulation. (b) Numérisation par le système <i>A</i> , avec une manipulation plus difficile, des manques d'information peuvent se produire.	118
5.5	Pièce métallique emboutie utilisée pour nos tests.	119
5.6	Comparaisons des modèles numérisés à la surface théorique. (a) Système <i>A</i> (b) Système <i>B</i> (c) Notre système tenu à la main	120
5.7	Distribution de l'erreur entre les modèles 3D et la surface CAO	121
5.8	Comparaison de nos résultats avec le scan <i>B</i> . (a) Avant finalisation. (b) Après finalisation	122
5.9	Erreur finale et scores de reconstruction sur une seule vue. Les zones grises sont les parties du modèle CAO où aucun point ne se projette. (a) Carte de scores de reconstruction. (b) Carte d'erreurs par projection sur le modèle CAO selon deux points de vue différents afin de voir l'ensemble du cylindre intérieur. . .	123
6.1	Visuels de la pièce « Moule »	127
6.2	Evolution de la carte de scores par aggrégation successive des vues. En direction du vert : amélioration du score. En direction du rouge : dégradation du score.	128
6.3	Types de voxels [Restrepo Specht 05]	130
6.4	Fonction d'utilité atomique pour un type de voxel et un point de vue donnés. La fonction est maximale pour un pourcentage α de ce type de voxel observables dans la vue candidate.	132

Introduction

LE cadre de notre étude est la conception d'un capteur de numérisation 3D tenu à la main et simple d'utilisation pour une personne non spécialiste en perception 3D. Le projet OptiNumTM a été initié pour répondre à la généralisation de l'usage de la modélisation 3D dans de nombreux domaines. Le recours à cette technologie s'est grandement développé depuis quelques années, tant dans les processus industriels ou le design graphique que dans la conservation du patrimoine ou la santé. Il a donc été nécessaire de proposer de nouvelles méthodes d'acquisition 3D afin de s'adapter à un public d'utilisateurs peu spécialistes souhaitant trouver un gain substantiel de productivité et d'efficacité. Ainsi, les problématiques essentielles de la numérisation 3D (recalage, calibrage, reconstruction, modélisation ...) ont dû être repensées pour s'intégrer dans un système le plus simple possible pour un utilisateur.

Dans un premier temps, il convient de présenter ce projet, les principaux choix structurants pour la conception ainsi que son positionnement par rapport aux applications et aux solutions existantes.

1 La numérisation 3D et ses applications

Le problème de la modélisation 3D à partir d'un capteur mobile a trait à l'acquisition de données 3D à partir de positions distinctes de ce capteur, à leur agencement cohérent avec la réalité de la scène à modéliser, puis à leur représentation sous une forme géométrique (points, arêtes, surfaces ...) dépendante de l'exploitation prévue pour le modèle final.

La phase d'acquisition est une étape sensorielle permettant d'obtenir une représentation partielle et bruitée de la scène réelle à modéliser. On exploite généralement des méthodes basées sur des capteurs télémétriques laser ou vision (avec ou sans projection de lumière). Dans notre cas, il s'agit de travailler sur notre propre capteur basé vision doté d'un système optoélectronique de numérisation qui permet l'acquisition de la géométrie et de la texture d'une scène sous la forme de nuages de points 3D. Ce capteur, décrit au chapitre 1, est conçu pour être déplacé à la main.

Une seule acquisition 3D produit une représentation parcellaire de l'espace qu'il faudra combiner avec d'autres acquisitions, prises sous d'autres points de vue. Fournir une modélisation réaliste passe donc par une étape de recalage de ces différentes vues. Cela impose de déterminer la position du capteur dans un repère global, appelé « repère du monde », au moment de chaque prise de vue.

Parmi les applications importantes de la numérisation 3D, on distingue celles liées à l'industrie. Les domaines de la mécanique et de l'ingénierie ont utilisé très tôt les technologies 3D, pour la conception particulièrement. Le recours à la numérisation dans ce contexte a plusieurs objectifs. Pour la veille technologique ou le « reverse engineering », la numérisation 3D offre un gain de temps et de résultats dans les processus de rétro-conception. Le contrôle dimensionnel pour la production permet d'analyser les performances d'un procédé de fabrication de manière rapide. Dans ce domaine, le couplage à la robotique est fréquent afin d'automatiser le contrôle.

D'autres secteurs industriels ont désormais généralisé l'usage de la numérisation 3D dans leurs processus. En dehors de notre étude, basée sur la modélisation de surfaces, il convient de

citer les méthodes de modélisation volumétriques, comme la tomographie, largement utilisées pour l'analyse des matériaux par exemple.

On peut également évoquer les applications médicales et la conservation du patrimoine, pour lesquelles des produits commerciaux spécifiques voient le jour. Ceci témoigne de l'importance qu'ont ces applications dans le monde de la numérisation 3D. L'imagerie médicale utilise depuis longtemps maintenant la tomographie [Fetita 04], par mesure à rayons X ou résonance magnétique, pour la numérisation 3D, mais il existe également des applications surfaciques en modélisation endoscopique ou pour l'orthoprothèse notamment.

Pour les applications dans le domaine du patrimoine, les méthodes non destructives de numérisation permettent l'analyse de sites ou d'objets historiques précieux ou fragiles sans manipulation de la pièce. Pour cela, le couplage aux techniques d'impression 3D est fréquent afin de réaliser des reproductions d'objets. De nombreux travaux scientifiques de modélisation 3D ont été menés autour de projets de conservation du patrimoine, tels que le projet Michelangelo de l'Université de Stanford [Levoy 00] par exemple.

Pour conclure sur les applications de la numérisation 3D, nous pouvons évoquer l'industrie cinématographique et celle du jeu vidéo. L'utilisation de systèmes de numérisation permet un gain de temps considérable comparée à la modélisation manuelle d'objets destinés à être inclus dans les films en images de synthèse, les trucages vidéos ou les jeux. On a également recours à la perception 3D pour les applications en réalité virtuelle ou réalité augmentée.

2 Principales techniques de numérisation

Il existe aujourd'hui un grand nombre de techniques permettant de réaliser la numérisation d'une scène par modélisation 3D. On classe généralement ces techniques en distinguant celles qui nécessitent un contact du capteur avec la pièce à numériser, et celles sans contact (Figure 1). Les capteurs à *contact* peuvent nécessiter la destruction de la pièce ou font l'acquisition de la géométrie en faisant parcourir la surface par un outil dont on mesure ensuite la position (palpage, machine à mesurer tridimensionnelle ...).

On sépare souvent [Curless 00] les capteurs *sans contact* selon qu'ils nécessitent l'ajout d'énergie à la scène (capteur *actif*) ou qu'ils effectuent leurs mesures en fonction de l'énergie déjà présente (capteur *passif*) dans l'environnement.

Une méthode classique de numérisation active consiste à utiliser la *triangulation optique* par illumination laser. On projette à l'aide d'un laser une primitive facilement détectable (point, ligne, croix ...). La scène illuminée est observée par une caméra dans laquelle on détecte la projection de la primitive. La mesure 3D est obtenue par triangulation, les positions relatives entre la caméra et le projecteur étant connues. On peut également acquérir l'information multispectrale pour la reconstruction [Mansouri 07].

Dans les systèmes par *temps de vol*, le numériseur émet une impulsion laser et mesure le temps qu'elle met à revenir au capteur. La vitesse de la lumière dans l'air étant connue, on obtient la distance parcourue par l'impulsion, soit le double de la distance du capteur au point mesuré.

Les systèmes qui exploitent la *modulation d'amplitude* remplacent les impulsions par une émission continue d'un faisceau modulé en amplitude, par une sinusoïde par exemple. Le signal reçu est également une sinusoïde déphasée en fonction du temps de parcours, et donc de la distance.

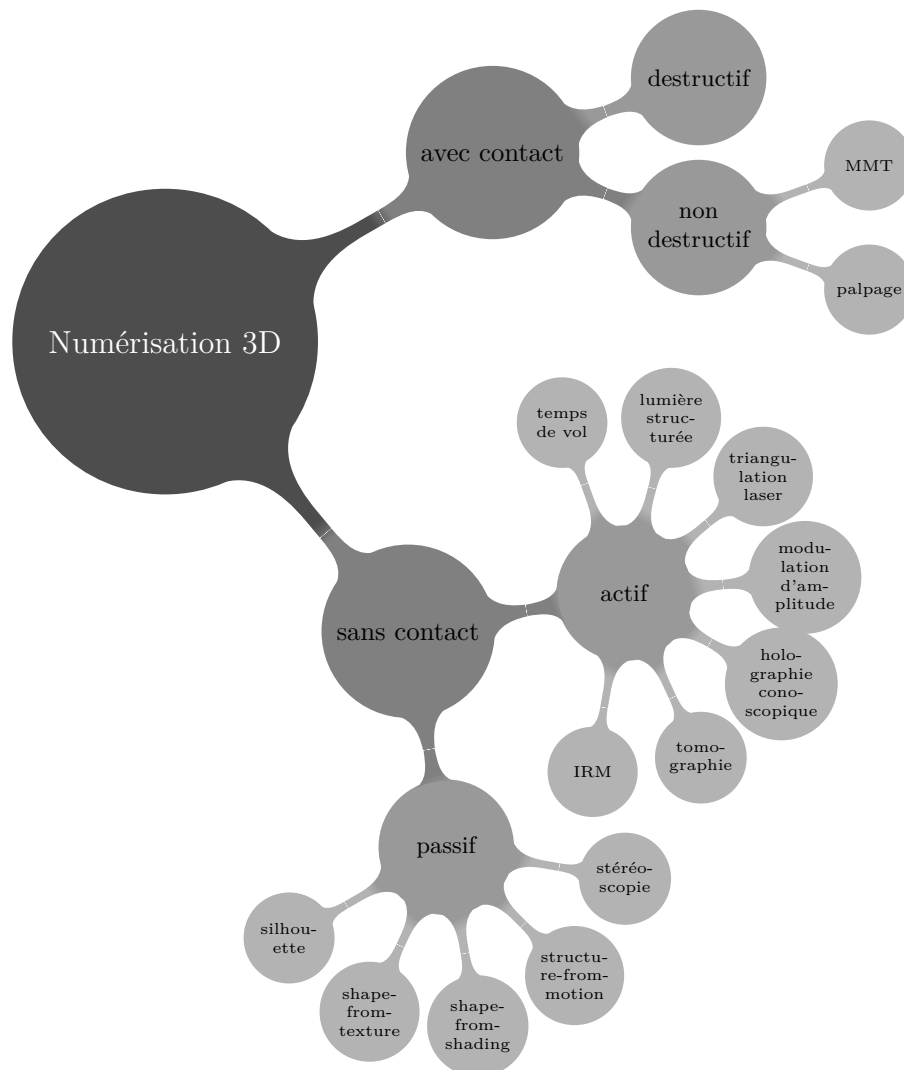


FIGURE 1 – Proposition de classification des méthodes de numérisation 3D

D'autres techniques courantes remplacent l'émission laser par un motif de lumière structurée. En observant la déformation de ce motif projeté sur la scène, il est possible d'en inférer la géométrie. L'émission lumineuse peut être faite à l'aide d'un projecteur LCD, une projection de diapositive ou un balayage laser.

Le recours à l'holographie conoscopique permet également des applications de mesure tridimensionnelle. Il existe, de plus, diverses autres méthodes utilisant l'émission de rayons X (tomographie) ou la résonance magnétique (IRM). Celles-ci sont très utilisées en imagerie médicale pour la mesure volumétrique.

Les méthodes *passives* n'ont pas recours à l'émission de lumière ou de quelque forme de radiation que ce soit mais utilisent la lumière ambiante. Une grande majorité des systèmes optiques utilisent l'imagerie en lumière visible mais il existe des applications exploitant la lumière infrarouge, notamment en bande III [8, $12\mu m$] permettant de percevoir dans l'obscurité.

L'utilisation de plusieurs caméras, appelée *stéréovision* ou *vision multi-oculaire*, ramène le

problème de la mesure 3D à la mise en correspondance entre deux ou plusieurs images acquises selon différents points de vue [Hartley 04]. L'intersection des rayons tirés depuis les centres optiques et passant par les points images correspondants permet de calculer la mesure 3D par triangulation. En n'utilisant qu'une seule caméra que l'on déplace, le problème reste similaire mais le nombre de rayons peut être très grand. De plus, le mouvement entre les points de vue est alors inconnu et doit être estimé en même temps que les mesures 3D des points. On parle, pour ce type de méthodes, de *structure-from-motion*, et de *bundle adjustment* pour la technique d'optimisation permettant de trianguler et de trouver les points de vue [Triggs 00].

D'autres méthodes exploitent l'observation de connaissances supplémentaires. Dans le cas du *shape-from-shading*, on infère l'information sur l'orientation des normales à la surface observée dans une image en regardant l'ombrage que le relief provoque [Prados 05]. Cette méthode suppose que les variations sur la texture de l'objet mesuré sont négligeables face aux variations de réflectance. La technique du *shape-from-texture* observe les distorsions de la texture lorsqu'elle est projetée d'un espace 3D au plan image 2D pour en extraire les paramètres de variations de la surface [Forsyth 02]. La méthode du *shape-from-silhouette* utilise la silhouette d'un objet extraite par segmentation entre la forme et le fond, supposés contrastés. En observant l'objet selon plusieurs points de vue successifs, on peut déterminer un volume englobant de l'objet. Celui-ci est une approximation plus ou moins précise de la géométrie de l'objet selon la résolution de la *voxel map* utilisée pour discrétiser le volume [Szeliski 93].

3 Quelques systèmes existants

Face à une demande croissante dans les domaines de la rétro-conception et de la mesure mécanique pour le contrôle dimensionnel, de nombreux systèmes de numérisation 3D ont vu le jour ces dernières années. Parmi les technologies développées, on trouve notamment les systèmes de mesure optique sans contact. Fondés sur le traitement d'images et la Vision par Ordinateur, beaucoup de ces systèmes reposent sur des méthodes de reconstruction 3D par projection de lumière structurée.

Depuis peu, quelques systèmes de numérisation 3D portables exploitant une technologie laser sont apparus. Ces systèmes nécessitent pour se localiser de positionner dans l'environnement de numérisation une référence magnétique ou d'équiper la pièce de cibles.

Par la suite, nous allons présenter succinctement un certain nombre de systèmes de numérisation 3D, sans chercher à être exhaustif tant le nombre de systèmes est important.

Polhemus FastSCAN™

Le Polhemus FastSCAN™ (Figure 2) est un système commercial laser de type profilométrie, tenu à la main. La localisation utilise un suivi magnétique depuis une référence placée dans la scène. L'incertitude de mesure annoncée par le constructeur est de $178 \mu m$ à une distance de mesure de $200 mm$.

Creaform HandyScan 3D™

La société Creaform propose une gamme de numériseurs tenus à la main. La figure 3 montre le système adapté aux applications de rétro-ingénierie. Les autres appareils de leur gamme sont également dédiés à des applications particulières. Il s'agit de systèmes laser,



FIGURE 2 – Polhemus FastSCAN™



FIGURE 3 – Creaform HandyScan 3D™ REVscan™

essentiellement adaptés aux pièces de petite à moyenne taille. La localisation se fait en plaçant des cibles photogrammétriques sur l'objet. L'incertitude annoncée est de $50\text{ }\mu\text{m}$.

Leica® T-Scan®

Le numériseur commercial Leica® T-Scan® (Figure 4) est un système laser portable tenu à la main pour la numérisation de moyens volumes. Il permet une acquisition à haute fréquence et une localisation automatique. De plus, l'intensité du laser est adaptative afin de rendre l'appareil indépendant aux conditions d'illumination. L'incertitude annoncée est de $30\text{ }\mu\text{m}$ pour une distance de mesure de 80 cm , avec une profondeur de champ de 8 cm et une largeur de balayage de 9 cm . Ce capteur nécessite des marqueurs optimisés pour calculer les positions de balayage.

3DMo du DLR Institute of Robotics and Mechatronics

Le *DLR Institute of Robotics and Mechatronics* a fondé plusieurs de ses travaux sur la mise en œuvre d'un système de numérisation multisensoriel modulaire. Le 3D-Modeller (Figure 5) intègre un capteur laser à balayage, un système stéréoscopique à deux caméras et deux plans laser. Il embarque un microcontrôleur et un écran. Selon l'application, il peut être déplacé manuellement ou fixé sur un robot pour des traitements automatiques. Le numériseur a notamment été intégré au robot Justin, également développé au DLR.



FIGURE 4 – Leica® T-Scan® TS50



FIGURE 5 – 3D-Modeller (3DMo)

GOM ATOS

Le système commercial ATOS de GOM (Figure 6) utilise la projection de franges pour la mesure 3D. Il s'agit d'un système fixe, couplé au système de photogrammétrie GOM TRITOP. Par conséquent, il est nécessaire d'équiper préalablement la scène de cibles. Le constructeur ne donne pas explicitement de valeur d'incertitude dans ses communications.

Konica Minolta VIVID 9i

Le système commercial VIVID 9i de Konica Minolta (Figure 7) est basé sur la profilométrie laser. Il s'agit d'un système fixe pour la numérisation de petits à moyens volumes. L'objet



FIGURE 6 – GOM ATOS III



FIGURE 7 – Konica Minolta VIVID 9i

est déplacé devant le capteur : le recalage se fait en exploitant un plateau rotatif ou en couplant l'appareil à un système photogrammétrique en équipant l'objet de cibles. L'incertitude annoncée est de $50\ \mu\text{m}$.

4 Spécifications de notre projet

Au regard des systèmes que nous avons évoqués, nous pouvons voir les inconvénients majeurs de ces appareils. Tout d'abord, ces systèmes sont généralement coûteux. Les systèmes fixes, bien que très performants, sont encombrants et nécessitent une mise en œuvre complexe qui impose de les fixer sur un trépied durant la phase de capture. Ces caractéristiques limitent leur utilisation à des lieux équipés spécifiquement, ce qui réduit considérablement le champ des applications. Les systèmes nécessitant des références externes ou un équipement particulier pour le recalage des points de vue peuvent être contraignants. Ces contraintes de mise en œuvre sont souvent problématiques et rédhibitoires pour certaines applications. C'est notamment le cas pour la numérisation de statues ou d'objets précieux.

Nous avons donc orienté la conception de notre système afin de répondre à ces limitations particulières.

Modélisation de la géométrie 3D et de l'apparence

La fonctionnalité essentielle du système est de permettre une acquisition 3D fiable et exploitable d'une scène. La modélisation 3D permet de donner une représentation mathématique de la géométrie et de l'apparence d'un objet ou d'une scène physique. Dans le cadre du projet, nous avons choisi les représentations en nuages de points, depuis lesquels sont générés des maillages triangulaires. Les nuages de points restent une représentation acceptable pour nombre d'applications mais il est impératif qu'ils soient construits de sorte à permettre un passage au maillage facile et en minimisant les dégradations lors de cette transformation.

Un système peu onéreux ...

Afin d'élargir les applications de la numérisation 3D à des domaines plus vastes, un des objectifs du projet était de concevoir un système en veillant à ce qu'il soit le moins onéreux possible. Par conséquent cela a induit des choix architecturaux simples et une optimisation du choix des gammes de composants. En contrepartie il a été nécessaire de travailler sur des composants sensibles au bruit, ou à dynamique réduite par exemple. Il s'agit donc d'une orientation très impactantes sur les méthodes et les travaux à mener pour la réalisation du système.

... Tenu à la main ...

Le développement des systèmes de numérisation déplacés par un opérateur a montré l'intérêt de ce type de solution. En effet, légers et maniables, ils permettent aux opérateurs de facilement se déplacer pour une numérisation sur site, ou encore d'accéder à des zones de tailles réduites. Notre système s'inscrivant dans cette catégorie de numériseurs tenus à la main, nous l'avons pensé de sorte à réduire le plus possible son encombrement et son poids.

Le fait que notre système soit tenu à la main est un paramètre important à considérer. Le mouvement de la main est complexe à prédire et peut être très bruité (tremblements, mouvements peu assurés ...). Par conséquent, il faudra considérer cette difficulté lors de la mise en œuvre de méthodes d'estimation de mouvement.

... Facile d'utilisation ...

Toujours dans le but d'ouvrir la numérisation au plus grand nombre, nous avons jugé important de rendre le fonctionnement très simple afin de donner la possibilité à des opérateurs non spécialistes d'exploiter le système. Pour cela, nous avons dû travailler selon deux axes :

- l'interaction : proposer à l'opérateur des interactions simples avec le système et le guider au cours du processus de numérisation ;
- automatisation : afin de permettre à l'opérateur d'intervenir au minimum dans le processus, le système doit être capable de réaliser ses tâches de manière totalement automatique.

... Sans équipement préalable

Pour étendre le point précédent, la facilité d'utilisation passe également par la simplification du processus de numérisation dans son intégralité. Par conséquent, nous avons souhaité fonctionner dans tout type d'environnement sans l'avoir balisé, équipé de marqueurs, ou

appris à l'avance. Cela nécessite la mise en œuvre de méthodes permettant de s'adapter à l'environnement ainsi que de s'y localiser.

5 Principales contributions

Au sein du projet que nous venons de décrire visant à concevoir un nouveau capteur pour la modélisation 3D, l'étude présentée dans cet ouvrage explore les problématiques de recalage en ligne, d'optimisation de poses hors ligne et de planification perceptuelle.

Les principales contributions de ce travail sont :

- une méthode de recalage de deux acquisitions. En se basant sur l'algorithme ICP nous proposons une méthode utilisant les images malgré une projection de mouchetis mobile. Notre capteur utilise la projection d'un motif pour la reconstruction 3D. Ce projecteur étant fixé au système, il se déplace avec ce dernier. De ce fait, entre deux acquisitions successives, la projection change dans la scène, rendant impossible une mise en correspondance directe de points entre des images successives. Notre méthode propose une mise en correspondance 3D basée sur la reprojection de points dans une paire stéréoscopique. Nous décrivons également la stratégie que nous mettons en œuvre à chaque itération de l'algorithme de recalage.
- une stratégie de recalage multi-vue. Nous couplons notre méthode de recalage de deux acquisitions avec un système de sélection de vue. Pour chaque nouvelle acquisition on détermine la meilleure vue précédente pour un recalage utilisant notre méthode. Nous proposons ensuite une optimisation en ligne en détectant les fermetures de boucles et en créant des cycles dans un graphe de scène. Nous optimisons alors un critère de consistance sur ces cycles. En fin de processus, une optimisation globale est réalisée par minimisation d'un critère de reprojection dans l'ensemble des acquisitions.
- l'intégration avec le capteur. Nos méthodes sont conçues pour le fonctionnement avec notre capteur spécifique. De ce fait une part du travail a été d'orienter la conception du système matériel. Cela concerne particulièrement l'ajout de la mesure inertielle au système.

Cette étude a donné lieu à trois publications. La confidentialité liée à notre projet, du fait de son implication industrielle, ne nous a pas permis de communiquer plus largement.

- B. Coudrin, M. Devy, L. Brèthes et J.J. Orteu. 3D Modelling of objects from a visual sensor moved manually by an operator. ORASIS'09 Congrès des jeunes chercheurs en vision par ordinateur, Juin 2009 [Coudrin 09].
- B. Coudrin, M. Devy, J.J. Orteu et L. Brèthes. Registration Strategies of 3D Images Acquired From a Hand-held Visual Sensor. 3D-IMS Conference on 3D-Imaging of Materials and Systems, Septembre 2010 [Coudrin 10].
- B. Coudrin, M. Devy, J.J. Orteu et L. Brèthes. An innovative hand-held vision-based digitizing system for 3D modelling. Optics and Lasers in Engineering, 2011 [Coudrin 11].

6 Organisation du manuscrit

L'étude que nous présentons dans ce manuscrit est centrée sur la conception d'un nouveau capteur de numérisation 3D et sur la méthode de recalage des acquisitions, adaptée au fonctionnement de ce système basé vision tenu à la main.

Le chapitre 1 présente le système que nous avons mis en œuvre. Nous y décrivons les choix de conception et les composants de notre appareil ainsi que leur fonctionnement. Ce chapitre décrit également la manière dont on réalise une acquisition, les données générées et les stratégies d'utilisation correspondant à l'usage que nous aurons dans toute la suite du manuscrit.

Le chapitre 2 revient sur les méthodes employées pour la mesure 3D dans notre système. Nous présentons les concepts de base de la vision stéréoscopique et des méthodes de reconstruction 3D. Ce chapitre sera l'occasion de présenter les modèles et les conventions que nous employons lorsque nous traitons de méthodes basées vision dans cette étude.

Le chapitre 3 traite du recalage de deux acquisitions. Nous introduisons donc ici le début de notre stratégie de recalage en étudiant le moyen d'estimer la transformation entre deux vues acquises successivement. Ce chapitre est divisé en deux parties afin d'aborder les problématiques de recalage initial puis de recalage précis. Nous présentons nos travaux ainsi qu'une évaluation de nos méthodes en comparaison avec des méthodes concurrentes.

Dans le chapitre 4 est proposée notre stratégie globale pour le recalage durant un processus de modélisation incrémentale. Nous présentons nos méthodes pour le recalage rapide en ligne et le raffinement des transformations aux fermetures de boucles. Nous introduisons ensuite notre méthode de réestimation globale hors-ligne. Une évaluation est faite en réalisant une modélisation sur un très grand nombre de vues.

Le chapitre 5 est une évaluation quantitative de notre système de numérisation et de nos méthodes. Nous avons fait numériser nos objets de tests par deux systèmes commerciaux et nous utilisons les modèles générés pour réaliser une évaluation des résultats de notre propre système. Les évaluations portent sur trois points. Le premier porte sur la précision de reconstruction, le second est une évaluation qualitative sur une application de conservation du patrimoine et le dernier évalue quantitativement les résultats sur une application de contrôle industriel. Nous introduisons, de plus, une évaluation de notre critère de qualité de reconstruction par comparaison à un modèle CAO.

Le chapitre 6 ouvre la voie à de futurs développements en présentant une stratégie pour la sélection d'un prochain point de vue pertinent afin de guider l'opérateur lors de la phase d'acquisition. Nous étudions comment le score de reconstruction des points peut être utilisé pour qualifier la qualité de reconstruction dans une approche incrémentale. Nous présentons ensuite une approche pour la sélection de point de vue utilisant ce critère pour prendre en compte la qualité du modèle dans la planification perceptuelle.

Description du capteur et méthodologie d'utilisation

Table des matières

1.1	Conception mécanique et électronique du système	26
1.1.1	Description de la conception et des composants	26
1.1.2	Mécanismes d'acquisition	29
1.2	Mode d'acquisition nominal	32
1.2.1	Description de la séquence	32
1.2.2	Revue des données	33
1.3	Cas d'utilisation	34
1.3.1	Interactions avec l'appareil	34
1.3.2	Analyse opérationnelle	37
1.3.3	Problématiques	39

P OUR répondre aux besoins qui ont motivé notre projet, et que nous avons présentés plus haut, nous avons conçu un nouveau capteur de numérisation 3D. C'est sur la base de cet appareil que nous avons fondé nos méthodes, particulièrement concernant le recalage, qui constitue le cœur de cette étude.

Nous avons souhaité nous orienter vers une technologie vision. En effet, les caméras sont des capteurs extrêmement riches au regard de leur coût. Outre la reconstruction 3D par stéréovision ou par *structure-from-motion*, elles permettent des applications de reconnaissance, de localisation ou de réalité augmentée. Dans notre application, elles nous offrent la possibilité d'intégrer l'information de texture des objets aux modèles finaux.

Ce chapitre décrit le capteur et nos choix techniques. Nous présenterons d'abord la conception et les composants du système puis nous décrirons les mécanismes et modes d'acquisition. Nous proposerons ensuite un scénario de manipulation. Ce scénario permettra de couvrir l'ensemble des cas d'utilisation utiles pour notre étude.

1.1 Conception mécanique et électronique du système

Nous décrivons ici le système que nous proposons. Dans un premier temps, nous présentons sa conception et les divers matériels qui le composent. Dans un second temps, nous aborderons la description des mécanismes d'acquisition qui coordonnent le fonctionnement des composants selon l'interaction avec l'opérateur.

La figure 1.1 montre le capteur, industrialisé et commercialisé par la société Noomeo, en fonctionnement. Ce système est de petite dimension ($220mm \times 240mm \times 90mm$) et pèse $1.8kg$.



FIGURE 1.1 – Le capteur OptiNum™

1.1.1 Description de la conception et des composants

Le système OptiNum™ est composé de plusieurs sous-systèmes. La figure 1.2 montre cette composition sur une coupe transversale de l'appareil.

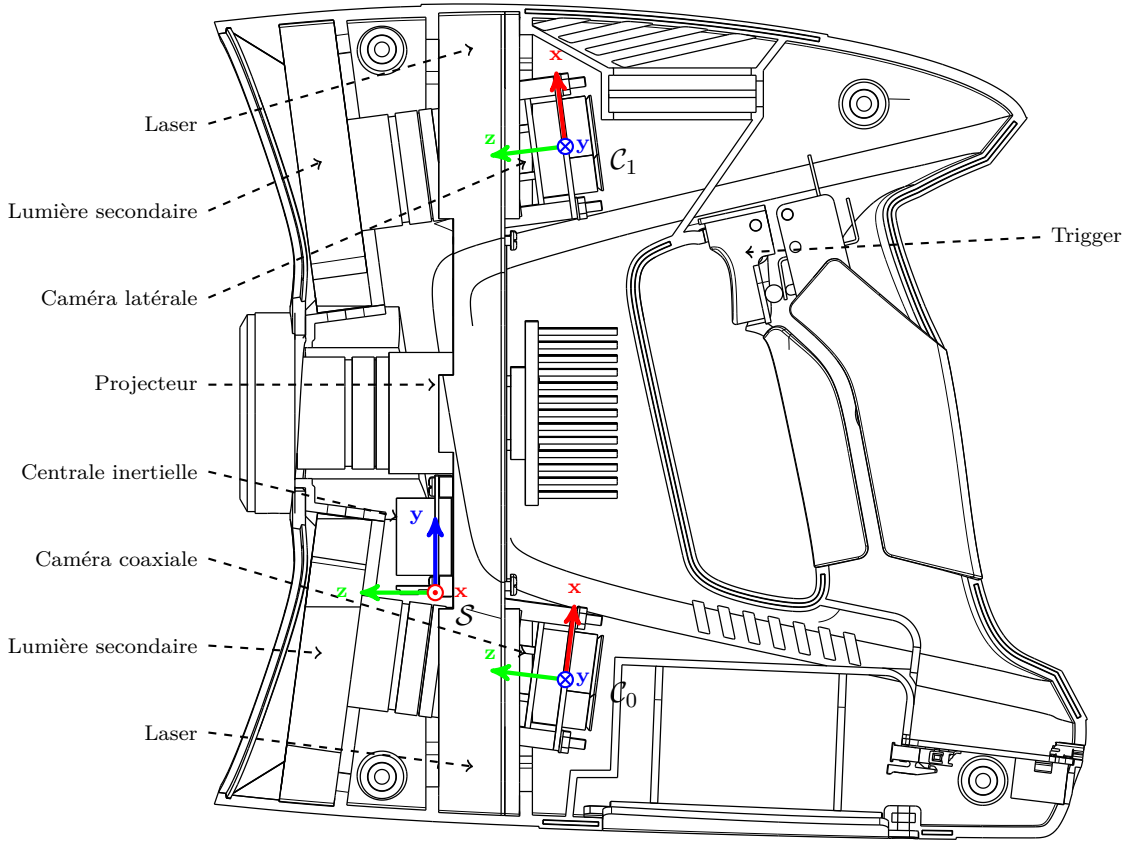


FIGURE 1.2 – Plan transversal d'OptiNum™

Caméras

Le capteur est basé vision. Les éléments essentiels de notre système sont les deux caméras dont il est équipé. La première, située au bas de l'appareil, est appelée *caméra coaxiale* car elle partage son repère avec le repère absolu que nous avons choisi pour l'appareil. Dans la suite de ce manuscrit, on notera $\mathcal{C}_0$ son repère.

La seconde caméra, située en haut du montage, est appelée *caméra latérale*. Son repère est noté $\mathcal{C}_1$. On suppose que la transformation permettant de passer du repère de la caméra latérale à la caméra coaxiale est connue et notée $\mathbf{H}^{\mathcal{C}_1\mathcal{C}_0}$

$$\begin{bmatrix} p^{\mathcal{C}_1} \\ 1 \end{bmatrix} = \mathbf{H}^{\mathcal{C}_1\mathcal{C}_0} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} \quad (1.1)$$

où $\mathbf{R}$ et $\mathbf{t}$ représentent respectivement la matrice de rotation et le vecteur de translation permettant le passage du repère de la caméra coaxiale au repère de la caméra latérale.

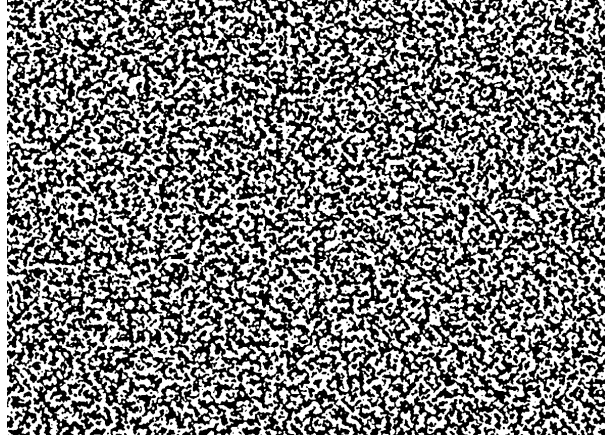


FIGURE 1.3 – Le motif aléatoire projeté sur la scène est basé sur un bruit de Perlin [Orteu 06]

Centrale inertielle

Le système est équipé d'une *centrale inertielle*, ou *IMU* (Inertial Motion Unit). Elle est composée de trois accéléromètres, trois gyromètres et trois magnétomètres. On bénéficie ainsi à haute fréquence ($\sim 100Hz$) d'informations sur la dynamique de notre système :

- L'accélération $\mathbf{a} = [a_x \ a_y \ a_z]^T$
- La vitesse angulaire $\omega = [\omega_x \ \omega_y \ \omega_z]^T$
- La direction du champ magnétique environnant $\mathbf{b} = [b_x \ b_y \ b_z]^T$

Par composition de ces mesures à haute dynamique, la centrale inertielle est capable de renvoyer une estimation de son attitude instantanée [Marins 01, Lefferts 82].

On note $\mathcal{S}$ le repère de la centrale inertielle. La transformation permettant de passer de ce repère à celui de la caméra coaxiale $\mathcal{C}_0$ est notée $\mathbf{H}^{SC_0}$.

$$\begin{bmatrix} p^{\mathcal{S}} \\ 1 \end{bmatrix} = \mathbf{H}^{SC_0} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R}^{SC_0} & \mathbf{t}^{SC_0} \\ \mathbf{0} & 1 \end{bmatrix} \begin{bmatrix} p^{\mathcal{C}_0} \\ 1 \end{bmatrix} \quad (1.2)$$

avec $\mathbf{R}^{SC_0}$ et $\mathbf{t}^{SC_0}$ respectivement la matrice de rotation et le vecteur de translation permettant le passage du repère de la caméra coaxiale au repère de la centrale inertielle.

Illumination

Le but du système est de permettre la numérisation 3D dans un maximum de conditions possibles. Par conséquent, il est important de se doter de mécanismes pour gagner de l'indépendance face aux variations de l'environnement. C'est pourquoi le capteur est équipé d'une projection d'un motif visant à texturer la scène observée et faciliter la phase de mise en correspondance dans les images. Ce projecteur est une matrice de LED couplée à une diapositive en verre sur laquelle est imprimé le motif (figure 1.3).

De plus, des anneaux de LED sont ajoutés autour de chaque caméra afin de servir d'éclairage secondaire dans des conditions d'illumination faible.

Pointeurs laser

Deux pointeurs laser sont fixés proches de chaque caméra et sont réglés pour que leurs faisceaux convergent à la distance de travail optimale du système. Ils n'ont d'autre rôle que d'aider l'opérateur à positionner le capteur devant l'objet pour minimiser le flou sur la projection du motif ou sur l'acquisition image.

Electronique

Ces composants sont reliés à un microcontrôleur. Ce dernier assure la synchronisation lors de leur pilotage et garantit la consistance des trames de données renvoyées via une connexion série USB à un ordinateur.

1.1.2 Mécanismes d'acquisition

Deux mécanismes essentiels permettent de définir le comportement du système pour l'acquisition. Il s'agit du *séquenceur* et du *trigger*. Le séquenceur est un mécanisme du microcontrôleur pour le pilotage du matériel et le trigger est le bouton avec lequel l'utilisateur va pouvoir interagir avec le système. En composant ces deux outils, il est possible de définir des modes de fonctionnement pour décrire une acquisition. Les mesures des capteurs intégrés sont assemblées dans des *structures de trames*, transmises par liaison série à un ordinateur.

Le système interagit aussi avec un programme exécuté sur l'ordinateur. Ce programme peut envoyer des commandes, soit au séquenceur, soit directement aux composants. Cette interaction se limite généralement à la phase d'initialisation.

Le séquenceur

Le mécanisme de séquencement permet de spécifier de manière synchrone les déclenchements des composants du numériseur (flash, lumières secondaires, caméras, centrale inertielle et lasers). Ces composants ont des comportements propres selon le signal envoyé par le séquenceur sur leur entrée de synchronisation. Le tableau 1.1 liste ces comportements.

Composant	Front montant	Front descendant
<i>Flash</i>	allumage	extinction
<i>Lumières</i>	allumage	extinction
<i>Caméras</i>	début de l'exposition	fin de l'exposition
<i>IMU</i>	mesure ponctuelle	-
<i>Lasers</i>	allumage	extinction

TABLE 1.1 – Effets des fronts du séquenceur sur les divers matériels

Une séquence est décomposée en une suite de *scénarios* et peut être rejouée en boucle un nombre imposé de fois ou à l'infini. Un scénario est une portion de temps qui dure de 1 à 250 *ms* dans laquelle on détermine le déclenchement et l'arrêt des composants pilotés. On peut en créer jusqu'à douze différents que l'on peut enchaîner et boucler. Certaines règles sont à observer lorsque l'on crée un scénario :

- Chaque scénario ne peut contenir qu'une seule paire déclenchement-arrêt par composant. Si l'on souhaite déclencher à nouveau un composant utilisé, il faut le faire dans un nouveau scénario joué à la suite.
- La première milliseconde de tout scénario ne peut contenir aucun événement. En effet, le séquenceur réserve ce temps au démarrage de chaque scénario pour réaliser les sauvegardes de contexte et changements d'états nécessaires.
- La durée des événements – c'est à dire la durée entre un déclenchement et un arrêt – est définie numériquement avec un pas minimal de $1/250ms$.
- Les deux caméras ne peuvent pas être déclenchées de manière séparée. Pour améliorer la synchronisation de ces appareils, elles partagent la même ligne de contrôle.
- De même, les lumières secondaires partagent leur cablage de contrôle. Elles ne peuvent donc être déclenchées séparément.
- Pour éviter la surchauffe du circuit de puissance, il n'est pas possible d'avoir les lumières secondaires et la lumière du projecteur principal allumées en même temps.

On construit donc une séquence en créant une série de scénarios et en définissant la manière dont ils sont chaînés ou répétés. La figure 1.4 illustre une séquence simple composée de deux scénarios.

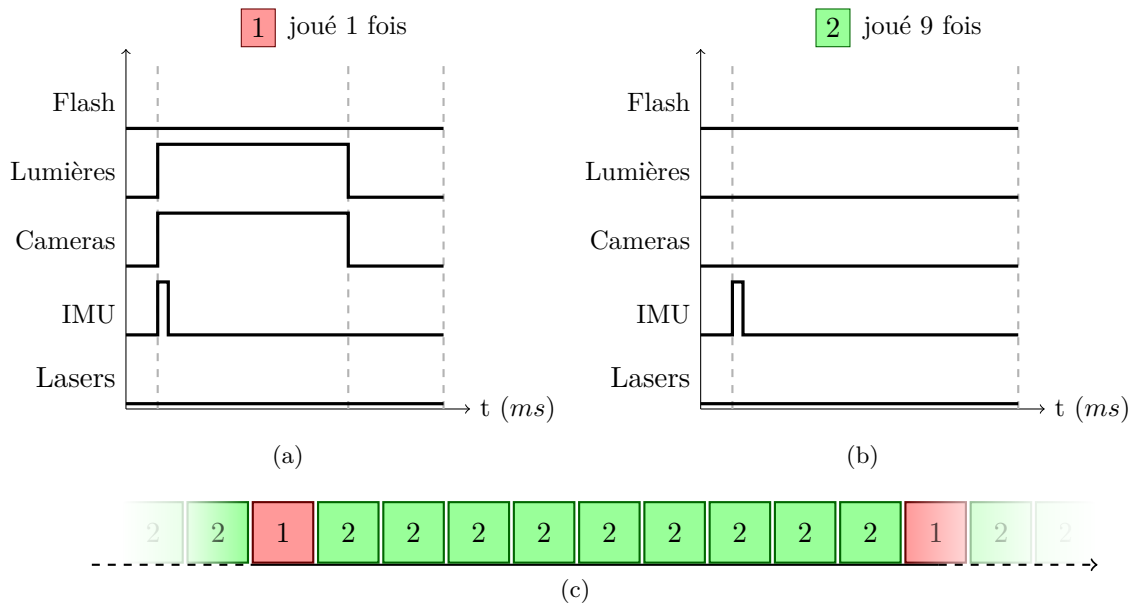


FIGURE 1.4 – Construction d'une séquence contenant des scénarios différents. (a) scénario 1 : il dure 10 ms ; la première est utilisée par le séquenceur, puis on déclenche une mesure inertielle synchronisée avec une exposition des caméras et l'illumination secondaire durant 6 ms. (b) scénario 2 : il dure 10 ms ; la première est toujours utilisée par le séquenceur, puis on déclenche une mesure inertielle. (c) Composition des scénarios dans la séquence : le scénario 1 est joué une fois puis le scénario 2 est répété 9 fois ; on obtient ainsi une mesure inertielle toutes les 10 ms (100 Hz) et une capture des caméras toutes les 100 ms (10 Hz) avec un temps d'exposition de 6 ms par capture. Ce type de scénario permet, par exemple, le suivi du mouvement par intégration de la mesure inertielle dans un programme exécuté sur l'ordinateur.

Il reste possible de se passer du mécanisme de séquençement pour réaliser le pilotage directement depuis un programme, mais dans ce cas la synchronisation n'est plus garantie. Par contre, les autres paramètres des composants ne sont accessibles que par commande. Typiquement, on pourra définir les intensités et courants des lumières par programmation mais pas via le séquenceur.

Le trigger

L'appareil OptiNumTM est muni d'un bouton – le trigger – qui constitue la seule interaction offerte à l'utilisateur directement disponible sur l'appareil.

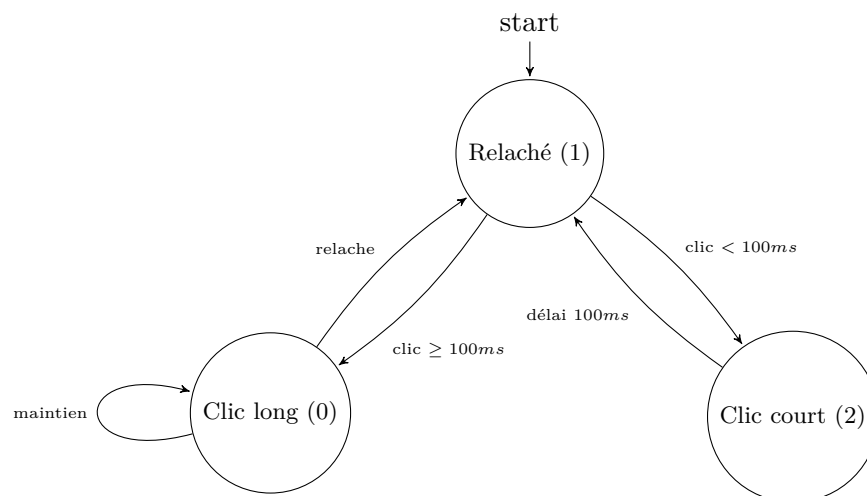


FIGURE 1.5 – États et transition du trigger

Le trigger est vu par le système selon son état : relaché, clic long, clic court. La figure 1.5 illustre les transitions entre ces modes. On considère que lorsque le bouton est relaché, le trigger est à son état de repos. Un front montant va déclencher un comptage. Si le front descendant correspondant survient dans les 100 *ms* suivantes, le clic est considéré *court* et le trigger change d'état, sinon le clic est vu comme *long*, le trigger change d'état et y restera jusqu'au front descendant suivant. Pour revenir de l'état court à l'état de repos, on impose un délai de 100 *ms* supplémentaires et le changement d'état se fait automatiquement. Ainsi on notera que l'interaction que l'opérateur a avec le système se fait selon des périodes de 100 *ms*, ce qui est acceptable pour un référentiel humain.

On peut dès lors définir les commandes en observant les changements d'états successifs sur une portion de temps donnée.

Structures de trames

Lorsqu'un composant sensoriel (caméra, centrale inertielle) est déclenché et a terminé son fonctionnement, il renvoie sa mesure au microcontrôleur. Ce dernier a pour mission de construire des trames de données à renvoyer à l'opérateur sur son ordinateur. Ces trames sont organisées en structures de données qui constituent des conteneurs hétérogènes. Ils sont composés de deux lignes de données, une pour les images et l'autre pour les mesures inertielles.

Chaque mesure se voit attribuer un *timestamp* permettant de le replacer dans son contexte temporel.

De plus, le rangement des données dans ces structures est accompagné d'un test de consistance des trames. Afin d'éviter toute désynchronisation en cas de donnée manquante ou corrompue, le microcontrôleur se charge de mener les actions nécessaires – suppression, décalage temporel – et de notifier le manque dans la structure.

1.2 Mode d'acquisition nominal

En exploitant les mécanismes de séquençement et d'interaction décrits plus haut et en les composant, on peut imaginer des *modes d'acquisition*, c'est à dire des commandes permettant le déclenchement de séquences particulières.

1.2.1 Description de la séquence

Le premier mode que nous allons présenter est le mode que nous avons utilisé principalement dans cette étude. D'autres pourront être abordés et seront introduits au fur et à mesure que nous les rencontrerons.

Le mode, que nous appellerons mode « normal », démarre par une action de *clic court* par l'opérateur. Cela déclenche en premier lieu une phase de calibrage des paramètres lumineux. Plusieurs acquisitions sur les caméras sont réalisées en faisant varier le temps d'exposition et les intensités du projecteur et des lumières secondaires. On réalise ainsi un réglage de ces paramètres adapté à la luminosité ambiante. Notons que pour pouvoir profiter du système de séquençement au mieux, il est préférable de ne pas fonctionner sur des temps d'exposition trop longs. Une bonne pratique est de régler les paramètres d'intensité et d'exposition en contraignant cette dernière pour qu'elle n'excède pas 10 *ms*.

Une fois ce calibrage réalisé, une séquence est lancée et bouclée sans fin. La figure 1.6 illustre les scénarios qui la composent.

Ce mode propose donc de réaliser une prise de vue des caméras avec projection du motif (1.6a) – vues exploitables pour calculer les points 3D – puis, immédiatement à la suite, une seconde prise de vue sans motif mais en utilisant les lumières secondaires (1.6b) – afin de disposer des informations de texture vraie de la scène. Chaque acquisition est synchronisée à une mesure inertielle de l'attitude du capteur.

Enfin, les lasers sont activés pour le reste de la durée de la période d'acquisition afin de servir d'aide à l'opérateur pour son positionnement lors de la capture (1.6c). La période d'acquisition est ici fixée à un minimum de 500*ms* (2*Hz*).

Le système renvoie donc à une fréquence de 2*Hz* un paquet de données appelé *Quad* – car composé, en plus de deux mesures d'attitude, de quatre images : deux avec projection de motif et deux sans.

L'ordinateur exploite ces données pour fournir un retour visuel à l'opérateur mais pour le moment les mesures 3D ne sont pas conservées dans le modèle. L'opérateur peut alors réaliser un *clic long* pour ordonner au logiciel de commencer à agréger les mesures 3D au modèle final. La modélisation dure tant que l'opérateur maintient le trigger en pression. Lorsqu'il le relâche, les données ne sont plus agrégées et ne sont à nouveau utilisées que pour un retour visuel de positionnement, jusqu'à ce que l'opérateur décide de presser à nouveau le trigger.

Un nouveau *clic court* lorsque le système est en marche provoque l'arrêt de la séquence et la mise en repos du capteur.

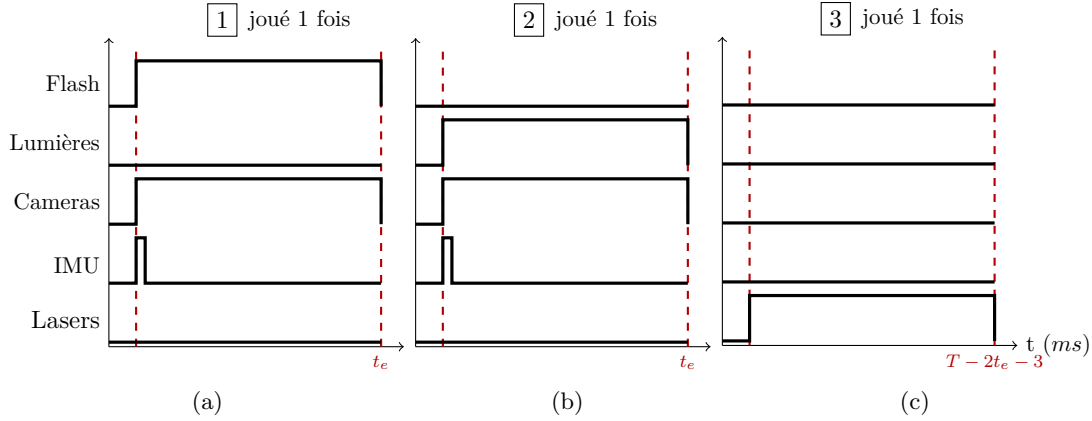


FIGURE 1.6 – Description de la séquence du mode « normal » de fonctionnement. Ici, t_e correspond au temps d'exposition des caméras, et T est la période d'acquisition. Le temps final est égal à la période d'acquisition tronquée du temps d'exposition pour chacune des deux acquisitions sur les caméras ainsi que les trois millisecondes nécessaires aux changements de contexte du séquenceur. Dans ce scénario, la centrale inertielle réalise matériellement l'intégration de ses mesures pour le suivi de mouvement. On récupère donc directement le résultat du suivi.

1.2.2 Revue des données

Conformément au mode d'acquisition nominal que nous venons de décrire, chaque acquisition par le capteur permet donc de fournir :

- une paire d'images avec projection du motif;
- une paire d'images sans projection de motif;
- pour chaque paire d'image, une mesure d'attitude synchronisée est fournie par la centrale inertielle.

La synchronisation est réalisée matériellement par l'électronique embarquée du capteur. On peut ainsi garantir la simultanéité de déclenchement des caméras et de la centrale inertielle.

Comme nous l'avons évoqué plus haut, et comme nous le décrirons au cours du chapitre 2, les images illuminées par le projecteur de motif permettent la génération des nuages 3D. Les images sans projection de motif ont deux intérêts. Tout d'abord, elles permettent d'observer la texture de la scène afin de reporter cette information sur le modèle final. D'autre part, elles permettent de réaliser un suivi de points ou des traitements de mise en correspondance de points d'intérêt entre deux acquisitions. Nous aborderons ce second usage au chapitre 3.

Il est important de noter que les deux captures de paires d'images sont réalisées séquentiellement. Entre ces deux acquisitions distinctes s'écoule un court temps (environ le temps d'exposition des caméras) durant lequel l'opérateur bouge, même légèrement. Réaliser des tâches de reconnaissance dans les images sans illumination n'est donc pas parfaitement équivalent, spatialement, à le faire dans les images avec illumination. On introduit donc ici un biais variable qu'il faut considérer lorsque l'on souhaitera utiliser conjointement les deux paires d'images – pour assigner l'information de texture ou associer les pixels entre les deux types d'images.

La définition d'un mode d'acquisition ne suffit pas à décrire le fonctionnement du système. Il est nécessaire d'étudier les cas d'utilisation et, à partir de cela, dresser la liste des méthodes qu'il va être nécessaire de mettre en œuvre pour remplir toutes les fonctionnalités du système.

1.3 Cas d'utilisation

Pour illustrer les cas d'utilisation du capteur, nous allons nous mettre à la place de l'opérateur souhaitant numériser un objet. Nous ne présenterons donc pas ici un mode opératoire imposé, mais uniquement une proposition de manipulation permettant de mettre en exergue toutes les problématiques qui vont nous concerner dans cette étude. En premier lieu, nous présenterons les cas en nous focalisant sur les actions de l'opérateur et les réactions du système. Puis nous en extrairons une liste d'opérations essentielles réalisées par le système pour dresser la liste des problématiques techniques qui nous concerneront.

1.3.1 Interactions avec l'appareil

Pour cette numérisation type, on suppose que l'opérateur veut modéliser un objet de petite taille (par exemple : $40cm \times 40cm \times 40cm$). L'objet est posé sur une table. La face de contact n'est donc pas observable sans que la pièce soit pivotée. De plus, on suppose que l'opérateur peut se déplacer autour de la table mais sans nécessairement pouvoir – ou vouloir – en faire le tour. Une face (le dos de la pièce) sera donc accessible préférentiellement en faisant pivoter la pièce également.

Comme nous l'avons abordé plus haut, l'opérateur démarre le système par un clic court sur le trigger. Nous entrons alors dans une phase de calibrage des paramètres lumineux que nous ne détaillerons pas ici.

Une fois cette phase achevée et le système paramétré, la séquence principale – telle que décrite en section 1.2 – est démarrée. En suivant cette séquence, le capteur va renvoyer des séries de *quads* à l'ordinateur à une fréquence de $2Hz$. Chaque *quad* va être utilisé pour la reconstruction d'un nuage de points 3D de prévisualisation (reconstruction rapide) qui ne sera pas conservé (figure 1.7).

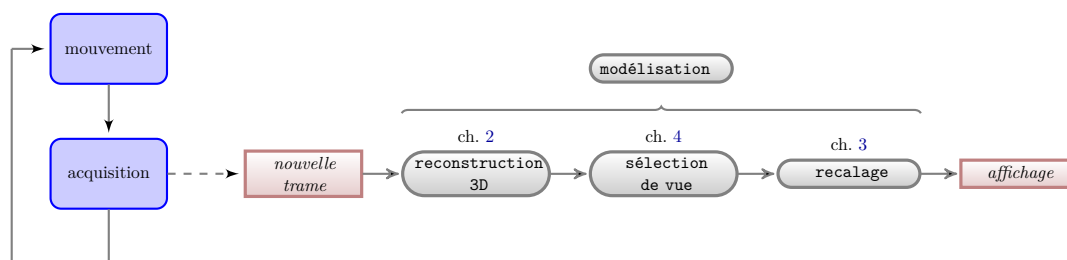


FIGURE 1.7 – Mode de prévisualisation temps-réel

En effet, le mode de prévisualisation a pour vocation d'afficher à l'écran la modélisation, sous forme de points 3D, obtenue à partir de la dernière trame acquise, et superposée au modèle courant, mais sans l'y agréger ni conserver les données acquises ou les résultats. Cela nécessite donc, en plus de la reconstruction, un recalage avec tout ou partie du modèle courant. L'ensemble des opérations permettant la création d'un nuage de points 3D recalé sur le modèle courant est appelé la phase de *modélisation*.

Au démarrage du système, aucun modèle n'est encore créé. Par conséquent, la phase de modélisation n'aura aucun recalage à effectuer.

Après s'être positionné, et avoir validé par le retour visuel que la vue 3D générée est satisfaisante, l'opérateur souhaite commencer à construire le modèle de sa pièce. Par un clic maintenu sur le trigger, le système passe en phase de modélisation incrémentale (figure 1.8).

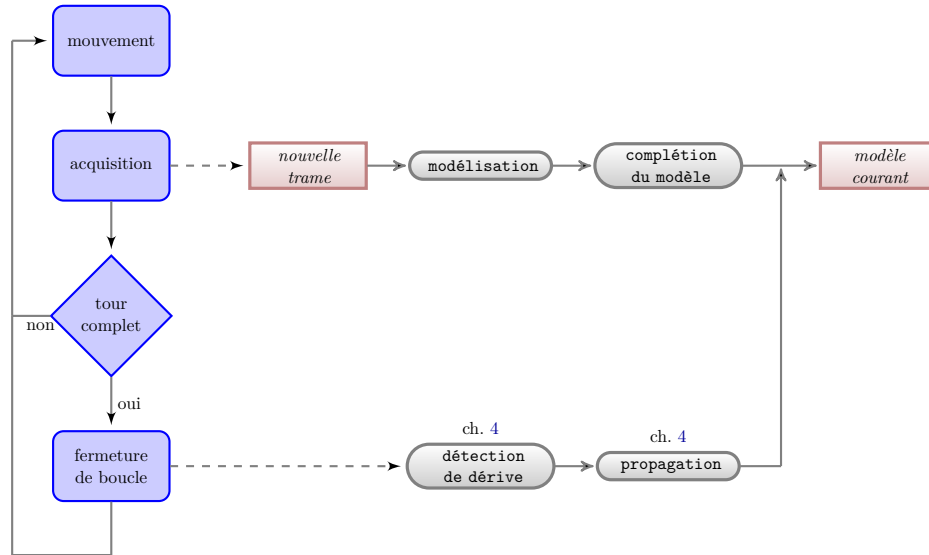


FIGURE 1.8 – Mode de modélisation incrémentale

Dans ce mode, l'opérateur va déplacer le capteur devant l'objet. A chaque nouvelle acquisition, le processus de modélisation est lancé afin de créer le nuage de points 3D et le recalcr dans un repère cohérent avec la scène. A la différence de la prévisualisation, ces résultats vont être agrégés au modèle courant. Ce dernier est donc complété de la dernière acquisition. Il se construit donc incrémentalement au fur et à mesure des acquisitions.

Au cours de la modélisation incrémentale, une dérive peut apparaître, du fait de l'imprécision des méthodes et du bruit de mesure. Il est donc important de pouvoir détecter lorsque l'on réobserve une partie de la scène dont nous avons déjà fait l'acquisition dans le modèle courant. Ainsi, il est possible de corriger la dérive en contraignant la consistance globale du modèle. On nomme ce procédé la *fermeture de boucle* : si le système a dérivé, le modèle ne se ferme pas convenablement et l'on doit corriger et propager la correction de l'erreur à tous les points de vue acquis.

Comme nous l'avons vu, l'opérateur va avoir besoin de déplacer la pièce à modéliser, soit pour éviter de trop se déplacer lui-même, soit pour observer des faces cachées (face de contact avec la table par exemple). Le repère du modèle courant déjà construit n'est alors plus cohérent avec la pièce et les mesures du capteur, du fait de la centrale inertielle particulièrement. Il faut donc recourir à une phase de relocalisation (figure 1.9).

L'opérateur doit d'abord se repositionner devant une partie déjà modélisée de la pièce. On génère ensuite un nuage de points que l'on cherchera à placer dans le modèle courant, comme on le ferait pour une modélisation classique. Il faudra cependant prendre en compte qu'il existe une transformation inconnue, correspondant au mouvement de l'objet, rendant la mesure inertielle inconsistante. Cette transformation inconnue est estimée par le recalage lors de la fonction de modélisation.

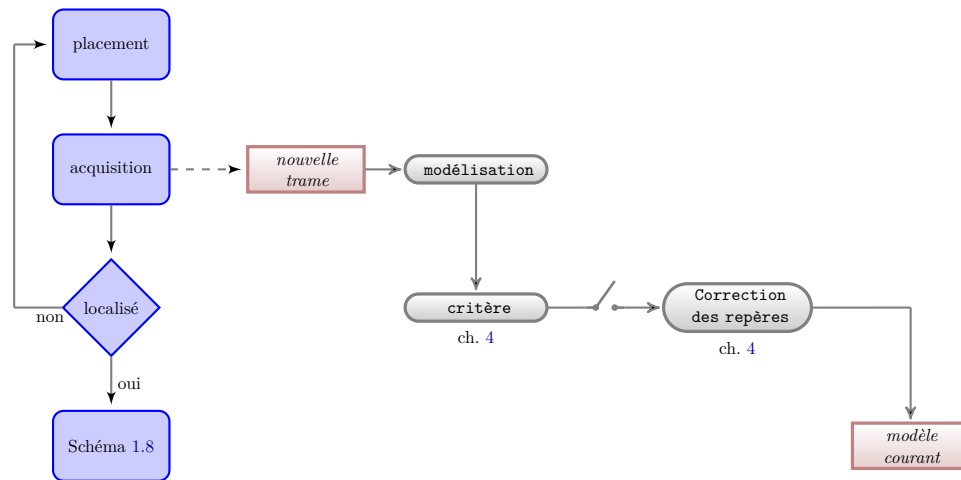


FIGURE 1.9 – Mode de relocalisation du capteur

Le résultat de cette étape nous donne une position candidate pour notre acquisition dans le modèle courant, et donc une pose candidate pour le capteur. Un critère nous permet alors de déterminer si cette pose est suffisamment bonne. Si ce n'est pas le cas, il faudra repositionner l'objet devant une vue plus proche d'une partie déjà modélisée de la pièce et recommencer. Si le critère est atteint, on peut rectifier le repère du capteur ou du modèle afin de prendre en compte le déplacement de la pièce que nous venons de trouver.

Après s'être relocalisé, on peut reprendre la modélisation incrémentale telle que décrite plus haut et dans la figure 1.8.

L'opérateur va ainsi compléter à sa guise le modèle. Une fois qu'il s'estime satisfait de son travail, il va demander au système d'analyser le modèle (figure 1.10). C'est à dire que le système va déterminer si le modèle est effectivement complet ou s'il présente des manques.

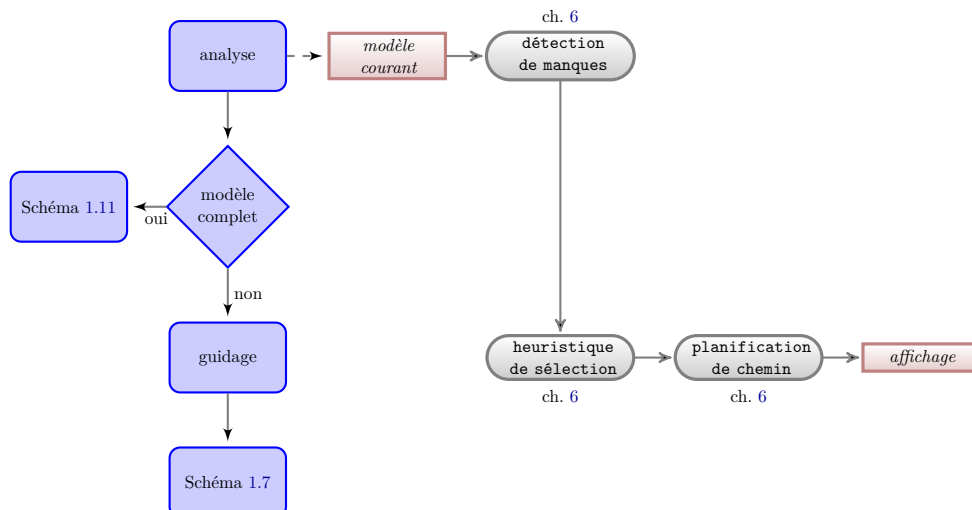


FIGURE 1.10 – Mode d'aide à l'opérateur pour compléter le modèle

Parmi les manques détectés, une heuristique doit permettre de déterminer celui qui doit

être comblé en premier. Un chemin doit être créé afin de guider l'opérateur pour qu'il se positionne devant ce trou ou cette zone ne contenant pas assez d'informations.

Une fois que le chemin ou la position finale est indiquée à l'opérateur, celui-ci utilise le mode de prévisualisation pendant son positionnement. Une fois bien placé, il pourra reprendre la modélisation incrémentale pour compléter son modèle.

Si le modèle est complet – donc si la liste de manques est vide – la numérisation est terminée et l'on démarre la finalisation du modèle (figure 1.11).

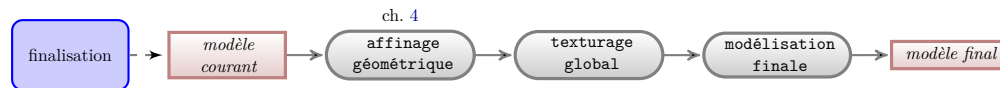


FIGURE 1.11 – Mode de finalisation

Afin de gagner du temps lors de l'acquisition et optimiser la productivité de l'opérateur, certaines opérations plus longues peuvent être déportées après l'acquisition. Généralement il s'agira des étapes visant à améliorer la précision quantitative ou les traitements dits « cosmétiques » visant à améliorer le rendu qualitatif.

En tout premier lieu, on cherchera donc à améliorer la géométrie du modèle : améliorer les points de vue estimés, améliorer les points 3D construits, corriger les dérives ...

Puis on s'attachera à rendre le modèle exploitable selon l'application à laquelle on le destine : texturage, reconstruction de surfaces, reconstruction de volumes ... Ces dernières étapes ne nous concernent pas dans le cadre de cette étude et ne seront donc pas abordées.

1.3.2 Analyse opérationnelle

De l'analyse des cas d'utilisation décrits plus haut, on peut dresser une liste des opérations essentielles que notre système doit pouvoir effectuer.

Reconstruction 3D

Nous appelons reconstruction 3D le processus permettant de créer la représentation de type nuage de points 3D à partir des images des caméras. Nous l'avons vu en section 2, il existe quantité de techniques permettant de réaliser une mesure 3D d'une scène. Dans notre cas un appariement de pixels est réalisé afin de calculer la position 3D d'un point par triangulation.

Recalage

Dans notre processus de numérisation, le modèle 3D est créé par agglomération de mesures successives. Chaque nuage de points est exprimé, après reconstruction dans le repère de la caméra coaxiale. Le recalage est l'étape permettant de déterminer la transformation rigide qui amène ce nuage dans le référentiel global du modèle courant. On la divise ici en deux étapes : le choix d'une vue précédente avec laquelle recalculer, et le calcul de la transformation entre notre vue courante et la vue candidate.

On notera qu'ici, l'opérateur peut procéder de deux manières pour réaliser l'acquisition du modèle. Dans un premier cas, il peut réaliser de longues séquences en faisant un balayage de la pièce. Ici, chaque nouveau point de vue est proche du précédent et permet de dessiner un mouvement relativement continu. Dans ce cas, la meilleure vue pour le recalage sera probablement celle acquise immédiatement avant. Dans un second cas, l'opérateur peut préférer

réaliser un ensemble d'acquisitions courtes – un seul *quad* typiquement – éclatées spatialement. Ici, la meilleure vue pour chaque acquisition séparée peut être n'importe quelle vue du modèle, et c'est l'usage de la centrale inertielle qui permettra de discriminer parmi les vues candidates.

Il faut donc différencier cette étape selon que l'on dispose d'information *a priori* sur la position du capteur au moment de la prise de vue (mouvement séquentiel, intégration des mesures inertielles) ou non. Dans le cas où l'on n'en dispose pas – après un mouvement de la pièce – on parlera plutôt de méthodes de *localisation*.

Localisation

Déterminer le point de vue d'acquisition sans *a priori* peut être vu comme un problème de reconnaissance, ou encore comme un problème de minimisation globale. Il est aussi possible d'envisager des méthodes stochastiques exploitant, par exemple, les acquisitions de prévisualisation non sauvegardées.

À mi-chemin entre une modélisation avec un fort *a priori* et une localisation plus globale, l'estimation du mouvement appliqué à un objet considère que l'on dispose d'une information sur la position du capteur mais que celle-ci est fautive. Ici encore, des approches stochastiques type minimax ou des approches par minimisation locale peuvent être envisagées.

Fermeture de boucle

Passer par des points de vue déjà visités permet d'obtenir de nouvelles mesures d'une zone déjà observée et donne la possibilité de confronter les résultats aux deux instants. Typiquement, un système devant déterminer sa position incrémentalement va être sujet à l'accumulation d'erreurs. Le mécanisme de fermeture de boucle permet de détecter ces dérives et donc d'y proposer une correction.

Propagation, corrections

Afin de capitaliser toute information pouvant apporter un gain de qualité à l'ensemble du modèle, il faut être capable de mettre en lien les différents instants d'acquisition ainsi que les mesures. Ainsi peut-on répercuter convenablement un apport d'information à tout niveau du modèle.

Détection de manques

La détection de manques va chercher principalement des trous dans le modèle mais aussi des zones où l'information 3D est trop éparse ou mal répartie. Le critère de détection sera particulièrement critique dans les cas où l'on ne dispose pas d'information sur la pièce observée (comme un modèle CAO simplifié par exemple).

Heuristique de sélection, planification de chemin

Nous cherchons ici à décider où l'opérateur doit réaliser sa prochaine acquisition et le guider pour qu'il aille se positionner convenablement.

Affinage géométrique

En fin de processus, et afin de rendre le modèle précis, propre et fiable, mais surtout exploitable, on procèdera donc à l'affinage des mesures 3D. Contrairement aux opérations précédentes devant être réalisées rapidement pour faciliter le processus de numérisation pour l'opérateur, cette dernière opération pourra être abordée sans contrainte de temps, ou du moins avec une contrainte de temps largement moins prioritaire qu'une contrainte de qualité du résultat.

Contrairement aux approches de modélisation incrémentale ou de localisation, qui exploitent essentiellement deux vues, ici on traite nécessairement d'un problème d'optimisation globale sur l'ensemble de l'information acquise durant le processus de numérisation.

1.3.3 Problématiques

On peut synthétiser les opérations essentielles abordées en section 1.3.2 pour dresser les axes principaux de cette étude. Ce mémoire s'attachera à traiter ces problématiques, à expliquer comment nous les avons résolues pour répondre aux fonctionnalités du système et, le cas échéant, les contributions que nous avons apportées.

Nous nous intéresserons tout d'abord à la manière dont nous pouvons *créer la mesure 3D* sur la base des images acquises depuis notre système. Cette problématique ne concerne pas directement le sujet de notre étude, traitant principalement de recalage, comme nous le verrons plus bas. C'est pourquoi nous n'aborderons que succinctement cette phase au cours du chapitre 2.

La problématique essentielle qui nous concerne est celle du *recalage* et de la *localisation*. En effet, le cœur des opérations décrites en section 1.3.2 aborde le propos du positionnement du capteur par rapport à la scène observée et le modèle qu'il a déjà acquis précédemment. Nous pouvons faire une distinction entre des *approches locales*, qui chercheront à déterminer la position relative optimale de deux acquisitions indépendantes, et des *approches globales* qui auront pour but d'exploiter l'ensemble de la mesure. Dans le premier cas on parlera de *recalage paire-à-paire*, ce qui constituera le propos du chapitre 3 et dans le second de *recalage global* ou multi-vues, que nous traiterons au chapitre 4.

Enfin, savoir comment *guider l'opérateur* suppose de savoir analyser le modèle et choisir où il est pertinent d'aller chercher de l'information supplémentaire. Cette problématique porte aussi la notion d'interaction homme-machine, puisqu'une attention particulière doit être portée à la manière dont le système guide l'opérateur par des consignes visuelles simples, conçues pour être faciles à interpréter et bien acceptées. Ainsi l'opérateur doit ressentir que le guidage est pertinent et qu'il est facile de suivre les consignes. Nous présenterons une piste vers ces travaux dans le chapitre 6.

Nous présenterons des résultats et des évaluations de précision de notre système au chapitre 5.

CHAPITRE 2

Stéréovision

Table des matières

2.1	Bases pour la stéréovision	42
2.1.1	Modèle de caméra	42
2.1.2	Modèle d'une paire stéréoscopique	45
2.1.3	Géométrie épipolaire	46
2.1.4	Rectification	47
2.2	Mise en correspondance dans les images stéréoscopiques	47
2.2.1	Mise en correspondance dense	47
2.2.2	Corrélation précise	49
2.2.3	Mise en correspondance de primitives	51
2.3	Reconstruction 3D	52
2.3.1	Chaine de reconstruction	52
2.3.2	Triangulation	53

AVANT de nous intéresser aux problématiques de recalage qui concernent notre étude, il faut nous arrêter sur les concepts géométriques permettant la reconstruction 3D par stéréovision.

Dans un premier temps, nous introduirons les modèles et les principales fonctions théoriques classiquement mis en œuvre en vision par ordinateur.

Nous aborderons également les concepts de la mise en correspondance de points entre images, qui interviennent dans le processus de reconstruction.

Il nous sera alors possible de présenter les méthodes que nous avons exploitées pour créer la mesure 3D sur la base des images et des points mis en correspondance.

2.1 Bases pour la stéréovision

La mise en œuvre de méthodes mathématiques pour la résolution de problèmes de reconstruction géométrique et de recalage nécessite un choix de modèle pour représenter le monde et les capteurs. Nous présentons tout d'abord les modèles pour la représentation des systèmes de caméras puis nous introduirons les simplifications que nous pourrions en tirer.

2.1.1 Modèle de caméra

Une caméra est une combinaison d'un capteur photoélectrique et d'un système optique de focalisation. Le modèle courant pour la représentation de ce système est le *modèle sténopé*. Il s'agit d'un modèle représentant le cas d'un système supposé parfait. On peut ensuite augmenter ce modèle en tenant compte des perturbations induites par les imperfections du système optique ou le mauvais alignement de ce système avec le capteur.

Modèle sténopé

Le modèle que nous utilisons pour représenter le système parfait nous permet d'établir une relation entre un point du monde et sa projection dans l'espace des pixels de l'image résultante.

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix}_{\mathcal{I}} = \mathbf{MatProj} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{\mathcal{W}} \quad (2.1)$$

Dans l'équation (2.1), $(u, v, 1)_{\mathcal{I}}$ représentent les coordonnées pixel homogènes du point projeté dans le repère image, $(X, Y, Z, 1)_{\mathcal{W}}$ étant les coordonnées homogènes du point dans le repère du monde, s est un facteur d'échelle, inconnu *a priori*. La matrice **MatProj** est notre modèle. On peut le décomposer en trois transformations permettant les passages successifs dans les repères figurés sur le schéma 2.1.

La transformation $[\mathbf{R}^{CW} | \mathbf{t}^{CW}]$ est la composition d'une rotation et d'une translation permettant le passage du repère monde $\mathcal{W}$ au repère de la caméra $\mathcal{C}$. On appelle cette matrice la *matrice des paramètres extrinsèques* ou la *pose de la caméra dans le repère $\mathcal{W}$* . De fait, cette matrice donne la position et l'orientation du repère monde vis à vis du repère caméra.

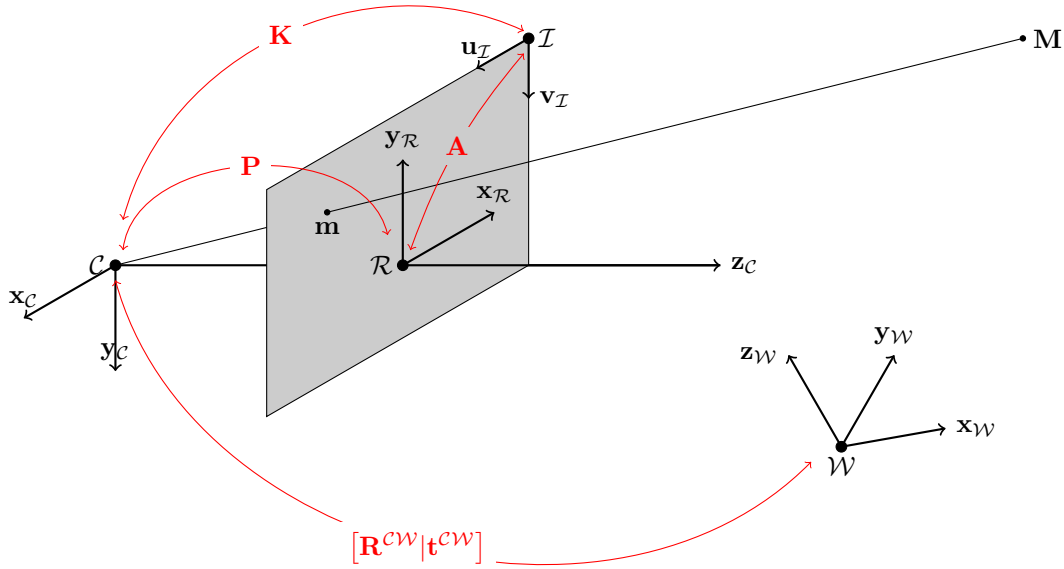


FIGURE 2.1 – Représentation du modèle sténopé et de la projection d'un point dans l'image

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_c = [\mathbf{R}^{cW} | \mathbf{t}^{cW}] \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_W$$

La matrice $\mathbf{P}$ est une projection perspective d'un point exprimé dans le repère de la caméra sur le plan rétinien de la caméra. On exprime cette matrice en fonction de la distance focale f . On note $\mathbf{M} = (X, Y, Z)$ le point projeté.

$$x = f \frac{X}{Z}; y = \frac{Y}{Z} \longrightarrow \begin{bmatrix} sx \\ sy \\ s \end{bmatrix}_c = \begin{pmatrix} f & 0 & 0 \\ 0 & f & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_c = \mathbf{P}\mathbf{M}_c$$

On voit ici que le facteur d'échelle s n'est autre que la profondeur du point 3D dans le repère caméra.

Le passage dans le repère image $\mathcal{I}$ se fait au moyen d'une transformation affine permettant d'exprimer les coordonnées du point projeté en fonction de la dimension des pixels (k_x^{-1}, k_y^{-1}) et des coordonnées image du centre de projection (x_0, y_0) .

$$\mathbf{A} = \begin{pmatrix} k_x & 0 & x_0 \\ 0 & k_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Les matrices $\mathbf{P}$ et $\mathbf{A}$ sont souvent composées et désignées par une unique matrice notée $\mathbf{K}$ appelée *matrice des paramètres intrinsèques*.

$$\mathbf{K} = \mathbf{A}\mathbf{P} = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Les paramètres $\alpha_x = fk_x$ et $\alpha_y = fk_y$ représentent la distance focale exprimée selon les dimensions des pixels.

On peut donc détailler l'équation (2.1) pour décrire notre modèle parfait.

$$s\mathbf{m} = \begin{bmatrix} su \\ sv \\ s \end{bmatrix}_{\mathcal{I}} = \underbrace{\mathbf{K} [\mathbf{R}^{CW} | \mathbf{t}^{CW}]}_{\text{MatProj}} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}_{\mathcal{W}}$$

Distorsions

Le modèle que nous venons d'introduire nécessite à présent d'être étendu afin de prendre en compte certaines perturbations par rapport au cas parfait. La présence d'imprécisions dans la construction et le montage de l'optique devant le capteur provoque certaines distorsions venant perturber la formation de l'image et fausser notre modèle. En les prenant en compte, on est capable de corriger la projection d'un point dans notre modèle.

En notant $\mathbf{m}_r = (x, y)$ un point parfait du plan rétinien, exprimé en coordonnées métriques issue de notre modèle parfait et $\mathbf{m}_{corrected}$ le résultat tenant compte des distorsions, on introduit le vecteur Δ qui décrit la relation entre eux.

$$\mathbf{m}_{corrected} = \mathbf{m} + \Delta$$

Ce vecteur Δ peut être décomposé en deux distorsions élémentaires, la *distorsion radiale* Δ_R et la *distorsion tangentielle* Δ_T [Weng 92].

$$\Delta = \Delta_R + \Delta_T$$

Distorsion radiale On modélise la composante radiale des distorsions de sorte qu'elle s'annule au centre de l'image et varie lorsque la distance au centre augmente. En notant $r = \sqrt{x^2 + y^2}$ la distance au centre et $\{k_i\}$ les coefficients radiaux pour $i = [1..\infty[$, on applique cette composante au point projeté $\mathbf{m} = [x \ y]^T$ afin d'obtenir le point $\mathbf{m}_{rad} = [x_{rad} \ y_{rad}]^T$.

$$\begin{aligned} x_{rad} &= x + \sum_{i=1}^{\infty} x k_i r^{2i} \\ y_{rad} &= y + \sum_{i=1}^{\infty} y k_i r^{2i} \end{aligned}$$

On simplifie généralement ce modèle en ne tenant compte que des trois premiers termes de la série.

$$\mathbf{m} + \Delta_R = \begin{cases} u_{rad} = u (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ v_{rad} = v (1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{cases}$$

Distorsion radiale et tangentielle Le mauvais alignement entre l'optique et le capteur est modélisé principalement par la composante tangentielle. On peut également modéliser cette distorsion par une série infinie dont on ne considère que les premiers termes, dont on note les coefficients p_1 et p_2 .

$$\mathbf{m} + \Delta_R + \Delta_T = \begin{cases} x_{corrected} = x_{rad} + [2p_1y + p_2(r^2 + 2x^2)] \\ y_{corrected} = y_{rad} + [p_1(r^2 + 2y^2) + 2p_2x] \end{cases}$$

2.1.2 Modèle d'une paire stéréoscopique

En liant rigidement deux caméras, on obtient un capteur composite appelé *paire stéréoscopique*. La figure 2.2 illustre les relations entre les repères de ce capteur.

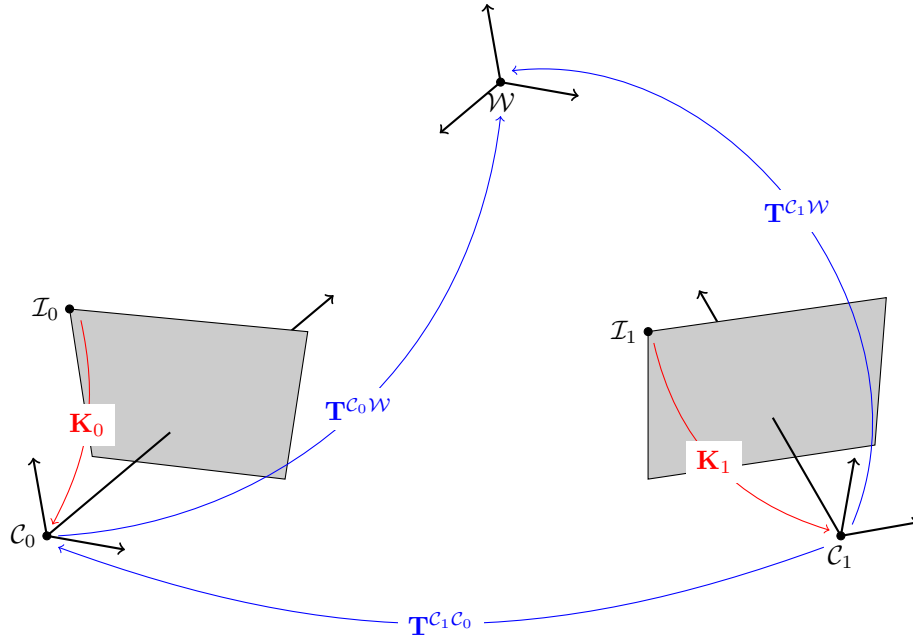


FIGURE 2.2 – Organisation des repères d'un capteur stéréoscopique

Ainsi un point du monde aura deux projections, une pour chaque image. On notera les projections du point $\mathbf{M}$ dans les images $\mathcal{I}_0$ et $\mathcal{I}_1$ par, respectivement, $\mathbf{m}_0$ et $\mathbf{m}_1$. On suppose que les distorsions sont absentes ou qu'elles ont été corrigées préalablement.

$$\begin{aligned} \mathbf{m}_0 &= \mathbf{K}_0 \mathbf{T}^{C_0W} \mathbf{M} \\ \mathbf{m}_1 &= \mathbf{K}_1 \mathbf{T}^{C_1C_0} \mathbf{T}^{C_0W} \mathbf{M} \end{aligned}$$

Le modèle de ce capteur inclut donc les paramètres intrinsèques des deux caméras, $\mathbf{K}_0$ et $\mathbf{K}_1$, les paramètres extrinsèques de la première caméra, $\mathbf{T}^{C_0W}$, ainsi que les paramètres de la transformation rigide entre les deux caméras, $\mathbf{T}^{C_1C_0}$.

2.1.3 Géométrie épipolaire

Jusqu'à présent nous avons étudié comment un point du monde se projette dans les repères images. Dans un processus de reconstruction 3D par stéréovision, notre objectif sera le problème inverse, c'est à dire calculer les points du monde correspondant à leurs projetés dans les images. Cela impose de pouvoir appairer les points d'une image avec les points de l'autre image étant les projections des mêmes points du monde. On appelle ces paires de points des *stéréo-correspondants*.

Pour un pixel $\mathbf{m}_0$ donné dans une des images, il existe une infinité de points du monde dont il puisse être la projection, alignés sur le rayon optique liant le centre optique $\mathbf{C}_0$ au point $\mathbf{m}_0$. Il est possible de déterminer le lieu géométrique du stéréo-correspondant $\mathbf{m}_1$ de $\mathbf{m}_0$ dans l'autre image selon la position du point du monde $\mathbf{M}$ supposé. En effet, il est notable (figure 2.3a) que les projections dans chaque image sont coplanaires avec le point $\mathbf{M}$ considéré et les centres optiques des deux caméras [Hartley 04].

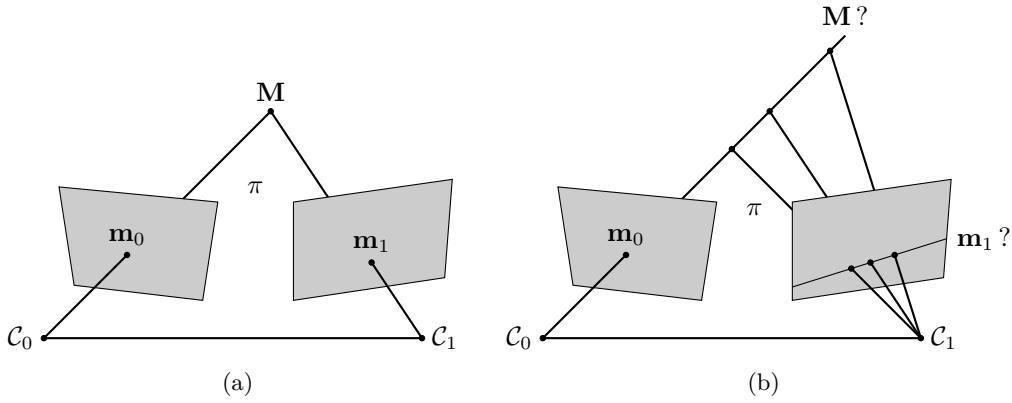


FIGURE 2.3 – (a) Les projections d'un point du monde dans les deux images sont coplanaires avec ce point et les centres optiques. (b) Le lieu géométrique du stéréo-correspondant d'un point image est une ligne épipolaire.

Par conséquent, où que se trouve le point $\mathbf{M}$ sur le rayon $[\mathbf{C}_0, \mathbf{m}_0]$, sa projection dans la seconde image se trouvera sur le plan π (Figure 2.3b). On peut donc déterminer que le lieu géométrique recherché pour le stéréo-correspondant de $\mathbf{m}_0$ est une ligne dans la seconde image appelée *ligne épipolaire correspondant au point $\mathbf{m}_0$* .

Cette droite peut être exprimée en fonction des points images $\mathbf{m}_0$ et $\mathbf{m}_1$ par le biais de la *matrice fondamentale $\mathbf{F}$* .

$$\mathbf{m}_1^T \mathbf{F} \mathbf{m}_0 = 0 \quad (2.2)$$

On dispose ainsi d'une expression de la ligne épipolaire dans $\mathcal{I}_1$ correspondant au point $\mathbf{m}_0$ de $\mathcal{I}_0$.

On note aussi que, tous les plans contenant ces lignes épipolaires ont en commun la base formée par les centres optiques. On prouve, de plus, que toutes les lignes épipolaires d'une image intersectent la base stéréoscopique au même point que l'on appelle *épipole*. On notera $\mathbf{e}_0$ l'épipole formée par l'intersection des lignes épipolaires de l'image $\mathcal{I}_0$, et $\mathbf{e}_1$ l'épipole formée par l'intersection des lignes épipolaires de l'image $\mathcal{I}_1$.

2.1.4 Rectification

L'équation (2.2) nous fournit une contrainte forte, appelée *contrainte épipolaire*, permettant de simplifier le processus de recherche de stéréo-correspondant. En effet, un point dans l'image $\mathcal{I}_0$ ne peut avoir de correspondant que sur la ligne épipolaire définie par la matrice fondamentale $\mathbf{F}$ dans l'image $\mathcal{I}_1$.

Pour simplifier encore la recherche d'appariements, on peut procéder à la rectification de l'image. Ce procédé cherche à rendre parallèles les lignes épipolaires dans les deux images et les placer de sorte que chaque droite épipolaire se trouve sur la même ligne image que son point correspondant dans l'autre image. On définit donc un *espace rectifié* pour chaque image au moyen de deux homographies $\mathbf{H}_0$ et $\mathbf{H}_1$.

$$\begin{aligned}\mathbf{m}'_0 &= \mathbf{H}_0 \mathbf{m}_0 \\ \mathbf{m}'_1 &= \mathbf{H}_1 \mathbf{m}_1\end{aligned}$$

Dans l'équation précédente, $\mathbf{m}'_0$ et $\mathbf{m}'_1$ sont les coordonnées des points projetés dans l'espace des images rectifiées. On obtient alors une contrainte épipolaire modifiée.

$$\mathbf{m}_1'^T \mathbf{H}_1^T \mathbf{F} \mathbf{H}_0 \mathbf{m}'_1 = 0$$

Il existe une infinité de solutions pour le choix des matrices $\mathbf{H}_0$ et $\mathbf{H}_1$. On les détermine généralement en cherchant à minimiser la déformation qu'elles induisent dans les images [Hartley 04, Bugarin 10].

À l'aide de ce procédé de rectification, on réduit l'espace de recherche de stéréo correspondant d'un point (x_0, y_0) de l'image rectifiée $\mathcal{I}'_0$ à la ligne de coordonnées $y = y_0$ dans l'image rectifiée $\mathcal{I}_1$.

2.2 Mise en correspondance dans les images stéréoscopiques

Comme nous l'avons vu, la technique d'imagerie est une projection d'un espace de dimension 3 dans un espace de dimension 2. Il s'agit d'une transformation non bijective par laquelle on perd l'information de profondeur. Afin de réaliser une mesure 3D par imagerie, il faut disposer de deux observations du même point selon des points de vue différents. On peut ainsi retrouver la position du point 3D en fonction des positions des deux caméras et des coordonnées image de leurs projetés (Figure 2.3a).

L'approche qu'il nous faut adopter est donc de déterminer des paires de points constituant les projections du même point du monde. On parle de *mise en correspondance*. Dans un premier temps, nous aborderons la mise en correspondance grossière de tous les pixels d'une image avec ceux de son homologue stéréoscopique. Puis nous étudierons le moyen de déterminer un affinage de ces appariements.

Nous évoquerons enfin une autre méthode de mise en correspondance exploitant l'information que l'on projette dans la scène.

2.2.1 Mise en correspondance dense

Nous avons déterminé en section 2.1.4 que l'on pouvait réduire l'espace de recherche de stéréo-correspondant d'un point d'une image rectifiée à l'ensemble des points sur la même ligne dans l'image rectifiée homologue.

La mise en correspondance stéréoscopique consiste à déterminer un point $\mathbf{p}_i$ de l'image $\mathcal{I}_0$ rectifiée et un point $\mathbf{q}_i$ de l'image $\mathcal{I}_1$ rectifiée définis comme les projections perspectives exprimées dans l'espace rectifié d'un même point $\mathbf{X}_i$ de la scène observée par les deux caméras distinctes $\mathcal{C}_0$ et $\mathcal{C}_1$.

La mise en correspondance dense va consister à déterminer, lorsqu'elles existent, les paires de correspondants $(\mathbf{p}_i, \mathbf{q}_i)$, $\forall i \in [1..\text{card}(\mathcal{I}_0)]$. Ces correspondants sont choisis en fonction de critères de similitude sur le signal de luminance au voisinage d'un point. Notons $\delta(\mathbf{p}_i, \mathbf{q}_i)$ la mesure de similitude entre les voisinages des points $\mathbf{p}_i$ et $\mathbf{q}_i$.

Dans les faits, selon le type de scène et la configuration des caméras, tous les points de $\mathcal{I}_0$ n'ont pas de correspondant dans l'image $\mathcal{I}_1$. Ce phénomène est appelé l'*occultation* (figure 2.4).

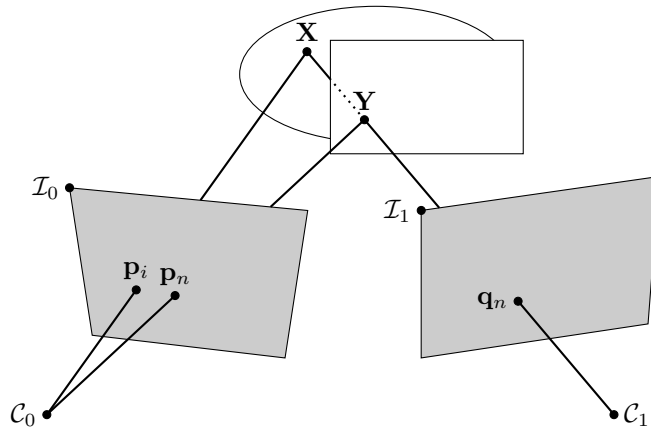


FIGURE 2.4 – Le point $\mathbf{X}$ n'est pas observable par la caméra $\mathcal{C}_1$ bien qu'il admette $\mathbf{q}_n$ comme projeté. Ainsi la paire $(\mathbf{p}_i, \mathbf{q}_n)$ est valide pour la triangulation mais aura une mesure de similitude $\delta(\mathbf{p}_i, \mathbf{q}_n)$ très faible puisque les caméras observent des parties différentes de la scène.

L'algorithme mis en œuvre se décompose de la manière suivante :

- Pour chaque point de l'image de référence $\mathcal{I}_0$, on va déterminer l'espace de recherche – le segment de ligne épipolaire, et le pas de recherche – dans l'autre image,
- On détermine la corrélation du voisinage du point dans l'image de référence avec le voisinage des points candidats dans l'image cible (figure 2.5),
- L'homologue est déterminé par la recherche d'extremum sur le score de corrélation

La recherche peut être accélérée en ayant recours à une approche multi-résolutions. Tout d'abord, on divise la taille des images pour réduire l'espace de recherche. Ensuite on revient progressivement à la taille nominale en limitant la recherche au voisinage de la position trouvée au facteur d'échelle précédent.

De plus, on peut rendre l'appariement plus robuste aux occultations et faux appariement par le biais de diverses contraintes.

- Unicité : deux points de l'image de référence ne peuvent avoir le même homologue,
- Ordre : on peut proposer l'hypothèse que l'ordre des points appariés dans l'image de recherche doit être similaire à celui des points de l'image de référence,
- Symétrie : en réalisant l'appariement des points de $\mathcal{I}_0$ dans $\mathcal{I}_1$ puis de $\mathcal{I}_1$ dans $\mathcal{I}_0$, on peut valider la consistance d'un appariement lorsqu'il existe dans les deux sens.

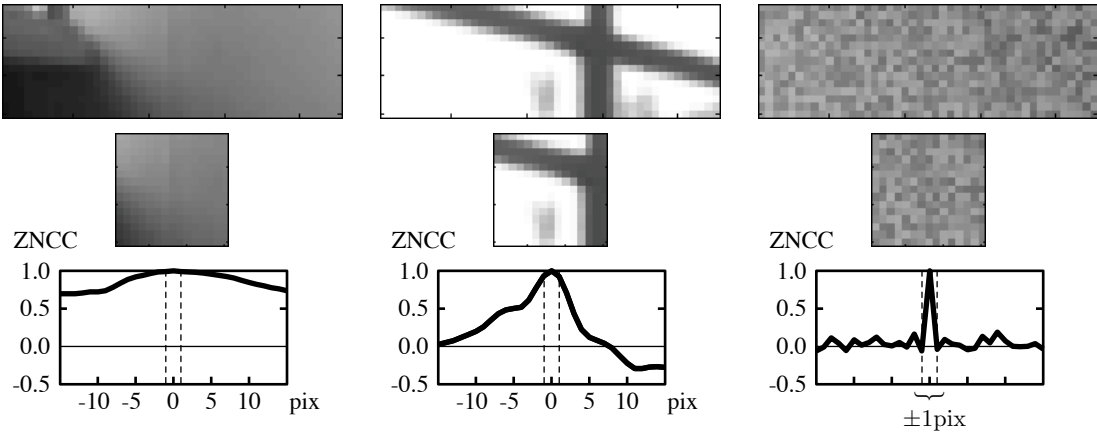


FIGURE 2.5 – Autocorrélation par le score ZNCC (Zero-mean Normalized Cross-Correlation) [Solà Ortega 07]. On corrèle une image de 15×15 pixels avec son voisinage horizontal. *Gauche* : une texture pauvre induit une autocorrélation forte dans tout le voisinage et de ce fait une mauvaise discrimination. *Centre* : dans une zone d'information riche la discrimination est facile et offre un pic de corrélation large permettant une mise en correspondance correcte lors de variations raisonnables d'apparence. *Droite* : une texture de type *bruit blanc* ou approchant fait tendre la fonction d'autocorrélation vers un pic de Dirac, rendant difficile sa détection dès lors que des légers changements d'apparence se produisent.

Le résultat de cette phase permet de construire une *carte de disparité* (figure 2.6). La disparité est le vecteur, dans un repère image normalisé, liant un point à son correspondant. Dans le cas rectifié, on peut se limiter à un scalaire décrivant le déplacement horizontal dans la mesure où l'on sait que le déplacement vertical sera toujours nul.

$$d = x_{\mathbf{m}_1} - x_{\mathbf{m}_0}$$

Dans l'équation précédente, d est la disparité pour un point image donné, $x_{\mathbf{m}_0}$ et $x_{\mathbf{m}_1}$ sont les coordonnées de deux points stéréo-correspondants dans les images rectifiées $\mathcal{I}_0$ et $\mathcal{I}_1$.

2.2.2 Corrélation précise

La méthode de mise en correspondance précédente constitue essentiellement une phase d'appariement initial. Sa précision est limitée à la résolution de l'espace de recherche, par exemple le pixel, et à la précision de la rectification. Il est nécessaire de déterminer avec une plus grande granularité le correspondant de chaque point pour lequel on dispose d'une disparité grossière.

On note que la mise en correspondance d'une région carrée de l'image de référence avec une région de même forme dans l'image de recherche suppose que la surface observée n'apporte pas de déformation perspective d'une caméra à l'autre [Garcia 01]. Cette hypothèse n'est valide que lorsque les caméras sont toutes deux fronto parallèles à un plan parfait. Dans un cas général de surface et de point de vue, cette hypothèse conduit à une approximation générant une imprécision de mise en correspondance.

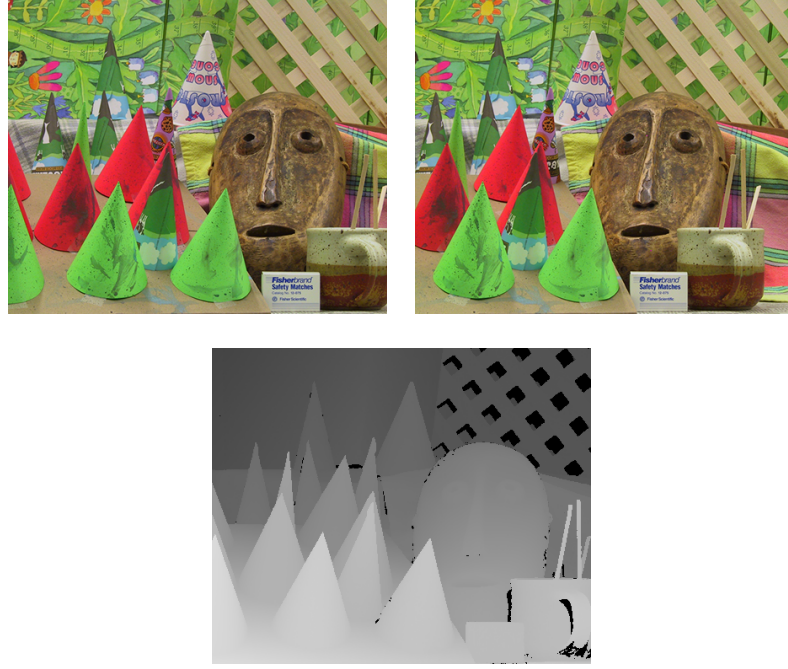


FIGURE 2.6 – Carte de disparité illustrant la mise en correspondance de points dans une paire d'images

Adaptation de la fenêtre de voisinage

En considérant une région W autour d'un point $\mathbf{p}_0 = (x, y)^T$ de l'image de référence. Tout point $\mathbf{p}_i$, $\forall i \in [1..card(W) - 1]$ peut être caractérisé par sa position relative au point $\mathbf{p}_0$ par les grandeurs $\delta_x(i)$ et $\delta_y(i)$ telles que $\mathbf{p}_i = (x + \delta_x(i), y + \delta_y(i))^T$.

La déformation à appliquer à la région W dans l'image de recherche est définie par une fonction $\Delta(x, y, \delta_x, \delta_y)$. Cette fonction peut être approximée par une décomposition en série de Taylor au point $\mathbf{p}_0$. Devernay [Devernay 97] limite l'utilisation de cette série au premier ou second ordre (figure 2.7).

À l'ordre 0, la transformation est équivalente au déplacement de la région selon la disparité $d(x, y)$ sans déformation.

$$\begin{pmatrix} x + \delta_x \\ y + \delta_y \end{pmatrix} \mapsto \begin{pmatrix} x + \delta_x + d(x, y) \\ y + \delta_y \end{pmatrix}$$

Au premier ordre, l'expression de la déformation est exprimée au moyen des dérivées de la disparité $d(x, y)$ selon x et y .

$$\begin{pmatrix} x + \delta_x \\ y + \delta_y \end{pmatrix} \mapsto \begin{pmatrix} x + \delta_x + d(x, y) + \frac{\partial d}{\partial x} \delta_x + \frac{\partial d}{\partial y} \delta_y \\ y + \delta_y \end{pmatrix}$$

Enfin, l'expression au second ordre utilise les dérivées secondes de la disparité.

$$\begin{pmatrix} x + \delta_x \\ y + \delta_y \end{pmatrix} \mapsto \begin{pmatrix} x + \delta_x + d(x, y) + \frac{\partial d}{\partial x} \delta_x + \frac{\partial d}{\partial y} \delta_y + \frac{1}{2} \frac{\partial^2 d}{\partial x^2} \delta_x^2 + \frac{1}{2} \frac{\partial^2 d}{\partial x \partial y} \delta_x \delta_y + \frac{1}{2} \frac{\partial^2 d}{\partial y^2} \delta_y^2 \\ y + \delta_y \end{pmatrix}$$

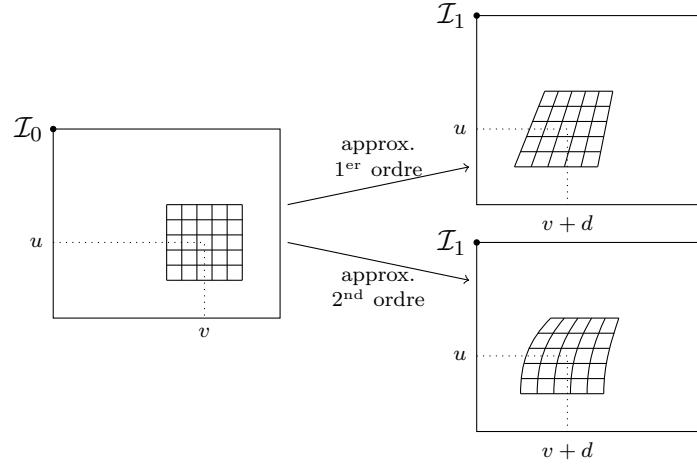


FIGURE 2.7 – Approximations de la déformation de la région de référence au premier et au second ordre dans le cas rectifié.

Trouver les valeurs de la disparité et de ses dérivées est résolu en minimisant le critère de corrélation entre la région de référence et la région déformée dans l'image de recherche.

$$s = \sum_W [\mathcal{I}_0(x + \delta_x, y + \delta_y) - \mathcal{I}_1(x + \Delta(x, y, \delta_x, \delta_y), y + \delta_y)]^2 \quad (2.3)$$

Interpolation de niveaux de gris

Après application de la transformation, la fenêtre de voisinage nous amène à considérer les intensités de coordonnées non entières dans l'image $\mathcal{I}_1$. Par conséquent, il est nécessaire de réaliser une interpolation de niveaux de gris sur l'image.

On peut réaliser cette étape par interpolation bilinéaire. Bien que rapide, cette méthode reste très imprécise et ne permet pas d'atteindre des résultats suffisants. On préférera s'orienter vers des interpolations polynomiales ou par splines de degrés supérieurs à 3 [Thévenaz 00].

2.2.3 Mise en correspondance de primitives

La méthode précédente permet d'atteindre une très bonne précision de mise en correspondance, mais nécessite un temps de calcul important. En utilisant les caractéristiques du motif projeté, il est possible de mettre au point une méthode rapide et relativement précise – bien que moins précise que la corrélation décrite plus haut – pour la mise en correspondance stéréoscopique.

Le motif projeté est composé d'un grand nombre de petites taches noires provoquant des successions de pics et de creux (figure 2.8) dans le signal image. On utilise le détecteur de Blais-Rioux [Blais 85] pour réaliser l'extraction des pics dans chaque image de la paire.

La mise en correspondance grossière permet de bénéficier d'une estimation initiale de la région candidate pour trouver le stéréo-correspondant d'un pic extrait de l'image de référence. On effectue alors une recherche dans le voisinage pour trouver la position exacte du pic par interpolation.

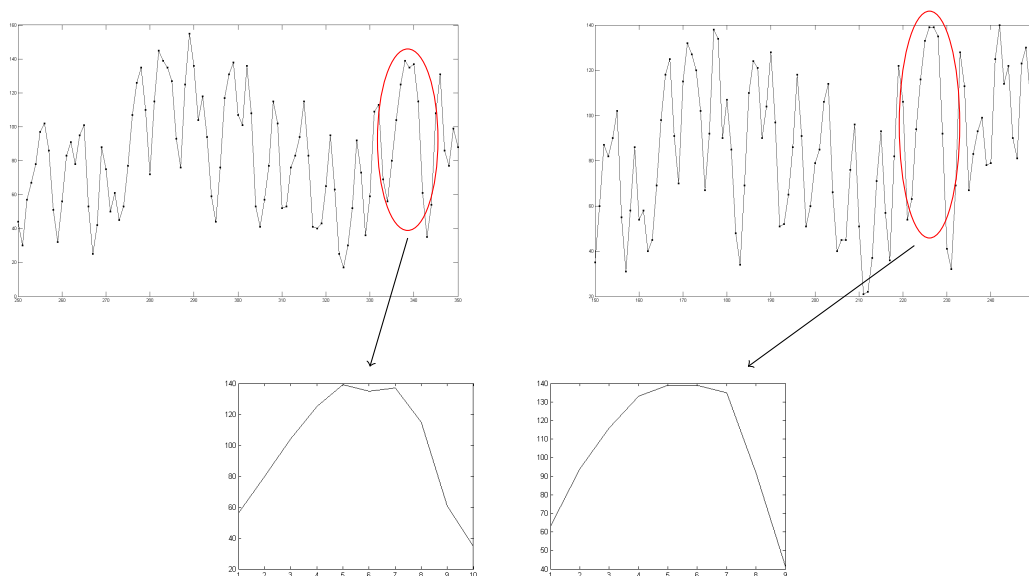


FIGURE 2.8 – Variations de luminance sur une ligne d'une paire d'images. Les pics de luminance sont extraits et mis en correspondance.

2.3 Reconstruction 3D

La mise en correspondance de points dans les images stéréoscopiques nous permet de réaliser la projection inverse afin de retrouver l'information 3D pour chaque paire de points appariés. On parle alors de triangulation. Nous exposons ici les étapes de la reconstruction ainsi que le mécanisme de triangulation à proprement parler.

2.3.1 Chaîne de reconstruction

Lorsque la mise en correspondance de points n'est pas parfaite, les rayons optiques liant les centres des caméras à ces points n'ont aucune intersection. On doit alors procéder à une optimisation ou une heuristique de sélection quelconque.

L'utilisation de la rectification permet, en plus de simplifier l'espace de recherche, d'apporter une contrainte supplémentaire. Dans l'espace rectifié, les points sur la ligne épipolaire sont tous coplanaires, par conséquent il existera toujours une intersection entre les deux rayons optiques.

La figure 2.9 illustre la chaîne de reconstruction mise en œuvre. Les paramètres des modèles sont supposés connus.

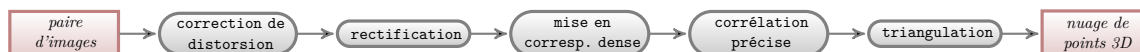


FIGURE 2.9 – Chaîne de reconstruction par stéréovision.

2.3.2 Triangulation

Nous avons vu plus haut les relations entre un point du monde et chacun de ses correspondants dans l'espace rectifié.

$$\mathbf{m}'_0 = \mathbf{H}_0 \mathbf{K}_0 \mathbf{T}^{C_0 \mathcal{W}} \mathbf{M} = \mathbf{P}_0 \mathbf{M} \quad (2.4)$$

$$\mathbf{m}'_1 = \mathbf{H}_1 \mathbf{K}_1 \mathbf{T}^{C_1 C_0} \mathbf{T}^{C_0 \mathcal{W}} \mathbf{M} = \mathbf{P}_1 \mathbf{M} \quad (2.5)$$

Puisque nous sommes dans l'espace rectifié, les stéréo-correspondants ont la même ordonnée y et les matrices $\mathbf{P}_0$ et $\mathbf{P}_1$ ont leurs deuxième et troisième lignes égales à un facteur multiplicatif λ non nul près.

$$\mathbf{p}_2^T = \lambda \mathbf{p}_2'^T ; \mathbf{p}_3^T = \lambda \mathbf{p}_3'^T$$

$$\mathbf{P}_0 = \begin{pmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{pmatrix}$$

$$\mathbf{P}_1 = \begin{pmatrix} \mathbf{p}_1'^T \\ \lambda^{-1} \mathbf{p}_2^T \\ \lambda^{-1} \mathbf{p}_3^T \end{pmatrix} = \begin{pmatrix} \lambda \mathbf{p}_1'^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{pmatrix}$$

On peut alors réexprimer les équations (2.4) et (2.5).

$$\mathbf{m}'_0 = \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{pmatrix} \mathbf{M}$$

$$\mathbf{m}'_1 = \begin{pmatrix} x + d(x, y) \\ y \\ 1 \end{pmatrix} = \begin{pmatrix} \lambda \mathbf{p}_1'^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{pmatrix} \mathbf{M}$$

Ces deux expressions nous permettent d'établir une relation unifiée permettant de lier les coordonnées d'un point du monde à celles d'un point image, connaissant la disparité qui le lie à son stéréo-correspondant.

$$\begin{pmatrix} x \\ y \\ d(x, y) \\ 1 \end{pmatrix}_{\mathcal{I}_0} = \begin{pmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \lambda \mathbf{p}_1'^T - \mathbf{p}_1^T \\ \mathbf{p}_3^T \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}_{\mathcal{W}} = \mathbf{Q}^{-1} \mathbf{M} \quad (2.6)$$

La relation de triangulation est l'opération inverse de celle de l'équation (2.6).

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}_{\mathcal{W}} = \mathbf{Q} \begin{pmatrix} x \\ y \\ d(x, y) \\ 1 \end{pmatrix}_{\mathcal{I}_0} \quad (2.7)$$

La matrice $\mathbf{Q}$ est appelée *matrice de reconstruction*.

Recalage paire à paire

Table des matières

3.1	Recalage grossier	56
3.1.1	État de l'art	56
3.1.2	Estimation de pose par appareillage	57
3.1.3	Estimation de pose par mise en correspondance de descripteurs locaux	58
3.1.4	Estimation de pose par approche globale	64
3.1.5	Méthode hybride mise en œuvre pour le recalage rapide	67
3.2	Recalage précis	69
3.2.1	État de l'art	69
3.2.2	Iterative Closest Points	71
3.2.3	Point-to-plane	74
3.2.4	Point-to-surface	75
3.2.5	Interpolation dans les images, approche pyramidale	75
3.3	Évaluations	78
3.3.1	Pièce « Bunny »	78
3.3.2	Pièce « Étalon »	81
3.4	Conclusions	82

EN considérant l'approche incrémentale qu'impose le besoin de modélisation « temps réel » de notre capteur, nous avons tout d'abord examiné les méthodes de recalage « paire à paire ». Ces méthodes cherchent à recaler notre dernière vue avec une vue précédente ou une composition de plusieurs vues déjà acquises.

Ce type d'approche consiste donc à déterminer la meilleure transformation rigide à appliquer entre deux modèles 3D partiels afin de retrouver la cohérence de la scène en les fusionnant. On peut considérer deux cas à ce problème. Le premier suppose que les deux vues sont exprimées dans des repères quelconques de l'espace, il s'agit donc d'un problème de recherche globale d'un jeu de paramètres. La seconde est une approche locale supposant que les repères des deux modèles sont proches. Il faut donc, dans cette seconde approche, disposer d'un *a priori* sur la position des repères.

De plus, on divise souvent le recalage en deux phases complémentaires : l'approche initiale et le recalage précis. L'approche initiale cherchera à fournir l'estimée initiale de position ou un rapprochement grossier des repères pour permettre l'exploitation de méthodes locales précises en vue de finaliser le recalage.

3.1 Recalage grossier

Les méthodes de recalage grossier cherchent à estimer une transformation rigide grossière entre deux vues 3D sans information *a priori* en se basant sur des algorithmes d'estimation globale ou des systèmes de mesure de pose directs.

Nous présenterons donc ici un état de l'art des méthodes d'approche initiale et nous détaillerons les méthodes les plus pertinentes pour notre projet ainsi que celle que nous proposons et que nous avons intégrée au système final.

3.1.1 État de l'art

Une classification courante [Li 07, Matabosch Geronès 07] de méthodes de recalage initial découple les approches selon que l'on cherchera à créer des descripteurs locaux à mettre en correspondance ou que l'on exploitera une vision globale des vues 3D. Afin d'étendre cette vision, nous ajouterons une classe de méthodes exploitant un équipement particulier pour donner l'information initiale de pose.

Les méthodes de mise en correspondance cherchent à reconnaître des primitives (points, lignes, surfaces ...) particulières dans chaque vue afin de les apparier pour estimer la transformation rigide permettant de les rapprocher.

Les approches de recalage basées sur la mise en correspondance de points ont rapidement proposé de résoudre le problème de l'initialisation en adaptant la méthode afin d'élargir le bassin de convergence [Besl 92, Zhang 92]. Ces approches restent cependant conçues pour fonctionner lorsque les vues ne sont pas trop éloignées initialement. Par conséquent, d'autres méthodes exploitant comme primitive le point 3D ont été proposées, mais en cherchant à construire des descripteurs en fonction de la structure du voisinage. Chua [Chua 97] a proposé une méthode de description de points basée sur la surface locale autour de celui-ci pour créer un vecteur signature. Dans le même type d'approche, Johnson [Johnson 97] a basé son descripteur sur une signature image, étendu plus tard [Carmichael 99] afin de rendre la méthode plus robuste à l'échantillonnage de points. Aujourd'hui encore, ces approches basées sur la mise en correspondance de points 3D restent très utilisées et donnent de bons résultats. Cependant elles ont un coup important pour le calcul des descripteurs et restent très dépendantes de la

résolution des modèles à recalcr. Notons que d'autres méthodes remplacent les descripteurs de points par des lignes [Wyngaerd 02] ou les inflexions de la surface [Krsek 97].

La nature de notre méthode de reconstruction 3D fait que nous pouvons exploiter des images. En effet, il est possible de réaliser une détection et un appariement de points dans l'espace 2D des images puis de réaliser l'estimation de transformation sur leurs équivalents reconstruits en 3D. La mise en correspondance de points dans une séquence d'images peut être réalisée à chaque image, on parle alors de suivi ou de *tracking*. On peut exploiter des points d'intérêt rapides types coins [Harris 88], étendus plus tard par Shi et Tomasi [Shi 94]. Des méthodes de recherche et de description de points sont également applicables dans l'espace des images. Lowe [Lowe 04] a proposé un descripteur de points invariants au changement d'échelle, les *Scale-Invariant Feature Transform* (SIFT). Les évolutions récentes dans ce domaine ont surtout porté l'attention sur la mise en place de descripteurs présentant plus d'invariances et de rapidité, tels que les *Speed-Up Robust feature* (SURF) [Bay 06] ou les détecteurs FAST [Taylor 09].

Les approches globales cherchent à donner une représentation de l'ensemble d'un nuage ou explorent l'espace de toutes les poses ou toutes les combinaisons de mise en correspondance de points en se donnant une heuristique ou un moyen de discriminer ces espaces.

La méthode de Chen [Chen 98] est basée sur une approche hypothèse-test, en cherchant à mettre en correspondance un triplet de points caractérisés par leur distance à l'aide de tirages aléatoires. Cette méthode nécessite un nombre important de tirages pour être valide et nécessite donc un long temps de calcul.

Les méthodes les plus rapides dans cette catégorie cherchent plutôt à décrire le modèle par des primitives simples. L'usage de l'*Analyse en Composantes Principales* [Kim 03] permet de réduire l'ensemble des données à une base de trois vecteurs constituant un descripteur invariant aux transformations indépendant des problématiques de mise en correspondances de points. Une idée similaire avait été proposée par Feldmar [Feldmar 96] qui proposait de réduire l'ensemble des données de chaque modèle selon la courbure principale autour d'une seule mise en correspondance de point. Cette approche a été étendue par Tarel [Tarel 98] dont la méthode estime des surfaces polynomiales implicites pour chaque nuage de points et détermine le mouvement entre elles. Ces approches nécessitent cependant un chevauchement important des vues à recalcr et échouent donc régulièrement.

Les travaux récents exploitant des approches globales pour résoudre ce problème tendent à exploiter des statistiques robustes et le filtrage stochastique pour décrire l'espace de recherche par une densité de probabilité et ainsi le discriminer. Ma [Ma 04] puis Moghari [Moghari 07] ont proposé des approches exploitant le filtrage particulaire ou le *Unscented Kalman Filter* et exploitent un critère de plus proche voisin pour la mesure *a posteriori*. Sandhu [Sandhu 10] étend leurs approches en exploitant un filtrage particulaire couplé à une optimisation locale pour la mesure *a posteriori*.

3.1.2 Estimation de pose par appareillage

Une approche directe pour obtenir une première estimation de la transformation 3D est l'utilisation de systèmes de mesure de pose à chaque instant d'acquisition. Ces méthodes nécessitent un appareillage particulier et ne sont donc adaptées qu'à des applications particulières.

Une méthode courante sur les appareils basés vision tenus à la main est l'usage de centrales inertielles, comme dans notre système. Une centrale inertielle mesure la dynamique du mouvement. En intégrant ces mesures au cours du temps, on peut disposer d'une estimée de pose ou de déplacement sur le capteur. Cependant, si l'on peut envisager qu'une centrale inertielle très haut de gamme peut fournir une mesure à très faible dérive, l'utilisation de centrales bas coût nécessite quelques précautions. En effet, l'importance du bruit et du biais sur ces gammes d'appareils provoque une rapide dérive à l'intégration, rendant les mesures inutiles. L'exploitation de ces mesures bruitées se fait en général selon deux approches. La composition des mesures tant dynamiques (accélération et vitesse angulaire) que statiques (champ magnétique) permet l'estimation de l'attitude du capteur de manière presque directe. Un léger retard est introduit si l'on lisse l'estimation au moyen d'une fenêtre glissante sur les dernières mesures. La seconde approche est de combiner ces mesures dynamiques à des mesures extéroceptives, type GPS ou vision. Dans le premier cas, l'estimée initiale est incomplète car aucune information sur la translation n'est fournie. Dans le second cas il est nécessaire de recourir à des algorithmes ne relevant pas exclusivement de méthodes de mesure directe par appareillage.

Lors d'exploitation de mesures 3D dans des applications robotiques, l'estimée initiale de pose peut être donnée directement par le robot. Encore une fois, il faut considérer cette méthode avec plus ou moins de prudence selon la précision que peut apporter le système. Dans le cas d'un robot fixe, type bras, les phénomènes de dérive sont négligeables et la précision de l'estimée fournie sera liée à la précision des codeurs du robot. Dans le cas d'un système mobile, au contraire, la dérive existe de façon conséquente du fait des aléas du terrain : glissements, heurts contre des obstacles mobiles ... ainsi que du fait de l'absence de base fixe de référence. Ici l'odométrie du robot pourra ne pas suffire ou risquera de devenir rapidement inexploitable. Une correction au moyen de mesures GPS ou vision par exemple peut alors être envisagée pour corriger la dérive.

Une autre approche, souvent utilisée avec les numériseurs 3D fixes, est l'exploitation d'un plateau tournant. Soit par lecture de codeurs sur le plateau ou par commande directe d'orientation, on est en mesure de connaître le déplacement de l'objet devant le capteur. A nouveau, ce type d'approche fournit une estimée incomplète puisque la composante de translation ne peut être déterminée directement.

Certains systèmes ont recours à une référence magnétique pour se positionner. La référence est placée sur la pièce à numériser ou à proximité et le capteur détermine sa position et son attitude relativement à elle. Ce type de méthode a l'avantage de proposer une lecture directe complète sans dérive de la pose du capteur par rapport à la pièce à numériser. Cependant il s'agit de méthode très sensibles aux perturbations électromagnétiques, fréquentes dans les environnements de travail (ordinateurs, télécommunications ...). Cela rend, de plus, la numérisation de pièces métalliques problématique.

3.1.3 Estimation de pose par mise en correspondance de descripteurs locaux

Les méthodes présentées ici utilisent des descripteurs 3D ou 2D pour réaliser des mises en correspondance, sans information autre au départ.

Il est à noter que les méthodes par descripteurs 3D sont plus génériques, elles exploitent le résultat de la reconstruction sans autre information. Les descripteurs 2D proposés se basent sur le traitement d'image et la mise en correspondance de descripteurs basés luminance. Pour

autant, elles sont regroupées dans cette section puisqu'il s'agit de méthodes « locales » dans le sens où l'on n'exploite pas l'intégralité d'un nuage mais des points d'intérêt et leurs voisinages.

Lorsque l'on dispose d'ensembles de paires 3D, il est possible de déterminer la transformation rigide (R, t) permettant d'approcher les deux ensembles de points au moyen de méthodes de minimisation au sens des moindres carrés [Arun 87, Lorusso 95].

Point Signature

Introduite par Chua [Chua 97], cette méthode crée un vecteur signature dont le but est de décrire la surface au voisinage d'un point 3D (figure 3.1).

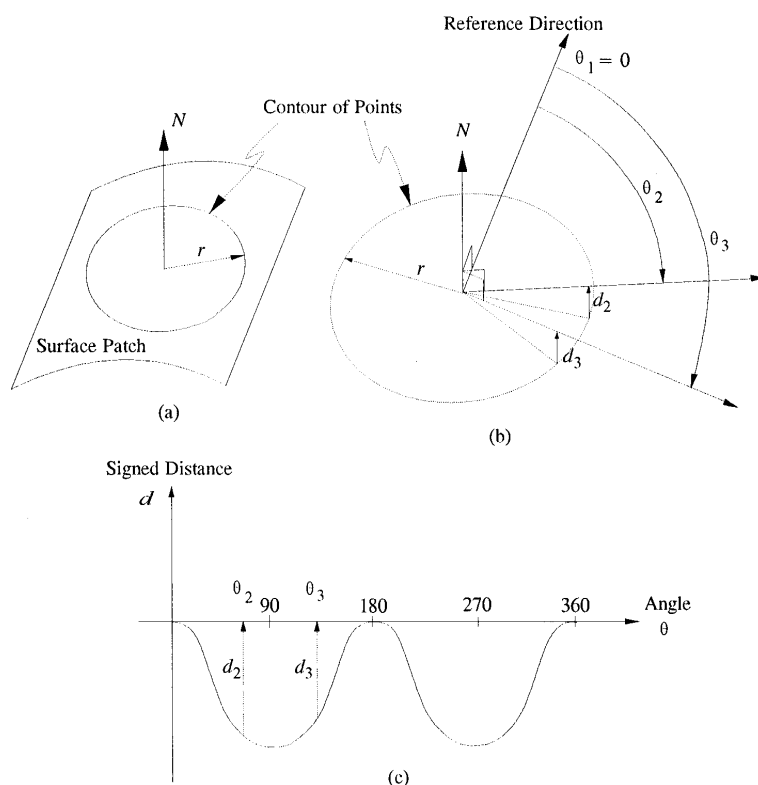


FIGURE 3.1 – Point signature [Chua 97]. (a) Définition du contour de surface (b) Expression des points du contour dans la nouvelle référence (c) Signature du point.

Au moyen d'une sphère centrée sur le point p à décrire, on délimite un contour C à l'intersection de cette sphère avec la surface du modèle. L'auteur approxime le plan tangent au point p en calculant le plan moyen P au contour C translaté dans la direction de la normale n_1 à P afin d'intersecter le point p . Le plan translaté, approximation du plan tangent, est noté P' .

Le contour C est projeté orthogonalement sur P' , déterminant le contour C' . Pour chaque point de C , apparaît la distance d avec son projeté dans C' .

Le point le plus éloigné de p sur la courbe C' est choisi comme premier point du descripteur. On note n_2 le vecteur de p à ce point. On a donc n_2 inclut dans le plan P' et n_1 perpendiculaire à ce plan.

La signature d'un point est une expression du contour C dans le repère formé par les vecteurs (n_1, n_2) . On exprime chaque point par l'angle que forme le vecteur entre son projeté et p avec n_2 et par la distance d de projection.

Ces descripteurs sont indépendants du repère du modèle. Cela permet la mise en correspondance de points par corrélation des signatures.

La méthode est cependant coûteuse. Particulièrement lorsque l'on travaille sur des données de type points, le calcul de surfaces est une étape pouvant grever sérieusement les performances d'un algorithme. De plus, l'indépendance du descripteur au repère est uniquement lié au choix du point de départ du descripteur, dont le critère est très sensible au bruit ou aux erreurs provoqués par une mauvaise ou incomplète interpolation de surface.

Spin images

La méthode des *Spin images* a été proposée par Johnson [Johnson 97] et étendue par Carmichael [Carmichael 99]. Les auteurs proposent de segmenter l'espace autour d'un point p selon deux grandeurs : la distance α à la normale au point p et la distance β au plan tangent au point p .

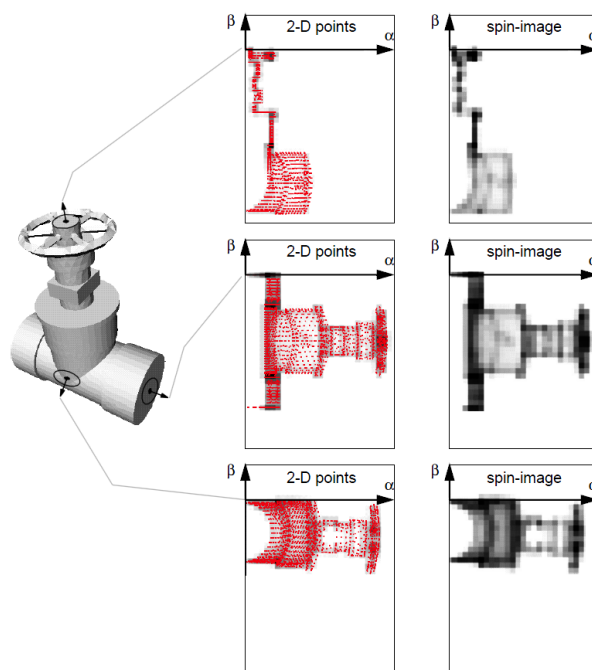


FIGURE 3.2 – Spin-images de quelques points d'un modèle [Johnson 97]

On commence donc par estimer le plan tangent à l'aide des points voisins puis on délimite une région autour du point. Les grandeurs α et β sont calculées pour chaque point de la région. On constitue une table indexée par ces deux grandeurs où chaque cellule contient le nombre de point contenus dans le segment correspondant. Cela constitue une « image 2D » de la géométrie autour du point, que l'on peut ensuite corréler aux spin-images des autres points.

En n'exploitant que les points du nuage on se rend très dépendants de la résolution de ce dernier. Par conséquent on s'attachera à réaliser un échantillonnage uniforme au préalable en interpolant des points dans chaque maille du modèle.

La figure 3.2 illustre les descripteurs de quelques points.

On obtient ainsi un descripteur très informatif et facile à mettre en correspondance par corrélation. La méthode reste cependant très liée à la précision du maillage, et donc à la résolution du nuage de points sous-jacent.

Descripteurs 2D

Ces méthodes exploitent la nature « image » du numériseur. En effet, l'information 3D est calculée sur la base d'images 2D de luminance. Il est donc possible de réaliser une mise en correspondance de pixels dans ces images, les points 3D issus de la reconstruction de ces pixels seront appariés de même. Il sera alors possible de calculer la transformation rigide comme nous l'expliquons plus haut.

Suivi par corrélation de voisinage Si on exploite une séquence d'images, la méthode de Kanade-Lucas-Tomasi [Shi 94] exploite le calcul du flot optique pour le suivi de points d'intérêt. Cette méthode cherche à minimiser un critère de corrélation en exploitant l'hypothèse de conservation de la luminance au voisinage d'un point.

En considérant un point $\mathbf{x}$ de l'image $I(\mathbf{x}, t)$ acquise à l'instant t , cette hypothèse peut être formalisée de la façon suivante :

$$I(\mathbf{x}, t) = I(\delta(\mathbf{x}), t + \tau) \quad \forall \mathbf{x} \in W \quad (3.1)$$

où W est le voisinage du point considéré et $\delta(\mathbf{x})$ exprime la transformation linéaire assurant le passage d'une image acquise au temps t à celle acquise au temps $t + \tau$. Cette transformation est supposée affine :

$$\delta(\mathbf{x}) = \mathbf{A}\mathbf{x} + \mathbf{d}$$

où $\mathbf{d}$ est la translation du centre de la fenêtre et

$$\mathbf{A} = \mathbf{1} + \begin{bmatrix} d_{xx} & d_{xy} \\ d_{yx} & d_{yy} \end{bmatrix}$$

avec $\mathbf{1}$ la matrice identité.

Du fait du bruit et de l'imperfection du modèle de mouvement, on ne peut satisfaire l'équation (3.1), $\mathbf{A}$ et $\mathbf{d}$ sont donc déterminés par la minimisation d'une mesure de similarité :

$$\epsilon = \sum_W [I(\delta(\mathbf{x}), t + \tau) - I(\mathbf{x}, t)]^2 \quad (3.2)$$

Considérant que les déplacements sont faibles, donc que l'acquisition des images est suffisamment rapide, on peut approximer $I(\delta(\mathbf{x}), t + \tau)$ par un développement en série de Taylor au premier ordre. En remplaçant son expression dans l'équation (3.2) et en différenciant par rapport aux inconnues $\mathbf{A}$ et $\mathbf{d}$, on obtient le système linéaire suivant :

$$\mathbf{Tz} = \mathbf{a} \quad (3.3)$$

avec $\mathbf{z} = [d_{xx} \ d_{yx} \ d_{xy} \ d_{yy} \ d_x \ d_y]^T$, le vecteur collectant les inconnues de la déformation $\mathbf{A}$ et du déplacement $\mathbf{d}$ et, en notant I_t , I_x , I_y les gradients temporels et spatiaux calculés au point $\mathbf{x}$, $\mathbf{T}$ et $\mathbf{a}$ sont donnés par :

$$\begin{aligned} \mathbf{a} &= -\tau \sum_W I_t [xI_x \ xI_y \ yI_x \ yI_y \ I_x \ I_y]^T \\ \mathbf{T} &= \sum_W \begin{bmatrix} \mathbf{U} & \mathbf{V} \\ \mathbf{V}^T & \mathbf{G} \end{bmatrix} \\ \mathbf{U} &= \begin{bmatrix} x^2 I_x^2 & x^2 I_x I_y & xy I_x^2 & xy I_x I_y \\ x^2 I_x I_y & x^2 I_y^2 & xy I_x I_y & xy I_y^2 \\ xy I_x^2 & xy I_x I_y & y^2 I_x^2 & y^2 I_x I_y \\ xy I_x I_y & xy I_y^2 & y^2 I_x I_y & y^2 I_y^2 \end{bmatrix} \\ \mathbf{V}^T &= \begin{bmatrix} xI_x^2 & xI_x I_y & yI_x^2 & yI_x I_y \\ xI_x I_y & xI_y^2 & yI_x I_y & yI_y^2 \end{bmatrix} \\ \mathbf{G} &= \begin{bmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{bmatrix} \end{aligned}$$

Dans l'équation (3.3), $\mathbf{z}$ peut être obtenu à l'aide de l'algorithme de Levenberg-Marquardt. Le modèle affine est utilisé pour comparer le point considéré et son voisinage entre la première image et l'image courante afin de jauger la qualité du tracking.

Cette méthode a le défaut de nécessiter une séquence continue et à relativement haute fréquence d'images. De plus les mises en correspondance sont peu précises et peu robustes.

Scale Invariant Feature Transforms (SIFT) La méthode de Lowe [Lowe 04] permet de détecter et de décrire des points clefs dans une image qu'il sera possible de détecter et apparier dans d'autres images prises selon des points de vue différents.

Les propriétés de ces descripteurs sont d'être invariant selon la translation, le changement d'échelle et la rotation. Dans une certaine mesure, ils sont robustes au changement d'illumination et à une déformation géométrique locale.

La détection se fait au moyen de différences-de-Gaussiennes calculées entre divers niveaux de lissages consécutifs d'une même image. On applique à l'image des lissages gaussiens en élargissant incrémentalement le noyau de convolution Gaussien. La différence entre deux images consécutives nous fournit une image d'une fonction différence-de-Gaussienne. Pour chacune de ces fonctions, les extremums locaux sont détectés pour constituer les points clefs candidats.

Pour chaque point, une relocalisation précise est effectuée en interpolant la position de l'extremum. Il est alors réalisé un filtrage, en commençant par rejeter les points à faible contraste, considérés instables. De plus, on note que de nombreux points seront détectés le long des arêtes car la fonction différence-de-Gaussienne a une forte réponse dans ce type de régions. On filtre donc les points ayant une faible courbure principale dans la direction de la perpendiculaire à l'arête. Pour cela on applique une variante de la méthode de Harris [Harris 88] en observant le rapport des valeurs propres d'une matrice Hessienne calculée au point et facteur d'échelle de l'extremum considéré.

Afin de se rendre indépendant à la rotation, l'orientation principale du point est déterminée à l'aide du gradient local dans l'image. Le descripteur est alors formé en composant

les amplitudes et orientations des gradients dans une région autour du point aggregées en histogrammes d'orientation.

La mise en correspondance des descripteurs d'une image à l'autre est réalisée par une recherche du plus proche voisin dans l'espace des descripteurs. Pour cela on adopte une approche rapide de la recherche par *k-d tree*, l'algorithme BBF (*Best-Bin-First*).

Cette méthode est très largement utilisée pour la mise en correspondance de points entre deux images. Les résultats sont fiables mais restent sensibles de façon importante aux déformations perspectives.

Speeded Up Robust Features (SURF) Dans une approche similaire à celle de Lowe [Lowe 04], Bay [Bay 06] propose un descripteur plus rapide permettant d'accroître la robustesse aux déformations affines. Pour cela, l'étape de détection et l'information contenue dans le descripteurs sont basées sur des méthodes différentes. L'usage d'images intégrales est également généralisé.

La détection est faite en convoluant l'image par des dérivées au second ordre de noyaux Gaussiens. Les points d'intérêt extraits sont directement calculés dans les images convoluées en ajoutant un critère maximisé localement dans l'image et entre les facteurs d'échelle voisins.

Le descripteur est composé des sommes des réponses aux ondelettes de Haar $\sum d_x$ et $\sum d_y$ calculées selon les direction, respectivement, x et y , ainsi que les sommes des valeurs absolues $\sum |d_x|$ et $\sum |d_y|$ de ces mêmes réponses (figure 3.3), dans une région autour du point d'intérêt.

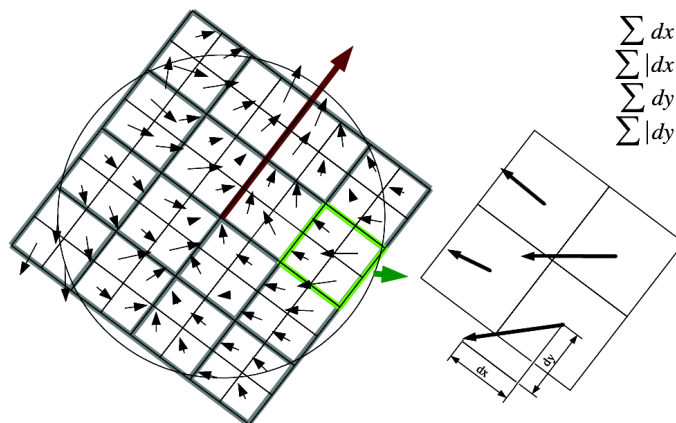


FIGURE 3.3 – Descripteur SURF [Bay 06], Le voisinage est décomposé en 4×4 régions pour le calcul des réponses d_x et d_y , chaque région est décrite par les sommes des réponses et les sommes des valeurs absolues des réponses.

Cette méthode permet d'atteindre des résultats comparables à ceux de l'approche des points *SIFT*. Il est à noter qu'elle permet un gain de temps significatif ainsi qu'un gain de robustesse face aux déformations perspectives, auxquelles elle reste cependant sensible.

Estimation de la transformation liant deux ensembles appariés

Nous détaillons ici la méthode de [Arun 87] permettant de déterminer la transformation rigide $(\mathbf{R}|\mathbf{t})$ entre deux ensembles de N paires de $(\mathbf{p}_i^k, \mathbf{p}_i^{k+l})$. On estime la pose par minimisation au sens des moindres carrés de l'erreur d'appariement.

$$\epsilon = \sum_{i=1}^N \left[\mathbf{p}_i^k - (\mathbf{R}\mathbf{p}_i^{k+l} + \mathbf{t}) \right]^2 \quad (3.4)$$

On résout le problème en deux temps en séparant la rotation et la translation. On peut se rendre indépendant à la translation en exprimant chaque point par rapport au centre de masse du nuage dont il est issu.

$$\begin{aligned} \mathbf{p}_m^k &= \frac{1}{N} \sum_{i=1}^N \mathbf{p}_i^k & \mathbf{p}_m^{k+l} &= \frac{1}{N} \sum_{i=1}^N \mathbf{p}_i^{k+l} \\ \mathbf{q}_i^k &= \mathbf{p}_i^k - \mathbf{p}_m^k & \mathbf{q}_i^{k+l} &= \mathbf{p}_i^{k+l} - \mathbf{p}_m^{k+l} \end{aligned}$$

Ainsi on peut ré-exprimer l'équation (3.4).

$$\begin{aligned} \epsilon &= \sum_{i=1}^N \left[\mathbf{q}_i^k - \mathbf{R}\mathbf{q}_i^{k+l} \right]^2 \\ \mathbf{t} &= \mathbf{p}_m^k - \mathbf{R}\mathbf{p}_m^{k+l} \end{aligned}$$

On calcule la matrice de corrélation $\mathbf{H}$ basée sur ces paires de points centrés. On prouve que la matrice de rotation optimale $\mathbf{R}$ peut être trouvée par la méthode dite *Procrustes orthogonale*, basée sur la décomposition SVD de la matrice $\mathbf{H}$.

$$\begin{aligned} \mathbf{H} &= \sum_{i=1}^N \mathbf{q}_i^{k+l} \mathbf{q}_i^{kT} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T \\ \mathbf{R} &= \mathbf{V}\mathbf{U}^T \end{aligned}$$

Si $\det(\mathbf{R}) = -1$, la matrice $\mathbf{R}$ est une réflexion. On corrige ce problème en inversant la colonne de la matrice $\mathbf{V}$ correspondant à la valeur singulière nulle de $\mathbf{H}$. En notant cette matrice modifiée $\mathbf{V}'$, on réexprime la matrice de rotation optimale par $\mathbf{R} = \mathbf{V}'\mathbf{U}^T$.

3.1.4 Estimation de pose par approche globale

La notion d'approche globale englobe les méthodes qui exploitent l'information du nuage de points 3D dans son ensemble et non plus une observation de descripteurs épars.

Analyse en composantes principales

Le principe de cette méthode est de regrouper l'information du nuage en axes principaux par le biais d'une *Analyse en Composantes Principales* sur l'ensemble des points. En alignant ensuite les axes principaux on réalise un rapprochement des vues.

Cette analyse se base sur le calcul de la matrice des moments centrés d'ordre deux de la distribution spatiale des points d'un nuage décomposée en valeurs singulières.

$$\begin{aligned}\bar{\mathbf{M}} &= [\mathbf{p}_1 - \bar{\mathbf{p}} \dots \mathbf{p}_N - \bar{\mathbf{p}}] \\ \Sigma_{\mathbf{p}} &= \frac{1}{N} \bar{\mathbf{M}} \bar{\mathbf{M}}^T \\ \Sigma_{\mathbf{p}} &= \mathbf{U} \mathbf{S} \mathbf{U}^T\end{aligned}$$

où $\mathbf{U}$ est la matrice des vecteurs propres de $\Sigma_{\mathbf{p}}$, la matrice des moments centrés d'ordre deux.

Le calcul de la transformation rigide entre deux nuages de points se fait en calculant un produit de leurs matrices de vecteurs propres en rapprochant leurs centres de masse. On note $\mathbf{U}_i$ et $\mathbf{U}_{i+1}$ les matrices correspondant aux deux nuages de points que l'on souhaite recalculer et $\bar{\mathbf{p}}_i$ et $\bar{\mathbf{p}}_{i+1}$ leurs centres de masse. Alors on trouve la transformation $(\mathbf{R}, \mathbf{t})$ comme suit :

$$\begin{aligned}\mathbf{R} &= \mathbf{U}_i \mathbf{U}_{i+1}^{-1} \\ \mathbf{t} &= \bar{\mathbf{p}}_{i+1} - \mathbf{R} \bar{\mathbf{p}}_i\end{aligned}$$

Cette méthode est très simple et particulièrement rapide. Cependant elle nécessite que les modèles recalés soient très recouvrants l'un par rapport à l'autre. En effet, si les vues sont trop éloignées les axes principaux seront trop différents pour permettre un calcul de transformation significatif.

Prédiction-vérification

Cette méthode se base sur les travaux de Chen [Chen 98] et de Feldmar [Feldmar 96]. Il s'agit de trouver des candidats possible à la mise en correspondance de primitives (points, courbures) entre les deux vues.

Le principe de cette méthode est de sélectionner trois points de la première vue p_1, p_2, p_3 et de déterminer les distances d_{12}, d_{13} et d_{23} entre eux. On va successivement regarder un échantillonnage de points de la seconde vue et le considérer comme un correspondant de p_1 possible. Les correspondants de p_2 et p_3 sont recherchés de sorte qu'ils respectent les contraintes de distances d_{12}, d_{13} et d_{23} caractérisant le triplet (p_1, p_2, p_3) .

Pour trouver les triplets correspondants candidats, on cherchera les correspondants possible de p_2 sur la surface d'une sphère centrée autour de p_1 et de rayon d_{12} . Si aucun point ne peut être trouvé, un nouveau correspondant pour p_1 doit être choisi. Puis, pour chaque candidat à la correspondance de p_2 , on cherchera les correspondants de p_3 sur le cercle formé par l'intersection des surfaces des sphères centrées en p_1 et p_2 et de rayons respectifs d_{13} et d_{23} .

Si un triplet de correspondant valide n'est pas trouvé on sélectionne trois nouveaux points. Sinon, chaque triplet candidat ayant pu être trouvé donne lieu au calcul d'une transformation candidate. La meilleure transformation est celle qui minimise les distances de chaque point d'une vue avec son plus proche voisin dans l'autre vue.

Cette méthode nécessite un nombre réduit de points par vue. Sans cela la complexité devient trop importante et le temps de calcul s'en trouve sérieusement affecté. Cependant,

si l'on diminue la résolution des vues, la possibilité de mise en correspondance de triplets s'amenuise et la précision du recalage s'en trouve affecté.

Filtrage stochastique

Ce type de méthodes se base sur l'estimation d'une densité de probabilité *a posteriori* d'un modèle dans l'espace des poses 3D. On exploite pour cela les techniques de filtrage stochastique.

Le vecteur d'état est donc une pose 3D, pouvant être composé comme suit :

$$\mathbf{x} = [x \ y \ z \ \phi \ \theta \ \psi]^T$$

où (x, y, z) sont les composantes de la translation et (ϕ, θ, ψ) sont les angles de la rotation.

L'approche simple pour résoudre le recalage comme un problème de filtrage stochastique est de propager l'état sans dynamique et de fonder l'observation sur la mise en correspondance de points au plus proche voisin.

$$\mathbf{x}_{t+1} = \mathbf{F}(\mathbf{x}_t, \mathbf{u}_t, \mathbf{w}_t) = \mathbf{x}_t + \mathcal{N}(\mathbf{0}, \mathbf{Q}_t) \quad (3.5)$$

Notre problème n'est pas dépendant d'un temps « physique », nous cherchons simplement une convergence d'un modèle vers un autre. De ce fait, on n'introduit pas de dynamique à proprement parler, l'état est propagé inchangé. Nous tenons compte d'un bruit Gaussien servant à agréger les erreurs de modélisation et introduire la possibilité d'un mouvement à notre estimation afin qu'elle puisse être corrigée par filtrage. La covariance $\mathbf{Q}_t$ de ce bruit convergera vers $\mathbf{0}$ à l'approche de la solution.

Les observations réalisables sur notre processus sont les positions du modèle que l'on souhaite recalculer. On applique la pose donnée par l'*a priori* courant de l'état, donné par l'équation (3.5).

$$\mathbf{z}_t = \mathbf{H}(\mathbf{x}_t, \mathbf{u}_t, \mathbf{v}_t) = [\mathbf{z}_t^0 \ \mathbf{z}_t^1 \ \dots \ \mathbf{z}_t^N]$$

$$\mathbf{z}_t^i = \mathbf{\Omega}(\phi, \theta, \psi) \mathbf{u}_t^i + [x \ y \ z]^T + \mathcal{N}(\mathbf{0}, \mathbf{R}_t^i)$$

La fonction $\mathbf{\Omega}(\phi, \theta, \psi)$ est la matrice de rotation calculée sur la base des angles ϕ , θ et ψ . Le vecteur $\mathbf{u}_t = [\mathbf{u}_t^0 \ \mathbf{u}_t^1 \ \dots \ \mathbf{u}_t^N]$ est constitué de l'ensemble des points du modèle à recalculer. La matrice $\mathbf{R}_t^i$ est la covariance du bruit que l'on ajoute au modèle liée à chaque point.

L'erreur d'estimation est alors la distance de l'observation avec la surface du modèle sur lequel on souhaite se recalculer. Pour cela on peut réaliser un appariement au plus proche voisin entre les deux ensembles de points.

Sandhu [Sandhu 08] se base sur ce type d'approche et l'étend en proposant, particulièrement, une observation non plus basée sur l'ensemble des points mais sur une optimisation locale afin d'en déduire une pose *a posteriori*. L'erreur d'estimation est alors la simple différence entre la pose estimée et cette pose « mesurée ».

L'inconvénient principal de ce type d'approche est son temps de calcul. En effet, afin d'éviter d'être contraint par un bassin de convergence trop étroit, il est nécessaire d'avoir recours à des distributions multi-modales ou des filtres multi-hypothèses. Par conséquent, la complexité du filtre augmente de façon importante, d'autant plus si le nombre de points est important dans les modèles.

3.1.5 Méthode hybride mise en œuvre pour le recalage rapide

Les méthodes de recalage peuvent nécessiter un temps de calcul important selon l'approche envisagée. Notre besoin de fonctionnement temps-réel impose de réduire ce temps le plus possible. Pour cela, il convient d'envisager une solution exploitant au mieux l'ensemble des possibilités que nous offrent le matériel et les ressources que nous mettons en œuvre.

Le recalage initial que nous proposons cherche à coupler des informations très rapidement accessibles provenant de la centrale inertielle aux informations provenant des caméras. En première approche, nous avons étudié l'utilisation d'un tracking de points dans le flux d'images des caméras. Cependant, cela impose un fonctionnement en acquisition continue, ce qui est plus contraignant que ce que propose le mode de fonctionnement nominal. On se référera à la section 1.2 pour les détails du mode nominal. Nous nous sommes donc orienté vers le couplage des mesures inertielles avec une mise en correspondance de descripteurs SURF (cf section 3.1.3).

Notre méthode propose donc d'exploiter au mieux les données disponibles afin de déterminer par un calcul rapide la composante de rotation au moyen des mesures inertielles puis un calcul robuste de translation par une amélioration de la mise en correspondance de descripteurs SURF dans les images.

Retour sur les données

Nous avons décrit en section 1.1.2 le mode d'acquisition nominal ainsi que les données générées. Ainsi, à chaque acquisition, le système génère une paire d'images avec projection de motif et une paire sans. A chaque paire, une mesure d'attitude synchronisée est également disponible.

Nous avons également noté que les paires d'images sont acquises séquentiellement. Entre les deux prises de paires d'images, l'opérateur peut avoir bougé, même légèrement. Les poses du capteur à ces deux instants n'est donc pas strictement la même. Utiliser conjointement les images avec et sans projection de motif conduit donc à réaliser une approximation, et donc l'ajout d'une erreur.

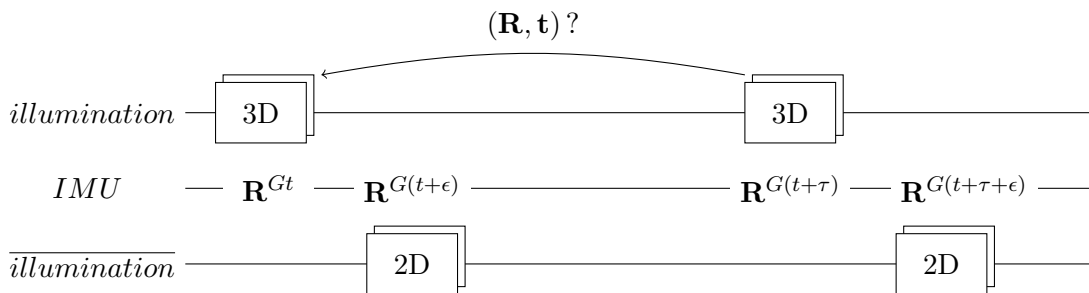


FIGURE 3.4 – Chaque acquisition d'images – avec ou sans illumination – donne lieu à l'estimation d'une attitude par la centrale inertielle. En composant les informations on détermine la pose de recalage recherchée

La figure 3.4 illustre le séquençage des données sur deux acquisitions distinctes. Notre problème est de donner une estimation grossière de la pose $(\mathbf{R}, \mathbf{t})$ permettant de passer du repère d'acquisition des caméras au temps t à celui au temps $t + \tau$:

$$(\mathbf{R}, \mathbf{t}) : \mathbf{X}_i^t = \mathbf{R}\mathbf{X}_i^{t+\tau} + \mathbf{t} \quad (3.6)$$

où $\mathbf{X}_i^t$ est un point du modèle exprimé dans le repère des caméras à l'instant d'acquisition t .

Composante de rotation

Les mesures inertiellees fournies permettent de connaitre l'attitude du repère de la caméra coaxiale aux instants d'acquisition par rapport à un repère global défini à l'initialisation des capteurs. La position de ce repère n'a pas besoin d'être connue tant que l'on utilise ces mesures relativement les unes aux autres. Cependant aucune information concernant la translation ne peut être déterminée directement par ce biais.

On réexprime l'équation (3.6) en passant par ce repère global :

$$\begin{aligned} \mathbf{X}_i^G &= \mathbf{R}^{G(t+\tau)}\mathbf{X}_i^{t+\tau} + \mathbf{t}^{G(t+\tau)} \\ \mathbf{X}_i^t &= (\mathbf{R}^{Gt})^{-1}(\mathbf{X}_i^G - \mathbf{t}^{Gt}) \\ &= \mathbf{R}^{tG}\mathbf{R}^{G(t+\tau)}\mathbf{X}_i^{t+\tau} + \mathbf{R}^{tG}(\mathbf{t}^{G(t+\tau)} - \mathbf{t}^{Gt}) \end{aligned} \quad (3.7)$$

où $\mathbf{R}^{tG}$ est la transformation inverse de $\mathbf{R}^{Gt}$. On note l'apparition des vecteurs $\mathbf{t}^{G(t+\tau)}$ et $\mathbf{t}^{Gt}$, translations introduites par les deux changements de repères successifs.

Par identification de l'équation (3.7) dans l'équation (3.6) on obtient :

$$\mathbf{R} = \mathbf{R}^{tG}\mathbf{R}^{G(t+\tau)} \quad (3.8)$$

Les mesures inertiellees étant bruitées, on obtient une approximation de la composante de rotation. Il faut à présent déterminer la composante de translation.

Composante de translation

La détection et la mise en correspondance de points d'intérêt dans les images sans illumination permet de manière simple de calculer cette translation. Comme nous l'avons vu, les paires de pixels trouvées peuvent être mises en correspondance avec des paires de points 3D par reconstruction depuis les images avec illumination. Pour cela on néglige le mouvement entre les vues avec et sans illumination.

Cette approximation, couplée aux erreurs de mise en correspondance des descripteurs SURF, peut conduire à faire échouer la méthode. Notre méthode doit donc chercher à rendre plus fiable la mise en correspondance et réaliser une estimation robuste de translation sur la base de ces mises en correspondance.

Les descripteurs SURF sont invariants affine mais ne sont que partiellement robustes aux déformations perspectives. Connaissant les matrices de calibrage de chaque caméra ainsi que la rotation entre le repère de la caméra coaxiale à l'instant t et ce même repère à l'instant $t+\tau$, on peut réaliser une rectification perspective des acquisitions à l'un des instants. Les points rectifiés sont donc ramenés à leurs correspondants à l'autre instant à une transformation affine près.

À présent, pour simplifier le propos, nous considérerons le traitement uniquement sur l'image « coaxiale » à chaque instant d'acquisition. Réaliser l'opération d'appariement sur les

deux images ne permet que de générer plus de paires en fin de processus. De plus on considère que l'on a corrigé la distorsion des images afin de se ramener à un problème de projection linéaire.

On note $\mathbf{K}$ la matrice de calibrage de la caméra coaxiale. On utilise l'expression de la matrice $\mathbf{R}$ de l'équation (3.8) donnant la rotation effectuée entre les repères aux temps t et $t + \tau$. Pour rectifier l'image au temps $t + \tau$ on applique une transformation aux pixels de l'image :

$$\begin{bmatrix} sx' \\ sy' \\ s \end{bmatrix} = \mathbf{K}\mathbf{R}\mathbf{K}^{-1} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \quad (3.9)$$

La figure (3.5a) illustre la mise en correspondance de points après cette rectification. Il reste maintenant à filtrer les mises en correspondance afin d'éliminer les paires incohérentes.

Ce filtrage est réalisé en deux temps. Le premier exploite directement la rectification perspective qui nous permet d'aligner les points correspondants selon une même direction [Hartley 04]. Au moyen d'un tri par histogramme (Figure 3.6), on peut éliminer les paires dont la direction de la droite les reliant est trop éloignée de la direction principale des appariements. La figure (3.5b) présente un ensemble de paires après filtrage.

Par reconstruction depuis les images avec illumination des paires de pixels restantes, on dispose d'un ensemble de paires de points 3D pour l'estimation de la translation. On utilise pour cela un RANSAC rapide afin de réaliser conjointement l'estimation et la seconde phase de filtrage de paires incohérentes.

3.2 Recalage précis

L'approche initiale permet d'exprimer les composantes d'un recalage selon le même repère. La recherche de précision demande d'affiner l'agencement de ces vues dans le modèle. On favorise souvent l'optimisation par approche locale pour ce type de problème.

Dans la suite de ce chapitre nous allons étudier les principales variantes de ce recalage que nous avons testées ainsi que celle que nous avons mis au point et adapté à notre système.

3.2.1 État de l'art

Le recalage précis se fait classiquement au moyen de la méthode ICP (*Iterative Closest Points*), introduite par Besl et McKay [Besl 92] ainsi que par Chen et Medioni [Chen 92]. Cette méthode a depuis connu de nombreuses variantes [Rusinkiewicz 01].

Parmi les variantes importantes, le calcul de la transformation a été largement étudié. Alors que la méthode initiale considère la minimisation de la somme des distances entre points correspondants au moyen d'une approche par moindres carrés linéaires, d'autres auteurs ont proposé l'utilisation du formalisme des quaternions et des quaternions doubles [Lorusso 95, Restrepo Specht 05].

Le filtrage des paires a également été le sujet de nombreux travaux [Dalley 02]. Parmi les premiers, et plus notables, Zhang [Zhang 94] propose un filtrage statistiques des appariements. D'autres approches exploitent une mesure de similarité globale ou locale basée sur des descripteurs [Keller 03, Sun 03] pour valider la mise en correspondance.

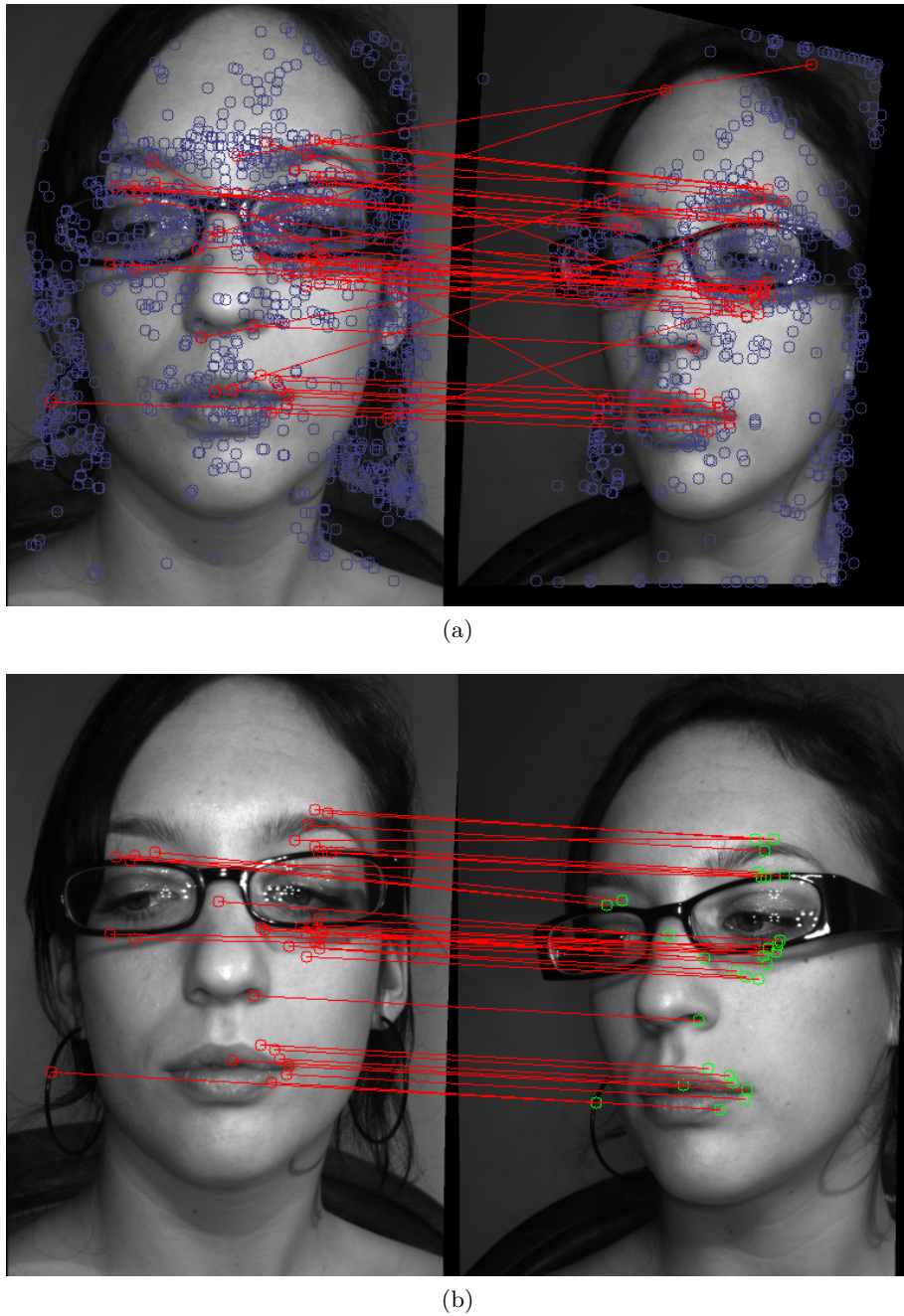


FIGURE 3.5 – Mise en correspondance de points d'intérêt entre les images coaxiales aux temps t et $t + \tau$ après rectification perspective de l'image au temps $t + \tau$. (a) Avant filtrage. (b) Après filtrage.

Des optimisations de performances ont été proposées, particulièrement en exploitant le calcul parallèle et distribué [Langis 01].

L'extension du recalage de Chen ou Besl au cas multi-vues a été proposé par Blais [Blais 95]. Ce type d'approche a connu un vaste développement dans la communauté vision.

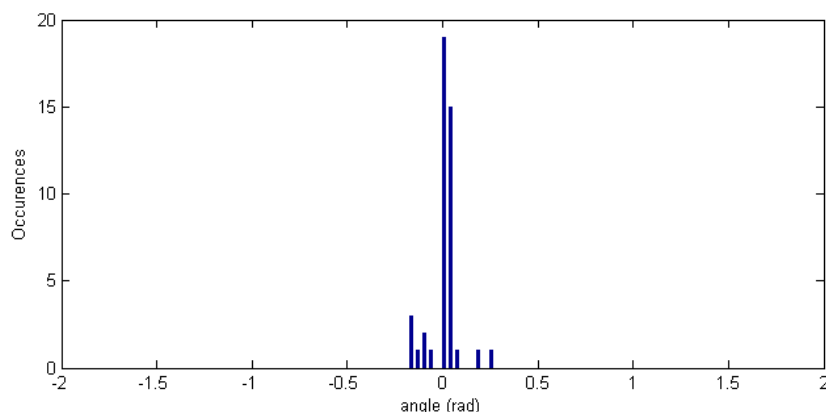


FIGURE 3.6 – La distribution des directions d’appariements permet d’éliminer les paires les plus incohérentes

On peut citer en particulier les travaux de Eggert ainsi que ceux de Robertson [Eggert 96, Robertson 02] qui ont contribué au projet de capteur 3D tenu à la main de l’université d’Édimbourg.

La majeure partie des contributions concerne la phase de mise en correspondance. La méthode ICP, telle que présentée par Besl et McKay, apparie entre eux les points d’un nuage de points 3D avec leur plus proche voisin dans l’autre nuage. On parle de mise en correspondance *point à point*. La méthode de Chen et Medioni introduit un autre mode de correspondance puisque les appariements sont réalisés selon les plans tangents à la surface. On parle alors de mise en correspondance *point à plan* ou plus globalement *point à surface*. Cette dernière approche est celle qui a été la plus largement étudiée du fait des bons résultats qu’elle propose [Rusinkiewicz 01]. Park [Park 03] propose une méthode combinant la projection de points sur la surface de destination et une mise en correspondance par plan. D’autres approches utilisent des approximations plus précises de la surface au moyen de NURBS [Ristic 97] ou de B-Splines [Fan 01]. Parmi les travaux récents, on peut évoquer ceux de Xie [Xie 10] combinant une mise en correspondance par surfaces duales avec un filtrage par courbure.

Les approches dites par *point à projection* [Blais 95, Neugebauer 97] dont s’est inspiré Park pour ses travaux permettent une mise en correspondance rapide mais sont généralement délaissées au profit des méthodes par surfaces dont il a été démontré qu’elles permettent d’atteindre une meilleure précision [Rusinkiewicz 01].

On pourra également citer les travaux de Morency [Morency 03] pour lesquels une paire d’images stéréo est utilisée afin d’effectuer un recalage combinant le tracking de points et le calibrage inverse vers les images de profondeur. D’autres travaux basés sur le traitement d’image exploitent l’information de luminance ou de couleur pour la mise en correspondance [Weik 97, Kim 04].

3.2.2 Iterative Closest Points

L’algorithme ICP cherche à minimiser un critère de distance point à point pour recalculer deux modèles 3D, notés V_k et V_{k+l} , préalablement rapprochés.

Il s'agit d'un algorithme itératif utilisant un ensemble de points $\{\mathbf{p}_i^k\}$ sélectionnés dans la vue V_k . A chaque itération on va chercher à :

- appairer les points $\mathbf{p}_i^k$ avec leur plus proche voisin dans la vue V_{k+l} . On utilise pour cela une structure k - d tree,
- pondérer et rejeter des appariements,
- estimer la transformation rigide par minimisation d'un critère SSD sur les points appariés.

On applique alors la transformation à la vue V_{k+l} pour la rapprocher de V_k . On itère à nouveau jusqu'à satisfaction d'un critère de convergence de distance entre les vues ou en fixant une limite sur le nombre d'itérations.

Filtrage des paires

La sélection des appariements se fait au moyen de deux filtres. Le premier assure l'unicité d'appariement pour chaque point [Restrepo Specht 05]. Un point de V_{k+l} ne peut être apparié qu'à un seul point de V_k . La phase d'appariement produit une liste de candidats dans V_{k+l} pour chaque point $\mathbf{p}_i^k$ de V_k . Ces listes sont ensuite parcourues afin de garantir à la fois l'unicité des appariements et le choix des paires optimales selon la distance 3D.

Le second filtre est le critère statistique basé sur la distribution des distances entre les points appariés de Zhang [Zhang 94]. On met à jour itérativement un seuil D_{max}^{k-1} et on supprime les appariements présentant une distance supérieure à ce seuil. La valeur initiale D_{max}^0 est fixée à une valeur élevée. On modifie le seuil en fonction de la moyenne μ des distances et de l'écart-type σ :

- si $\mu < D_{max}^{k-1}$ alors $D_{max}^k = \mu + 3\sigma$,
- sinon si $\mu < 3D_{max}^{k-1}$ alors $D_{max}^k = \mu + 2\sigma$,
- sinon si $\mu < 6D_{max}^{k-1}$ alors $D_{max}^k = \mu + \sigma$,
- sinon $D_{max}^k = \xi$.

ξ est la valeur médiane sur toutes les distances.

Estimation de la transformation

Nous disposons donc d'un ensemble de N paires de points $(\mathbf{p}_i^k, \mathbf{p}_i^{k+l})$. Les appariements n'étant pas parfaits, estimer la transformation rigide entre ces deux ensembles revient à minimiser une mesure de similarité :

$$\epsilon = \sum_{i=1}^N \|\mathbf{p}_i^k - (\mathbf{R}\mathbf{p}_i^{k+l} + \mathbf{t})\|^2 \quad (3.10)$$

Arun *et al.* [Arun 87] proposent une méthode de minimisation au sens des moindres carrés basée sur la décomposition SVD d'une matrice représentant l'ensemble des appariements.

La méthode propose de résoudre le problème en deux temps, en séparant la rotation et la translation. On peut se rendre indépendant à la translation en exprimant chaque point par rapport au centre de masse du nuage dont il est issu.

$$\begin{aligned} \mathbf{p}_m^k &= \frac{1}{N} \sum_{i=1}^N \mathbf{p}_i^k & \mathbf{p}_m^{k+l} &= \frac{1}{N} \sum_{i=1}^N \mathbf{p}_i^{k+l} \\ \mathbf{q}_i^k &= \mathbf{p}_i^k - \mathbf{p}_m^k & \mathbf{q}_i^{k+l} &= \mathbf{p}_i^{k+l} - \mathbf{p}_m^{k+l} \end{aligned}$$

Ainsi on peut exprimer l'équation (3.10) en fonction de $\mathbf{R}$ uniquement :

$$\epsilon = \sum_{i=1}^N \left[\mathbf{q}_i^k + \mathbf{p}_m^k - \mathbf{R}\mathbf{q}_i^{k+l} - \mathbf{R}\mathbf{p}_m^{k+l} - \mathbf{t} \right]^2$$

Par construction, $\mathbf{p}_m^k - (\mathbf{R}\mathbf{p}_m^{k+l} + \mathbf{t}) = 0$, donc on peut exprimer la quantité à minimiser par :

$$\epsilon = \sum_{i=1}^N \left[\mathbf{q}_i^k - \mathbf{R}\mathbf{q}_i^{k+l} \right]^2 \quad (3.11)$$

La méthode permet de déterminer $\mathbf{R}$ par minimisation du critère (3.11). La translation $\mathbf{t}$ est ensuite obtenue par

$$\mathbf{t} = \mathbf{p}_m^k - \mathbf{R}\mathbf{p}_m^{k+l}$$

On développe l'équation (3.11) :

$$\begin{aligned} \epsilon &= \sum_{i=1}^N \left(\mathbf{q}_i^k - \mathbf{R}\mathbf{q}_i^{k+l} \right)^T \left(\mathbf{q}_i^k - \mathbf{R}\mathbf{q}_i^{k+l} \right) \\ &= \sum_{i=1}^N \left(\mathbf{q}_i^{kT} \mathbf{q}_i^k + \mathbf{q}_i^{k+lT} \mathbf{R}^T \mathbf{R} \mathbf{q}_i^{k+l} - \mathbf{q}_i^{kT} \mathbf{R} \mathbf{q}_i^{k+l} - \mathbf{q}_i^{k+lT} \mathbf{R}^T \mathbf{q}_i^k \right) \\ &= \sum_{i=1}^N \left(\mathbf{q}_i^{kT} \mathbf{q}_i^k + \mathbf{q}_i^{k+lT} \mathbf{q}_i^{k+l} - 2\mathbf{q}_i^{kT} \mathbf{R} \mathbf{q}_i^{k+l} \right) \end{aligned}$$

Par conséquent, minimiser ϵ est équivalent à maximiser

$$\begin{aligned} F &= \sum_{i=1}^N \mathbf{q}_i^{kT} \mathbf{R} \mathbf{q}_i^{k+l} \\ &= \text{Trace} \left(\sum_{i=1}^N \mathbf{R} \mathbf{q}_i^{k+l} \mathbf{q}_i^{kT} \right) = \text{Trace}(\mathbf{R}\mathbf{H}) \end{aligned} \quad (3.12)$$

avec

$$\mathbf{H} = \sum_{i=1}^N \mathbf{q}_i^{k+l} \mathbf{q}_i^{kT}$$

Soit la décomposition SVD de $\mathbf{H}$:

$$\mathbf{H} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T$$

Dans cette égalité, les matrices $\mathbf{U}$ et $\mathbf{V}$ sont orthonormales et $\mathbf{\Lambda}$ est une matrice diagonale positive contenant les valeurs propres (ou singulières). Soit

$$\mathbf{X} = \mathbf{V}\mathbf{U}^T \rightarrow \mathbf{X}\mathbf{H} = \mathbf{V}\mathbf{U}^T \mathbf{U}\mathbf{\Lambda}\mathbf{V}^T = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$$

$\mathbf{XH}$ est symétrique positif. Or, pour toute matrice $\mathbf{B}$ orthonormale 3×3 :

$$\text{Trace}(\mathbf{XH}) \geq \text{Trace}(\mathbf{BXH})$$

Dans ces conditions, $\mathbf{X}$ maximise la fonction (3.12). Si $\det(\mathbf{X}) = +1$, $\mathbf{X} = \mathbf{R}$. Si $\det(\mathbf{X}) = -1$, $\mathbf{X}$ est une réflexion et deux cas se présentent :

- une des valeurs singulière de $\mathbf{H}$ est nulle. Dans ce cas, la rotation recherchée est :

$$\mathbf{X}' = \mathbf{V}'\mathbf{U}^T$$

où $\mathbf{V}'$ est obtenu en changeant le signe de la troisième colonne de $\mathbf{V}$.

- aucune des valeurs singulières de H n'est nulle. Dans ce cas, la méthode proposée ne convient pas, on préférera alors s'orienter vers des solutions de type RANSAC, par exemple.

3.2.3 Point-to-plane

La méthode proposée par Chen [Chen 92] s'apparente à une variante de mise en correspondance pour ICP. On cherchera ici à mettre en correspondance un ensemble de points de la première surface avec un ensemble de plans tangents à la seconde.

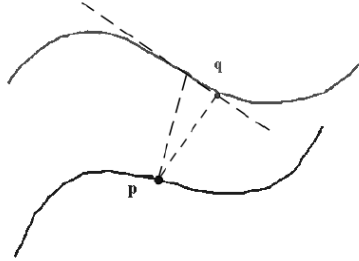


FIGURE 3.7 – Point-to-plane

Pour chaque point sélectionné dans la première surface, on détermine l'intersection de sa normale avec la seconde surface. La distance de mise en correspondance minimisée est la distance entre le plan tangent à la seconde surface en ce point et le point sélectionné dans la première surface.

Il a été mis en avant [Rusinkiewicz 01] qu'une telle approche permet une convergence plus facile et plus rapide de l'algorithme ICP. De plus la recherche à partir de projection de normale depuis la première surface permet un procédé de mise en correspondance plus robuste lorsque la première surface n'est pas intégralement recouvrante avec la seconde. En effet, un point hors d'une zone de recouvrement ne verra pas sa normale intersecter la seconde surface.

Le principal inconvénient de cette méthode est la dépendance au calcul des normales. A moins de disposer de ces informations depuis le capteur, leur estimation utilise le voisinage spatial, dont la recherche est coûteuse et peut mener à un résultat relativement peu précis. D'autre part, il n'existe pas, avec cette approche, de solution analytique pour le calcul de la transformation rigide. Celle-ci doit être déterminée à l'aide d'une méthode d'optimisation non-linéaire ou par linéarisation du problème.

La recherche d'intersection de la normale avec la seconde surface peut s'avérer être un processus complexe également. Park *et al.* [Park 03] ont proposé une méthode rapide basée sur une approche de calibrage inverse itératif.

3.2.4 Point-to-surface

La méthode de Chen présentée plus haut cherche à affiner l'espace de mise en correspondance en interpolant la surface de destination par un ensemble de plans tangents. Ce type d'approche peut être étendu à des interpolations d'ordres plus élevés.

Une approche commune consiste à interpoler localement la surface par B-splines uniformes bicubiques. On considère un ensemble de points de contrôle définissant une matrice de dimension 4×4 .

$$\mathbf{P} = \begin{bmatrix} P_{00} & P_{01} & P_{02} & P_{03} \\ P_{10} & P_{11} & P_{12} & P_{13} \\ P_{20} & P_{21} & P_{22} & P_{23} \\ P_{30} & P_{31} & P_{32} & P_{33} \end{bmatrix}$$

L'élément de surface local $\mathbf{S}(u, v)$ ($0 \leq u, v \leq 1$) est calculé à partir des seize points de contrôle.

$$\mathbf{S}(u, v) = \begin{bmatrix} 1 & u & u^2 & u^3 \end{bmatrix} \mathbf{MPM}^T \begin{bmatrix} 1 \\ v \\ v^2 \\ v^3 \end{bmatrix}$$

où $\mathbf{M}$ est la matrice des coefficients.

$$\mathbf{M} = \frac{1}{6} \begin{bmatrix} 1 & 4 & 1 & 0 \\ -3 & 0 & 3 & 0 \\ 3 & -6 & 3 & 0 \\ -1 & 3 & -3 & 1 \end{bmatrix}$$

Cette méthode, pour permettre un apport sur la convergence de l'algorithme ICP nécessite cependant une densité importante de points pour caractériser les surfaces.

3.2.5 Interpolation dans les images, approche pyramidale

Les méthodes précédentes sont génériques et ne permettent pas la prise en compte des données brutes issues du capteur. Pourtant, nous l'avons vu, l'exploitation des images permet l'accès à une information très riche. Ainsi, l'exploitation de la nature « vision » du numériseur permet d'envisager des méthodes basées sur la projection de points dans les images.

Les points 3D sont générés à chaque instant d'acquisition en utilisant la stéréo-corrélation dans une paire d'image avec projection de mouchetis (cf. Section 2.3). On dispose donc de deux paires d'images stéréoscopiques (C_t, L_t) , pour les caméras à l'instant t , et $(C_{t+\tau}, L_{t+\tau})$, pour les caméras à l'instant $t + \tau$. Les images C_t et $C_{t+\tau}$ sont issues de la caméra coaxiale, les images L_t et $L_{t+\tau}$ de la caméra latérale. On connaît une transformation approximative (R_0, t_0) de la transformation (R, t) permettant de passer du repère caméra à l'instant $t + \tau$ au repère caméra à l'instant t . On cherche à affiner notre estimée $(\hat{R}, \hat{t})$ de cette transformation. Au démarrage de l'algorithme, on a donc $(\hat{R}, \hat{t}) = (R_0, t_0)$.

Recherche de correspondant par calibrage inverse

A partir des images à l'instant $t + \tau$, on construit un ensemble de points 3D $\{p_i^{t+\tau}\}$. On approxime l'expression de ce set de point dans le repère du capteur à l'instant t à l'aide de la transformation estimée courante $(\hat{R}, \hat{t})$.

La recherche d'un correspondant $\mathbf{q}$, construit depuis les images (C_t, L_t) , pour tout point $\mathbf{p}$, construit depuis la paire $(C_{t+\tau}, L_{t+\tau})$ et exprimé dans le repère du capteur à l'instant t , peut être faite rapidement en utilisant l'approche du *calibrage inverse* [Blais 95] (figure 3.8).

En suivant cette approche on projette le point $\mathbf{p}$ dans l'image L_t .

$$\mathbf{p}_C^t = \mathbf{K}_0 \mathbf{p}^t \quad (3.13)$$

Dans l'équation (3.13), $\mathbf{K}_0$ est la matrice des paramètres intrinsèques de la caméra coaxiale, $\mathbf{p}_C^t$ est la projection dans l'image C_t d'un point $\mathbf{p}^t$, exprimé dans le repère du capteur à l'instant n . On peut exprimer cette relation en fonction du point original, exprimé dans le repère du capteur à l'instant $t + \tau$.

$$\mathbf{p}_C^t = \mathbf{K}_0 [\hat{\mathbf{R}} | \hat{\mathbf{t}}] \begin{bmatrix} \mathbf{p}^{t+\tau} \\ 1 \end{bmatrix} \quad (3.14)$$

Dans l'équation (3.14), le point $\mathbf{p}^{t+\tau}$ est exprimé dans le repère capteur à l'instant $t + \tau$ et la matrice $[\hat{\mathbf{R}} | \hat{\mathbf{t}}]$ est l'estimation courante de la transformation permettant le passage du repère capteur à l'instant t au repère capteur à l'instant $t + \tau$.

Par mise en correspondance stéréoscopique (section 2.2) du point image $\mathbf{p}_L^t$ puis triangulation (section 2.3.2), on détermine le point $\mathbf{q}^t$.

Métrie et estimation de la transformation

Avec cette méthode on parvient à réaliser rapidement la mise en correspondance de points entre les deux vues à recalculer. De plus, le travail en résolution sous pixelique dans les images permet d'atteindre une interpolation de surface de grande précision. Cependant cette mise en correspondance est valide dans une bande de convergence assez étroite. Il est nécessaire de prendre en compte des contraintes supplémentaires au moyen d'une fonction de pondération dans le critère à minimiser.

$$\epsilon = \sum_{i=1}^N \phi(i) \|\mathbf{q}_i^t - (\hat{\mathbf{R}} \mathbf{p}_i^{t+\tau} + \hat{\mathbf{t}})\|^2 \quad (3.15)$$

La fonction $\phi(i)$ de l'équation (3.15) est la fonction de pondération appliquée à chaque paire de points.

La pondération choisie dans notre méthode s'inspire de la méthode de recherche par intersection de normales de Chen [Chen 92]. Pour chaque point $\mathbf{q}_i^t$, on lui associe une distance à la normale ponctuelle au point $\mathbf{p}_i^t$ (figure 3.8).

$$\phi(i) = 1 - \frac{d_i}{d_{max}}$$

La mesure d_i est définie comme la distance euclidienne entre un point $\mathbf{q}_i^t$ et la normale $\vec{n}_i$ au point correspondant $\mathbf{p}_i^t$, d_{max} est la distance maximale dans l'ensemble des paires $(\mathbf{p}_i^t, \mathbf{q}_i^t)$.

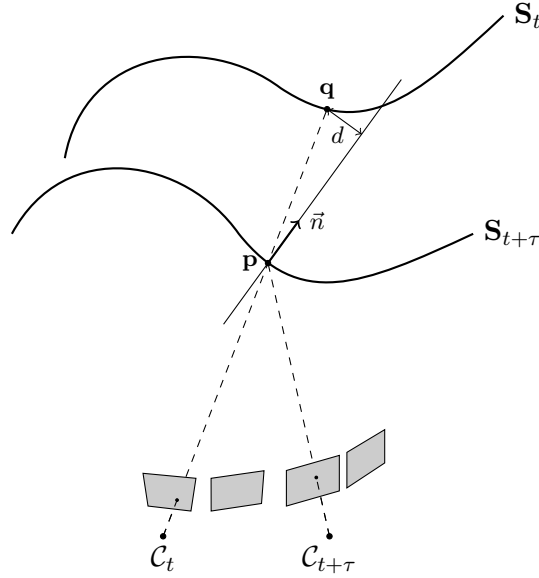


FIGURE 3.8 – Mise en correspondance par calibrage inverse. La surface S_t peut être reconstruite depuis la paire à l’instant t et la surface $S_{t+\tau}$ peut être reconstruite depuis la paire à l’instant $t + \tau$. Le correspondant q du point p est la projection de ce dernier sur la surface S_t selon le point de vue du capteur à l’instant t . La fonction de pondération appliquée à cette paire est liée à la distance du point q à la normale ponctuelle au point p .

Approche pyramidale

Pour accélérer la convergence et se rendre moins dépendant de la qualité de l’estimée initiale, nous avons proposé une approche pyramidale pour cette méthode. Après avoir réalisé la projection d’un point p^t dans l’image C_t , on ne reconstruit pas uniquement le point reconstruit depuis le point image p_C^t et son stéréo correspondant, mais également un échantillonnage du voisinage de ce point dans l’image C_t (figure 3.9). Enfin, le correspondant est choisi dans l’ensemble des voisins en fonction de la distance point à point entre lui et le point du point p^t que l’on cherche à apparier.

Cette méthode introduit deux paramètres : la taille du voisinage et la résolution de l’échantillonnage dans ce voisinage. Afin de permettre une convergence plus rapide lorsque l’estimée initiale est lointaine, on pourra choisir un grand voisinage. Il n’est pas nécessaire de choisir à ce moment là une résolution très fine, ce qui causerait la reconstruction d’un nombre important de points et donc un temps plus important.

Au contraire, lorsque l’on se trouve proche de la solution, en fin d’algorithme par exemple, on cherchera à atteindre la meilleure précision possible. Pour cela on va souhaiter privilégier une résolution très fine pour la reconstruction. Dans ce cas, la taille du voisinage peut être réduite, ce qui permet à nouveau de ne pas reconstruire un nombre trop important de points.

Au cours de l’algorithme on va donc chercher à modifier ces paramètres au fur et à mesure que l’on atteint une convergence de notre recalage. Cette approche est dite *pyramidale* car il introduit une stratégie visant à partir d’une fenêtre de voisinage très large et éparse pour la réduire et augmenter sa résolution progressivement.

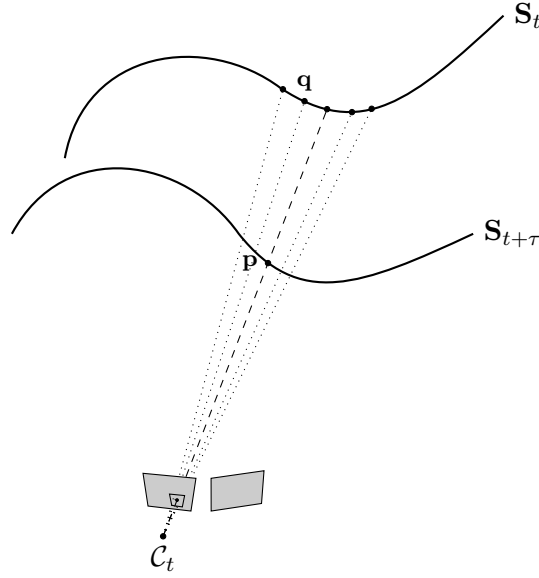


FIGURE 3.9 – Lors de la recherche de correspondant on reconstruit une partie du voisinage du point image déterminé. Le point 3D que l'on apparie finalement est déterminé selon une distance point à point.

3.3 Évaluations

Pour l'évaluation des méthodes de recalage, nous avons réalisé des tests sur deux séquences : en données simulées et en données réelles. La première scène est un modèle 3D de l'Université de Stanford [3Dscanrep 10] appelé « Bunny ». Le second est une pièce d'étalonnage numérisée par notre capteur. Cette scène est appelée « Étalon ».

Pour la génération des données de synthèse, nous avons utilisé un programme réalisé par nos soins. À partir du modèle STL d'un objet, nous avons simulé la génération d'images par lancer de rayon. Le modèle de caméra utilisé est conforme à celui présenté en section 2.1 et correspond à la configuration – paramètres et placement des caméras – de notre capteur, à un facteur d'échelle près. Les caméras ne provoquent pas de distorsion. Les nuages sont générés depuis le point de vue de la caméra simulant la caméra coaxiale. Un point 3D est généré pour chaque pixel visible dans les deux caméras sans filtrer en fonction de la direction de normale. On obtient donc des points exacts dans les zones peu front-parallèles à la caméra principale, ce qui constitue un cas très favorable pour les algorithmes en comparaison avec un cas réel, où ces points ne peuvent être reconstruits avec précision.

3.3.1 Pièce « Bunny »

L'évaluation en simulation sur la scène du « Bunny » de Stanford a été faite en comparant la mise en œuvre de trois méthodes de recalage précis : l'algorithme ICP original, la méthode de Chen, et notre méthode. Nous avons également comparé les résultats obtenus sur une variante de notre algorithme ne réalisant que l'appariement par calibrage inverse. Nous souhaitons ainsi montrer l'influence de l'étape de pondération et de l'approche pyramidale.

Pour cette évaluation, les deux vues à recaler sont un même ensemble de points auxquels sont appliqués deux poses différentes. Les deux nuages sont donc parfaitement recouvrants

et les correspondants exacts de chaque point existent. L'erreur de correspondance peut alors devenir nulle et un recalage parfait existe.

On applique au premier nuage une transformation identité et au second une transformation faible détaillée en équation (3.16). La configuration des nuages de points aux poses initiales est montrée en figure 3.12a.

$$[\mathbf{R}|\mathbf{t}] = \left[\begin{array}{ccc|c} 0.9999 & -0.0058 & -0.0096 & 0.0100 \\ 0.0050 & 0.9968 & -0.0800 & -0.3000 \\ 0.0100 & 0.0799 & 0.9968 & -0.0002 \end{array} \right] \quad (3.16)$$

Pour notre méthode, nous fixons le voisinage initial à une taille de 15 pixels centrés sur le point projetés, et une résolution de 2 pixels. La limite de convergence est fixée à 10^{-4} mm pour la dérivée de l'erreur. Lors d'une détection de convergence, le voisinage et la résolution sont divisés par deux, et la limite de convergence par dix.

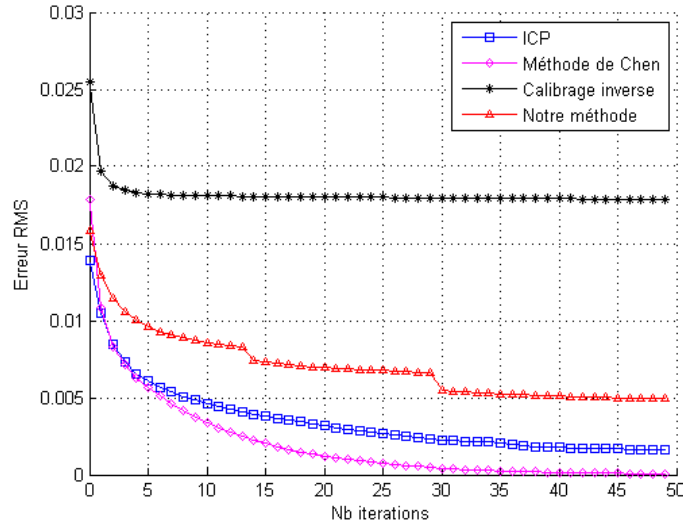


FIGURE 3.10 – Erreur RMS de recalage pour la pièce « Bunny »

L'évolution de l'erreur de mise en correspondance est montrée en figure 3.10. Les méthodes de Chen et ICP, convergent rapidement vers des valeurs faibles. La méthode de Chen atteint même une erreur de mise en correspondance nulle. En effet, ces méthodes mettent directement en correspondance les points des nuages, elles bénéficient donc du cas favorable de l'« auto-recalage » réalisé ici. Notre méthode, au contraire reconstruit par corrélation les correspondants. La corrélation, comme nous l'avons vu en section 2.2 réalise une approximation de la transformation perspective engendrée par la pose des caméras et l'orientation de la surface. Par conséquent les points mis en correspondance ne sont pas parfaits comme le sont ceux générés directement par le simulateur. On note donc que notre méthode ajoute une erreur systématique liée à l'erreur de reconstruction. Malgré tout, l'algorithme converge rapidement vers une valeur faible. On observe les apports de l'approche pyramidale, provoquant des sauts dans l'évolution de l'erreur. La méthode par calibrage inverse ne permet pas d'atteindre le niveau de précision des trois autres méthodes et trouve assez rapidement une limite.

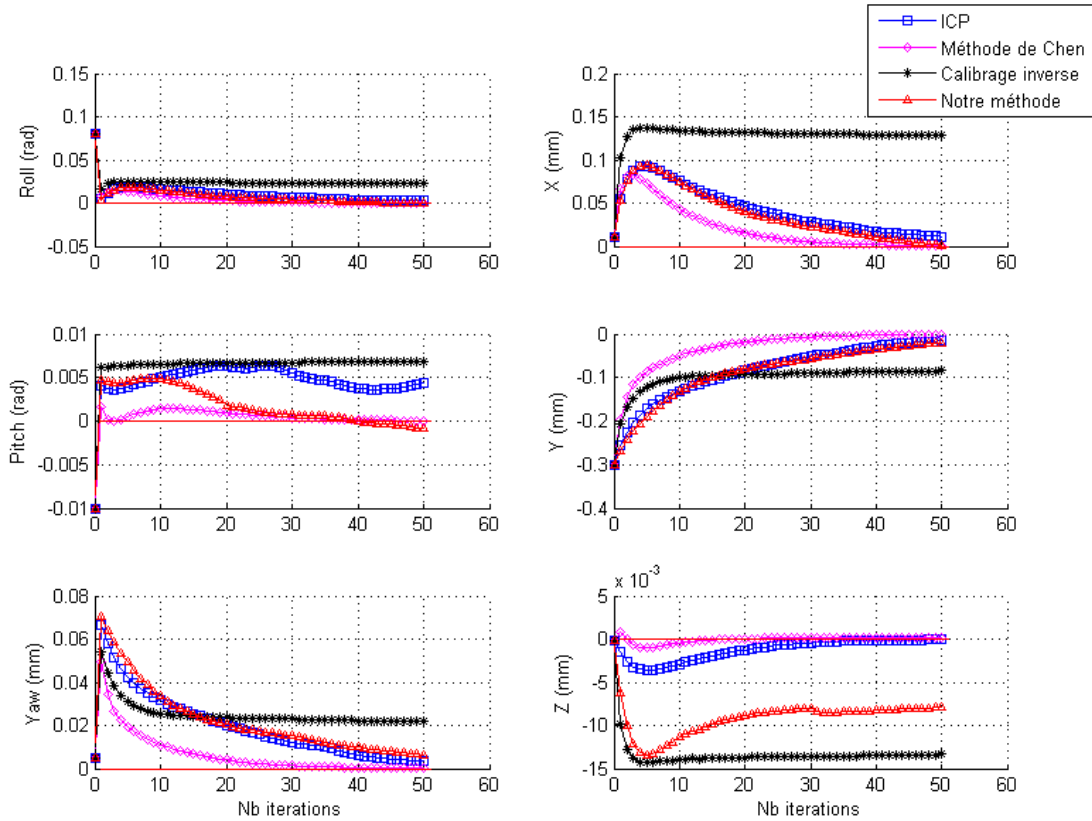


FIGURE 3.11 – Évolution de la pose estimée. La ligne rouge représente la vérité terrain.

L'évolution de la pose pour chaque méthode est proposée en figure 3.11, la vérité terrain est identifiée dans chaque graphique par une ligne rouge. On peut ainsi observer que notre méthode provoque essentiellement un décalage sur l'axe Z . L'algorithme ICP conserve une faible erreur de rotation sur un des axes.

Les nuages dans leurs configurations finales apparaissent en figure 3.12. La méthode de Chen ayant convergé vers la pose exacte, les nuages sont confondus, nous n'en voyons donc apparaître qu'un seul, le dernier dessiné à l'écran par l'outil de visualisation utilisé. L'algorithme ICP a montré une erreur légère de rotation sur un axe (Figure 3.12) que l'on voit apparaître ici. Pour notre méthode, le décalage a lieu selon l'axe de visualisation, le nuage à recaler se trouve donc légèrement derrière le nuage de destination.

Nous avons vu que notre méthode permet d'atteindre une précision de recalage bornée par la qualité de la reconstruction. L'erreur de reconstruction est aussi une limite pour toutes les autres méthodes, basées sur la géométrie uniquement, puisqu'elles sont dépendantes de la qualité des points 3D générés. Ce test propose donc des conditions particulièrement défavorables pour notre méthode, qui ne bénéficie pas de points 3D parfaits comme les autres méthodes.

Le test suivant, en données réelles va lever cette différence car les points de tous les nuages auront été générés avec la même méthode, la mise en correspondance par corrélation présentée

en section 2.2.

3.3.2 Pièce « Étalon »

Pour une évaluation sur des données réelles issues de notre capteur, nous avons réalisé le recalage de deux vues d'une pièce d'étalonnage. La figure 3.13 présente les images avec illumination utilisée pour la génération des points 3D.

L'étape de reconstruction dense sur ces vues permet la création de nuages d'environ 120 000 points chacun. Nous avons initialisé nos algorithmes de recalage précis en utilisant la méthode hybride exploitant les données inertielles et l'appariement dans les images. Le résultat de cette étape est donné en figure 3.15a.

Après recalage initial les vues sont partiellement recouvrantes : les deux plans principaux sont totalement recouvrants, les percages ne le sont que partiellement et les faces latérales sont totalement indépendantes.

Nous avons mis en œuvre trois méthodes : l'algorithme ICP original, la méthode de Chen, et notre méthode. Afin de montrer l'apport de notre pondération et de l'approche pyramidale, nous avons également réalisé les tests sur notre méthode sans ces deux étapes. Dans les figures qui suivent, cette variante est identifiée par le nom de « Calibrage inverse ».

La stratégie adoptée pour notre méthode dans le premier test fixe le voisinage à une taille de 15 pixels centrés sur le point projetés, et une résolution de 2 pixels. La limite de convergence est fixée à 10^{-4} mm pour la dérivée de l'erreur. Lors d'une détection de convergence, le voisinage et la résolution sont divisés par deux, et la limite de convergence par dix. Il s'agit de la même stratégie que pour la pièce de simulation.

La figure 3.14 montre l'évolution de l'erreur de mise en correspondance sur 50 itérations pour les quatre méthodes. On peut observer ici que la méthode de Chen propose une convergence rapide, comme cela a été démontré dans [Rusinkiewicz 01] et meilleure comparée à l'algorithme ICP original. On note, de plus, que notre méthode, dans sa première stratégie réalise une convergence rapide également. Le premier changement de stratégie, après la quatorzième itération, permet d'atteindre une erreur semblable à celle de la méthode de Chen. Le second changement de stratégie, après la quarantième itération, donne une résolution d'un demi pixel, ce qui permet d'augmenter la précision de mise en correspondance et donc de réduire l'erreur. Sans l'approche pyramidale et la fonction de pondération, on voit que le calibrage inverse ne permet pas d'atteindre une erreur équivalente à celle des autres méthodes.

Les nuages recalés par chaque méthode sont montrés en figure 3.15 avec l'estimée initiale. Comme nous l'avons vu plus haut, la méthode par calibrage inverse ne permet d'atteindre qu'un recalage peu précis. Les méthodes ICP et Chen permettent d'obtenir un recalage de bonne qualité mais on note un biais. En effet, ces méthodes vont tendre principalement à rapprocher les zones où les points sont les plus nombreux, ou les plus uniformes. Ici, elles permettent un rapprochement essentiellement selon l'un des plans dominants. Notre méthode permet d'atteindre un recalage plus homogène du fait de la contrainte qu'impose notre pondération mais aussi car notre méthode peut nécessiter moins de points pour fonctionner (Table 3.1) et se rend donc moins sensible aux zones dominantes.

Pour une étude qualitative des résultats, nous avons comparé les modèles recalés au modèle CAO de la pièce. Ce modèle CAO a été réalisé après production en remesurant la pièce en atelier. Ce dernier est incomplet, l'opérateur ayant eu des difficultés à produire une mesure fiable du coin creusé. Par conséquent nous avons supprimé manuellement les points de nos modèles

	A	B	C	D
Nb points	119 906	119 906	119 906	21 378
Moyenne (mm)	−0.017	−0.027	0.041	0.024
Écart-type (mm)	0.277	0.213	0.262	0.208
Erreur max (mm)	1.922	1.315	1.732	1.264
Erreur min (mm)	−1.645	−1.724	−1.549	−1.864

TABLE 3.1 – Evaluation des méthodes de recalage après comparaison au modèle théorique. (A) ICP (B) Méthode de Chen (C) Interpolation image seule (D) Notre méthode.

sur cette zone avant de les comparer au modèle théorique. L’alignement et la comparaison au théorique ont été réalisés à l’aide du logiciel commercial Geomagic[®] Qualify[®].

Ce qu’il faut observer sur ce type de comparaison est l’homogénéité des zones. Un mauvais recalage entraîne un biais dans l’alignement avec le modèle théorique résultant en une hétérogénéité dans la carte de déviation. Le phénomène est très observable pour la méthode du calibrage inverse. Pour les autres méthodes, les résultats sont très comparables. On notera cependant que dans les percages, notre méthode montre une déviation plus faible et plus homogène. Cela semble corroborer l’observation précédente sur la qualité de cette méthode de recalage. La distribution des erreurs pour chaque méthode est proposée en figure 3.17, et la table 3.1 résume les principales mesures d’erreur que nous avons extrait de la comparaison au modèle théorique.

3.4 Conclusions

Dans ce chapitre nous avons abordé les méthodes pour l’alignement de deux acquisitions 3D. Nous avons présenté un état de l’art des méthodes classiques en différenciant les approches initiales, sans estimée initiale de pose, et les méthodes précises, utilisant des informations *a priori* pour chercher un alignement localement optimal.

Différentes approches pour le recalage initial ont été présentées parmi les plus courantes de la communauté. Selon l’application et les données disponibles certaines méthodes peuvent être plus rapidement mises en œuvre ou plus appropriées. Pour nos besoins nous avons mis en place une méthode combinant une centrale inertielle et une mise en correspondance de descripteurs locaux, afin de tirer le meilleur parti des données que propose notre capteur à chaque acquisition. La méthode actuelle extrait les descripteurs des images sans illumination pour réaliser une mise en correspondance avec le point 3D généré au même pixel dans l’image avec illumination. Une extension de cette méthode peut être envisagée mettant en œuvre un descripteur exclusivement géométrique pour les cas où la scène ne présente assez de points d’intérêt dans les images.

Nous avons ensuite réalisé une étude sur le recalage précis de deux vues. Cette étude a été orientée autour de la méthode ICP développée par Besl *et al.* [Besl 92]. Plus particulièrement nous avons abordé les variantes de mise en correspondance. Pour le développement de notre méthode, nous avons cherché à tirer parti de la richesse d’information disponible dans les images. Les méthodes par mise en correspondance de fonctions de luminance entre les instants

d'acquisition n'étaient pas envisageables. En effet, le motif projeté sur la scène se déplaçant avec les caméras, la scène n'est plus consistante entre deux instants d'acquisition. Nous avons donc proposé une méthode hybride utilisant les images pour la mise en correspondance de points 3D, se ramenant ainsi à un schéma itératif géométrique de type ICP.

L'étude menée démontre la capacité de notre méthode à réaliser un recalage de précision, pouvant dépasser les méthodes classiques de recalage purement géométrique. Elle reste cependant très dépendante de la qualité de la reconstruction, et ne permet pas de raffiner les points comme une méthode de type *Bundle Adjustment* le permettrait. Cette méthode pourra donc être étendue en réalisant un raffinement conjoint des poses et des points 3D.

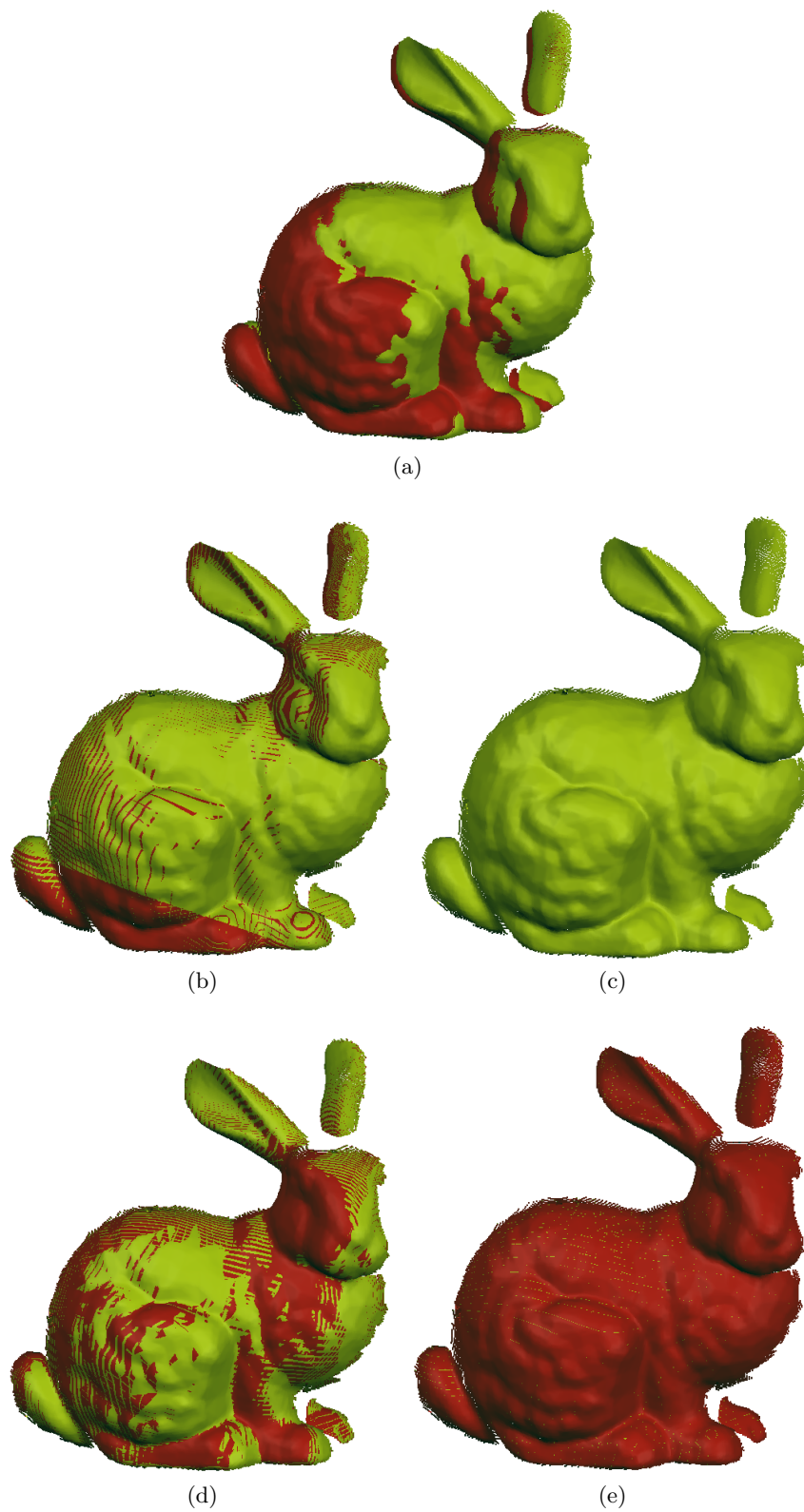


FIGURE 3.12 – Nuages de points de la scène test « Bunny ». (a) Estimée initiale (b) ICP (c) Méthode de Chen (d) Calibrage inverse (e) Notre méthode.

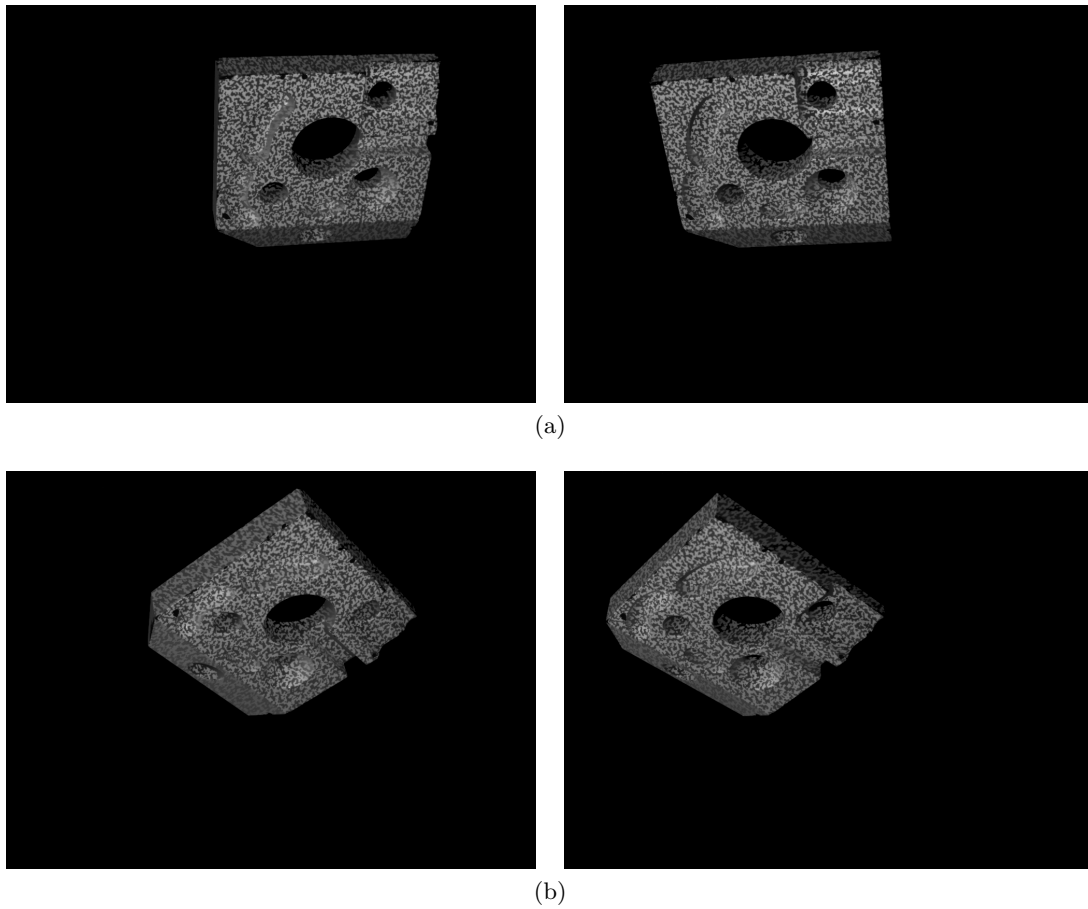


FIGURE 3.13 – Paires d'images sources de la pièce de test « Étalon » (a) Vue n (b) Vue $n + 1$

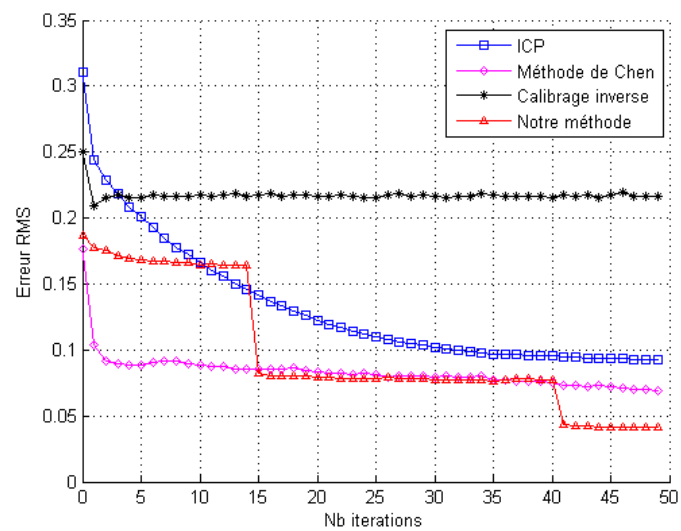


FIGURE 3.14 – Erreur RMS de recalage pour la pièce « Étalon »

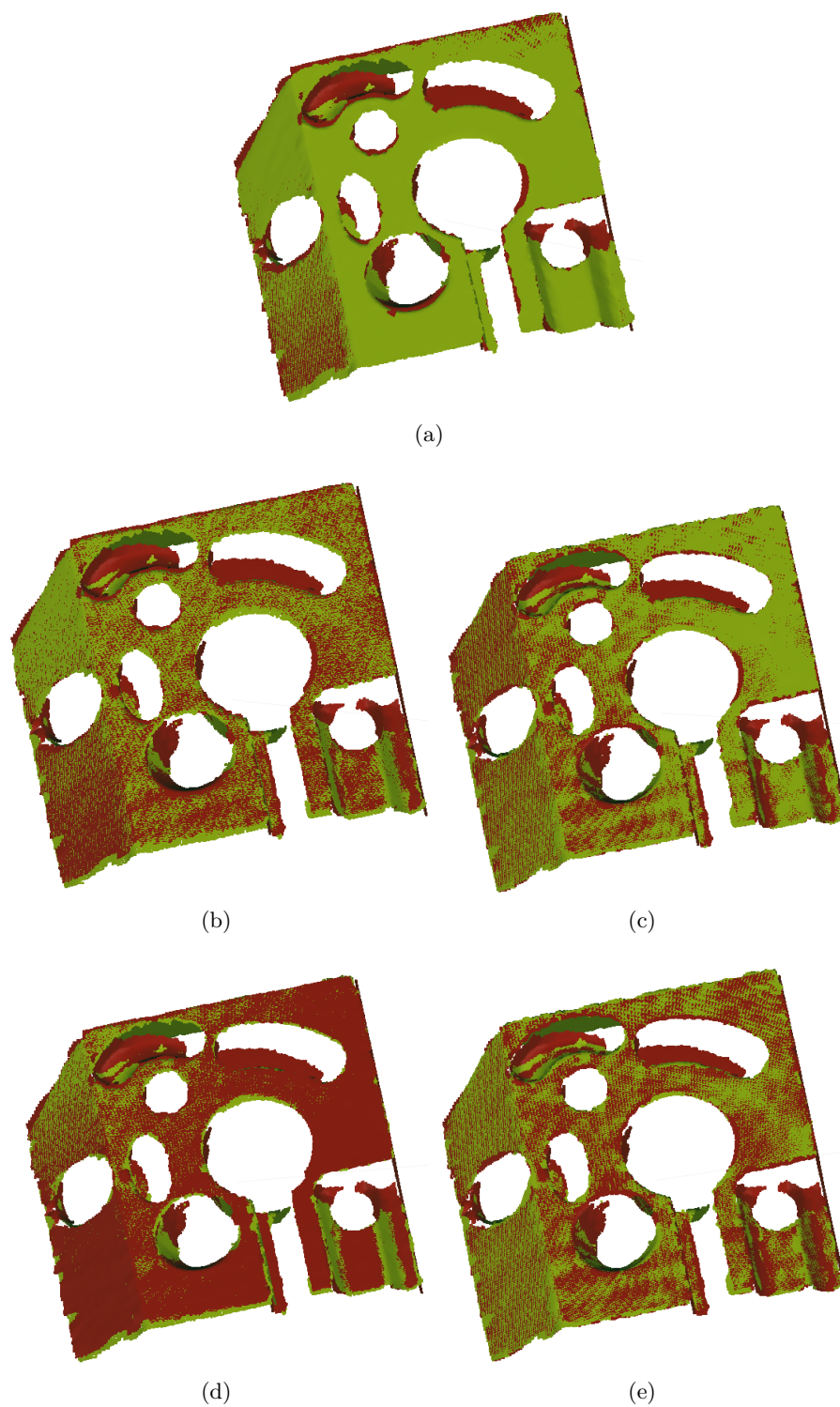


FIGURE 3.15 – Nuages de points de la scène test « Étalon ». (a) Estimée initiale (b) ICP (c) Méthode de Chen (d) Calibrage inverse (e) Notre méthode.

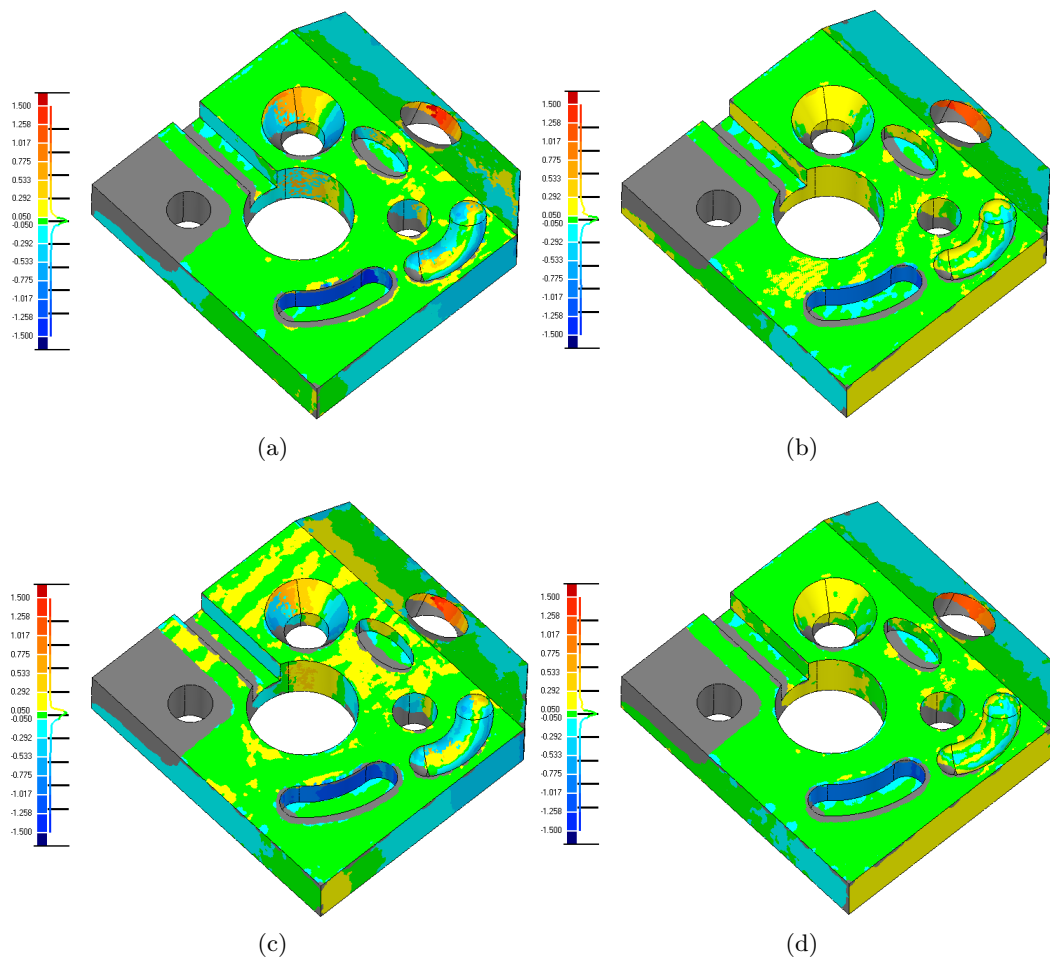


FIGURE 3.16 – Comparaison des nuages recalés au modèle CAO. (a) ICP (b) Méthode de Chen (c) Calibrage inverse (d) Notre méthode.

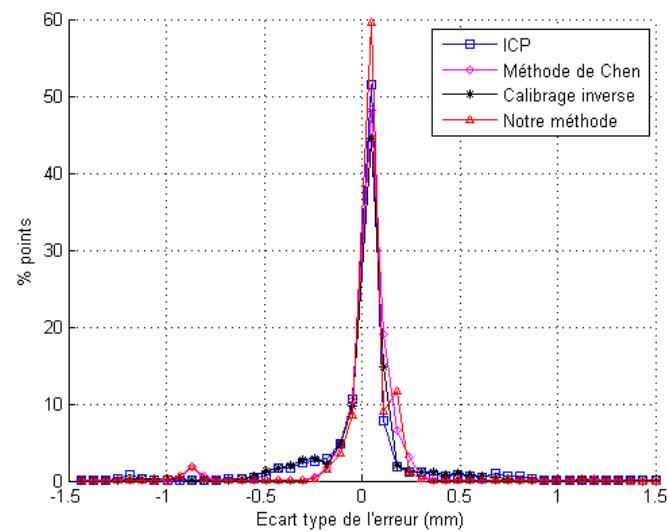


FIGURE 3.17 – Distributions des erreurs par comparaison au modèle théorique

Approche globale pour le recalage

Table des matières

4.1	État de l'art	90
4.2	Présentation liminaire	92
4.2.1	Problématique	92
4.2.2	Données et pré-traitements	93
4.3	Champs de rotations	93
4.3.1	Sélection de vue	94
4.3.2	Score de reprojection	96
4.3.3	Erreur inertielle et correction	97
4.3.4	Algorithme complet	98
4.4	Propagation par graphe	99
4.4.1	Détection de fermeture de boucle	99
4.4.2	Optimisation à la fermeture	101
4.5	Optimisation hors-ligne	102
4.5.1	Stratégie multi-vues	102
4.5.2	Critère	102
4.5.3	Sélection de points et pondération	103
4.6	Évaluations	104
4.7	Conclusion	105

LA modélisation d'un objet complet nécessite l'acquisition de plusieurs vues successives. Pour cela, l'opérateur peut soit déplacer le capteur devant l'objet, soit déplacer l'objet devant le capteur. Chaque nouvelle vue doit alors être placée dans un repère cohérent avec celui de l'objet.

Une approche globale de ce problème va chercher à exploiter l'ensemble de l'information utile déjà disponible au moment d'une acquisition. Le principal apport de ce type d'approche est d'éviter au maximum les incohérences, d'améliorer la précision, et de lever des hypothèses lors de certaines phases de la modélisation.

Ce chapitre s'attache donc à présenter et évaluer les stratégies que nous avons mises en œuvre pour le recalage multi-vues incrémental appliqué à notre capteur. Après avoir introduit l'état de l'art des stratégies pour la modélisation incrémentale et le recalage multi-vues, nous présenterons les approches que nous avons proposées et nous les évaluerons.

4.1 État de l'art

Les avantages, en maniabilité et en accessibilité, des technologies tenues à la main pour la modélisation 3D, ainsi que l'intérêt scientifique de ce type d'approche ont motivé de nombreux travaux récents.

Parmi les travaux notables, Pollefeys *et al.* [Pollefeys 04] ont proposé une méthode de *Structure-From-Motion* pour la modélisation hors ligne d'une scène acquise depuis une seule caméra mobile non calibrée. Cette méthode réalise un suivi de points dans toute la séquence pour déterminer globalement les poses de la caméra, ses paramètres intrinsèques et la géométrie des points clefs suivis. Dans un second temps la scène 3D est densifiée par reconstruction stéréoscopique multi-vues. Une évolution récente [Newcombe 10] de ce type d'approche combine un schéma probabiliste pour le *Structure-From-Motion* temps-réel et une densification par ajustements de faisceaux locaux [Triggs 00] intervenant sur une dynamique plus faible. Le principal inconvénient de ces méthodes est la nécessité d'observer une scène fortement texturée.

Pour pallier à ce problème, le recours à l'illumination ou l'apport d'énergie à la scène sont souvent adoptés. La projection de lumière structurée ou la triangulation laser sont des solutions régulièrement utilisées. Notons les travaux particuliers de [Matabosch Geronès 07] sur un système de modélisation 3D mono-vue basé sur la triangulation de plusieurs raies laser. Par ailleurs, le projet de numériseur 3D du DLR Institute of Robotics and Mechatronics [Strobl 09] a également proposé la mise en œuvre d'un système de reconstruction par triangulation laser. Le recalage dans cette approche est résolu par une combinaison de suivi inertiel et visuel.

Les premières approches globales pour le recalage cherchant à résoudre le problème multi-vues sont liées à la mise en œuvre des méthodes type ICP paire-à-paire, comme extensions de celles-ci. Ainsi [Chen 92] propose un recalage successif de toutes les vues et l'aggrégation de chaque nouvelle vue recalée à un *méta-modèle* composé de toutes les vues précédentes. L'inconvénient majeur de cette méthode réside dans la propagation et l'accumulation de chaque erreur de recalage. Il ne s'agit pas exactement, d'autre part, d'une méthode globale puisqu'une vue agrégée ne bénéficie pas de l'apport d'information des recalages suivants. Une extension de ce travail est proposée par [Pulli 99] déterminant au préalable l'ensemble des recalages possibles avant de déterminer une combinaison de paires pour le recalage multi-vues. La contribution récente de [Toldo 10] apporte une méthode originale pour le recalage

simultané de toutes les vues dans un schéma itératif de type ICP. Les points sélectionnés sont regroupés en grappes avec leurs plus proches voisins dans les autres vues recouvrantes. L'appariement d'un point se fait avec le centroïde de la grappe. Cette méthode permet une précision améliorée et une meilleure robustesse aux mauvaises initialisations.

Afin de rendre plus efficace le recalage d'un ensemble de vues, de nombreux travaux ont cherché à déterminer une organisation de la scène permettant la mutualisation et la propagation d'information pour une optimisation globale des transformations. Les travaux de [Bergevin 96] formalisent le problème en représentant les vues et les transformations rigides entre elles au moyen d'un graphe. L'exploitation de telles structures permet de rejeter les correspondances fausses mais localement correctes [Huber 03]. Cependant la résolution de tels problèmes met en œuvre des méthodes complexes et coûteuses comme l'exploration de solutions par algorithmes génétiques [Silva 07].

Malgré tout, la connexité d'un ensemble de vues reliées par un ensemble de transformations rigides permet l'exploitation de contraintes pour rendre l'optimisation plus robuste. Afin de décomposer le graphe de correspondances liant les vues à leurs voisinages, [Sharp 04] propose une décomposition en composantes connexes. Les transformations optimales sont alors déterminées en considérant que la composition successive des transformations sur une composante connexe doit être une identité. L'approche a depuis été étendue pour s'adapter à un fonctionnement de modélisation incrémentale [Matabosch Geronès 07, Batlle Subirós 08]. Pour cela, les vues sont recalées en paire-à-paire jusqu'à la détection de cycles permettant une optimisation multi-vues.

En exploitant les images, dans le cas de systèmes visuels, le recours à l'ajustement de faisceaux [Triggs 00] sur l'ensemble des vues permet de raffiner à la fois les poses du capteur et les points 3D reconstruits. Les points sont projetés dans les vues dans lesquels ils sont visibles et on cherche alors à minimiser globalement leur erreur de reprojection. De plus, les travaux récents sur le sujet ont permis d'améliorer grandement l'efficacité et le temps de calcul nécessaire, particulièrement en ayant recours à un fonctionnement incrémental [Mouragnon 06]. Toutefois ces approches restent très dépendantes de la qualité de l'estimée initiale.

La prise en compte de l'incertitude de chaque pose peut être abordée en représentant les transformations spatiales comme des variables aléatoires. Les travaux de [Lu 97] introduisent une telle approche stochastique pour une application à la navigation robotique. Les travaux de Nüchter *et al.* [Nüchter 04, Nüchter 06] proposent un cadre pour résoudre le problème de la localisation et modélisation simultanée d'environnement pour la navigation de robots d'exploration en terrain difficile. D'autres auteurs ont exploré ce type d'approche en se basant sur le filtrage stochastique [Nieto 05] par filtre de Kalman étendu par exemple.

Des travaux récents sur le SLAM (*Simultaneous Localization and Mapping*) se sont orientés vers l'exploitation de données visuelles. La localisation par techniques photométriques a été explorée pour introduire une odométrie optique [Olson 03, Pan 09] mais subit une dérive rendant le système imprécis rapidement. Davison *et al.* [Davison 03] ont proposé une approche de localisation et cartographie éparse simultanées à l'aide d'une seule caméra mobile et basée sur le filtrage de Kalman. On peut bénéficier de l'apport de la fédération d'information et de la fermeture de boucle lors de la réobservation de points de vues précédents. Ces travaux ont par ailleurs été étendus pour proposer des implémentations temps-réel [Royer 07] exploitant le principe de recherche active dans les images [Davison 07]. Les travaux de [Montiel 06] ont permis de résoudre le problème de l'initialisation retardée des amers visuels en proposant une

paramétrisation utilisant la distance inverse entre l’amer et le centre optique. Cette solution permet également de maintenir une incertitude acceptable pour la linéarisation tout en proposant une représentation des points jusqu’à l’infini.

L’utilisation de systèmes stéréoscopiques pour le SLAM basé vision a également été abordée [Paz 08]. Dans ce cas, les amers visuels sont rapidement initialisés ce qui permet de travailler sur une information globalement moins incertaine. De ce fait on réduit la dérive et l’augmentation d’incertitude pouvant rendre un système de navigation inconsistant. Cependant le champ de vue d’un système stéréoscopique est largement réduit par rapport à une approche mono-caméra. On ne bénéficie donc plus de la vision « à l’infini » qu’offre la paramétrisation de Montiel *et al.* [Montiel 06]. Il a été mis en avant par [Solà 07] que l’exploitation de chaque caméra d’un système stéréoscopique indépendamment plutôt qu’en *banc* donne de meilleures performances pour la navigation. En effet on bénéficie d’une initialisation et convergence rapide de l’estimation des amers sans restreindre le champ de leur observabilité.

Dans les applications aérospatiales, le recours à la mesure inertielle de précision est courant pour la navigation [Lefferts 82, Roumeliotis 02], et le couplage à la mesure visuelle a été abordé assez tôt. Ce type de technique a été adapté pour l’intégration de capteurs inertiels bas coût pour augmenter et améliorer les systèmes de suivi ou de navigation basés vision [Bleser 09]. Souvent le recours à un filtre stochastique est utilisé pour la fusion de données multi-sensorielles [Gemeiner 07, Pinies 07]. Cependant ces approches se basent sur une intégration continue de données. L’exploitation de ces mesures dans un système discontinu n’a pas fait l’objet d’importants travaux dans la communauté de la navigation ou de la localisation et cartographie simultanées.

4.2 Présentation liminaire

Afin de bien aborder les stratégies et méthodes de ce chapitre, il convient de présenter les données dont on dispose. Le lecteur se référera au chapitre 1 concernant le fonctionnement du système et le déroulement de la modélisation. Nous allons tout de même rappeler ici les informations dont nous disposons ainsi que ce que l’on cherche à obtenir.

4.2.1 Problématique

Dans le chapitre 3 nous nous sommes attachés à présenter le principe du recalage de deux vues, c’est à dire déterminer la transformation permettant d’exprimer un point du repère de la vue $n + 1$ dans le repère de la vue n .

La problématique essentielle de ce chapitre est de déterminer la pose du capteur à chaque instant d’acquisition par rapport au repère de l’objet. C’est à dire que l’on souhaite connaître la transformation $[\mathbf{R}_n | \mathbf{t}_n]$ permettant d’exprimer un point $\mathbf{p}$ exprimé dans le repère caméra à l’instant d’acquisition n dans le repère global de l’objet.

$$\mathbf{p}_{objet} = \mathbf{R}_n \mathbf{p}_n + \mathbf{t}_n$$

Notons que notre repère global est lié à l’objet. Ainsi, déplacer l’objet modifie son repère. Cette précision est importante car nous exploitons les mesures d’une centrale inertielle, dont le référentiel est fixe et lié, grossièrement, à la Terre. Par conséquent, le mouvement de l’objet rend inconsistente la mesure inertielle.

Dans l'optique d'une modélisation incrémentale en ligne, donc temps réel, il faut s'attacher à proposer un fonctionnement et un traitement rapide. Nous avons donc divisé le travail en deux phases. La première est la modélisation incrémentale en ligne à proprement parler, qui va essentiellement chercher à cartographier l'acquisition sans proposer la précision de recalage nominale. La seconde phase est une optimisation hors ligne exploitant la cartographie de l'acquisition afin d'atteindre un niveau de précision permettant l'exploitation des résultats dans des applications telles que le contrôle dimensionnel ou la rétro-conception par exemple.

Ainsi, on peut diviser notre problématique selon les axes suivants :

- Chaque nouvelle acquisition va être prioritairement recalée avec une vue précédente. Une heuristique de sélection doit donc être déterminée, puis la pose est estimée par recalage paire-à-paire (Figure 1.7),
- Le recours à une cartographie permet de détecter des cycles, ou *fermetures de boucles* dans la modélisation, donnant la possibilité d'une propagation et d'une optimisation prioritaire des poses sur ces cycles (Figure 1.8),
- Lors du déplacement de l'objet, une correction doit être apportée aux mesures inertielles afin de se ramener à nouveau dans le repère de l'objet (Figure 1.9),
- En fin de traitement, il convient d'exploiter la cartographie et l'ensemble des données pour une optimisation globale des poses (Figure 1.11).

4.2.2 Données et pré-traitements

À chaque acquisition, nous disposons d'images et de mesures inertielles en provenance du capteur. On considère comme disponibles certains pré-traitements sur ces données, tels que la correction de distorsions ou la rectification stéréoscopique des images, la reconstruction 3D ou l'extraction de points d'intérêt. La Figure 4.1 illustre les pré-traitements et ce dont on dispose pour nos algorithmes.

Le capteur fournit, à chaque acquisition, deux paires d'images. La première avec projection de motif, la seconde sans. Pour chaque paire d'image, une mesure d'attitude venant de la centrale inertielle est ajoutée.

Comme nous l'avons vu au chapitre 2, les images sont préalablement corrigées et rectifiées afin de supprimer l'effet des distorsions optiques et pour aligner les droites épipolaires. Les images avec illumination étant utilisées pour la génération des nuages de points 3D, la mise en correspondance dense et la corrélation sont réalisées avec elles. On génère ainsi un ensemble de points $\{\mathbf{p}_i^n\}$ pour la vue n .

Notons que les images sans illumination peuvent également être prétraitées afin de mettre à jour une base de points d'intérêt extraits de ces images à des fins d'indexation. Nous aborderons cette utilisation plus loin dans ce chapitre.

4.3 Champs de rotations

Nous avons introduit dans le chapitre 3 nos méthodes pour le recalage paire-à-paire. Ces méthodes exploitent les données inertielles et les images afin d'accélérer l'estimation initiale ou le recalage de deux vues.

Lors d'une modélisation incrémentale en ligne, l'exploitation de ces méthodes pour le recalage de la vue courante avec le modèle déjà acquis nécessite la sélection préalable d'une bonne vue candidate. Une fois la vue sélectionnée, l'estimée initiale est obtenue par la méthode proposée en section 3.1.5.

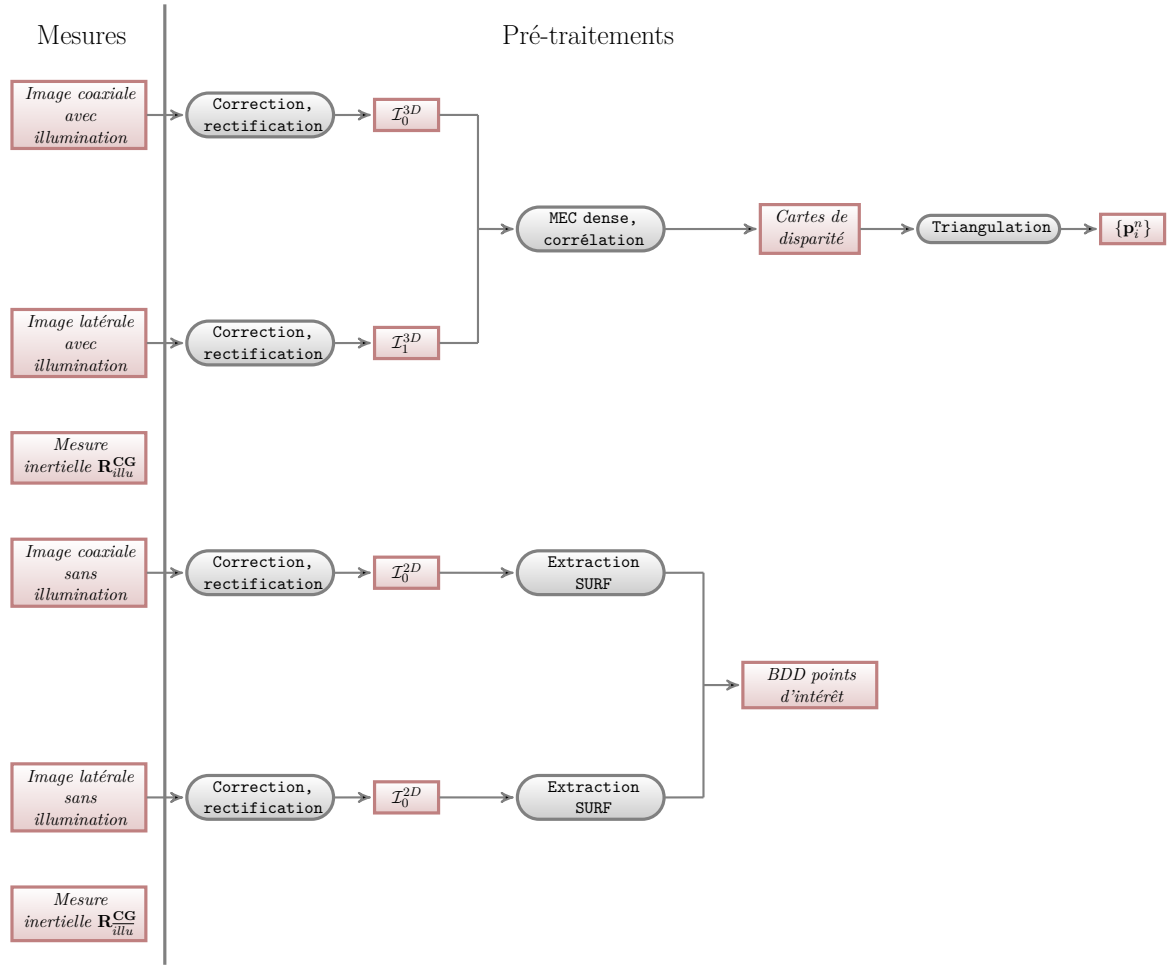


FIGURE 4.1 – Données et pré-traitements pour l'acquisition à l'instant n . Les informations stockées pouvant être utilisées sont identifiées en rouge.

4.3.1 Sélection de vue

Sans autre information *a priori* que la mesure inertielle, la sélection d'une bonne vue candidate n'est pas un problème trivial. Il est pourtant essentiel puisqu'il va induire le résultat de tous les recalages. Une mauvaise sélection peut donc conduire à un recalage imprécis voire mauvais pouvant se propager sur toute nouvelle acquisition que l'on voudra recalcr à cette vue aberrante.

On peut dès lors distinguer deux cas. Soit l'opérateur réalise une acquisition séquentielle, appliquant au capteur un mouvement continu. Soit la nouvelle acquisition fait suite à une reprise du capteur après interruption par l'opérateur – pour aller combler un trou, observer une nouvelle partie de l'objet ou suite à un déplacement de l'objet.

Dans le cas d'une acquisition séquentielle, un *a priori* simple est de considérer que la vue précédente est probablement la meilleure candidate pour un recalage puisqu'on peut supposer que la pose du capteur a peu changé dans le court instant séparant deux acquisitions. Cette hypothèse n'est bien entendu plus valide lorsque l'on interrompt puis reprend la capture. De plus elle peut être limitante si l'on considère que l'on réobserve une zone déjà numérisée.

Dans ce cas, la vue précédente n'est pas nécessairement le meilleur choix. Par conséquent il est nécessaire de disposer d'une méthode pour la sélection de vues plus générique, fonctionnant même si le recalage avec la vue précédente s'avère mauvais – en cas de mouvement brusque par exemple.

Dans notre approche, nous avons proposé une sélection par voisinage dans l'espace des rotations. En effet, rappelons que la méthode de recalage initial proposée réalise une rectification perspective dans les images avant de faire la mise en correspondance des descripteurs. Par conséquent, une vue voisine dans l'espace des rotations permettra une déformation moins importante de l'image et donc minimisera la dégradation d'information que la rectification peut apporter aux images.

Ainsi notre critère de sélection va permettre de choisir la vue précédente offrant une transformation minimale dans les images. Cette transformation est la relation permettant d'exprimer les points images de la vue $n + 1$ dans le repère image de la vue n .

$$\mathbf{M} = \mathbf{K} \mathbf{R}_{\mathcal{GC}_n}^{-1} \mathbf{R}_{\mathcal{GC}_{n+1}} \mathbf{K}^{-1} \quad (4.1)$$

$$w \begin{pmatrix} x' \\ y' \\ 1 \end{pmatrix} = \mathbf{M} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}$$

La matrice $\mathbf{K}$ est la matrice de calibrage de la caméra coaxiale. Les matrices $\mathbf{R}_{\mathcal{GC}_n}$ et $\mathbf{R}_{\mathcal{GC}_{n+1}}$ sont les mesures d'attitudes fournies par la centrale, respectivement à l'acquisition n et $n + 1$. Ces matrices sont des rotations vers un repère propre de la centrale – lié au champ magnétique terrestre. La composition $\mathbf{R}_{\mathcal{GC}_n}^{-1} \mathbf{R}_{\mathcal{GC}_{n+1}}$ exprime donc la rotation *mesurée* entre les repères de la caméra coaxiales aux acquisitions n et $n + 1$.

La distance entre les coordonnées images (x, y) et leur transformation (x', y') va servir à la mesure de déformation. Cette mesure n'a pas besoin d'être faite sur tous les points images. On choisira de l'observer uniquement sur les quatres coins de l'image (Figure 4.2). On choisit pour la translation celle permettant de ramener le centre transformé au centre de l'image initial. Cela permet d'avoir une référence fixe pour la comparaison de l'influence des transformations sur les images candidates. De plus cette configuration permet, la plupart du temps, de maximiser approximativement le recouvrement entre l'image initial et l'image déformée.

La mesure utilisée pour le choix de la vue candidate est la somme des distances euclidiennes entre chaque coin de l'image avant et après transformation. La vue minimisant cette mesure sera choisie comme meilleure candidate pour le recalage.

$$n = \underset{\mathbf{R}_{\mathcal{GC}_{n+1}}}{\operatorname{argmin}} \sum_i \|\mathbf{p}_i - \mathbf{M}(\mathbf{R}_{\mathcal{GC}_n}, \mathbf{R}_{\mathcal{GC}_{n+1}}) \mathbf{p}_i\|^2$$

L'ensemble des points $\{\mathbf{p}_i\}$ est constitué des coins de l'image. La matrice $\mathbf{M}$, fonction des rotations mesurées aux deux instants d'acquisition considérés, est détaillée dans l'équation (4.1).

Un critère de qualité peut être ajouté sur la vue sélectionnée. En appliquant un seuil sur la somme des distances on se donne la possibilité de ne choisir aucune vue candidate par la méthode utilisant la centrale inertielle.

Une fois la vue candidate sélectionnée, on propage la transformation à l'ensemble des points de l'image et on réalise l'interpolation de la fonction de luminance pour créer une

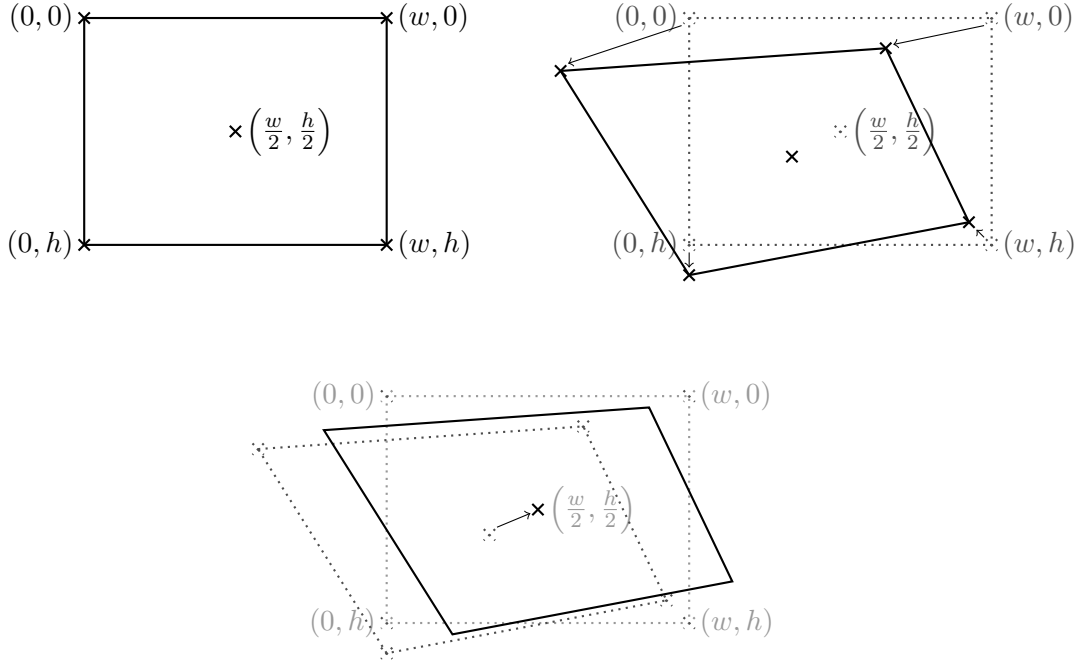


FIGURE 4.2 – Application de la déformation sur les coins de l'image et le centre. On applique une translation pour ramener le centre transformé au centre réel de l'image.

image rectifiée afin d'exploiter la méthode de recalage détaillée en section 3.1.5. A l'issue du recalage initial, on réalise un recalage précis paire-à-paire rapide – limité à quelques itérations.

Si aucun candidat n'est trouvé ou si le recalage précis n'a pas donné de bon résultat (voir section 4.3.2), on s'orientera plutôt vers des méthodes de sélection et recalage par indexation de points d'intérêts (section 3.1.3).

4.3.2 Score de reprojection

Pour jauger de la qualité de sélection et de recalage en fin de processus, on met en place un critère basé sur la reprojection des points 3D calculés à l'instant $n + 1$ recalés dans les images de la vue candidate n .

Les points 3D calculés à la vue $n + 1$ recalés par la transformation $[\mathbf{R}|\mathbf{t}]$ estimée dans le repère de la vue n sont projetés dans les deux images de la paire stéréoscopique de la vue n .

$$\begin{aligned}\mathbf{q}_c &= \mathbf{K}_c(\mathbf{R}\mathbf{q} + \mathbf{t}) \\ \mathbf{q}_l &= \mathbf{K}_l\mathbf{T}_{C_l C_c}(\mathbf{R}\mathbf{q} + \mathbf{t})\end{aligned}$$

Les points $\mathbf{q}_c$ et $\mathbf{q}_l$ sont les projections des points $\mathbf{q}$, exprimé dans le repère de la caméra coaxiale à l'instant d'acquisition $n + 1$, dans les images rectifiées – respectivement – de la caméra coaxiale et de la caméra latérale. Les matrices $\mathbf{K}_c$ et $\mathbf{K}_l$ sont les matrices des paramètres intrinsèques des caméras coaxiales et latérales, dans l'espace rectifié. La transformation $\mathbf{T}_{C_l C_c}$ est le changement de repère permettant d'exprimer les points du repère coaxial dans le repère latéral, toujours dans l'espace rectifié.

Pour tout point ainsi projeté on peut calculer, en observant son voisinage, un score de corrélation.

Soit le vecteur $\mathbf{y} = [y_0 \ y_1 \ \dots \ y_m]^T$ le vecteur des scores de corrélation sur les voisinages des projections des points $\mathbf{q} = [q_0 \ q_1 \ \dots \ q_m]^T$ reconstruits depuis l'acquisition à l'instant $n+1$. On considère que $\mathbf{q}$ est la réalisation d'une variable aléatoire $\mathbf{Q}$. On définit $\mathcal{H}_1$ l'hypothèse que le recalage est correct, et $\mathcal{H}_2$ qu'il ne l'est pas. Ces hypothèses sont définies par les distributions P_1 et P_2 , la réalisation $\mathbf{q}$ est issue de l'une d'elle.

$$\mathcal{H}_i : \{\mathbf{q} \sim P_i\}$$

Les distributions P_1 et P_2 admettent f_1 et f_2 pour densités. On peut alors définir le test $\tilde{\delta}$ du rapport de vraisemblance.

$$\tilde{\delta}(q_0, \dots, q_n) = \begin{cases} \mathcal{H}_1 & \text{si } \Lambda(q_0, \dots, q_n) = \frac{f_2(q_0, \dots, q_n)}{f_1(q_0, \dots, q_n)} < h \\ \mathcal{H}_2 & \text{si } \Lambda(q_0, \dots, q_n) \geq h \end{cases}$$

Le paramètre h est solution de l'équation suivante :

$$\mathbb{P}_1(\Lambda(q_0, \dots, q_n) \geq h) = \alpha \quad (4.2)$$

La probabilité $\mathbb{P}_1$ est l'erreur de première espèce (probabilité de fausse alarme). La valeur α est la borne supérieure acceptable de cette probabilité pour un test donné.

L'équation (4.2) nécessite de connaître la distribution du rapport de vraisemblance. On approxime généralement celle-ci en considérant que $-2\log(\Lambda)$ suit asymptotiquement une distribution du χ^2 .

En utilisant le test du rapport de vraisemblance on peut ramener l'analyse du score de reprojection, et la décision du rejet ou de l'acceptation du recalage avec la vue candidate, aux paramètres de la tolérance sur la pose choisie.

4.3.3 Erreur inertielle et correction

La méthode proposée nous permet également de propager une correction de dérive de la centrale inertielle. En effet, la centrale dérive lentement au cours du temps. Au bout d'un certain temps le recalage sur la seule base de cette information devient inconsistent. Par conséquent, il nous faut mettre à jour une matrice de rotation corrective R^+ . Lorsqu'un recalage a été accepté, on modifie cette matrice pour qu'elle représente la variation entre la rotation mesurée par la centrale $\mathbf{R}_{\mathcal{G}_{n+1}}$ et la rotation finale trouvée par le recalage $\mathbf{R}$.

$$\mathbf{R}^+ = \mathbf{R} \mathbf{R}_{\mathcal{G}_{n+1}}^{-1}$$

Cette correction permet donc de corriger la mesure inertielle afin de passer du repère dérivant de la centrale au repère réel de l'objet.

On peut donc gérer, par cette correction le mouvement de l'objet. En effet, lorsque l'opérateur déplace l'objet il rend la mesure inertielle inconsistante. La sélection de vue ne donne alors plus de résultats concluants en se servant de la mesure d'attitude et on a donc recours à des méthodes basées images uniquement. Lorsque les conditions d'un tel recalage – point de vue proche d'un point de vue déjà acquis – sont remplies, celui-ci peut être résolu et on peut donc déterminer une correction inertielle, qui va dès lors permettre de ramener la mesure d'attitude dans le repère de l'objet après déplacement.

La correction est appliquée à la rotation mesurée et cette composition remplace la mesure dans les algorithmes que nous avons décrits jusqu'à présent.

$$\mathbf{R}'_{\mathcal{G}_{n+1}} = \mathbf{R}^+ \mathbf{R}_{\mathcal{G}_{n+1}}$$

4.3.4 Algorithme complet

La figure 4.3 détaille l'algorithme de recalage en ligne.

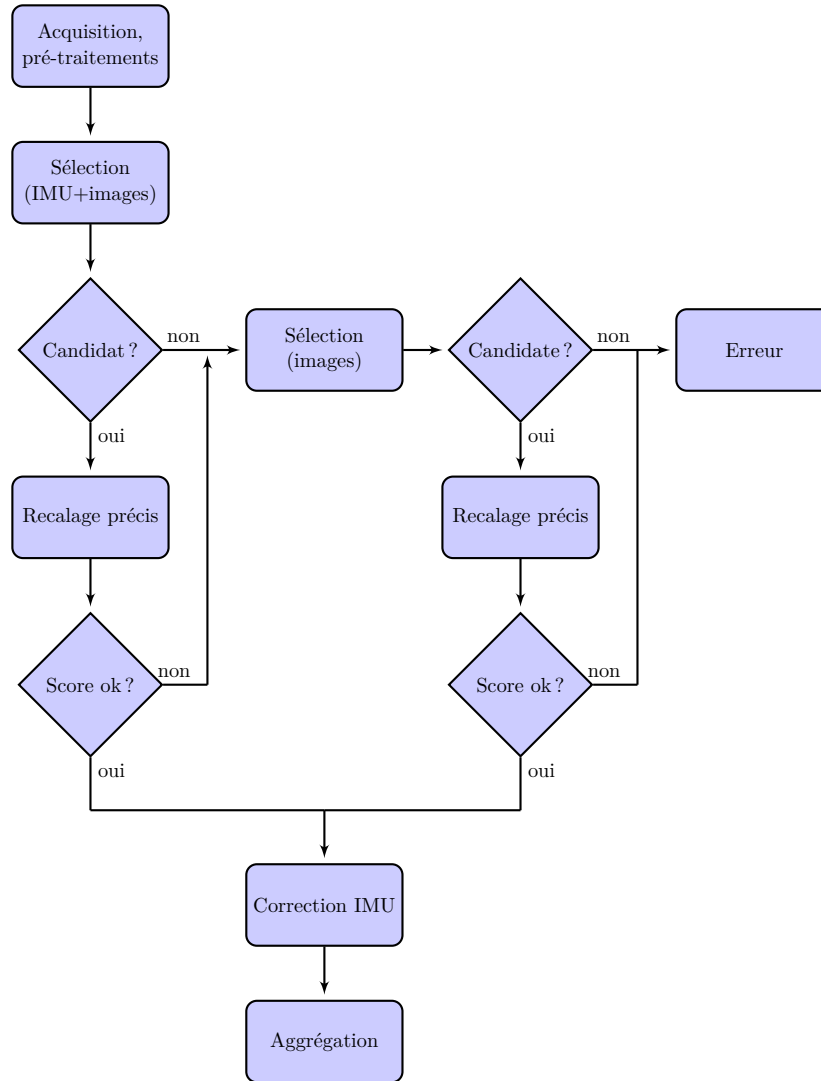


FIGURE 4.3 – Algorithme de sélection de vue et recalage

La branche principale de l'algorithme, à gauche dans la figure 4.3, utilise la mesure inertielle pour sélectionner la vue candidate et pour réaliser le recalage grossier. Ce mode est valide lorsque l'opérateur ne déplace pas la pièce, qu'il réalise une séquence continue d'acquisition ou non. En cas d'échec dans ce mode – à la sélection ou au score de reprojection final – la branche secondaire, à droite, est utilisée. Ce second mode va avoir lieu lorsque l'opérateur

a déplacé la pièce ou lorsque la centrale inertielle tombe en défaut – panne, perturbation magnétique.

Lorsque les deux branches sont en défaut, il n’y a plus de recalage possible. Ce cas survient lorsque l’acquisition a été mauvaise – erreur de trame, images floues, mouvement trop brusque – ou lorsque l’opérateur réalise l’acquisition d’une zone ne présentant pas assez de recouvrement avec le modèle déjà acquis.

4.4 Propagation par graphe

Après avoir sélectionné une vue candidate pour le recalage paire à paire, on ajoute cette vue au modèle. Pour permettre la prise en compte globale au niveau du modèle entier du problème de recalage, il faut mettre en place une structure vecteur pour une optimisation multi-vues.

En identifiant les vues par des nœuds et les transformations par des liaisons, plusieurs auteurs ont utilisé le formalisme des graphes pour la représentation du modèle numérisé. A chaque acquisition pour laquelle un recalage a été trouvé, un nœud est créé et une liaison est constituée entre ce dernier et le nœud correspondant à la vue avec laquelle on a recalé. On observe que lorsqu’un cycle existe dans une telle structure, la composition des transformations doit donner une identité. Il est donc nécessaire de détecter lorsqu’une telle fermeture de boucle est possible afin de l’utiliser pour réaliser une optimisation des vues concernées à l’aide de la contrainte de consistance par n -uplet sur les transformations entre elles.

4.4.1 Détection de fermeture de boucle

Dans la construction incrémentale d’un graphe de recalage, une fermeture va résulter en la création d’une liaison depuis une vue déjà stockée dans le modèle vers la dernière vue acquise. La figure 4.4 illustre ce principe. Par recalages successifs, on crée une chaîne : deux vues se trouvent liées par la transformation estimée. Créer un cycle revient à ajouter une transformation dans notre graphe entre une vue déjà incluse au modèle et la dernière vue acquise.

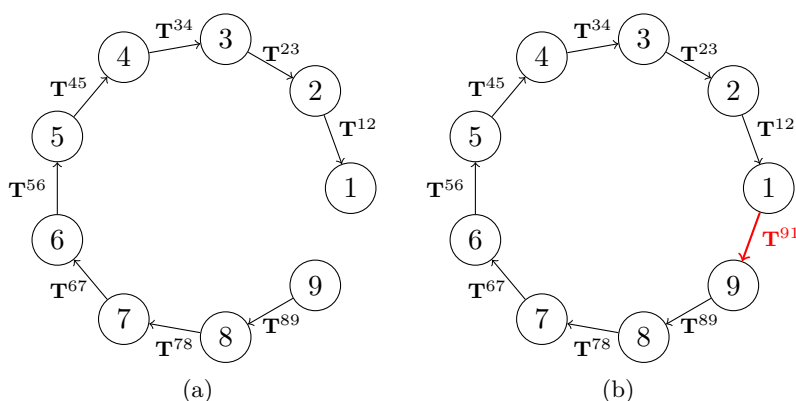


FIGURE 4.4 – Principe de la fermeture de boucle. (a) Lors de l’acquisition des neuf premières vues, on construit le graphe de relations en ligne. (b) Lorsque la possibilité de cycle est détectée, on fait l’ajout de la relation T^{91} pour effectivement créer ce cycle.

En détectant ces possibilités et en ajoutant une relation dans le graphe on rompt le fonctionnement séquentiel pour permettre une optimisation entre toutes les vues concernées. La détection doit cependant être un processus rapide pour être réalisable en ligne ainsi que l'optimisation du cycle créé. Une hypothèse courante consiste à supposer qu'une fermeture de boucle peut exister si deux noeuds non reliés constituent des positions de centres de caméras proches spatialement (Figure 4.5). Cette hypothèse est raisonnable dans la mesure où deux positions de visualisation proches sont probablement des points de vues proches sur une sphère d'observation (au sens de la position mais aussi de l'orientation).

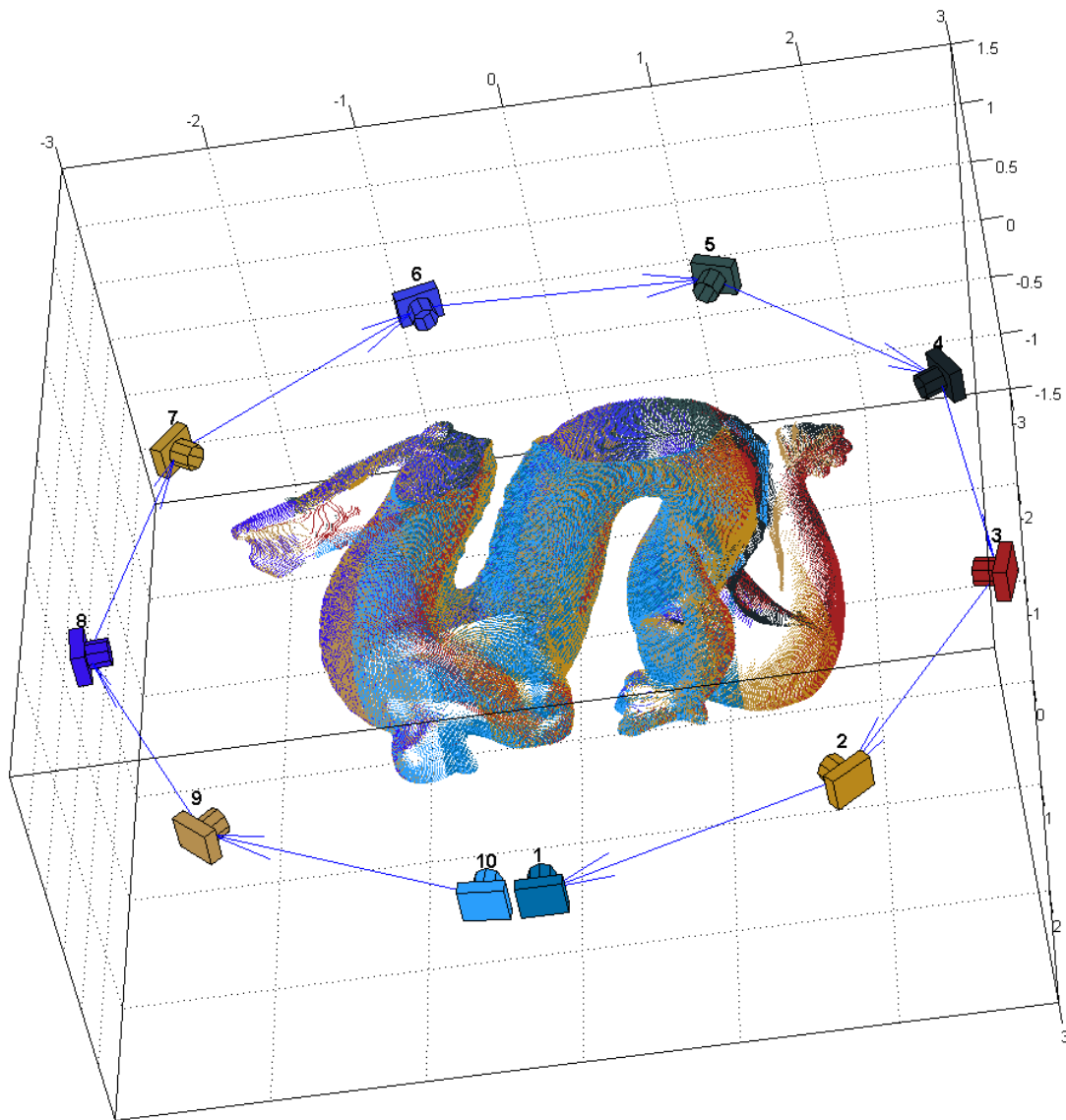


FIGURE 4.5 – Après recalages successifs, la position estimée de la dixième vue est proche de celle de la première. On va chercher la présence d'une fermeture de boucle sur ces vues.

La proximité spatiale n'est cependant pas une condition suffisante pour qu'un recalage soit possible, il faut que les acquisitions soient assez recouvrantes pour permettre un recalage.

Pour cela nous exploitons le critère de reprojection de la section 4.3.2. Dans ce cas la tolérance appliquée au critère doit aussi tenir compte de l'imprécision apportée par la propagation d'erreur pendant la construction de la chaîne. Si le test de vraisemblance est favorable, les vues sont recouvrantes et vont pouvoir être utilisées dans l'optimisation du cycle.

4.4.2 Optimisation à la fermeture

Lors de la création d'un cycle dans le graphe de recalage on a la possibilité de réduire significativement l'erreur de propagation en utilisant la contrainte de consistance par n -uplet (Figure 4.6).

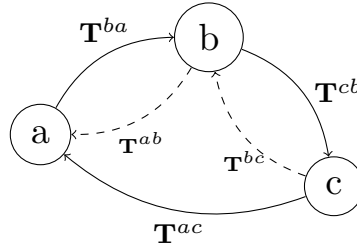


FIGURE 4.6 – Consistance par n -uplet. La composition de toutes les transformations d'un cycle dans le sens de l'orientation du graphe est une identité. De plus $\mathbf{T}^{ac} = \mathbf{T}^{ab}\mathbf{T}^{bc}$.

Le recalage multi-vues d'un cycle est une extension du critère paire-à-paire décrit en section 3.2.5. On va chercher à raffiner les poses pour chaque acquisition en minimisant un critère de distance entre points appariés en créant une contrainte pour tenir compte de la consistance des transformations sur le cycle. L'erreur d'appariement pour un couple de points est donc exprimée comme la somme de la distance entre les points par la transformation directe entre les vues concernées et la distance en utilisant la composition des inverses de toutes les autres transformations.

$$e = d_{i(i+1)} + d_{(i+1)i} \quad (4.3)$$

Les deux composantes de l'erreur s'écrivent en fonction des points appariés $\mathbf{p}^i$, exprimés dans le repère de la vue i et $\mathbf{q}^{(i+1)}$, exprimés dans le repère de la vue $(i+1)$.

$$\begin{aligned} d_{i(i+1)} &= \left\| \mathbf{p}^i - \mathbf{T}^{i(i+1)} \mathbf{q}^{(i+1)} \right\| \\ d_{(i+1)i} &= \left\| \mathbf{p}^i - \left[\mathbf{T}^{i(i-1)} \dots \mathbf{T}^{10} \mathbf{T}^{0(i+N)} \dots \mathbf{T}^{(i+2)(i+1)} \right] \mathbf{q}^{(i+1)} \right\| \end{aligned}$$

On renforce de plus la contrainte de cycle en maintenant que la composition de toutes les transformations est une identité.

$$\mathbf{E} = \left(\prod_{cycle} \mathbf{T} \right) - \mathbf{I}_{4 \times 4} \quad (4.4)$$

L'expression $\|\mathbf{E}\|_F$ est la norme de Frobenius de la matrice $\mathbf{E}$.

On construit maintenant le critère à minimiser à partir des équations (4.3) et (4.4).

$$\underset{\{\mathbf{T}^{ij}\}, \|\mathbf{E}\|_F^2=0}{\operatorname{argmin}} \sum_{n=0}^{N_{\text{paires}}} e^2 \quad (4.5)$$

4.5 Optimisation hors-ligne

Une fois l'acquisition terminée, le calcul précis des poses peut être réalisé. Pour cela nous procédons à une optimisation sur l'ensemble des vues acquises et pré-recalées lors du processus en ligne.

Le critère mis en œuvre exploite l'erreur de reprojection d'un point dans une paire de vues stéréoscopiques. Lorsqu'on projette un point dans une paire d'images, le score de reprojection de ce point sera le score de corrélation au voisinage des pixels projetés. Ce type de critère ne dispose que d'un bassin d'attraction étroit, il est donc nécessaire d'avoir réalisé les optimisations lors du processus en ligne afin de disposer d'une initialisation la plus précise possible.

4.5.1 Stratégie multi-vues

Un point 3D reconstruit est généralement observable dans plusieurs paires d'images de la séquence. La première étape de notre optimisation consiste à déterminer les vues dans lesquelles chaque point 3D sera projeté.

Pour être observable un point doit avoir son projeté inclus dans les deux images d'une acquisition, et ne pas être occulté par une autre surface. Afin de détecter les occultations, nous réalisons un maillage grossier sur le modèle complet recalé avant optimisation. De plus, on considère qu'un point n'est observable que si la surface qui le contient est suffisamment fronto-parallèle aux caméras. On utilise pour cela l'orientation relative des directions de visées et des normales locales aux points considérés.

La figure 4.7 illustre l'observabilité des points pour une caméra. On étend ce fonctionnement pour considérer les deux caméras de la paire stéréoscopique à chaque position d'acquisition.

L'observabilité de chaque point 3D pour chaque paire de caméras est déterminée une seule fois au démarrage du processus.

4.5.2 Critère

L'optimisation des poses se fait en minimisant un critère global de corrélation des points 3D projetés dans chaque vue qui l'observe. Pour un couple point 3D – paire d'images, le score de reprojection associé est le score de corrélation SSD, tel que décrit en section 2.2.2.

$$s_i^n = \sum_W [\mathcal{I}_0^n(\mathbf{m}_0 + \delta) - \mathcal{I}_1^n(\mathbf{m}_1 + \Delta')]^2 \quad (4.6)$$

L'équation (4.6) décrit le critère de corrélation d'un point i ($i \in [1..N_{3D}]$) pour une acquisition n ($n \in [1..N_V]$) où $\mathbf{m}_0$ est la projection du point 3D considéré dans l'image coaxiale et $\mathbf{m}_1$ est sa projection dans l'image latérale. Le vecteur δ est le déplacement dans la région W et Δ' est la déformation de la région W sans composante de translation, calculée à partir de la normale locale au point.

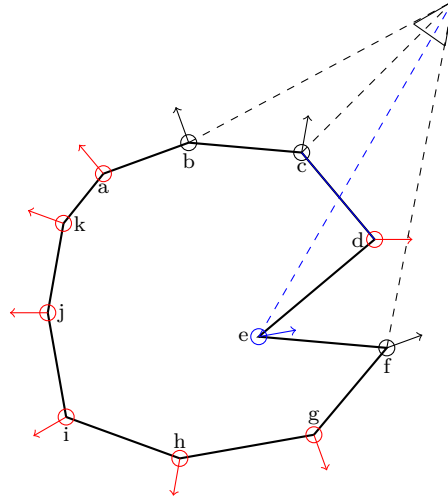


FIGURE 4.7 – Pour une position de caméra, les points intervenant dans l’optimisation de sa pose sont choisis en fonction du maillage global et des normales locales. *En rouge* : les points dont l’angle entre leur normale et la direction de visée est trop importante ne sont pas observables. *En bleu* : le rayon liant le centre optique et le point e est occulté par une maille du modèle, il n’est pas observable non plus.

On écrit donc le critère global pour chaque point projeté dans chaque pose candidate. On utilise la fonction $\phi(i, n)$ pour décrire l’observabilité d’un point i dans une acquisition n . La fonction prend une valeur 1 lorsque le point est observable et 0 sinon.

$$\operatorname{argmin} \sum_{i=1}^{N_{3D}} \sum_{n=1}^{N_V} \phi(i, n) \sum_W [\mathcal{I}_0^n(\mathbf{m}_0 + \delta) - \mathcal{I}_1^n(\mathbf{m}_1 + \Delta')]^2 \quad (4.7)$$

4.5.3 Sélection de points et pondération

Pour accélérer le processus, un échantillonnage de points peut être sélectionné. Dans ce cas, il est pertinent de sélectionner les points permettant d’apporter la meilleure information au processus. Le critère de corrélation que nous mettons en œuvre dans la reconstruction d’un point suppose que la surface est localement quasi-plane. Par conséquent, on choisira les points situés sur des zones favorisant une corrélation de bonne qualité.

Il s’agit d’un échantillonnage par courbure favorisant les faibles variations de courbures. La stratégie adoptée construit une *voxel map* avec une résolution fixée au départ. Chaque voxel est ensuite considéré. Si la distribution de l’orientation des normales des points contenus dans ce voxel est quasi-uniforme, on divise ce voxel en huit et on ajoute ces sub-divisions à la liste des voxels à considérer. Si la distribution est très disparate, ce voxel est retiré de la liste. On détermine si un voxel a une distribution uniforme ou non en observant si l’écart-type dépasse un seuil donné. Lorsque tous les voxels ont été considérés l’algorithme s’arrête, on sélectionne aléatoirement un point par voxel pour intégrer l’ensemble échantillonné.

Cette méthode nécessite le choix de deux paramètres. Le premier est la résolution initiale de la *voxel map*. Plus cette dernière sera grande, plus l’algorithme va retirer de points dans de grandes régions autour des zones de courbures non-uniformes. Au contraire lorsque ce paramètre sera petit, les zones non-uniformes seront gommées moins largement. Le second

paramètre est le seuil sur l'écart-type de la distribution d'orientation de normales. Plus ce seuil sera élevé, plus on sélectionnera de zones non planes.

Une autre adaptation de la méthode consiste à modifier la fonction $\phi(i, n)$ pour qu'elle devienne une fonction de pondération. La qualité de reconstruction d'un point par corrélation dépend de la courbure de la surface mais aussi de son orientation par rapport à la direction de visée. Plus une surface sera fronto-parallèle aux deux caméras, plus le point sera corrélé – et donc reconstruit – avec précision. Par conséquent, on choisit de réaliser la pondération en fonction de l'orientation de la normale d'un point relativement à la direction de visée de la caméra coaxiale de la vue considérée.

$$\phi(i, n) = \begin{cases} 0 & \text{si } i \text{ n'est pas observable dans } n \\ \frac{|\mathbf{n}_i \cdot \mathbf{z}_n|}{\|\mathbf{n}_i\| \times \|\mathbf{z}_n\|} & \text{si } i \text{ est observable dans } n \end{cases}$$

Le vecteur $\mathbf{n}_i$ est la normale locale au point i et le vecteur $\mathbf{z}_n$ est l'axe de visée de la caméra coaxiale à la vue n .

4.6 Évaluations

La scène que nous utilisons pour cette évaluation est la numérisation d'une personne assise sur un fauteuil. La numérisation a été réalisée en 27 séquences. Une séquence est une acquisition continue – dans un mouvement considéré quasi-continu – de vues. Au total cette manipulation représente 966 acquisitions à recaler. Cette numérisation a été réalisée par le photographe Florian Baudrain dans son atelier. Ainsi nous avons pu bénéficier d'une lumière relativement maîtrisée, bien que l'ombre de l'opérateur qui se déplace devant la scène reste un facteur perturbant pour l'acquisition et la fusion de textures. De plus l'opérateur étant un utilisateur indépendant et sans expérience de notre appareil, la manipulation n'est pas biaisée et aucune simplification dans les cas d'utilisation n'est introduite.

Nous abordons cette évaluation en deux temps. En premier lieu, nous traitons un détail de la numérisation afin d'évaluer nos méthodes de fermeture de boucle et d'optimisation. Dans un second temps nous traitons l'ensemble des données.

L'extrait exploité porte sur l'un des bras du fauteuil qui a été numérisé. Cette manipulation a été réalisée en une séquence continue. Elle a démarré à l'aplomb de l'avant du bras pour ensuite se diriger vers le dossier, descendre sur l'assise et remonter vers le haut de l'accoudoir pour fermer la séquence.

La figure 4.8 montre le résultat en nuages de points selon que l'on utilise la méthode de [Chen 92], ou la détection de fermetures de boucles que nous avons présentée. Sans réestimation, une erreur s'ajoute à chaque recalage et crée une dérive de l'estimation de pose. Par détection de cycle on permet la correction des poses du modèle afin de contraindre une correspondance entre les vues finale et initiale – recouvrantes – dans un recalage.

La figure 4.9 montre l'évolution des paramètres de pose estimés au long de la séquence. Nous comparons les résultats des séquences avec et sans fermeture. Dans le premier cas on constate bien une dérive légère de certains paramètres par rapport à l'estimation corrigée. Apparaît également l'évolution des paramètres après l'optimisation finale, permettant d'affiner le recalage global. L'influence de l'optimisation finale est visible sur la figure 4.10. On constate globalement une erreur moindre sur le début et la fin de la séquence. Il s'agit là d'un défaut

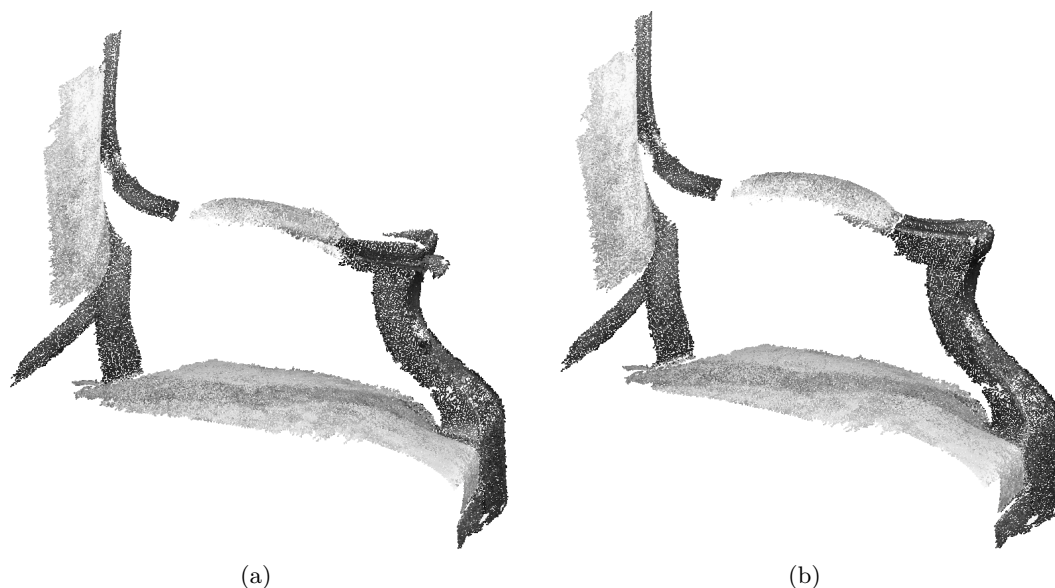


FIGURE 4.8 – Séquence du bras de fauteuil. (a) Sans fermeture de boucle, l’erreur se propage (b) Après optimisation, l’erreur a été corrigée sur l’ensemble du modèle

de la méthode en ligne. En effet, les vues sur lesquelles s’est faite la fermeture ont été mieux contraintes. Plus on s’éloigne temporellement de ces vues, plus l’effet de la contrainte semble se diluer.

Les trajectoires estimées du capteur sont présentées en figure 4.11 en projection sur plans 2D.

L’ensemble des données a été traité pour la reconstruction de la scène acquise. Le résultat est montré en figure 4.12. Le modèle est montré après traitement. Un filtrage de bruit et un échantillonnage uniforme par *voxel map* ont été appliqués afin de nettoyer le modèle et alléger la visualisation. La couleur n’a pas été corrigée. Cependant, afin d’éviter les effets de « couture » entre les prises de vues, les points les plus sombres pour chaque vue sont automatiquement supprimés.

La modélisation finale, après maillage est proposée en figure 4.13. Les figures 4.14 et 4.15 montrent des détails du modèle selon différentes représentations (points, mailles ...)

4.7 Conclusion

Dans ce chapitre nous avons présenté nos stratégies et méthodes pour la modélisation et le recalage 3D par un système tenu à la main. Nous avons présenté une approche pour la modélisation incrémentale en-ligne. Un état de l’art de la modélisation multi-vues a été présenté. Nous avons présenté les principales approches pour la reconstruction 3D utilisant des systèmes basés-vision et les approches de cartographie par graphe ainsi que des approches SLAM.

Lors de l’acquisition, notre méthode agrège les vues au modèle courant en construisant un graphe de relations. Lors de la détection de fermetures de boucles, une optimisation partielle

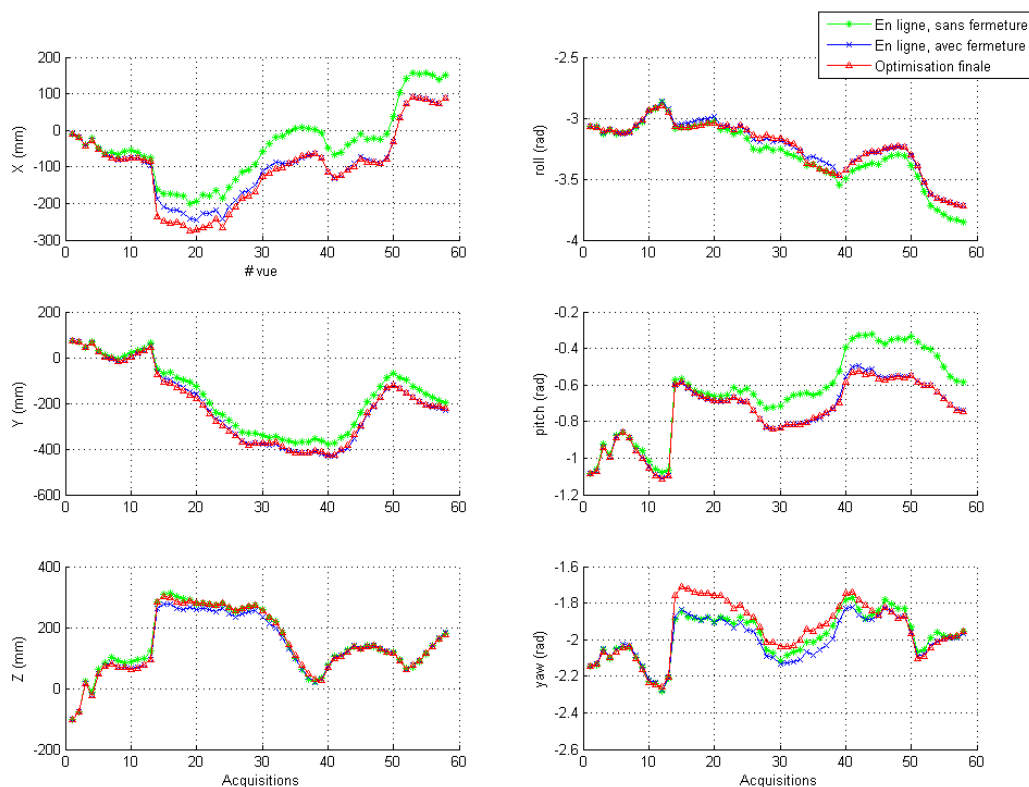


FIGURE 4.9 – Variations des paramètres de pose estimés à chaque acquisition.

est proposée. En fin de processus une optimisation globale a été proposée utilisant un critère de reprojection dans les images.

Nous avons évalué notre approche en réalisant une numérisation complète. Cela nous a permis de montrer la capacité de nos méthodes à réaliser un recalage multi-vues et à corriger la dérive apportée par une modélisation incrémentale simple.

Il convient cependant de caractériser les performances de notre approche au moyen d'une étude quantitative des résultats. Le prochain chapitre s'attache donc à proposer une évaluation quantitative de notre système.

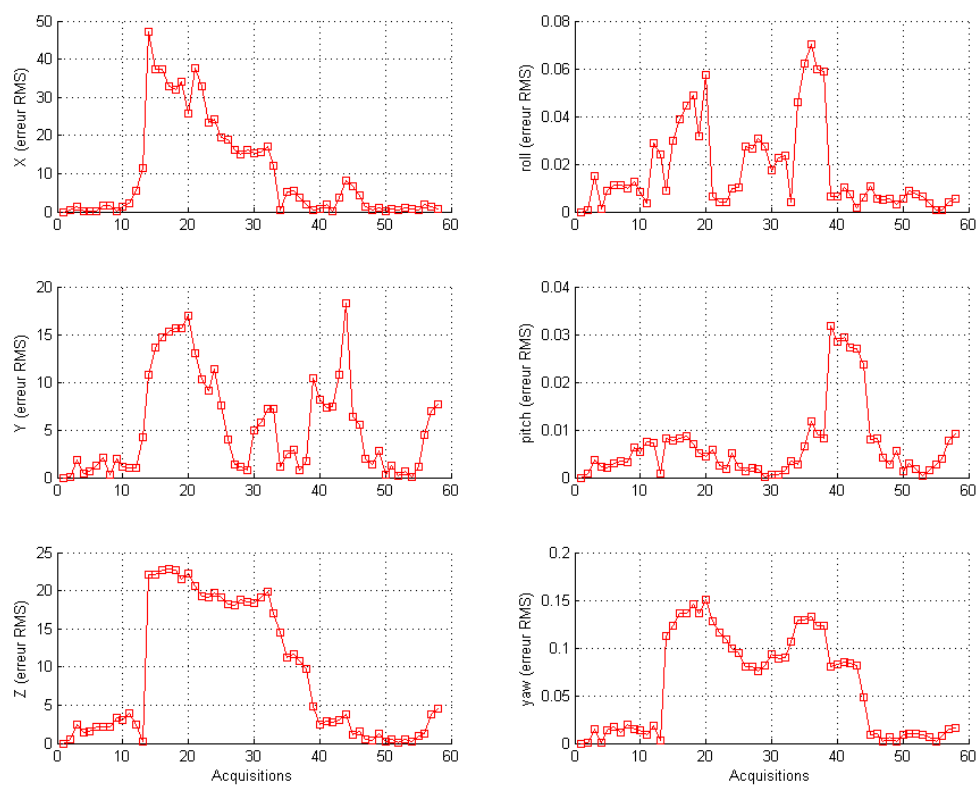
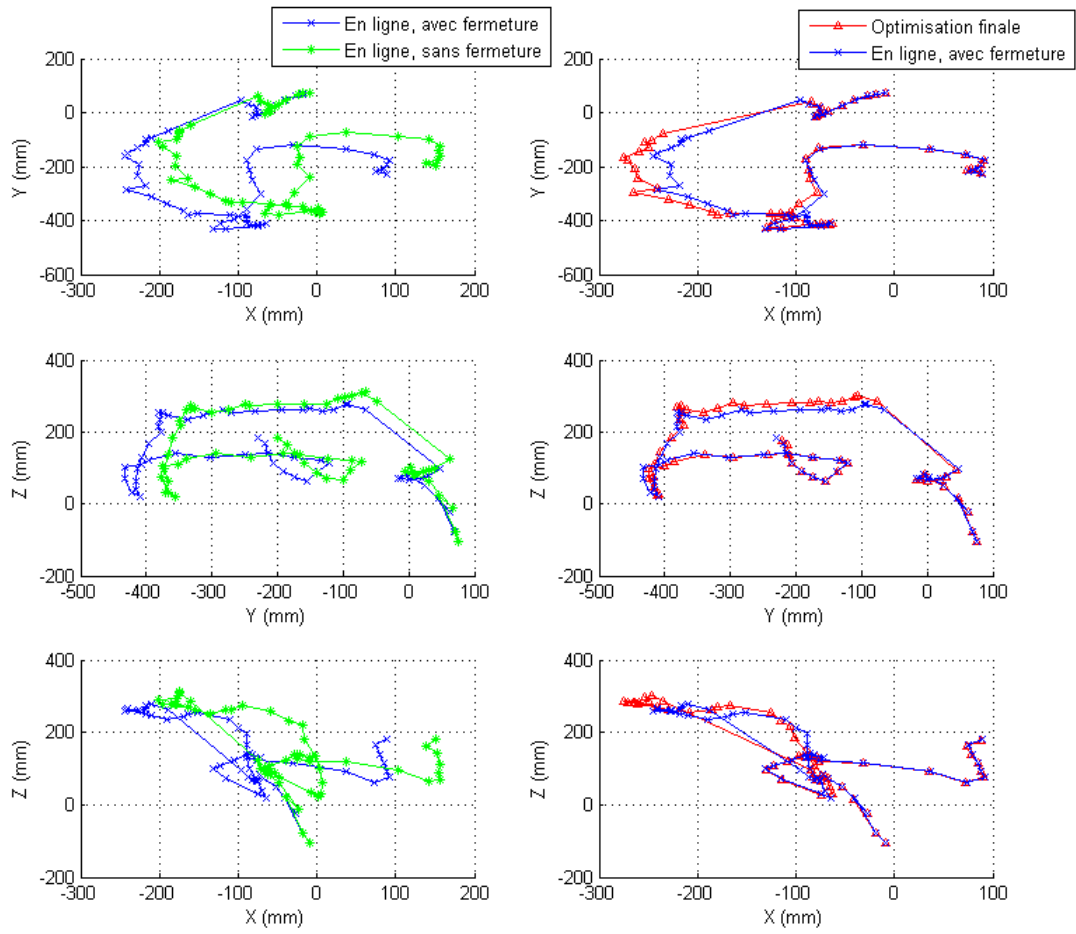


FIGURE 4.10 – Correction apportée par l'optimisation finale.

FIGURE 4.11 – Trajectoires du capteur projetées sur les plans (X, Y) , (Y, Z) , (X, Z)

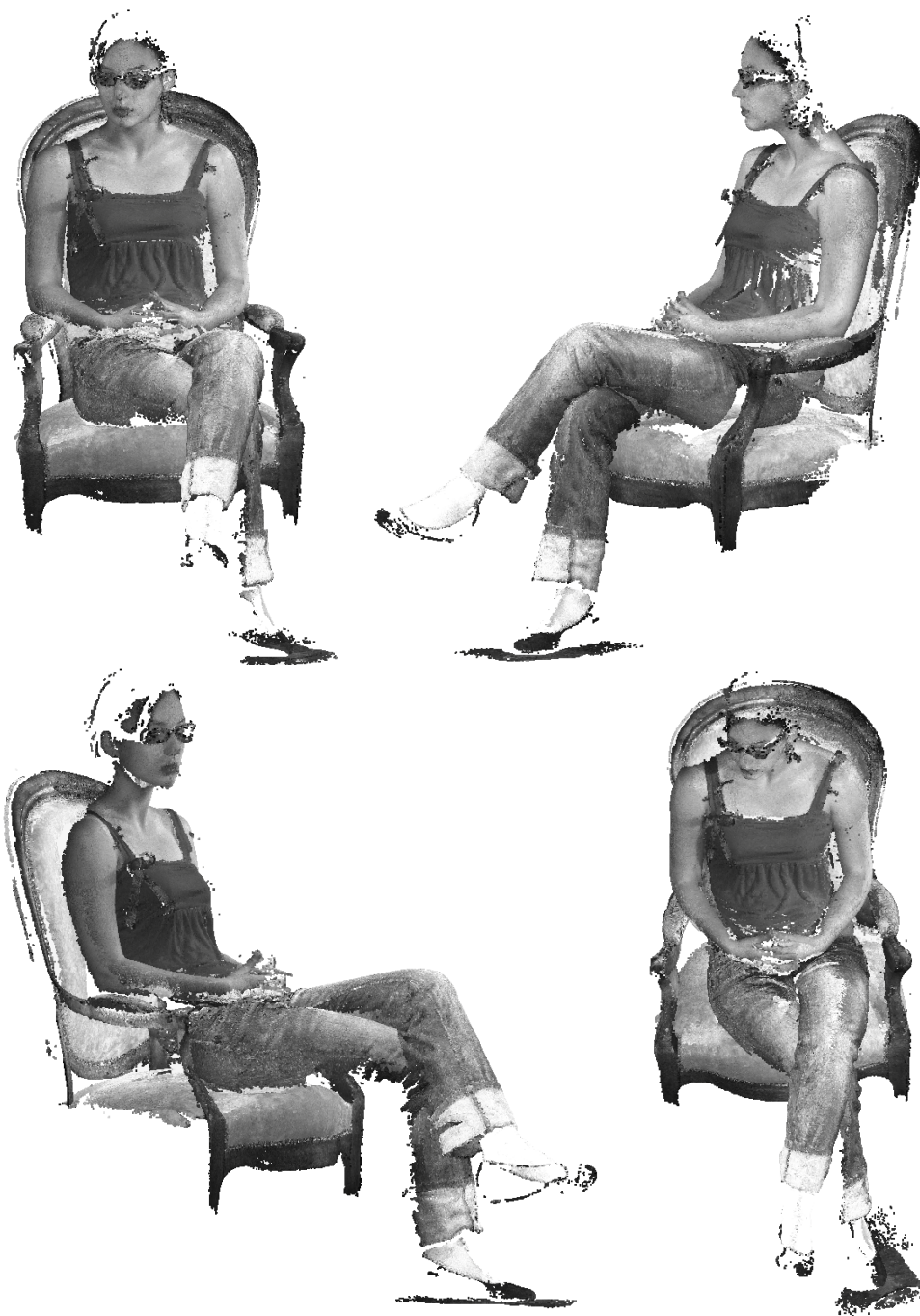


FIGURE 4.12 – Résultat sur la séquence complète (nuage de point en niveaux de gris)

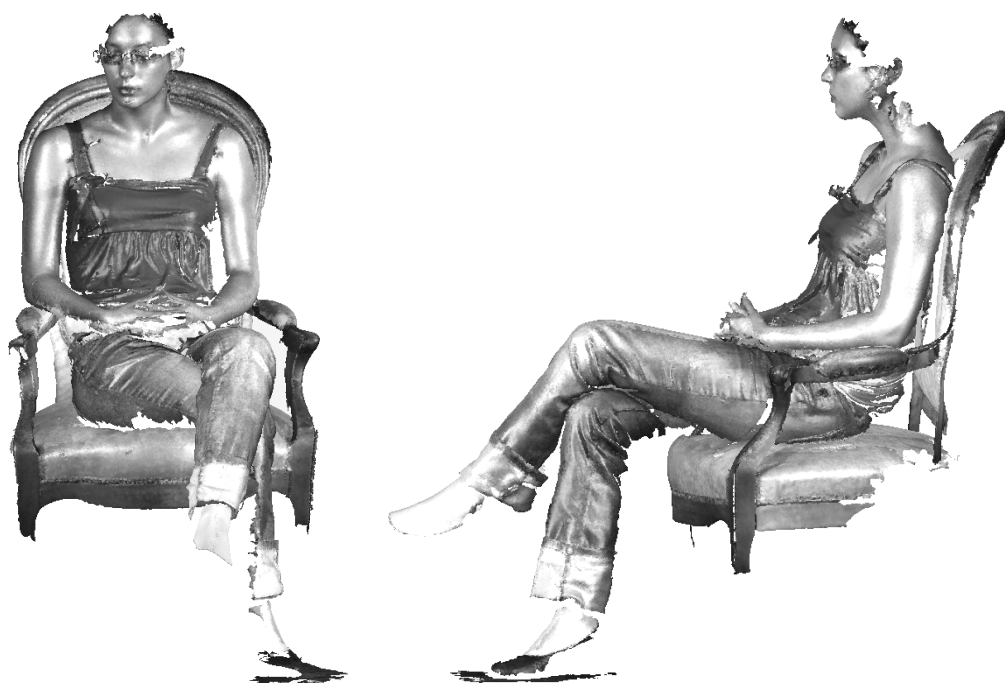


FIGURE 4.13 – Résultat sur la séquence complète (maillage en niveaux de gris)



FIGURE 4.14 – Détail des plis du vêtement (faces et arêtes du maillage)

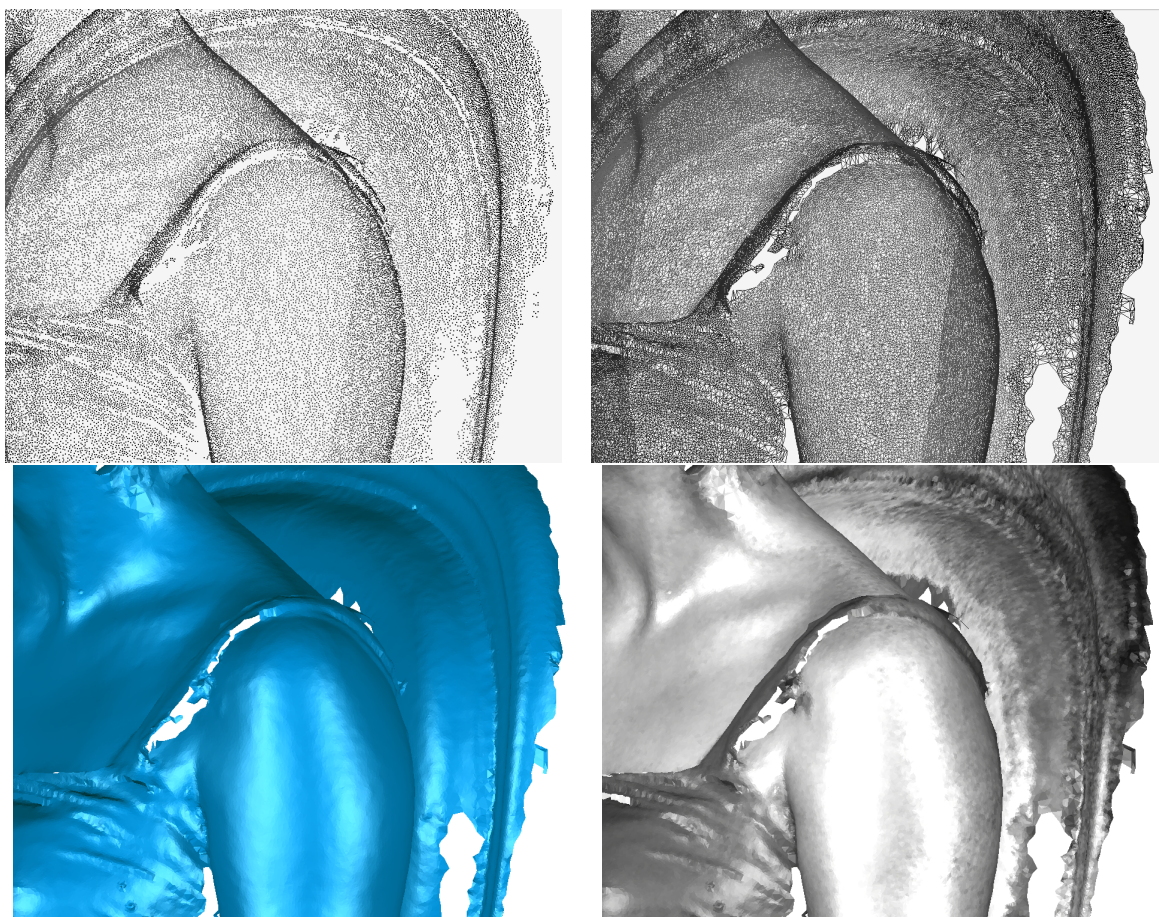


FIGURE 4.15 – Détail de l'épaule et du dossier du fauteil (points, arêtes, faces et faces en niveaux de gris)

Évaluations quantitatives

Table des matières

5.1	Protocole expérimental	114
5.2	Cylindre étalon	115
5.3	Statue	116
5.4	Pièce emboutie	117
5.4.1	Comparaisons au modèle CAO	119
5.4.2	Comparaison avec le modèle B	119
5.5	Lien avec la qualité de reconstruction	120
5.6	Conclusion	122

LES précédents chapitres nous ont permis de présenter notre capteur ainsi que les stratégies et méthodes pour la modélisation 3D et le recalage que nous avons mises en œuvre. Nous avons proposé des évaluations de ces méthodes afin de démontrer leur capacité à réaliser la modélisation d’une scène.

Dans ce chapitre nous proposons l’évaluation de notre capteur et de nos méthodes pour deux types d’analyses, qualitative et quantitative. De plus, nous avons fait numériser nos objets de test par des systèmes commerciaux faisant référence dans la communauté afin de comparer les résultats de notre système de numérisation 3D à des systèmes existants.

Dans un premier temps nous présenterons notre protocole expérimental et les systèmes utilisés, puis nous aborderons les manipulations en commençant par une comparaison sur une pièce étalon. Nous présenterons ensuite les évaluations à proprement parler sur deux pièces, un buste sculpté et une pièce emboutie.

5.1 Protocole expérimental

Pour ces évaluations nous avons scanné plusieurs objets avec trois systèmes : notre capteur tenu à la main, et deux scanners commerciaux largement utilisés dans les communautés industrielle et scientifique, que nous appellerons *A* et *B* dans la suite de ce chapitre. Les systèmes commerciaux sont des appareils fixes utilisant des technologies optiques pour la reconstruction 3D. Nous n’identifions volontairement pas ces systèmes afin de ne pas biaiser le propos vers un caractère commercial.

Les informations techniques concernant les trois systèmes sont fournis par le tableau 5.1.

	Système A	Système B	Notre système
Déplacement	Fixe	Fixe	Tenu à la main
Technique	Triangulation laser	Lumière structurée	Lumière structurée
Recalage	Manuel	Photogrammetrie	Automatique
Poids (kg)	15	7.4	1.8
Dimensions (mm)	$221 \times 412 \times 282$	$490 \times 300 \times 170$	$220 \times 240 \times 90$
Distance focale (mm)	14	16	8
Distance de travail¹ (m)	0.6 – 1.2	0.7 – 2	0.35 – 0.5
Volume de mesure (mm)	de $111 \times 83 \times 40$	de $135 \times 108 \times 108$	de 10cm^3 à 1m^3
	à $1196 \times 897 \times 800$	à $1700 \times 1360 \times 1360$	NC
Précision (X,Y,Z) (mm)	$\pm (0.22, 0.16, 0.10)$	NC	01
Espacement des points (mm)	NC	0.08 – 1.0	0.3
Bruit de mesure (mm)	NC	0.002 – 0.02	0.1

TABLE 5.1 – Spécifications des systèmes (les spécifications des systèmes commerciaux sont fournies d’après les données techniques des constructeurs, certaines données n’ont pas été communiquées)

1. Nécessite un changement de focale pour couvrir l’ensemble du champ de distances.

Dans ces évaluations nous avons utilisé le logiciel Geomagic® Qualify® v12 pour les comparaisons entre les numérisations et les modèles CAO. Le système *A* ne fournit pas de méthode de recalage, nous avons donc eu recours au logiciel Geomagic® Studio® v12 pour l'alignement des nuages de points.

Le Système *A* est fixe et utilise la triangulation laser pour la mesure 3D. Il s'agit d'un appareil lourd et difficile à manipuler. Le système *B* est également fixe, mais plus léger que le système *A*. Il utilise la projection de franges pour la mesure 3D et est couplé à un système photogrammétrique pour la localisation. Cela requiert de placer préalablement plusieurs cibles dans la scène observée. Les deux systèmes ont des optiques modifiables pour la modification de la longueur focale.

Les modes opératoires et les spécifications des scanners *A* et *B* ne sont pas vraiment comparables à ceux de notre système. Cependant ils ont été choisis pour les précisions de numérisation qu'ils annoncent et parce qu'il s'agit de systèmes largement utilisés dans les communautés scientifique et industrielle.

5.2 Cylindre étalon

Pour avoir une meilleure idée des précisions relatives des systèmes, nous avons commencé par scanner un cylindre étalon. Cette pièce est une pièce de précision utilisée pour l'étalonnage des micromètres. Le diamètre interne du cylindre est connu (70.004 mm) avec une précision de 1 μm .

L'évaluation est faite sur une seule acquisition. Pour chaque système, le nuage de points 3D qu'il génère est recalé au modèle CAO théorique d'un cylindre de diamètre 70.004 mm à l'aide du logiciel Geomagic® Qualify® v12. Les points sont projetés orthogonalement sur la surface et la distance de projection – l'erreur – est mesurée. Les analyses sont basées sur la moyenne et l'écart-type des distances d'erreur.

Le tableau 5.2 montre les résultats de ces tests.

	Système A	Système B	Notre système
Moyenne (mm)	-0.00115	0.00003	-0.00115
Écart-type (mm)	0.02764	0.00508	0.02480
Erreur max. (mm)	0.10238	0.02162	0.08649
Erreur min. (mm)	-0.09151	-0.02569	-0.09392

TABLE 5.2 – Résultats d'évaluation : numérisation avec les trois systèmes d'un étalon de dimensions connues

Ce premier test n'a pas pour but de réaliser une comparaison et un classement des scanners mais doit être utilisé comme base pour l'évaluation. Il tend à fournir une référence absolue concernant le processus de reconstruction 3D pour chaque appareil. On notera que notre système se place dans la même gamme de précision que le système *A* tandis que le système *B* fournit des résultats plus précis.

5.3 Statue

Pour la première évaluation nous avons numérisé une sculpture gallo-romaine (Figure 5.1). Cet objet est constitué de larges parties lisses (visage, face de support ...) et de zones gauches (la cape, principalement).



FIGURE 5.1 – Statue utilisée pour notre évaluation

Au démarrage du processus, la statue est posée sur sa face de support. On débute l'acquisition face au visage puis en tournant vers notre gauche. L'objet a été tourné selon son axe vertical plusieurs fois durant le processus afin d'acquérir le « cylindre » de la tête. Après un tour complet, nous avons numérisé le dessus de la tête. On a ensuite déplacé la statue pour la positionner visage vers le haut. Nous avons fait l'acquisition de la face de support en commençant une séquence depuis le bas du visage. Nous avons complété cette face en positionnant l'objet visage vers le bas et en recommençant une acquisition depuis l'arrière de la tête vers la face de support.

Le résultat de la numérisation est proposé en Figure 5.2.

Le modèle est bruité mais déjà reconnaissable. Il nécessite maintenant d'être finalisé. Nous avons affiné le recalage du modèle et nettoyé les points aberrants, puis réalisé le maillage de la surface. Le résultat final apparaît en figure 5.3. La texture fournie par les images a également été appliquée au modèle final.

Le mode opératoire de notre système permet une manipulation facile ce qui offre un grand intérêt pour ce type d'objet. La statue a été numérisée à l'aide des systèmes A et B mais la manipulation d'appareils fixes ajoute des difficultés lors de l'acquisition de détails partiellement cachés, tels que ceux sur la cape. La figure 5.4 illustre un détail de la cape où

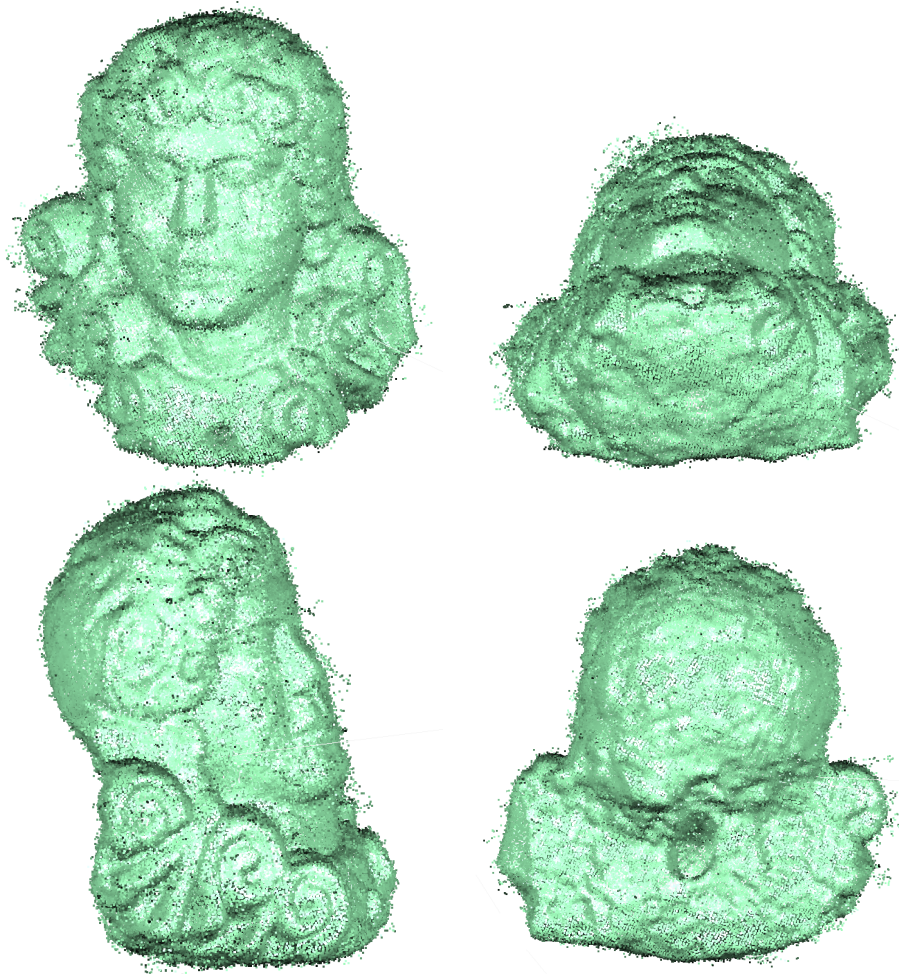


FIGURE 5.2 – Résultat de numérisation après la modélisation incrémentale et avant la finalisation.

l'on peut observer l'apparition de manques sur les nuages obtenus par les systèmes fixes (ici le système A).

5.4 Pièce emboutie

Pour une analyse plus quantitative, nous avons utilisé une pièce métallique issue d'un procédé d'emboutissage. Nous disposons d'un modèle CAO de cet objet. Pour éviter les réflexions spéculaires, l'objet a été matifié.

L'objet n'est numérisé que sur sa face supérieure. La plaque emboutie étant fine, l'intérêt de numériser la face opposée ainsi que les tranches est discutable si l'on considère la résolution et la précision des systèmes dont nous disposons.



FIGURE 5.3 – Modélisation finalisée et texturée de la statue à l'aide de notre système

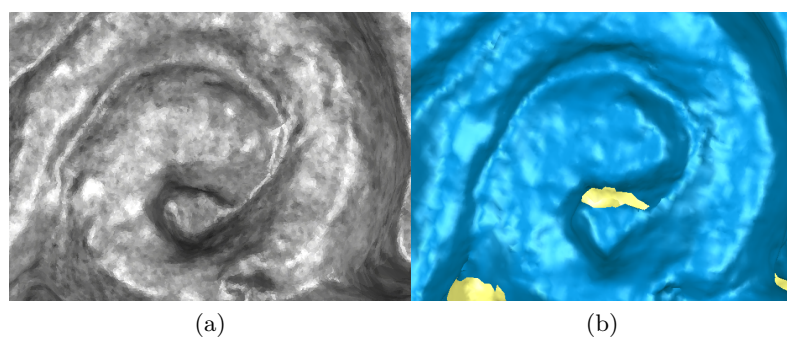


FIGURE 5.4 – Detail de l'objet. Dans les zones gauches, certaines faces peuvent être difficiles à observer du fait des occultations. (a) L'observation de surfaces gauches avec notre système est facilitée par son mode de manipulation. (b) Numérisation par le système *A*, avec une manipulation plus difficile, des manques d'information peuvent se produire.



FIGURE 5.5 – Pièce métallique emboutie utilisée pour nos tests.

5.4.1 Comparaisons au modèle CAO

Dans notre plan de tests, nous avons tout d’abord comparé les modèles issus de notre système et des systèmes *A* et *B* au modèle CAO de l’objet. Du fait des efforts mécaniques, particulièrement au relâchement du serre-flanc, l’objet a été largement déformé et ne correspond plus au modèle théorique. Nous avons recalé nos modèles de test à la référence CAO selon la même zone d’intérêt, favorisant la zone centrale de l’embouti, supposée moins déformée. Les résultats sont donnés en figure 5.6.

La métrique utilisée est basée sur la distance orthogonale directe d’un point numérisé à la surface de référence. La figure 5.7 propose la distribution de l’erreur pour chaque modèle.

Ce test nous permet de vérifier s’il existe un biais entre les différentes mesures du même objet. En comparant les résultats, notre première observation est la similitude des cartes d’erreur. Avec ce test, on voit également que notre méthode de comparaison est stable. Des données similaires conduisent à des recalages et projections similaires, en considérant que nous avons utilisé une référence de recalage commune. On peut donc utiliser cette méthode de comparaison pour nos prochains tests.

5.4.2 Comparaison avec le modèle *B*

L’évaluation suivante se concentre sur l’impact de l’étape de finalisation sur la précision. La surface CAO étant une mauvaise vérité terrain, nous avons décidé d’utiliser le modèle issu du système *B* – apparaissant comme le plus précis d’après les spécifications constructeur et le test du cylindre – comme référence.

La figure 5.8 compare la mesure 3D issue de notre capteur avant l’étape de post-traitement (recalage global et filtrage) et après cette dernière. Les nuages de points sont projetés sur le modèle de référence issu du système *B*.

Dans le nuage source, une grande quantité de bruit est présente et apparaît comme un phénomène oscillant d’un côté à l’autre de la surface moyenne. Une dérive importante est observable aux extrémités du modèle. Un effet de bordure apparaît, apportant une mesure largement biaisée dans ce type de zone.

Après l’étape de finalisation, le bruit a été largement réduit, au point qu’il n’est plus visible avec notre distribution de couleurs. Au regard de l’histogramme, on note que la dérive

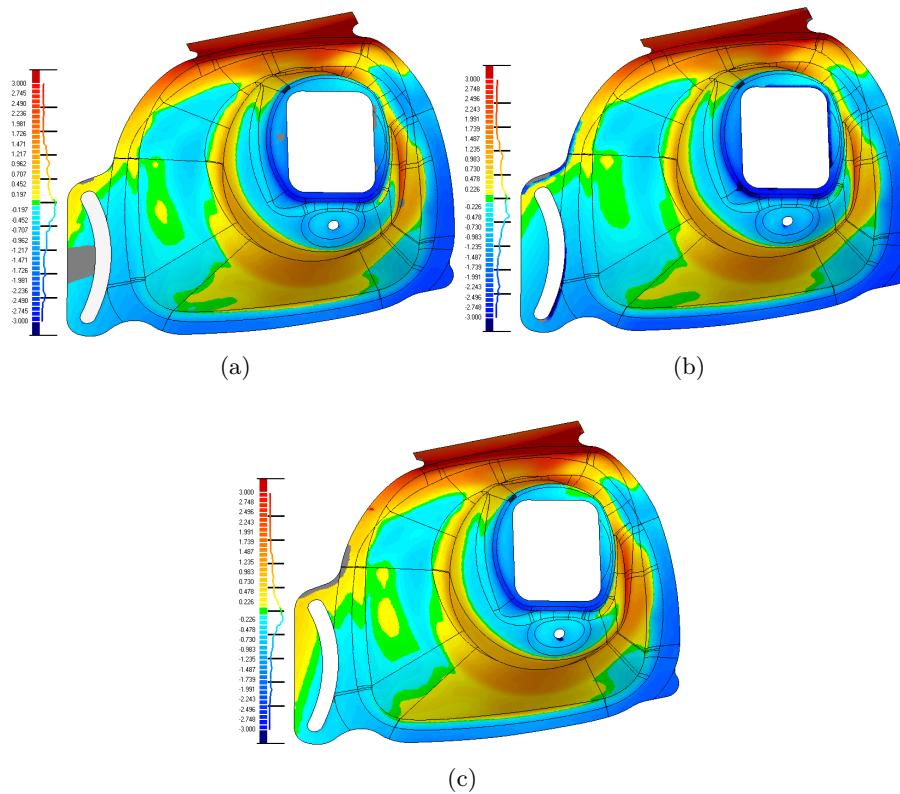


FIGURE 5.6 – Comparaisons des modèles numérisés à la surface théorique. (a) Système A (b) Système B (c) Notre système tenu à la main

apportée par les méthodes de recalage en ligne a été grandement réduite par l'étape de recalage de précision. Le modèle est désormais plus comparable au modèle B. On notera également que les effets aux bordures, s'ils ont été réduits, demeurent.

Les erreurs pour ces deux étapes sont données dans le tableau 5.3.

	Avant finalisation	Après finalisation
Erreur moyenne (mm)	-0.004	-0.001
Écart-type (mm)	0.373	0.125

TABLE 5.3 – Erreurs lors de la comparaison avec le modèle B

5.5 Lien avec la qualité de reconstruction

Lors de la reconstruction 3D d'un point, le résidu de la fonction de corrélation est une indication de la qualité du point. Un point disposant d'un mauvais score de corrélation est de mauvaise qualité. L'inverse n'est pas nécessairement toujours vrai – dans le cas d'ambiguïtés image par exemple – mais ce critère reste une bonne discrimination malgré tout.

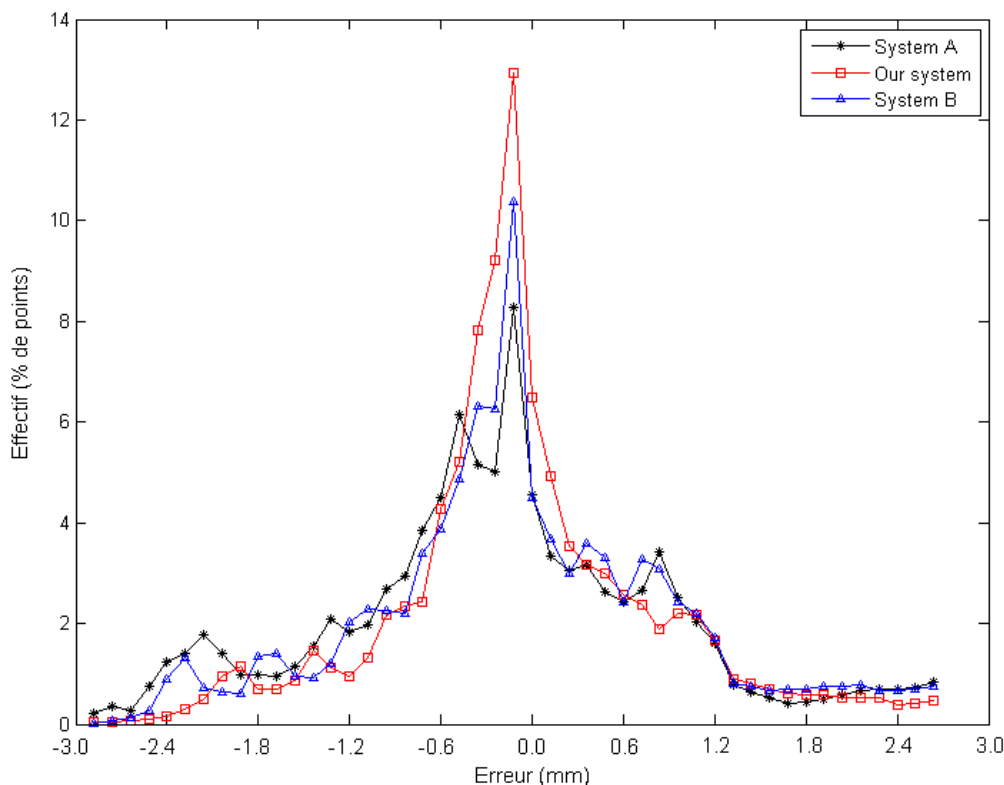


FIGURE 5.7 – Distribution de l'erreur entre les modèles 3D et la surface CAO

Pour terminer nos évaluations nous avons souhaité vérifier s'il existait effectivement un lien entre le score nous permettant de jauger la qualité de reconstruction et l'erreur effective finale.

L'erreur finale est celle que nous obtenons par comparaison au modèle CAO dans le logiciel Geomagic® Qualify® v12, comme nous l'avons fait au cours de ce chapitre. Nous avons réalisé une comparaison sur une acquisition du cylindre utilisé en section (5.2). On dispose, pour chaque point, du score de reconstruction associé.

La figure 5.9 nous permet de comparer sur une acquisition la carte des scores de reconstruction et la carte des erreurs par comparaison au modèle théorique.

Lors de la reconstruction, les zones de courbure du cylindre intérieur les moins fronto-parallèles aux caméras ont un score de corrélation plus faible. En effet, dans ces zones la déformation de la fenêtre de corrélation entre les deux caméras est importante. Elle est donc plus difficile à estimer par optimisation locale. On observe que ces zones sont effectivement sujettes à une erreur plus importante que la majorité du nuage lors de la comparaison au modèle théorique. Les points sur les arêtes du modèle sont également de plus mauvaise qualité, à la fois dans la carte de scores et la carte d'erreur. Notre modèle de corrélation suppose que le voisinage d'un point corrélié est quasi-plan. Cette hypothèse devient fautive à l'approche d'arêtes et conduit à une corrélation de moindre qualité.

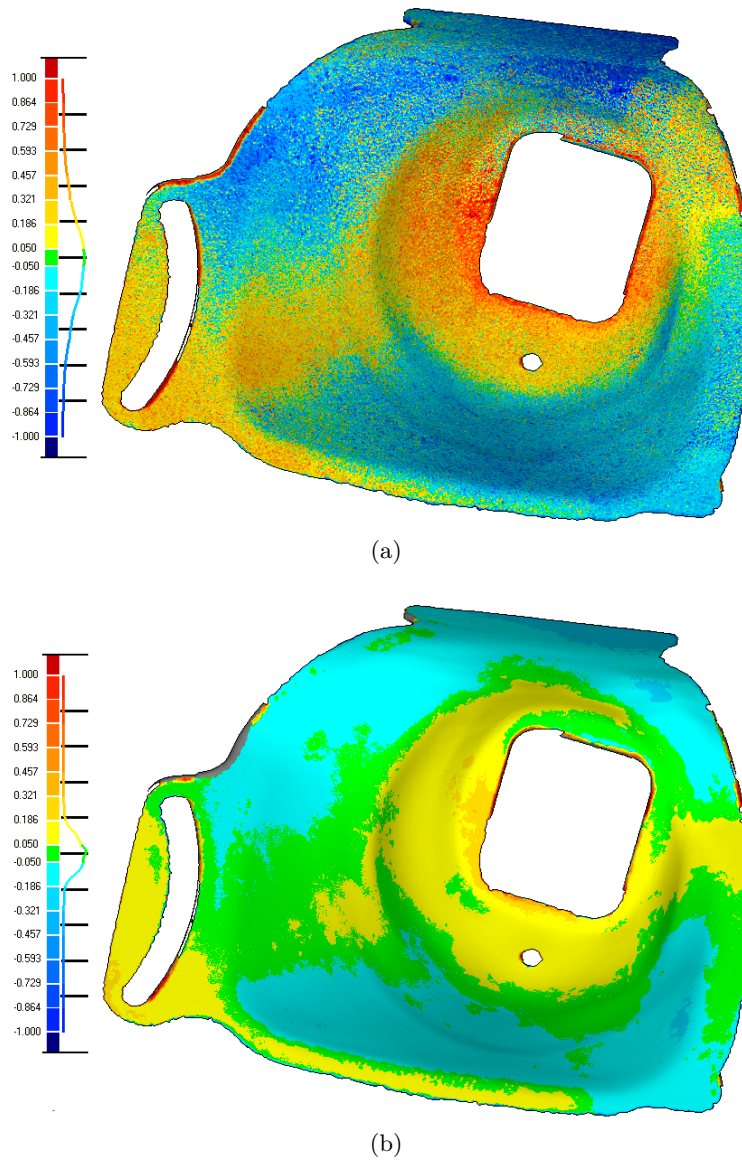


FIGURE 5.8 – Comparaison de nos résultats avec le scan *B*. (a) Avant finalisation. (b) Après finalisation

Cette analyse nous montre effectivement une certaine corrélation entre le score de reconstruction d'une acquisition et l'erreur « réelle » que l'on observe sur le nuage final.

5.6 Conclusion

Dans ce chapitre, nous avons évalué notre système de numérisation 3D basé vision tenu à la main et les méthodes de recalage que nous avons mises en œuvre dans cette étude. Nous avons évalué les performances sur plusieurs objets de test pour couvrir plus largement les applications de numérisation 3D. Nous avons basé nos évaluations sur les résultats obtenus en numérisant un même objet par des systèmes commerciaux connus et couramment utilisés.

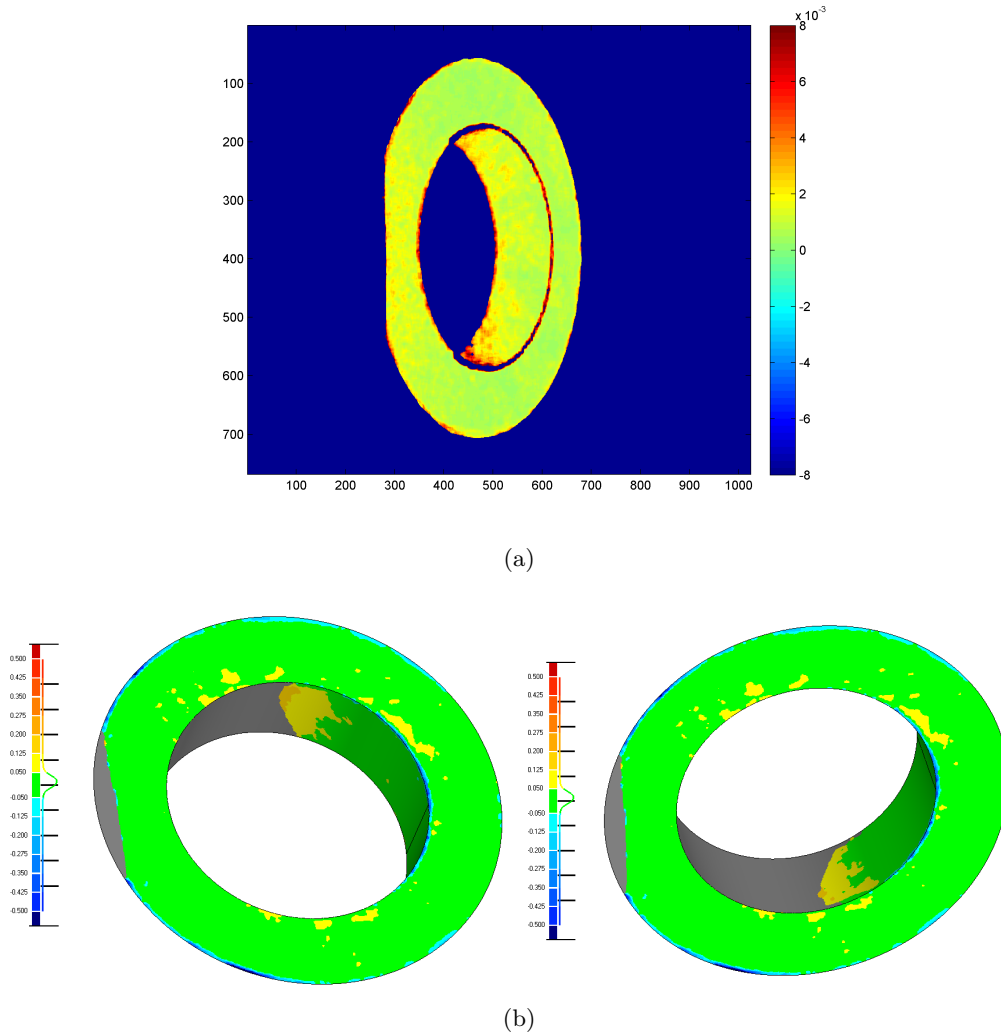


FIGURE 5.9 – Erreur finale et scores de reconstruction sur une seule vue. Les zones grises sont les parties du modèle CAO où aucun point ne se projette. (a) Carte de scores de reconstruction. (b) Carte d'erreurs par projection sur le modèle CAO selon deux points de vue différents afin de voir l'ensemble du cylindre intérieur.

Nos tests tendent à prouver que le système que nous proposons permet d'atteindre les performances et la qualité requise pour des applications classiques en termes qualitatif, de précision et d'utilisabilité mais avec un mode opératoire facilité. Cela valide que notre système répond à notre objectif initial : un système peu onéreux, de bonne précision et qui permet un processus de modélisation globalement facilité.

Nous avons de plus établi un lien entre la qualité finale du modèle et la qualité de reconstruction mesurée dans notre processus. Afin d'améliorer globalement la qualité du modèle, il est donc possible de se baser sur la cartographie des scores de reconstruction pour choisir des zones où l'acquisition de nouvelles vues permet un apport significatif d'information. Le prochain chapitre évoque le problème d'analyse et décision en vue de guider l'utilisateur lors de la numérisation.

Vers l'analyse de qualité pour la sélection de point de vue

Table des matières

6.1	Évolution de la qualité d'un modèle	126
6.2	Sélection de point de vue	128
6.2.1	Méthodes notables	128
6.2.2	Fonction d'utilité d'un point de vue	129
6.3	Conclusions	133

AFIN d'améliorer la qualité du modèle final, il est nécessaire d'avoir acquis un nombre suffisant de vues à des positions pertinentes. Un point de vue pertinent est un point de vue depuis lequel l'apport d'information de la mesure 3D est important au regard de l'information déjà disponible dans le modèle courant.

Nous avons vu précédemment (section 5.5) qu'il existe une corrélation entre le score de corrélation utilisé pour la reconstruction et la qualité finale du modèle. Au cours de ce chapitre nous commençons par étendre cette analyse afin d'étudier comment une nouvelle vue permet l'amélioration de la qualité globale d'un modèle. Nous évoquerons ensuite des pistes d'étude possible afin de réaliser une prédiction de vue pour le guidage de l'opérateur. Le but de telles méthodes est de suggérer à l'opérateur un prochain point d'acquisition pour permettre la complétion ou l'amélioration du modèle courant.

Cette étude a pour but d'ouvrir la voie à de futurs travaux permettant l'intégration d'un système de guidage à notre processus de modélisation quand le système est tenu à la main ; il pourrait aussi conduire à un processus de numérisation 3D automatique avec une version embarquée sur un robot.

6.1 Évolution de la qualité d'un modèle

Afin de réaliser la planification avec prise en compte de la qualité du modèle, nous avons souhaité étendre l'analyse faite au chapitre 5.5 en observant l'évolution du critère de reconstruction quand de nouvelles vues sont ajoutées. Dans cette approche, nous n'utilisons pas les nouvelles acquisitions pour réaliser une reconstruction multi-vues afin de gagner en précision. Nous nous sommes orientés vers l'utilisation d'une *voxel map* : pour chaque voxel on ne considère que le point offrant le meilleur score.

La figure 6.1 montre la pièce que nous utilisons dans l'analyse qui suit. Il s'agit d'un moule en métal. Il a été matifié préalablement.

Nous avons réalisé quatre acquisitions de l'objet en parcourant visuellement le galbe central. Pour chaque acquisition le score de reconstruction normalisé est attaché à chaque point. La figure 6.2 montre la cartographie des scores de reconstruction pour chaque acquisition et pour les différentes étapes de la modélisation incrémentale.

La carte de couleur choisie tend vers le vert pour les meilleurs scores et vers le rouge pour les moins bons. L'échelle de scores – donc l'échelle de couleur – est identique entre chaque image afin de permettre une comparaison directe des résultats.

La première acquisition propose des points de bonne qualité mais présente des manques importants sur les zones courbes. En effet, il s'agit d'un point de vue fronto-parallèle au plan dominant du moule. De plus, peu de points sont disponibles aux arêtes du modèle. Le point de vue ne permet pas d'observer suffisamment les deux faces constituant le voisinage des points sur la plupart des arêtes. Par conséquent ils n'ont pas, ou peu, été mesurés.

La seconde acquisition s'est attachée à observer le galbe central, depuis le bas de la pièce (dans le sens de l'image en figure 6.2). Les arêtes étant mieux observables elles ont été mesurées. L'aggrégation avec la première vue n'apporte globalement pas d'amélioration mais permet essentiellement de compléter des manques.

La troisième acquisition a été prise depuis le haut de la pièce (toujours dans le sens de l'image) permettant de compléter le galbe central ainsi que la rainure principale. L'aggrégation de cette vue avec le modèle précédent permet un gain de qualité. Si l'on observe, par exemple, le galbe central, on note qu'une partie de l'imprécision de mesure dans chaque vue

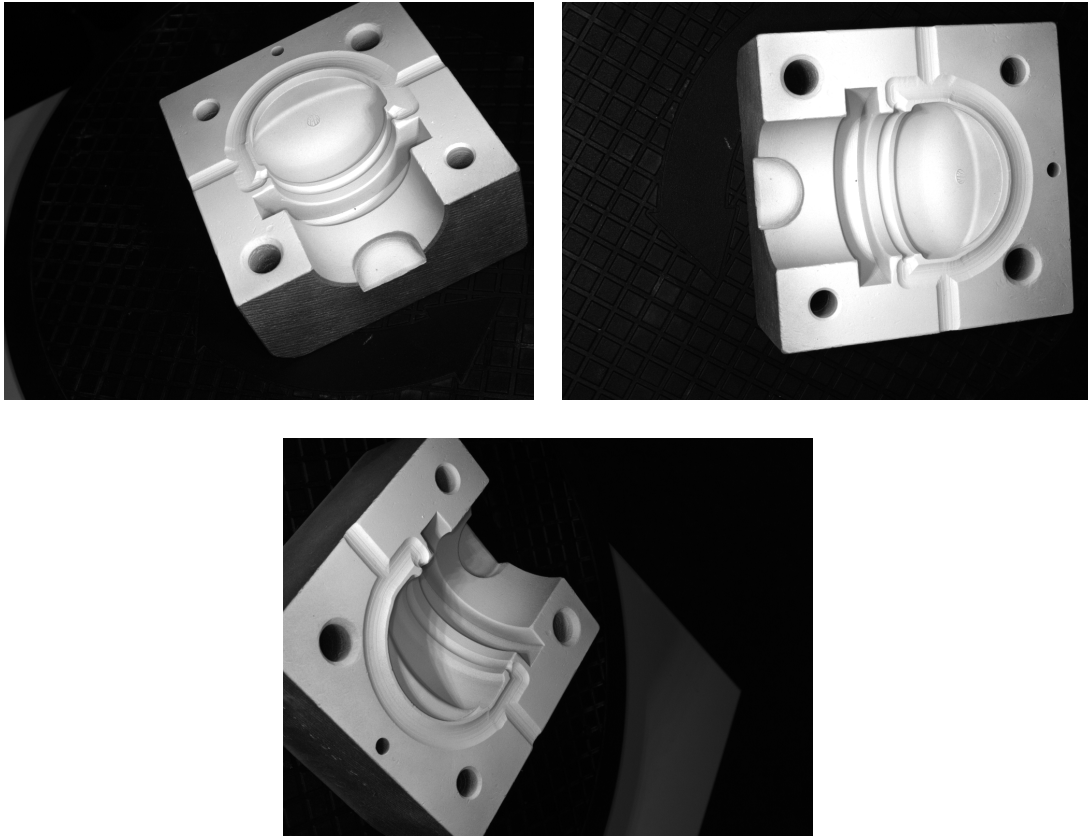


FIGURE 6.1 – Visuels de la pièce « Moule »

est compensée par les autres lorsque des points de meilleure qualité sont disponibles. On observe qu'il subsiste des points à faible score dans des zones où l'on a obtenu un apport de qualité. Cela se produit quand le point a été effectivement mal mesuré, et qu'il a été reconstruit avec un écart important par rapport à la surface réelle. Ce type de points se retrouvent donc généralement isolés dans leurs voxels : il s'agit alors d'un bruit de surface qui pourrait être filtré.

La quatrième vue est issue d'un point de vue situé à gauche dans le sens de nos images. Cette acquisition permet de terminer le galbe central et le cylindre du bas de la pièce. Le modèle complet, fusionnant les quatre acquisitions nous permet de constater l'apport global d'information.

La démarche que nous avons menée avec cette modélisation incrémentale pourrait être applicable à une manipulation à plus grande échelle. Nous avons cherché, à la fois, à combler les manques du modèle et à trouver des points de vue offrant des conditions plus favorables pour la reconstruction des points de mauvaise qualité du modèle courant. Cette approche empirique, ici conduite par un opérateur expérimenté, peut être rendue plus facile et accessible à tout utilisateur par un système de planification de point de vue automatique.

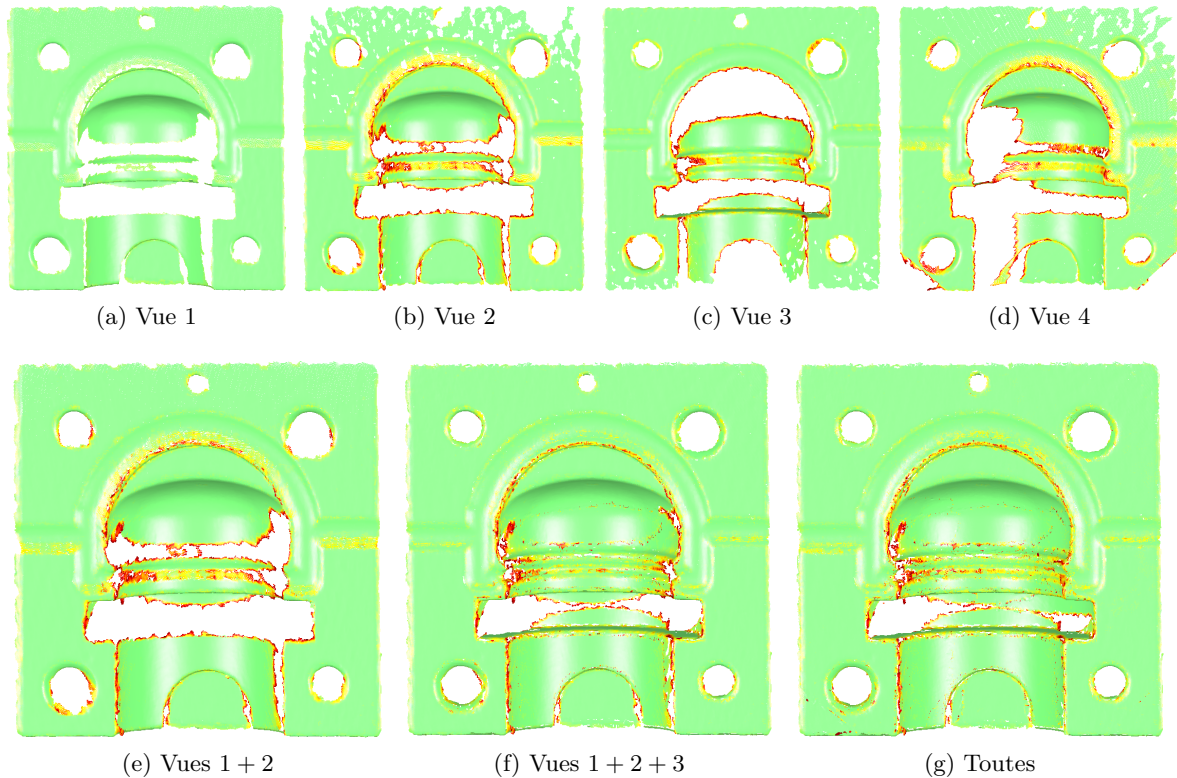


FIGURE 6.2 – Evolution de la carte de scores par aggrégation successive des vues. En direction du vert : amélioration du score. En direction du rouge : dégradation du score.

6.2 Sélection de point de vue

Le problème de la sélection de point de vue pour la modélisation 3D cherche à déterminer une pose de capteur depuis laquelle l'acquisition réalisée sera pertinente. La résolution de ce problème sera différente selon la définition que l'on donne pour le concept d'acquisition pertinente. L'application joue aussi un rôle important dans les possibilités de résolution. En effet, l'application amène des contraintes sur le mouvement et les positions atteignables à un instant donné. Par exemple, un robot sera limité par la complexité des trajectoires, en particulier s'il est non holonome, et aussi par la puissance de ses moteurs. L'encombrement de l'espace joue aussi un rôle important rendant certaines positions non atteignables.

Nous présentons ici une revue de quelques méthodes notables et une approche particulière tenant compte de la qualité de reconstruction présentée plus haut.

6.2.1 Méthodes notables

Le problème de sélection de point de vue a été très étudié, principalement, pour des applications robotiques. On parle de planification perceptuelle, ou problème de *Next Best View*. Le robot cherche à trouver une prochaine position ou une nouvelle trajectoire, dans des tâches d'exploration par exemple.

González-Baños *et al.* [González-Baños 00] utilisent l'approche de la *galerie d'art* – combien de gardiens faut-il et où les placer pour qu'ils puissent surveiller l'intégralité d'une galerie

d'art – pour résoudre ce problème. Au moyen d'une approche probabiliste, ils déterminent le nombre de capteurs et leurs positions afin de couvrir l'ensemble d'une carte polygonale 2D. Cette méthode a été étendue par [Nüchter 03], toujours pour la reconstruction d'environnements intérieurs pour la navigation 2D de robots mobiles.

D'autres auteurs ont abordé le problème du *Next Best View* pour des applications à la modélisation 3D. Une approche commune consiste à représenter l'espace des positions possibles pour le capteur comme une sphère discrète centrée sur l'objet modélisé [García 98]. La sphère est discrétisée en facettes triangulaires uniformes. Pour chaque facette, un score est attribué, basé sur la quantité d'information « vue » par cette face. La sélection de point de vue est alors faite en fonction de ces scores et de la quantité de mouvement à produire pour positionner le capteur au centre de la facette candidate comme meilleur point de vue. Cette méthode a été étendue par [Wong 99] et permet de mesurer la performance de l'algorithme. Parmi les travaux récents, [Pan 09] propose également une représentation d'espace par une discrétisation de scène et propose une intégration dans un système temps-réel de modélisation par une seule caméra.

Une autre approche courante pour les applications de modélisation est la représentation de l'espace par une grille d'occupation. Banta *et al.* [Banta 95] introduisent une méthode où la grille d'occupation permet la représentation des occultations et réalise une planification pour atteindre les zones cachées. Les travaux de [Sequeira 96] puis de [Klein 00] sont également fondés sur la détection d'occultations et l'intégration de leur algorithme sur une carte graphique pour accélérer le traitement. Ils introduisent également un critère de qualité de modèle dans le choix du point de vue optimal.

Par extension de la représentation des occultations, certains auteurs tels que [Massios 98] [Sanchiz 99] et [Lozano 02] représentent également l'espace de travail à modéliser dans lequel le capteur est déplacé par une carte de voxels, initialement tous étiquetés « inconnu ». À chaque itération, l'étiquetage des voxels est mis à jour : en fonction des observations, les voxels peuvent devenir « vide », « occupé », « occulté ». Ils proposent, de plus, une stratégie d'optimisation basée sur la construction de fonctions d'utilité multi-critères d'un point de vue. Les travaux de [Restrepo Specht 05] reprennent une approche similaire, dans la représentation de la scène et la stratégie d'optimisation. Ils ajoutent, de plus, un critère de qualité de modèle en considérant que les voxels en bordure du modèle sont de moindre qualité et nécessitent d'être mieux observés.

Notons que certains auteurs basent leur approche sur l'utilisation d'un modèle *a priori*. Dérigent *et al.* [Dérigent 07] exploitent l'information du modèle CAO de l'objet à modéliser. Leur capteur de triangulation laser est monté sur une MMT. Pour chaque face du modèle, un ensemble de positions est déterminé afin de constituer une carte de visibilité. Une position candidate est définie comme ne présentant d'occlusion ni pour la caméra ni pour la projection laser.

6.2.2 Fonction d'utilité d'un point de vue

La construction d'une fonction d'utilité pour un point de vue utilise la représentation en *voxel map* de l'espace modélisable. On définit l'utilité d'un point de vue selon la capacité qu'il aura d'augmenter l'information dans le modèle et la qualité de ce dernier. Une bonne vue candidate doit donc permettre d'observer des secteurs qui n'ont pas été explorés, d'obtenir

de l'information, si cela est possible, dans des zones occultées dans les vues précédentes mais doit aussi être partiellement recouvrante avec des secteurs déjà modélisés afin de garantir un bon recalage. D'autres critères tenant compte de la qualité du modèle sont aussi à prendre en compte, dans notre cas la qualité de reconstruction des points 3D.

Types de voxels

Dans ses travaux, [Restrepo Specht 05] propose une classification des types de voxels intéressants au sens de la planification perceptuelle. Cette classification sépare les voxels selon qu'ils ont déjà été explorés ou pas, ou qu'ils sont particuliers dans le modèle courant. La figure 6.3 résume cette classification.

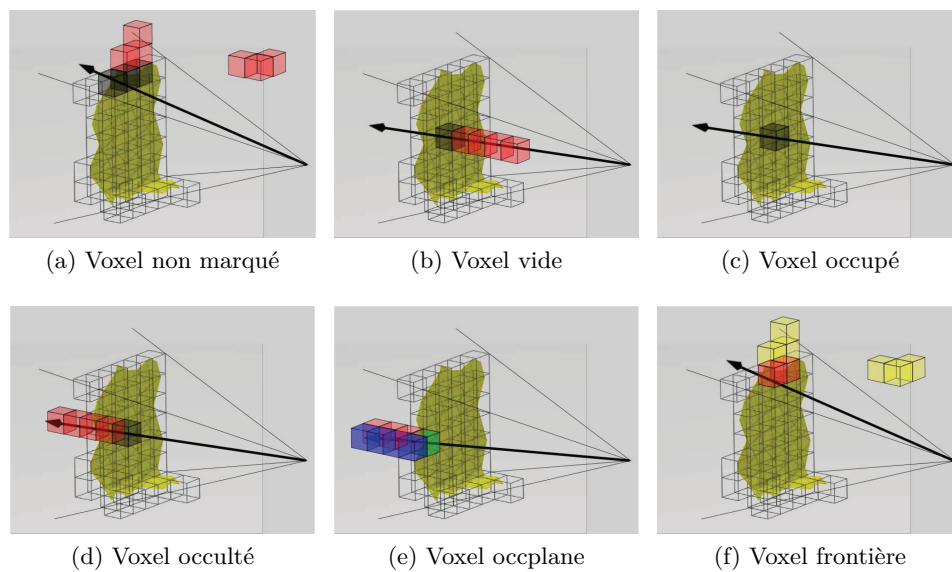


FIGURE 6.3 – Types de voxels [Restrepo Specht 05]

Les types de voxels identifiés sont :

- le *voxel non marqué*. Lorsque aucun point de vue précédent n'a créé de point dans un voxel, celui-ci est classé comme inconnu, ou non marqué. La planification perceptuelle a pour but de positionner le capteur pour observer ces éléments de volume.
- le *voxel vide*. Le voxel a été observé par au moins un point de vue précédent mais aucun point n'y a été construit.
- le *voxel occupé*. Lorsqu'un voxel contient des points 3D, il est marqué comme occupé. Lors de la planification, une quantité suffisante de voxels occupés doivent être visibles depuis un point de vue candidat pour permettre le recalage de l'acquisition qui en résulterait avec le modèle courant.
- le *voxel occulté*. Un voxel intersecté par un rayon optique d'une vue précédente mais situé derrière un voxel occupé. Ce type est à prendre avec précaution. En effet, un voxel occulté peut ne jamais être visible (intérieur d'un volume).
- le *voxel ocplane*. Il s'agit d'un voxel d'abord marqué comme occulté mais situé à côté d'un voxel vide sur une de ses six facettes. Ces voxels représentent des voxels occultés que l'on suppose observables dans au moins un point de vue, qu'il faudra définir. On

s'attachera donc à construire notre fonction d'utilité en tenant compte de ces voxels plutôt que les voxels occultés précédents.

- le *voxel frontière*. Lorsqu'un voxel occupé est adjacent à un voxel non marqué, on le note comme frontière. L'introduction de ce type de voxel est motivée par plusieurs raisons. La recherche de présence de voxels frontières lors de la planification d'une vue permet à la fois de garantir un recouvrement avec le modèle courant lorsqu'on souhaite résoudre des voxels inconnus, et permet aussi une première approche vers l'amélioration de la qualité du modèle. Les voxels observés en marge de visibilité sont, pour la plupart des capteurs, les points les moins fiables, et donc ceux qu'il conviendrait de réobserver d'un point de vue plus favorable.

Fonctions d'utilité atomiques

La fonction d'utilité d'un point de vue doit permettre de qualifier l'intérêt d'un point de vue selon qu'il permet l'observation et la résolution d'un certain nombre de voxels. En fonction du besoin, et donc de l'étape de modélisation, on définira de façon différente l'intérêt d'un point de vue. Lors des premières étapes d'une modélisation, on souhaite réaliser l'acquisition de la géométrie globale de la scène, on favorisera donc l'exploration, c'est à dire la résolution de voxels inconnus et occplane. Lors de phases ultérieures de modélisation on cherchera à augmenter la qualité du modèle. Par conséquent on s'attachera à observer les voxels encore occultés et les voxels occupés dont le score de qualité est faible.

La fonction d'utilité d'un point de vue doit donc qualifier que le point de vue incluera une quantité désirée de chaque type de voxel. On définit donc, pour chaque type de voxel, une fonction d'utilité atomique. Cette fonction est une simple courbe cubique par morceaux fonction de la quantité – plus précisément le pourcentage α – de voxels du type donné observables dans la vue candidate [Sanchiz 99, Lozano 02].

$$f(x)_{\mathbf{q}} = \begin{cases} A_1x^3 + B_1x^2 + C_1x + D_1 & \text{pour } x < \alpha \\ A_2x^3 + B_2x^2 + C_2x + D_2 & \text{pour } x \geq \alpha \end{cases}$$

Le vecteur $\mathbf{q}$ représente les paramètres de pose pour le point de vue candidat considéré. Les paramètres $(A_1, B_1, C_1, D_1, A_2, B_2, C_2, D_2)$ sont choisis d'après les hypothèses que l'on prend pour la construction de nos fonctions atomiques. On impose que la fonction soit continue, continuellement dérivable, monotone sur chacune des deux parties de la fonction, maximale en α et minimale en 0 et 1 (100%).

$$\begin{aligned} f(0)_{\mathbf{q}} &= 0 & \dot{f}(0)_{\mathbf{q}} &= 0 \\ f(\alpha)_{\mathbf{q}} &= 1 & \dot{f}(\alpha)_{\mathbf{q}} &= 0 \\ f(1)_{\mathbf{q}} &= 0 & \dot{f}(1)_{\mathbf{q}} &= 0 \\ f(x)_{\mathbf{q}} &> 0 & \forall x &\in [0, 1] \\ \dot{f}(x)_{\mathbf{q}} &> 0 & \forall x &\in [0, \alpha] \\ \dot{f}(x)_{\mathbf{q}} &< 0 & \forall x &\in [\alpha, 1] \end{aligned}$$

La figure 6.4 illustre la forme de $f(x)$ pour un α donné. Selon le type de comportement que l'on souhaite donner au système de planification, on attribuera des valeurs différentes aux paramètres α pour les différents types de voxels considérés. Les types utilisés pour la fonction globale sont, en proportions définies par l'affectation des paramètres α , les voxels *occupés*, *occplane*, *frontières*, et *non marqués*. Les types *vide* et *occultés* ne nous apportent pas

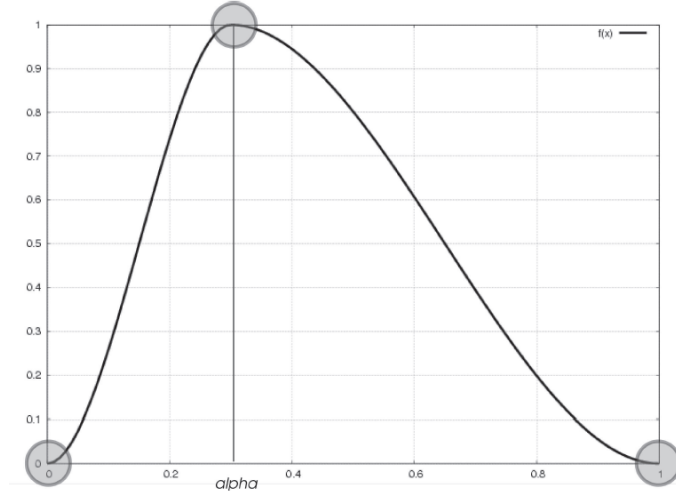


FIGURE 6.4 – Fonction d'utilité atomique pour un type de voxel et un point de vue donnés. La fonction est maximale pour un pourcentage α de ce type de voxel observables dans la vue candidate.

d'information et ne sont utilisés que pour marquer les voxels *non marqués* et peuvent également mener à la création de voxels *ocplane*. La fonction d'utilité globale que l'on cherchera à optimiser est définie comme une somme des fonctions d'utilité atomiques.

$$F(x)_{\mathbf{q}} = -f_{occupe}(x)_{\mathbf{q}} - f_{ocplane}(x)_{\mathbf{q}} - f_{inconnu}(x)_{\mathbf{q}} - f_{frontiere}(x)_{\mathbf{q}} \quad (6.1)$$

Prise en compte de la qualité

L'ajout de critères de qualité sur le modèle présenté par l'équation (6.1) peut se faire simplement par pondération des différentes fonctions atomiques qui composent la fonction d'utilité d'un point de vue.

On redéfinit donc la fonction d'utilité globale en pondérant les fonctions atomiques.

$$F(x)_{\mathbf{q}} = -\omega_{occupe} f_{occupe}(x)_{\mathbf{q}} - \omega_{ocplane} f_{ocplane}(x)_{\mathbf{q}} - \omega_{inconnu} f_{inconnu}(x)_{\mathbf{q}} - \omega_{frontiere} f_{frontiere}(x)_{\mathbf{q}}$$

Chaque pondération ω permet donc de minimiser ou augmenter l'importance relative d'un type de voxel par rapport aux autres. Dans le cas des voxels occupés et frontière, cette pondération peut être déterminée à partir des scores de corrélation des points observés dans la vue candidate. Selon le besoin on peut souhaiter favoriser la présence de points de bonne qualité dans la vue sélectionnée – pour l'exploration par exemple – ou favoriser la présence de points de mauvaise qualité – pour le raffinage.

$$\begin{aligned} \omega_{occupe,1} &= 1 - \frac{1}{N_{occupe}} \sum_{i=1}^{N_{occupe}} nssd(\mathbf{p}_i) \\ \omega_{occupe,2} &= \frac{1}{N_{occupe}} \sum_{i=1}^{N_{occupe}} nssd(\mathbf{p}_i) \end{aligned}$$

Ainsi, le gain $\omega_{occupe,1}$ maximisera la fonction atomique $f_{occupe}(x)$ pour les vues proposant un pourcentage α_{occupe} de voxels occupés dont le score associé est minimal – donc de meilleure qualité. Au contraire, le gain $\omega_{occupe,2}$ maximisera $f_{occupe}(x)$ lorsque les points occupés de la vue candidate seront majoritairement pauvres en qualité.

Pour la pondération des voxels *ocplane*, [Restrepo Specht 05] propose l'utilisation d'un critère lié à l'orientation du rayon d'observation de la vue candidate et la normale supposée en ce voxel. Lorsqu'on marque un voxel comme *ocplane*, on lui associe une « normale » supposée, dirigée orthogonalement par rapport au rayon qui aurait dû observer ce voxel.

$$\omega_{ocplane} = \frac{\sum_{i=1}^{N_{occupe}} \cos(\delta_i)}{N_{ocplane}}$$

Pour une position candidate et un voxel i donné, δ_i est l'angle entre le rayon optique qui l'intersecte et l'inverse de la normale supposée.

La pondération des voxels frontières est discutable, dans la mesure où l'on a pris l'hypothèse plus haut que les points en marge de visibilité sont de moins bonne qualité que les points au centre du champ de vue du capteur. De même, la pondération des voxels inconnus est contestable, car y attacher un critère de qualité supposerait que l'on propage l'information de voxels marqués aux zones inconnues avoisinantes sans avoir observé ces régions.

La fonction d'utilité globale finale peut donc être résumée en ne tenant compte des pondérations que pour les voxels occupés et *ocplane*.

$$F(x)_{\mathbf{q}} = -\omega_{occupe} f_{occupe}(x)_{\mathbf{q}} - \omega_{ocplane} f_{ocplane}(x)_{\mathbf{q}} - f_{inconnu}(x)_{\mathbf{q}} - f_{frontiere}(x)_{\mathbf{q}}$$

La pondération ω_{occupe} prenant pour valeur soit $\omega_{occupe,1}$ soit $\omega_{occupe,2}$.

La définition des paramètres α pour chaque type de voxel et le choix d'une pondération pour ω_{occupe} permet donc l'adaptation de la méthode de sélection et la définition de stratégies selon la phase de modélisation.

La résolution de la sélection de point de vue peut se faire par minimisation de la fonction d'utilité en utilisant le modèle courant et les caractéristiques du mouvement applicable au capteur pour définir les contraintes de placement. Signalons que de part sa construction, la fonction d'utilité n'est pas dérivable. On peut seulement l'évaluer en chaque point de vue potentiel. Dans leurs approches, [Sanchiz 99] et [Restrepo Specht 05] ont exploité la méthode du simplex : ils signalent des problèmes de convergence. Foissotte *et al.* [Foissotte 09] ont amélioré la phase d'optimisation grâce à des méthodes d'optimisation sans dérivée (NEWUOA ou BOBIQA).

6.3 Conclusions

La sélection de point de vue dans la modélisation 3D par un capteur tenu à la main offre plusieurs avantages. Tout d'abord cela facilite le travail de l'opérateur qui est guidé dans son processus, ouvrant les applications de numérisation 3D à une population d'utilisateurs totalement non experte. D'autre part, le choix du point de vue en tenant compte de l'intégralité de la scène permet une détection mieux garantie des manques et des défauts du modèle que lorsque l'opérateur doit le jauger visuellement lui-même. Enfin, le guidage de l'opérateur

définit la prochaine pose conseillée pour la prise de vue ; il est alors possible d'utiliser cette pose suggérée comme estimée initiale pour le recalage de la nouvelle vue, dans la mesure où l'on peut vérifier que l'opérateur suit bien notre conseil.

Nous avons proposé ici une première approche, basée sur les travaux de [Lozano 02] et [Restrepo Specht 05], pour résoudre le problème de la sélection de vue pour le guidage opérateur. Nous avons étudié comment adapter un critère de qualité propre à la technologie de reconstruction employée par notre capteur. Nous avons tout d'abord analysé comment ce critère pouvait se construire et vérifié sa pertinence, puis nous avons présenté l'approche envisagée.

Cette étude permet d'envisager de futurs travaux visant à concevoir et intégrer un système de sélection de point de vue et de guidage opérateur dans notre solution de numérisation utilisant notre capteur basé vision tenu à la main. Dans notre cas, le positionnement du capteur est fait par l'opérateur, mais le mouvement étant mesuré par les capteurs inertiels ou par odométrie visuelle, il serait possible de contrôler l'exécution du mouvement et de guider l'opérateur par une interface à définir.

Conclusions et perspectives

Nous avons souhaité, au lancement du projet, réaliser un système de modélisation 3D facile d'accès. Notre but était de concevoir un système plus simple à mettre en œuvre et à manipuler afin d'ouvrir la numérisation 3D aux utilisateurs non experts. Nous voulions aussi réaliser un système moins coûteux que les solutions existantes. Trois ans de travail plus tard, en collaboration avec nos collègues de la société Noomeo, nous avons conçu et mis en œuvre un capteur de numérisation 3D innovant, pleinement fonctionnel. Nous avons défini les enjeux clefs du projet :

- *Concevoir un système de modélisation 3D* capable de produire une mesure 3D fiable et exploitable. De l'acquisition d'images à un modèle final sous la forme d'un maillage texturé, nous avons effectivement réalisé un système complet de modélisation en mesure de répondre à des applications telles que la rétro-conception ou le contrôle dimensionnel par exemple.
- *Réaliser un appareil peu onéreux* en choisissant des composants bas coût et en simplifiant la conception matérielle. Aujourd'hui la société Noomeo propose ce capteur et sa solution de numérisation à un prix deux à trois fois inférieur à des solutions similaires concurrentes.
- *Choisir des composants bas coût* et s'affranchir des problèmes qu'ils posent. Une part de notre travail a donc été de tenir compte de la qualité des composants et de mettre en œuvre des méthodes peu sensibles aux bruits ou aux dégradations de mesure par exemple. Par une gestion fine de ces composants et des données générées, notre système est en mesure de générer un modèle 3D fiable même lors de désynchronisations matérielles, de perturbations magnétiques, ou de variations de température par exemple.
- *Proposer un système tenu à la main*. La gestion de l'incertitude et des approximations que génèrent l'opérateur – pas d'information *a priori* sur le mouvement, tremblements – a été un point clef de ce travail.
- *Faciliter l'utilisation* pour des utilisateurs non-experts. En concevant des mécanismes simples et un maximum d'automatisation, on permet de réduire la complexité du travail pour l'opérateur ; mais fiabiliser nos algorithmes et nos processus est également indispensable pour ne pas placer l'utilisateur dans des situations d'erreurs difficiles à résoudre.
- *Ne pas nécessiter de préparation de la scène* afin, également, de rendre simple le processus de modélisation. Dans toutes les étapes de la modélisation, le système n'exploite pas d'information *a priori* sur la scène. De plus, par un travail sur le réglage des composants et des critères de nos méthodes d'estimation, nous avons pu rendre le fonctionnement du système stable dans la plupart des situations.

Pour cette étude, notre contribution a porté une attention particulière au problème du recalage en-ligne et à son raffinement hors-ligne. Nous avons proposé une approche intégrée à un processus de modélisation incrémentale basée sur le recalage rapide paire-à-paire de deux vues. L'utilisation d'un système de mesure d'attitude couplé à une mise en correspondance de points d'intérêt permet de choisir une vue candidate pertinente et de réaliser le recalage paire-à-paire.

De manière plus globale, chaque recalage permet d'ajouter un nœud et une arête dans un graphe liant l'ensemble des acquisitions. En détectant des possibilités de cycles, on peut

remanier le graphe afin de créer ces cycles. On offre ainsi la possibilité de procéder aux premiers raffinage du modèle. Hors ligne, la minimisation d'un critère de reprojection global sur l'ensemble des vues nous permet de raffiner l'ensemble des poses estimées.

Nos principales contributions pour cette thèse sont :

- la proposition d'une chaîne de modélisation depuis le capteur jusqu'au modèle 3D. Par le choix des méthodes et la conception du capteur nous avons réalisé un système basé sur cette chaîne et évalué nos résultats. Nous nous appuyons sur des méthodes de vision stéréoscopique à deux caméras pour la reconstruction inspirés des travaux de [Garcia 01] notamment.
- une méthode de recalage basée sur l'algorithme ICP. Le mouchetis projeté se déplace avec notre capteur et rend impossible la mise en correspondance directe des points entre des images acquises à des instants successifs. Nous avons proposé une mise en correspondance basée sur l'information image malgré la présence du mouchetis mobile.
- une stratégie de recalage multi-vues. Nous utilisons la mesure inertielle et l'indexation d'image pour le recalage rapide. Nous avons présenté une approche basée sur la minimisation de cycles pour la fermeture de boucles et la minimisation d'un score de reprojection pour le raffinage du modèle.

Nous avons évalué les méthodes que nous avons proposées ainsi que le système complet. Pour cela nous avons fait numériser nos pièces de test par d'autres systèmes bien connus dans les communautés industrielle et scientifique. Nos évaluations ont porté sur plusieurs critères. A l'aide d'une cale étalon de dimensions connues, nous avons pu proposer une évaluation sommaire de l'incertitude de mesure de notre système.

La comparaison entre nos résultats et les modèles issus des autres numériseurs nous a permis de constater que nous atteignons les performances requises pour nos applications. Nous avons pu conclure que notre système permet d'atteindre les précisions et la fiabilité des autres systèmes tout en proposant une mise en œuvre ainsi qu'un mode opératoire largement simplifiés. La maniabilité de notre système permet beaucoup plus facilement d'observer certaines zones partiellement occultées ou gauches dans une scène.

Ne nécessitant pas d'équiper la scène ou d'intervenir manuellement dans le processus de recalage, le système et les méthodes que nous proposons permettent d'obtenir le modèle final beaucoup plus rapidement et à moindre effort. Cette amélioration est particulièrement précieuse dans les procédés industriels puisqu'elle permet de réduire largement les coûts d'étude ou de contrôle. Dans le cas du contrôle non destructif cela autorise, de plus, à généraliser l'usage de la numérisation vers des procédés de contrôle systématique par exemple.

Cette thèse a également permis la production d'outils logiciels et la participation à des projets connexes. Elle a été l'occasion de collaborer aux développements de méthodes de calibrage et de corrélation, notamment. Plusieurs outils permettant la reconstruction, le recalage, la modélisation temps-réel ou la modélisation hors ligne ont été développés. Une partie de nos méthodes et de ces outils ont permis la création d'une solution logicielle dédiée à la modélisation 3D basée sur notre capteur depuis l'acquisition jusqu'à l'exploitation. Ce logiciel constitue à la fois le pilote d'acquisition, offrant la visualisation directe de la modélisation incrémentale, et un outil pour la finalisation et le nettoyage du modèle.

Nous avons également eu l'occasion d'exploiter le capteur que nous avons développé dans des applications spécifiques. Par intégration à un système robotisé, nous avons collaboré à un projet de contrôle et d'inspection industrielle de pièces de grande taille. Le projet a nécessité

l'adaptation de l'appareil ainsi que l'étude de problématiques de commande afin de réaliser l'inspection automatique d'un assemblage. Le système a permis la détection automatique de défauts d'assemblage ou d'usinage par comparaison entre le modèle CAO et la numérisation de zones clefs.

Enfin, cette thèse a contribué au développement de la société Noomeo, qui produit et commercialise le capteur OptiNumTM et son logiciel NumiSoftTM.

Malgré tout, il reste à prolonger cette étude afin d'améliorer le travail que nous avons mené. La nécessité de disposer d'une très bonne estimée initiale pour démarrer la procédure de recalage globale reste une contrainte puisqu'elle nécessite que le système soit très robuste en amont. De plus cette approche ne permet pas de réestimer les points reconstruits. En retravaillant sur le critère minimisé, il serait envisageable de disposer d'une meilleure méthode de réestimation finale capable de recalculer points et poses conjointement, à la manière d'un ajustement de faisceaux mais fonctionnant sur des images avec un mouchetis mobile.

De plus, le calibrage et, particulièrement, le recalibrage de l'appareil mériteraient une étude plus poussée. En effet, nous avons porté nos efforts sur la conception d'un boîtier rigide afin de minimiser la déformation et le décalibrage du système. Cependant il faudrait disposer d'une méthode de détection de décalibrage et de recalibrage automatique afin d'assurer la meilleure précision de reconstruction possible.

Le guidage de l'opérateur nécessite encore une quantité importante de travail afin d'évaluer notre proposition et développer une solution de guidage pratique pour l'opérateur.

Au delà de cette étude, nous pourrions étendre nos recherches et notre capteur vers de nouveaux travaux. Plusieurs secteurs industriels ont des difficultés à exploiter la modélisation 3D faute de solutions simples. C'est le cas, particulièrement, pour des matériaux complexes pour la vision. Les composites carbone ou le verre, par exemple, ont des surfaces particulièrement défavorables pour les traitements en vision par ordinateur s'ils ne sont pas matifiés. On se ramène donc ici au besoin d'équiper préalablement la scène, complexifiant ainsi le processus. L'utilisation de filtres polarisants [Rantoson 09] permettraient, par exemple, de s'affranchir des réflexions spéculaires. L'exploitation de systèmes stéréoscopiques à plus de deux caméras [Harvent 10] peut constituer une autre approche. En effet, avec un nombre suffisant de vues on peut espérer disposer de deux vues sans reflets spéculaires pour la mise en correspondance et la reconstruction de chaque point.

Bibliographie

- [3Dscanrep 10] 3Dscanrep. *The Stanford 3D Scanning Repository*. <http://www-graphics.stanford.edu/data/3Dscanrep/>, 2010. 78
- [Arun 87] K.S. Arun, T.S. Huang & S.D. Blostein. *Least-squares fitting of two 3-D point sets*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 9, no. 5, pages 698–700, 1987. 59, 64, 72
- [Banta 95] J.E. Banta, Y. Zhién, X.Z. Wang, G. Zhang, M.T. Smith & M.A. Abidi. *A « Best Next View » algorithm for three-dimensional scene reconstruction using range images*. In SPIE Conf. on Intelligent Robots and Computer Vision XIV, volume 2588, pages 418–429, octobre 1995. 129
- [Batlle Subirós 08] E. Batlle Subirós. *Large-scale surface registration*. PhD thesis, Universitat de Girona, Girona, November 2008. 91
- [Bay 06] H. Bay, T. Tuytelaars & L. Van Gool. *Surf : Speeded up robust features*. In European Conference on Computer Vision, pages 404–417, 2006. 12, 57, 63
- [Bergevin 96] R. Bergevin, M. Soucy, H. Gagnon & D. Laurendeau. *Towards a general multi-view registration technique*. PAMI, vol. 18, no. 5, pages 540–547, 1996. 91
- [Besl 92] P.J. Besl & N.D. McKay. *A Method for Registration of 3-D Shapes*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, no. 2, pages 239–256, 1992. 56, 69, 82
- [Blais 85] F. Blais & M. Rioux. *Real-time numerical peak detector*. Signal Processing, vol. 11, no. 2, pages 145–155, 1985. 51
- [Blais 95] G. Blais & M.D. Levine. *Registering multiview range data to create 3D computer objects*. IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 17, pages 820–824, 1995. 70, 71, 76
- [Bleser 09] G. Bleser. *Towards Visual-Inertial SLAM for Mobile Augmented Reality*. PhD thesis, Technical University Kaiserslautern, 2009. 92
- [Bugarin 10] F. Bugarin, D. Henrion, T. Sentenac, J.B. Lasserre & J.J. Orteu. *Optimisation globale polynomiale appliquée à la rectification projective d'images non calibrées (in French)*. In Conférence en Reconnaissance des Formes et Intelligence Artificielle, January 2010. 47
- [Carmichael 99] O. Carmichael, D. Huber & M. Hebert. *Large data sets and confusing scenes in 3-D surface matching and recognition*. In International Conference on 3-D Digital Imaging and Modeling, pages 258–267, 1999. 56, 60
- [Chen 92] Y. Chen & G. Medioni. *Object modelling by registration of multiple range images*. Image Vision Comput., vol. 10, no. 3, pages 145–155, 1992. 69, 74, 76, 90, 104
- [Chen 98] C.S. Chen, Y.P. Hung & J.B. Cheng. *A Fast Automatic Method for Registration of Partially-Overlapping Range Images*. In International Conference on Computer Vision, 1998. 57, 65

- [Chua 97] C.S. Chua & R. Jarvis. *Point signatures : A new representation for 3d object recognition*. International Journal of Computer Vision, vol. 25, no. 1, pages 63–85, Octobre 1997. [12](#), [56](#), [59](#)
- [Coudrin 09] B. Coudrin, M. Devy, L. Brèthes & J.J. Orteu. *3D Modelling of objects from a visual sensor moved manually by an operator*. In ORASIS'09 Congrès des jeunes chercheurs en vision par ordinateur, Juin 2009. [23](#)
- [Coudrin 10] B. Coudrin, M. Devy, J.J. Orteu & L. Brèthes. *Registration Strategies of 3D Images Acquired From a Hand-held Visual Sensor*. In 3D-IMS Conference on 3D-Imaging of Materials and Systems, Septembre 2010. [23](#)
- [Coudrin 11] B. Coudrin, M. Devy, J.J. Orteu & L. Brèthes. *An innovative hand-held vision-based digitizing system for 3D modelling*. Optics and Lasers in Engineering, 2011. [23](#)
- [Curless 00] B. Curless. *From Range Scans to 3D Models*. SIGGRAPH Comput. Graph., vol. 33, no. 4, pages 38–41, 2000. [16](#)
- [Dalley 02] G. Dalley & P. Flynn. *Pair-Wise Range Image Registration : A Study in Outlier Classification*. Computer Vision and Image Understanding, vol. 87, no. 1-3, pages 104 – 115, 2002. [69](#)
- [Davison 03] A.J. Davison. *Real-time simultaneous localisation and mapping with a single camera*. In International Conference on Computer Vision, pages 1403–1410, 2003. [91](#)
- [Davison 07] A.J. Davison, I.D. Reid, N.D. Molton & O. Stasse. *MonoSLAM : Real-time single camera SLAM*. IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 29, page 2007, 2007. [91](#)
- [Derigent 07] W. Derigent, S. Rémy, E. Chapotot, A. Bernard & G. Ris. *3D digitizing strategy planning approach based on CAD model*. Journal of Computing and Information Science in Engineering, vol. 7, no. 10, pages 10–19, 2007. [129](#)
- [Devernay 97] F. Devernay. *Vision stéréoscopique et propriétés différentielles des surfaces*. PhD thesis, École Polytechnique, France, Février 1997. [50](#)
- [Eggert 96] D.W. Eggert, A.W. Fitzgibbon & R.B. Fisher. *Simultaneous Registration of Multiple Range Views For Use In Reverse Engineering of CAD Models*, 1996. [71](#)
- [Fan 01] K.C. Fan & T.H. Tsai. *Optimal shape error analysis of the matching image for a free-form surface*. International Journal of Robotic and Computer-Integrated Manufacturing, vol. 17, no. 3, pages 215–222, 2001. [71](#)
- [Feldmar 96] J. Feldmar & N. Ayache. *Rigid, affine and locally affine registration of free-form surfaces*. Int. J. Comput. Vision, vol. 18, no. 2, pages 99–119, 1996. [57](#), [65](#)
- [Fetita 04] C.I. Fetita, F.J. Prêteux, C. Beigelman-Aubry & P. Grenier. *Pulmonary airways : 3-D reconstruction from multislice CT and clinical investigation*. IEEE Trans. Med. Imaging, vol. 23, no. 11, pages 1353–1364, 2004. [16](#)

- [Foissotte 09] T. Foissotte, O. Stasse, P.B. Wieber, A. Escande & A. Kheddar. *A Two-Steps Next-Best-View Algorithm for Autonomous 3D Object Modeling by a Humanoid Robot*. In International Conference on Robotics and Automation (ICRA), Mai 2009. 133
- [Forsyth 02] D.A. Forsyth. *Shape from Texture without Boundaries*. In In Proc. ECCV, pages 225–239, 2002. 18
- [García 98] M.A. García, S. Valázquez & A.D. Sappa. *A Two Stage Algorithm for Planning the Next Best View From Range Images*. In British Machine Vision Conference, 1998. 129
- [Garcia 01] D. Garcia. *Mesure de formes et de champs de déplacements tridimensionnels par stéréocorrélation d'images (in French)*. PhD thesis, Ecole des Mines, Albi, 2001. 49, 136
- [Gemeiner 07] P. Gemeiner, P. Einramhof & M. Vincze. *Simultaneous Motion and Structure Estimation by Fusion of Inertial and Vision Data*. IJRR, vol. 26, 2007. 92
- [González-Baños 00] H. González-Baños & J.C. Latombe. *Robot Navigation for Automatic Model Construction using Safe Regions*. In PROC. of Int. Symposium on Experimental Robotics (ISER'01), décembre 2000. 128
- [Harris 88] C. Harris & M. Stephens. *A combined corner and edge detector*. In 4th Alvey Vision Conference, pages 147–151, 1988. 57, 62
- [Hartley 04] R.I. Hartley & A. Zisserman. *Multiple view geometry in computer vision*. Cambridge University Press, second edition, 2004. 18, 46, 47, 69
- [Harvent 10] J. Harvent. *Mesure de formes par corrélation multi-images : application à l'inspection de pièces aéronautiques à l'aide d'un système multi-caméras*. PhD thesis, Université de Toulouse, 2010. 137
- [Huber 03] D. Huber & M. Hebert. *Fully automatic registration of multiple 3D data sets*. Image and Vision Computing, vol. 21, no. 7, pages 637–650, 2003. 91
- [Johnson 97] A. Johnson. *Spin-images : A Representation for 3-D Surface Matching*. PhD thesis, Robotics Institute, Carnegie Mellon University, Pittsburgh, PA, Août 1997. 12, 56, 60
- [Keller 03] Y. Keller & A. Averbuch. *Implicit similarity : a new approach to multi-sensor image registration*. Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, vol. 2, page 543, 2003. 69
- [Kim 03] S. Kim, C. Jho & H. Hong. *Automatic registration of 3D data sets from unknown viewpoints*. In Workshop on Frontiers of Computer Vision, pages 155–159, 2003. 57
- [Kim 04] S. Kim, K. Kim & W. Who. *3D Registration for image-based virtual environment generation using color and depth information*. In IEEE International Conference on Image Processing, 2004. 71
- [Klein 00] K. Klein & V. Sequeira. *The view-cube : an efficient method of view planning for 3D modelling from range data*. In Fifth IEEE Workshop on Applications of Computer Vision (WACV 2000), décembre 2000. 129

- [Krsek 97] P. Krsek, T. Pajdla & V. Hlavac. *Estimation of differential structures on triangulated surfaces*. In 21st Workshop of the Austrian Association for Pattern Recognition, pages 157–164, 1997. [57](#)
- [Langis 01] C. Langis, M.A. Greenspan & G. Godin. *The Parallel Iterative Closest Point Algorithm*. In 3rd International Conference on 3-D Digital Imaging and Modeling, 2001. [70](#)
- [Lefferts 82] E.J. Lefferts, F.L. Markley & M.D. Shuster. *Kalman filtering for spacecraft attitude estimation*. Journal of Guidance, Control and Dynamics, vol. 5, no. 5, pages 417–429, 1982. [28](#), [92](#)
- [Levoy 00] M. Levoy, J. Ginsberg, J. Shade, D. Fulk, K. Pulli, B. Curless, S. Rusinkiewicz, D. Koller, L. Pereira, M. Ginzton, S. Anderson & J. Davis. *The Digital Michelangelo Project : 3D Scanning of Large Statues*. In Proc. of the 27th annual conference on Computer Graphics and Interactive Techniques, pages 131–144, 2000. [16](#)
- [Li 07] H. Li & R. Hartley. *The 3D-3D Registration Problem Revisited*. In IEEE International Conference on Computer Vision, pages 1–8, 2007. [56](#)
- [Lorusso 95] A. Lorusso, D.W. Eggert & R.B. Fisher. *A comparison of four algorithms for estimating 3-D rigid transformations*. In BMVC '95 : Proceedings of the 1995 British conference on Machine vision (Vol. 1), pages 237–246, 1995. [59](#), [69](#)
- [Lowe 04] D.G. Lowe. *Distinctive Image Features from Scale-Invariant Keypoints*. International Journal of Computer Vision, vol. 60, pages 91–110, 2004. [57](#), [62](#), [63](#)
- [Lozano 02] M. Lozano & M. Devy. *Perception planning for an exploration task of a 3D environment*. Rapport technique, LAAS-CNRS, Toulouse FRANCE, 2002. [129](#), [131](#), [134](#)
- [Lu 97] F. Lu & E. Milios. *Globally Consistent Range Scan Alignment for Environment Mapping*. Autonomous Robots, vol. 4, pages 333–349, 1997. [91](#)
- [Ma 04] B. Ma & Y.E. Ellis. *Surface-based registration with a particle filter*. In In Proceedings of Medical Image Computing and Computer Assisted Intervention, pages 566–573, 2004. [57](#)
- [Mansouri 07] A. Mansouri, A. Lathuiliere, F. Marzani, Y. Voisin & P. Gouton. *Toward a 3D Multispectral Scanner : An Application to Multimedia*. IEEE MultiMedia, vol. 14, no. 1, pages 40–47, 2007. [16](#)
- [Marins 01] J.L. Marins, X. Yun, E.R. Bachmann, R.B. McGhee & M.J. Zyda. *An Extended Kalman Filter for Quaternion-Based Orientation Estimation Using MARG Sensors*. In International Conference on Intelligent Robots and Systems, 2001. [28](#)
- [Massios 98] N.A. Massios & R.B. Fisher. *A Best Next View Selection Algorithm incorporating a Quality Criterion*. In British Machine Vision Conference, pages 780–789, 1998. [129](#)
- [Matabosch Geronès 07] C. Matabosch Geronès. *3D hand-held sensor for large surface registration*. PhD thesis, Universitat de Girona, Girona, Juin 2007. [56](#), [90](#), [91](#)

- [Maybeck 79] Peter S. Maybeck. Stochastic models, estimation, and control, volume 1 of *Mathematics in Science and Engineering*. Academic Press, Inc., 1979.
- [Moghari 07] M.H. Moghari & P. Abolmaesumi. *Point-Based Rigid-Body Registration Using an Unscented Kalman Filter*. IEEE Trans. Med. Imaging, vol. 26, no. 12, pages 1708–1728, 2007. [57](#)
- [Montiel 06] J.M.M. Montiel, J. Civera & A.J. Davison. *Unified Inverse Depth Parametrization for Monocular SLAM*. In RSS, 2006. [91](#), [92](#)
- [Morency 03] L.P. Morency & R. Gupta. *Robust Real-time Egomotion from Stereo Images*. In IEEE International Conference on Image Processing, 2003. [71](#)
- [Mouragnon 06] E. Mouragnon, M. Lhuillier, M. Dhome, F. Dekeyser & P. Sayd. *3D reconstruction of complex structures with bundle adjustment : an incremental approach*. In In ICRA 2006, 2006. [91](#)
- [Neugebauer 97] P. Neugebauer. *Geometrical cloning 3D objects via simultaneous registration of multiview range images*. In IEEE Conference on Shape Modeling and Applications, pages 130–139, 1997. [71](#)
- [Newcombe 10] R.A. Newcombe & A.J. Davison. *Live Dense Reconstruction with a Single Moving Camera*. In CVPR 2010, 2010. [90](#)
- [Nieto 05] J. Nieto, T. Bailey & E. Nebot. *Scan-SLAM : Combining EKF-SLAM and Scan Correlation*. In FSR, 2005. [91](#)
- [Nuchter 04] A. Nuchter, 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, volume 2, pages 1998–2003, 2004. [91](#)
- [Nüchter 03] A. Nüchter, H. Surmann & J. Hertzberg. *Planning Robot Motion for 3D Digitalization of Indoor Environments*. In In Proc. of the 11th International Conference on Advanced Robotics (ICAR, pages 222–227, 2003. [129](#)
- [Nüchter 06] A. Nüchter, K. Lingemann, J. Hertzberg & H. Surmann. *6-D SLAM-3D mapping outdoor environments*. J.Field Robot., vol. 24, 2006. [91](#)
- [Olson 03] C.F. Olson, L.H. Matthies, M. Schoppers & M.W. Maimone. *Rover navigation using stereo ego-motion*. Robotics and Autonomous Systems, vol. 43, no. 4, pages 215 – 229, 2003. [91](#)
- [Orteu 06] J.J. Orteu, D. Garcia, L. Robert & F. Bugarin. *A speckle texture image generator*. In Speckle06 Int. Conference, 2006. [11](#), [28](#)
- [Pan 09] Q. Pan, G. Reitmayr & T.W. Drummond. *Interactive model reconstruction with user guidance*. In Mixed and Augmented Reality, IEEE / ACM International Symposium on, pages 209–210, 2009. [91](#), [129](#)
- [Park 03] S.Y. Park & M. Subbarao. *A fast point-to-tangent plane technique for multi-view registration*. In Fourth International Conference on 3-D Digital Imaging and Modeling, pages 276–284, 2003. [71](#), [75](#)
- [Paz 08] L.M. Paz, P. Pinies, J.D. Tardos & J. Neira. *Large-Scale 6-DOF SLAM with Stereo-in-Hand*. IEEE Trans. on Robotics, vol. 24, 2008. [92](#)

- [Pinies 07] P. Pinies, T. Lupton, S. Sukkarieh & J.D. Tardos. *Inertial aiding of inverse depth SLAM using a monocular camera*. In ICRA, 2007. [92](#)
- [Pollefeys 04] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops & R. Koch. *Visual Modeling with a Hand-Held Camera*. International Journal of Computer Vision, vol. 59, pages 207–232, 2004. [90](#)
- [Prados 05] E. Prados & O. Faugeras. *Shape from shading*. In in Mathematical Models in Computer Vision : The Handbook. Springer, 2005. [18](#)
- [Pulli 99] K. Pulli. *Multiview registration for large data sets*. In 3-D Digital Imaging and Modeling, pages 160–168, 1999. [90](#)
- [Rantson 09] R. Rantson, C. Stolz, D. Fofi & F. Meriaudeau. *3D Reconstruction by Polarization Imaging Method based on Perspective Model*. In SPIE Europe Optical Metrology, Juin 2009. [137](#)
- [Restrepo Specht 05] J.A Restrepo Specht. *Modélisation d'objets 3D par construction incrémentale d'un maillage triangulaire, dans un contexte robotique*. PhD thesis, LAAS-CNRS, Université de Toulouse, Toulouse, France, Juin 2005. [14](#), [69](#), [72](#), [129](#), [130](#), [133](#), [134](#)
- [Ristic 97] M. Ristic & D. Brujic. *Efficient Registration of NURBS geometry*. Image and Vision Computing, vol. 15, no. 12, pages 925–935, 1997. [71](#)
- [Robertson 02] C. Robertson & R.B. Fisher. *Parallel Evolutionary Registration of Range Data*. Computer Vision and Image Understanding, vol. 87, pages 1–7, 2002. [71](#)
- [Roumeliotis 02] S.I. Roumeliotis, A.E. Johnson & J.F. Montgomery. *Augmenting inertial navigation with image-based motion estimation*. In in IEEE International Conference on Robotics and Automation, Washington D.C., 2002, pages 4326–4333, 2002. [92](#)
- [Royer 07] E. Royer, M. Lhuillier, M. Dhome & J.M. Lavest. *Monocular Vision for Mobile Robot Localization and Autonomous Navigation*. International Journal of Computer Vision, vol. 74, no. 3, 2007. [91](#)
- [Rusinkiewicz 01] S. Rusinkiewicz & M. Levoy. *Efficient Variants of the ICP Algorithm*. In Third International Conference on 3D Digital Imaging and Modeling (3DIM), Juin 2001. [69](#), [71](#), [74](#), [81](#)
- [Sanchiz 99] J.M. Sanchiz & R.B. Fisher. *A next-best-view algorithm for 3D scene recovery with 5 degrees of freedom*. In British Machine Vision Conference, 1999. [129](#), [131](#), [133](#)
- [Sandhu 08] R. Sandhu, S. Dambreville & A. Tannenbaum. *Particle Filtering for Registration of 2D and 3D Point Sets with Stochastic Dynamics*. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2008. [66](#)
- [Sandhu 10] R. Sandhu, S. Dambreville & A. Tannenbaum. *Point Set Registration Via Particle Filtering and Stochastic Dynamics*. IEEE Trans Pattern Anal Mach Intell, vol. 32, no. 8, pages 1459–73, Août 2010. [57](#)

- [Sequeira 96] V. Sequeira, J.G.M. Gonçalves & M.I. Ribeiro. *3D reconstruction of indoor environments*. In IRRR International Conference on Image Processing (ICIP'96), pages 405–408, septembre 1996. [129](#)
- [Sharp 04] G.C. 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, vol. 26, no. 8, pages 1037–1050, 2004. [91](#)
- [Shi 94] J. Shi & C. Tomasi. *Good Features to Track*. In IEEE Conference on Computer Vision and Pattern Recognition, pages 593–600, 1994. [57](#), [61](#)
- [Silva 07] L. Silva, O.R.P. Bellon & K.L. Boyer. *Multiview range image registration using the surface interpenetration measure*. Image and Vision Computing, vol. 25, no. 1, pages 114 – 125, 2007. [91](#)
- [Solà Ortega 07] J. Solà Ortega. *Towards visual localization, mapping and moving objects tracking by a mobile robot : a geometric and probabilistic approach*. PhD thesis, Université de Toulouse, Toulouse, Février 2007. [12](#), [49](#)
- [Solà 07] J. Solà, A. Monin & M. Devy. *BiCamSLAM : Two times mono is more than stereo*. In ICRA, 2007. [92](#)
- [Strobl 09] K.H. Strobl, E. Mair, T. Bodenmüller, S. Kielhöfer, W. Sepp, M. Suppa, D. Burschka & G. Hirzinger. *The Self-Referenced DLR 3D-Modeler*. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, pages 21–28, October 2009. [90](#)
- [Sun 03] Y. Sun, J. Paik, A. Koschan & M.A. Abidi. *Point fingerprint : A new 3-d object representation scheme*. IEEE transaction on Systems, Man, and Cybernetics, vol. 33, pages 712–717, 2003. [69](#)
- [Szeliski 93] R. Szeliski. *Rapid octree construction from image sequences*. CVGIP : Image Underst., vol. 58, no. 1, pages 23–32, 1993. [18](#)
- [Tarel 98] J.P. Tarel, H. Civi & D.B. Cooper. *Pose Estimation of Free-Form 3D Objects without Point Matching using algebraic surface Models*. In Proceedings of IEEE Workshop Model Based 3D Image Analysis, pages 13–21, Mumbai, India, 1998. [57](#)
- [Taylor 09] S. Taylor, E. Rosten & T. Drummond. *Robust feature matching in 2.3μs*. In In IEEE CVPR Workshop on Feature Detectors and Descriptors : The State Of The Art and Beyond, 2009. [57](#)
- [Thévenaz 00] P. Thévenaz, T. Blu & M. Unser. *Interpolation revisited*. IEEE transactions on medical imaging, vol. 19, no. 7, pages 739–758, 2000. [51](#)
- [Toldo 10] R. Toldo, A. Beinat & F. Crosilla. *Global registration of multiple point clouds embedding the Generalized Procrustes Analysis into an ICP framework*. In 3DPVT 2010 Conference, 2010. [90](#)
- [Triggs 00] B. Triggs, P. Mclauchlan, R. Hartley & A. Fitzgibbon. *Bundle Adjustment – A Modern Synthesis*, 2000. [18](#), [90](#), [91](#)
- [Weik 97] S. Weik. *Registration of 3D Partial Surface Models Using Luminance and Depth Information*. In International Conference on 3-D Digital Imaging and Modeling, 1997. [71](#)

- [Weng 92] J. Weng, P. Cohen & M. Herniou. *Camera Calibration with Distortion Models and Accuracy Evaluation*. IEEE Trans. Pattern Anal. Mach. Intell., vol. 14, pages 965–980, October 1992. [44](#)
- [Wong 99] L.M. Wong, C. Dumont & M.A. Abidi. *Next Best View System in a 3-D Object Modeling Task*. In IEEE International Symposium on Computational Intelligence in Robotics and Automation, pages 306–311, novembre 1999. [129](#)
- [Wyngaerd 02] J.V. Wyngaerd. *Automatic crude patch registration : Toward automatic 3D model building*. Computer Vision and Image Understanding, vol. 87, pages 8–26, 2002. [57](#)
- [Xie 10] Z. Xie, S. Xu & X. Li. *A high-accuracy method for fine registration of overlapping point clouds*. Image and Vision Computing, vol. 28, no. 4, pages 563–570, 2010. [71](#)
- [Zhang 92] Z. Zhang. *Iterative Point Matching for Registration of Free-Form Curves*. Rapport technique, INRIA Sophia Antipolis, Mars 1992. [56](#)
- [Zhang 94] Z. Zhang. *Iterative Point Matching for Registration of Free-Form Curves and Surfaces*. International Journal of Computer Vision, vol. 13, no. 2, pages 119–149, 1994. [69](#), [72](#)

The work of this thesis focuses on modeling 3D objects from small to medium sized (1 m^3 maximum) with an innovative sensor moved manually by an operator, developed by the NOOMEOTM company.

To acquire 3D data, the sensor consists of a camera system coupled with an illuminator. We infer information from a visual pattern projected on the scene, used to create the artificial texture on 3D surface, to produce a 3D image of the scene. Surfaces are represented as 3D points reconstructed from the current viewpoint. To completely model an object, the sensor must be moved around the object to acquire several 3D images. These images are successively registered and merged into a single 3D point cloud. A triangular mesh is then generated from the 3D point cloud. The appearance (color and texture) of the object can also be extracted from data acquisition to enrich the final model.

This thesis addresses the problem of registration of 3D views in two ways : the fast online registration and offline accurate optimization. For the first approach we proposed a variant of the ICP algorithm using information from images and inertial measurement. This method was validated by comparing it to other common geometric registration methods. For the second approach, we proposed a refinement strategy by minimization of cycles in a graph of relations between the 3D views ; and a global pose optimization method.

Our methods and their integration into a complete modeling system were validated by comparing our results with those from other scanning systems commonly used in scientific and industrial communities.

3D modeling, 3D image registration, Bundle adjustment, Next best view selection, Inertial measurement

AUTEUR : Benjamin COUDRIN

TITRE : Modélisation 3D d'objets par un capteur visuel déplacé par un opérateur

DIRECTEURS DE THÈSE : Michel DEVY, Jean-José ORTEU

LIEU ET DATE DE SOUTENANCE : LAAS-CNRS, 2 mars 2011

RÉSUMÉ

Les travaux de cette thèse portent sur la modélisation 3D d'objets de petite et moyenne tailles (1 m^3 maximum) par un système innovant, développé par la société NOOMEOTM, déplacé manuellement par un opérateur.

Pour acquérir les données 3D, le capteur est constitué d'un système de caméras couplé à un illuminateur. La projection sur la scène d'un motif visuel permet de créer sur les surfaces 3D une texture artificielle exploitée pour produire une image 3D de la scène, soit l'ensemble de points 3D reconstruits depuis le point de vue courant. Pour obtenir le modèle complet d'un objet, le capteur doit être déplacé autour de l'objet afin d'acquérir plusieurs images 3D. Ces images sont successivement recalées et fusionnées dans un seul nuage de points 3D. Un maillage triangulaire est ensuite généré à partir de ce nuage de points 3D. L'apparence (couleur et texture) de l'objet peut également être extraite des données d'acquisition afin d'enrichir le modèle final.

Cette thèse aborde le problème du recalage de vues 3D selon deux approches : le recalage rapide en ligne et l'optimisation précise hors ligne. La première approche nous a conduit à proposer une variante de la méthode ICP exploitant l'information des images et la mesure inertielle. Cette méthode a été validée en la comparant à d'autres méthodes courantes de recalage géométrique. Pour la seconde approche, nous avons proposé une stratégie de raffinement exploitant la minimisation de cycles dans un graphe de relations entre les vues 3D ; puis une méthode d'optimisation globale des poses.

Nos méthodes et leur intégration dans le système de modélisation complet ont été validés en comparant nos résultats avec ceux d'autres systèmes de numérisation couramment utilisés dans les communautés scientifique et industrielle.

MOTS-CLÉS

Modélisation 3D, Recalage d'images 3D, Ajustement de faisceaux, Sélection de points de vues, Mesure inertielle

DISCIPLINE : Systèmes embarqués