



HAL
open science

Relevé Volumétrique par Stéréovision à partir de Deux Vues non Calibrées – Application au Suivi Thérapeutique d’Escarres

Benjamin Albouy-Kissi

► **To cite this version:**

Benjamin Albouy-Kissi. Relevé Volumétrique par Stéréovision à partir de Deux Vues non Calibrées – Application au Suivi Thérapeutique d’Escarres. Vision par ordinateur et reconnaissance de formes [cs.CV]. Université d’Orléans, 2006. Français. ⟨NNT : 2006ORLE2035⟩. ⟨tel-01548940⟩

HAL Id: tel-01548940

<https://hal.science/tel-01548940v1>

Submitted on 28 Jun 2017

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.



HAL Authorization



THÈSE DE DOCTORAT DE L'UNIVERSITÉ D'ORLÉANS

Spécialité
Sciences et Technologies Industrielles

Présentée par
Benjamin ALBOUY

Pour obtenir le grade de
DOCTEUR DE L'UNIVERSITÉ D'ORLÉANS

Relevé Volumétrique par Stéréovision à partir de Deux Vues non Calibrées

Application au Suivi Thérapeutique d'Escarres

Soutenue le 11 décembre 2006

Devant le jury composé de :

Rapporteurs :

Jean-Marc LAVEST	Professeur à l'Université d'Auvergne
Peter STURM	Directeur de Recherche à l'INRIA Rhône-Alpes

Présidente :

Edwige PISSALOUX	Professeur à l'Université de Rouen
------------------	------------------------------------

Examineurs :

Rachid HARBA	Professeur à l'Université d'Orléans
Yves LUCAS	Maître de Conférences à l'Université d'Orléans
Rémy MALGOUYRES	Professeur à l'Université d'Auvergne
Sylvie TREUILLET	Maître de Conférences à l'Université d'Orléans

Remerciements

Je tiens tout d'abord à exprimer ma gratitude aux personnes qui m'ont fait l'honneur de participer au jury de cette thèse :

Monsieur Jean-Marc LAVEST, Professeur à l'Université Blaise Pascal de Clermont-Ferrand et Monsieur Peter STURM, Directeur de Recherche à l'INRIA Rhône-Alpes, qui ont accepté d'être les rapporteurs de mon travail.

Madame Edwige PISSALOUX, présidente du jury, et Messieurs Rachid HARBA et Rémy MALGOUYRES, qui ont accepté d'examiner ma thèse.

Monsieur Pierre MARCHE, Professeur à l'ENSI de Bourges, qui m'a accueilli au sein de l'ENSIB et m'a permis de travailler dans les meilleures conditions en assumant le rôle de Directeur de Thèse.

Je suis tout particulièrement reconnaissant à Madame Sylvie TREUILLET et Monsieur Yves LUCAS, tous deux Maîtres de Conférences à l'Université d'Orléans, pour avoir co-encadré mes travaux durant ces trois années. Je les remercie également très chaleureusement pour tous les merveilleux moments passés en leur compagnie.

Toute ma sympathie va aux membres du Laboratoire Vision et Robotique et au personnel de l'ENSI de Bourges.

Je tiens également à remercier les amis que j'ai rencontrés au cours de mon passage à Bourges, tout particulièrement Antonio, Benjamin, Hazem et Mehdi qui ont largement contribué à cette thèse par les moments de détente.

Enfin, je ne remercierai jamais assez mes parents et mes proches pour leur soutien inconditionnel, leurs encouragements dans les moments difficiles et leur disponibilité.

Table des matières

Introduction.....	9
-------------------	---

Première Partie - La Chaîne de Reconstruction 3D

Chapitre 1 Principe de la stéréovision non calibrée.....	21
1.1 Modèle géométrique de formation des images.....	23
1.1.1 Le modèle sténopé.....	25
1.2 Géométrie épipolaire.....	28
1.2.1 Relations géométriques.....	28
1.2.2 Relations analytiques.....	29
1.3 Calibrage des capteurs d'image.....	30
1.3.1 Méthodes de calibrage.....	30
1.3.2 Stratification des reconstructions 3D.....	31
1.3.3 Définition du problème de l'auto-calibrage et méthodes de résolution.....	34
1.3.4 Cas de deux vues.....	36
1.4 Triangulation.....	37
1.5 Optimisation par ajustement de faisceaux.....	37
1.6 Conclusion.....	38
Chapitre 2 Estimation de la matrice fondamentale.....	39
2.1 Méthodes d'estimation de la matrice fondamentale.....	41
2.1.1 Caractéristiques communes à tous les algorithmes.....	42
2.1.2 La méthode des moindres carrés.....	43
2.1.3 La technique des moindres carrés orthogonaux.....	43
2.1.4 La méthode de Bookstein.....	44
2.1.5 La méthode itérative des moindres carrés repondérés.....	45
2.1.6 L'algorithme LMedS.....	46
2.1.7 L'algorithme MAPSAC.....	47
2.2 Analyse des estimations de la matrice fondamentale.....	48
2.2.1 La chaîne de reconstruction 3D.....	48
2.2.2 Protocole d'évaluation.....	49
2.2.3 Simulations.....	51
2.3 Discussion sur le choix de la méthode pour la stéréovision non calibrée.....	56
2.4 Conclusion.....	58
Chapitre 3 Mise en correspondance des images.....	59
3.1 Contexte de prise de vue des escarres.....	61
3.2 Méthodes courantes d'appariement de points d'intérêt.....	63
3.2.1 Détecteur de coins de Harris et Stephens adapté à la couleur.....	63
3.2.2 Caractérisation locale des points d'intérêt.....	64
3.2.3 Méthodes d'appariement.....	65

3.3	Algorithme itératif de mise en correspondance.....	67
3.3.1	Images en entrée de l'algorithme.....	68
3.3.2	Mise en correspondance initiale.....	69
3.3.3	Détection de points d'intérêt.....	70
3.3.4	Raffinement de la mise en correspondance par un processus itératif.....	71
3.3.5	Détection et suppression des points 3D aberrants.....	75
3.3.6	Densification par appariement de triangles homologues.....	77
3.4	Évaluation de l'algorithme sur des images réelles.....	79
3.4.1	Protocole d'expérimentation.....	79
3.4.2	Résultats.....	81
3.4.3	Comparaison à une méthode de mise en correspondance dense.....	83
3.5	Conclusion.....	84
Chapitre 4	Reconstruction euclidienne de la structure 3D.....	87
4.1	Auto-calibrage.....	89
4.1.1	Détermination des paramètres intrinsèques.....	90
4.1.2	Détermination des paramètres extrinsèques.....	92
4.1.3	Validation expérimentale de l'auto-calibrage.....	93
4.2	Triangulation.....	100
4.3	Optimisation par ajustement de faisceaux.....	101
4.4	Détermination euclidienne du modèle 3D.....	105
4.5	Conclusion.....	105
Seconde partie - Évaluation des Performances		
Chapitre 5	Évaluation de la qualité de la reconstruction.....	109
5.1	Les systèmes de numérisation 3D.....	112
5.1.1	Le système d'Eotech.....	112
5.1.2	Le système de Kréon.....	113
5.1.3	Le système de Vidéometric.....	114
5.2	Évaluation de la qualité d'un modèle 3D.....	114
5.2.1	Méthodes bidimensionnelles.....	114
5.2.2	Méthodes tridimensionnelles.....	116
5.3	La méthode de comparaison.....	117
5.4	Résultats.....	119
5.4.1	Choix de la vérité terrain.....	119
5.4.2	Évaluation de notre chaîne de reconstruction.....	120
5.5	Conclusion.....	125
Chapitre 6	Choix d'un protocole de prise de vue.....	127
6.1	Protocole expérimental.....	129
6.2	Résultats.....	131
6.3	Conclusion.....	137
Chapitre 7	Évaluation de la précision volumétrique.....	139
7.1	Stratégie.....	140
7.2	Calcul du volume sous une surface.....	142
7.2.1	Calcul d'un volume élémentaire.....	143
7.2.2	Transformation du plan de référence.....	146
7.2.3	Calcul des normales aux triangles.....	149

7.3	Expérimentation sur des images réelles.....	151
7.3.1	Protocole de test.....	151
7.3.2	Résultats.....	153
7.4	Conclusion.....	155
Conclusion et perspectives.....		157

Annexes

Annexe A	Résultats du calibrage des appareils photos.....	163
A.1	Position du point principal.....	165
A.2	Distance focale.....	171
A.3	Distorsions optiques.....	175
Annexe B	Paramétrage de l'ajustement de faisceaux.....	179
B.1	Paramétrage de la rotation.....	181
B.1.1	De R vers r	181
B.1.2	De r vers R	181
B.2	Paramétrage de la translation.....	182
B.2.1	De t vers θ et ϕ	182
B.2.2	De θ et ϕ vers t	182

Bibliographie

Bibliographie.....	185
Bibliographie de l'auteur.....	193

Introduction

Ces travaux s'intègrent au projet RAVIR (Relevés Anatomiques par VIsion laseR) du Laboratoire Vision et Robotique de l'Université d'Orléans. Le laboratoire a acquis, grâce à ce projet, de sérieuses compétences dans les systèmes d'acquisition tridimensionnelle du corps humain à des fins thérapeutiques. Le domaine de l'imagerie médicale propose des systèmes d'imagerie interne très sophistiqués basés sur des phénomènes physiques très divers, comme l'Imagerie par Résonance Magnétique Nucléaire, la tomographie à rayons X ou l'échographie ultra sonore. L'imagerie de surface a longtemps été délaissée, alors que de nombreuses informations thérapeutiques peuvent être trouvées par ce moyen. C'est dans ce domaine que l'équipe du projet RAVIR oriente ses travaux de recherche.

L'équipe s'est premièrement intéressée au dépistage systématique des scoliozes chez le jeune adolescent par le dispositif SYDESCO (SYstème de DEpistage des SCOLioses). Ce système couplait un laser et une caméra afin de reconstruire en 3 dimensions le dos du patient. Un traitement informatique des données 3D permettait de détecter automatiquement une scoliose naissante chez l'adolescent. L'attention de l'équipe s'est ensuite portée sur la réalisation d'un podoscope numérique baptisé OPAL (Orthèse Plantaire Assistée par Laser) utilisant la même technique de vision laser. Ce podoscope était destiné à la conception de semelles orthopédiques pour les handicapés ou ergonomiques pour les compétiteurs. Un nouveau cap est franchi en abandonnant le laser et les systèmes calibrés avec le projet ESCALE (ESCarre Analyse Lisibilité Evaluation). Ce dernier traite de l'aide au suivi thérapeutique apportée par les méthodes d'analyse d'image. C'est au sein de ce projet que s'intègre cette thèse. L'introduction de ce mémoire permet de fixer le contexte dans lequel les travaux ont été menés, en définissant dans un premier temps les escarres et la problématique qu'elles soulèvent. Nous aborderons ensuite les techniques utilisées par les équipes médicales pour quantifier l'évolution de la cicatrisation, avant de présenter les défis à relever par le projet ESCALE et plus spécialement ceux qui font l'objet de cette thèse. Enfin, le plan de ce mémoire sera détaillé.

Définition des escarres

Formation d'une escarre

Une escarre est une plaie cutanée provoquée par une pression prolongée sur la peau qui bloque la circulation sanguine dans les tissus biologiques. Privés de l'oxygène normalement véhiculé par les vaisseaux sanguins, les tissus meurent par hypoxie, provoquant l'apparition de la plaie (voir figure 1). Cette hypoxie est d'autant facilitée que les vaisseaux sanguins sont en mauvais état ou que le sang est appauvri en oxygène. Ainsi, le diabète, l'hypertension, le tabagisme, les infections pulmonaires ou tout simplement le vieillissement naturel sont autant de facteurs aggravant le risque d'apparition d'escarres. Les conditions neurologiques du patient sont également un facteur déterminant. En effet, nous avons tous le réflexe de bouger régulièrement, même lorsque nous dormons, évitant ainsi les pressions prolongées et localisées. Une personne plongée dans le coma ou hémiparétique n'a pas cette capacité, et c'est pour cela que ces patients doivent être régulièrement changés de position afin de prévenir la formation d'escarres.

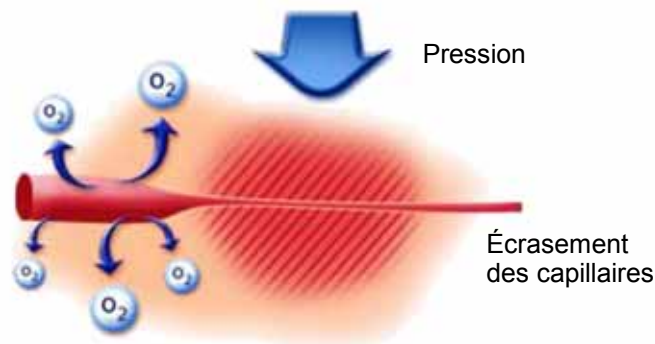


Figure 1 : Processus d'apparition d'une escarre par hypoxie.

Il existe différents types d'escarres, que l'on peut classer en quatre stades d'évolution en fonction de leur gravité. Le premier stade d'une escarre est l'érythème qui est une rougeur cutanée qui ne blanchit pas lors d'une pression du doigt. Cet état correspond au premier signe visible d'une escarre et peut être rapidement soigné de façon réversible, c'est-à-dire sans signe visible après guérison. Le deuxième stade atteint s'appelle la désépidermisation. Il s'agit d'un arrachement de l'épiderme et éventuellement du derme. Une variante de ce stade bien connue des marcheurs est la phlyctène, communément appelée « ampoule ». Le stade suivant d'une escarre est l'état de nécrose, où la plaie devient profonde avec un tissu mort de couleur noire la recouvrant. La guérison à partir de ce stade n'est plus réversible, il s'agit d'une cicatrisation. Enfin, le dernier stade d'évolution est l'état d'ulcère résultant le plus souvent du stade 3 après élimination des tissus nécrotiques. La dynamique d'apparition d'une escarre est très rapide. Une escarre de stade 1 peut évoluer jusqu'à l'état d'ulcère en seulement quelques heures si elle n'est pas traitée.

Les deux derniers stades d'évolution sont les plus douloureux. La souffrance est double pour les patients, à la fois physique et psychologique. La souffrance physique est évidemment liée à la plaie elle-même et elle entraîne un handicap au mouvement parfois très important. Du point de vue psychologique, l'escarre renvoie une image dégradante du patient à lui-même. De plus, les escarres ont malheureusement encore une image péjorative dans l'esprit de certains soignants. Par ailleurs, la dépendance vis-à-vis des soignants provoquée par la nature des soins est parfois difficilement vécue. Enfin, l'apparition d'escarres est perçue comme une confrontation à la mort chez certains patients, de par le dépérissement de leur propre chair.

En outre, les personnes atteintes d'escarres ne sont pas quelques cas isolés car 10 % des patients hospitalisés en long séjour en France développent des escarres. L'âge moyen des patients atteints est de 74 ans [escarre.fr]. Le vieillissement actuel de la population fait craindre une augmentation de la proportion de patients atteints pour les années à venir. Par ailleurs, l'escarre étant une pathologie à évolution lente, la durée des soins peut atteindre plusieurs mois pour les escarres les plus graves. Cela entraîne d'énormes coûts toujours croissants pour les hôpitaux puisque les frais d'hospitalisation peuvent atteindre 60 k€ pour un seul patient. Au delà de cet aspect purement économique, il est essentiel de soigner les patients au plus vite pour leur confort, voire mieux, de tout mettre en œuvre pour leur éviter l'apparition d'escarres. La prise en charge des escarres se fait donc à deux niveaux dans les hôpitaux, d'abord par la prévention puis par la thérapie.

Prévention des escarres

La plupart des hôpitaux intègrent aujourd'hui des processus de prévention des escarres. Cela passe tout d'abord par la formation et la sensibilisation du personnel hospitalier au sein de comités anti-escarres. L'action doit être interdisciplinaire comme pour la lutte contre les maladies nosocomiales. Leur formation porte sur la détection des patients à risque, c'est-à-dire les personnes présentant un ou plusieurs des facteurs d'apparition d'escarres. Dans le cas où il est impossible d'agir sur ces facteurs, les patients doivent être régulièrement surveillés afin de détecter les escarres naissantes de stade 1. Il faut particulièrement observer les points d'appui du patient, zones où les pressions sont les plus fortes et les plus longues (talon, sacrum...). Ces parties du corps doivent également faire l'objet d'effleurage régulier, technique qui consiste à masser sans appuyer sur la peau. Cette pratique permet de favoriser la microvascularisation cutanée et ainsi d'éviter l'hypoxie des tissus en favorisant l'apport d'oxygène.

La prévention passe également par le choix du support du patient. Ce terme désigne le matelas ou les coussins sur lesquels repose le malade. De nouveaux matériaux ont été mis au point pour éviter les zones de fortes pressions. On distingue les supports passifs des supports actifs. Les premiers sont des matériaux qui épousent la forme du corps, afin de répartir le poids sur des zones plus grandes, faisant baisser par conséquent la pression locale. Les seconds sont des systèmes qui permettent de contrôler la pression exercée localement par le support sur le corps. Régulièrement, les niveaux de pression locale sont modifiés, déplaçant ainsi les points d'appui du patient. Bien qu'apportant un progrès notable vis à vis des matelas classiques, ces supports ne dispensent pas des autres méthodes de prévention, et leur diffusion est limitée de par leur coût considérable.

Lorsque l'apparition d'une escarre de stade 3 ou 4 n'a pas pu être évitée, un suivi thérapeutique doit être mis en place durant toute la période de soin. Ce suivi permet d'adapter les soins en fonction de l'évolution de la plaie. Il est assuré par le relevé régulier de certains indicateurs. Ces observations médicales sont décrites dans la section suivante.

Caractérisation clinique des escarres

Des paramètres plus quantitatifs que le stade d'évolution permettent de préciser l'état instantané d'une escarre. Ces indices peuvent être classés en deux catégories : les indices surfaciques et volumétriques [Koeffler, 2004].

Les indices surfaciques

Les escarres sont constituées de quatre types de tissus de couleurs différentes : les parties nécrotiques de couleur noire, les tissus fibrineux de couleur jaune qui sont de la nécrose récente, les zones en bourgeonnement de couleur rouge vif qui correspondent aux parties en voie de cicatrisation et les tissus d'épidermisation de couleur rose, constituant la cicatrice. Le pourcentage respectif de ces tissus est un critère déterminant pour l'évaluation de l'état d'une escarre. Au cours du processus de guérison, les pourcentages de nécrose et de fibrine doivent diminuer alors que les pourcentages de bourgeonnement et de tissus d'épidermisation doivent augmenter. La surface globale de l'escarre est également un indicateur important de la gravité d'une escarre et doit évidemment diminuer au cours de la guérison.

Les indices volumétriques

Les escarres étant des plaies profondes voire cavitaires, la mesure de leur volume est déterminante. Alors qu'il est évident que l'aire d'une escarre doit diminuer au cours de la cicatrisation, l'analyse de son volume n'est pas aussi triviale. On définit le volume de l'escarre par l'espace compris entre la surface escarrifiée et la surface normale du corps non blessé. Bien que le volume doit diminuer de façon globale pendant le processus de cicatrisation, il doit également faire l'objet d'une analyse locale. En effet, certaines zones de l'escarre peuvent trop se régénérer et dépasser la surface normale. Il faut alors détecter ces parties de l'escarre et modifier les soins en conséquence.

Ces indicateurs sont mesurés par différentes techniques qui sont détaillées dans la section suivante.

Examen clinique des escarres

Les méthodes de suivi thérapeutique peuvent être partagées en deux groupes : les méthodes principalement manuelles et les méthodes basées sur la vision par ordinateur.

Méthodes manuelles

La méthode de suivi la plus naturelle repose sur la mémoire visuelle du médecin. Cependant, cette méthode souffre d'un manque d'objectivité lié à l'observateur. L'évolution d'une escarre étant un processus parfois très lent, il est difficile de constater une amélioration ou une détérioration à l'œil nu. De plus, l'observation n'est pas toujours réalisée par la même personne, introduisant ainsi une nouvelle source de subjectivité. Dans ces conditions, le suivi thérapeutique est très difficile. Des techniques plus ou moins évoluées ont alors été mises en place afin de pallier ce manque d'objectivité de la mesure.

La première de ces techniques est l'utilisation d'une simple règle graduée. En mesurant la longueur, la largeur et la profondeur de la plaie, il est possible d'estimer à la fois la surface et le volume de l'escarre. Cette technique souffre d'un manque de précision intrinsèque mais a l'avantage d'être simple et très peu onéreuse.

Une technique plus précise consiste à disposer un film transparent sur l'escarre afin de reproduire les contours de celle-ci. Le tracé est ensuite analysé afin d'estimer la surface. L'estimation peut être réalisée manuellement grâce à un quadrillage ou automatiquement par des outils informatiques onéreux. Cette technique présente des risques de contamination non négligeables dus au contact entre le film transparent et la plaie, et elle ne permet pas la mesure du volume. Pour obtenir le volume, il est possible de procéder à l'injection à l'aide d'une seringue de sérum physiologique sous un film plastique collé étanchement autour de l'escarre.

Vient ensuite la technique de moulage de la plaie. Le moulage est réalisé par injection d'alginate (pâte utilisée notamment pour les empreintes dentaires) dans la plaie. Cette technique est très précise mais son temps de réalisation est long et nécessite un apprentissage spécifique. De plus, son coût est important, constituant un obstacle à l'utilisation routinière de cette technique. Enfin, le contact nécessaire sur la plaie provoque des risques de contamination et une source de souffrance.

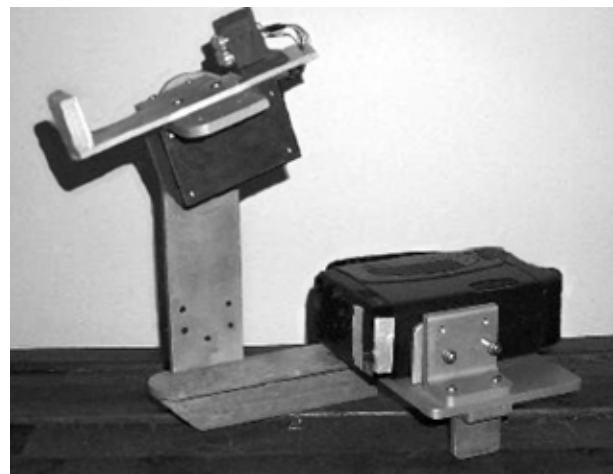
Afin d'obtenir des mesures précises et ne présentant aucun risque de contamination, des équipes de recherche ont tenté d'adapter des méthodes de vision par ordinateur à la problématique des escarres. Ces méthodes ont l'avantage de ne nécessiter aucun contact entre la plaie et le système de mesure.

Méthodes de vision par ordinateur

Les premiers travaux de vision par ordinateur appliquée à l'estimation du volume des escarres remontent au début des années 1990. Il s'agit de prototypes utilisant de la lumière structurée et un capteur d'image spécifique. Les travaux les plus anciens sont ceux de Jones et Plassmann [Jones, 1995] qui ont mis au point un système appelé MAVIS (Measurement of Area and Volume Instrument System) projetant une lumière structurée composée de codes couleurs (figure 2a). La projection de cette lumière sur la plaie est acquise par un capteur CCD solidaire du projecteur, ce qui permet par analyse des déformations des codes couleurs de reconstruire en trois dimensions la surface de la plaie. Leur système permet d'estimer le volume avec une précision de l'ordre de 5 %, satisfaisant ainsi les exigences de l'application. Un système similaire a été développé par Krouskop et al. en 2002 [Krouskop, 2002] (figure 2b) obtenant une précision légèrement supérieure de l'ordre de 3 %. Cependant, les performances de leurs systèmes se voient fortement dégradées pour les escarres de petites tailles, la résolution spatiale des éclairages structurés étant limitée.



(a)



(b)

Figure 2 : Système MAVIS de Jones et Plassmann (a) et système de Krouskop et al. (b).

S'attaquant à ce défaut, Malian et al. ont mis au point un système trinoculaire équipé d'un projecteur de texture [Malian, 2002] (figure 3). Baptisé MEDPHOS pour Medical Digital PHOtogrammetric System, cet appareil utilise les relations géométriques entre les 3 caméras pour mettre en correspondance les pixels des 3 images. A partir de ces correspondances et de la position connue des 3 caméras, la reconstruction 3D de la blessure est obtenue par une

simple triangulation. Leur méthode permet d'atteindre des précisions de l'ordre du micromètre sur la position des points 3D.

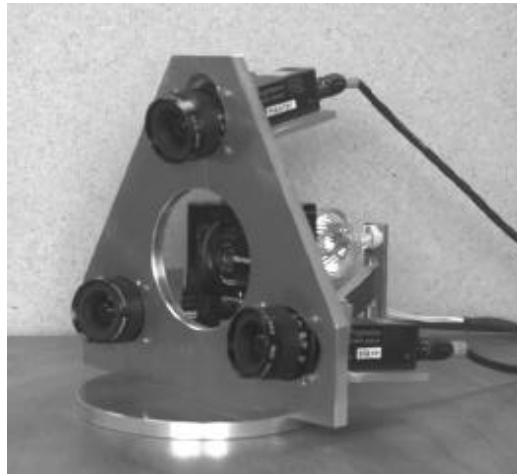


Figure 3 : Système MEDPHOS développé par Malian et al.

Tous les prototypes présentés précédemment partagent les mêmes inconvénients. Ce sont des systèmes calibrés qui nécessitent des recalibrages fréquents pour assurer la stabilité des mesures, et dont le champ de vue est limité par la configuration géométrique des capteurs. En outre, leur encombrement et leur poids rendent leur utilisation peu ergonomique. Enfin, le fait qu'ils soient des instruments spécifiques aux escarres empêche une production massive de ces appareils, qui restent donc chers.

Très récemment, pour pallier le manque d'ergonomie, l'équipe de Plassmann a mis au point un nouveau procédé qui se concrétise par un objectif particulier pour appareil photo numérique [MAVIS-II]. L'objectif transforme l'appareil photo en un capteur stéréoscopique à l'aide d'un jeu de miroirs et de lentilles (voir figure 4). Ainsi, en une pression sur le déclencheur de l'appareil photo, deux images sont acquises de deux points de vue légèrement différents. Un modèle 3D de l'escarre est obtenu ensuite par l'analyse des différences entre les deux images. Afin d'assurer une distance de prise de vue constante, et donc le bon recouvrement des demi-images gauche et droite, deux rayons lumineux convergents sont envoyés depuis l'appareil photo. Lorsque les faisceaux convergent en un point, la distance de prise de vue est bonne. L'ergonomie de ce système est clairement un avantage mais son prix d'environ 4000 € reste un frein à sa diffusion. De plus, sa précision est intrinsèquement limitée par le faible écartement entre les deux vues.



Figure 4 : Objectif développé pour le projet MAVIS-II.

ESCALE : une solution par la vision non calibrée

La critique des différentes techniques existantes de suivi thérapeutique permet de fixer le cahier des charges du projet ESCALE. Certaines contraintes d'utilisation doivent être respectées.

Tout d'abord, le coût de revient de la technique doit être le plus modeste possible. Un coût trop important freine l'adoption du système par les hôpitaux. Ensuite, le protocole de prise de vue doit être simple afin de ne pas imposer d'apprentissage aux utilisateurs, facilitant ainsi l'acceptation de la technique par le personnel de soin. Enfin, le système doit pouvoir réaliser toutes les mesures nécessaires au suivi thérapeutique, tant du point de vue colorimétrique et surfacique que du point de vue volumétrique. Les médecins estiment que l'erreur sur ces mesures ne doit pas dépasser 5 %. Tout cela doit être conçu dans un souci de confort et de sécurité pour le patient, impliquant que la mesure soit sans contact avec la plaie.

Au vu de ces contraintes, l'approche retenue est d'utiliser un appareil photo numérique standard. Associées à de puissants traitements informatiques, les images numériques permettent d'extraire les informations nécessaires au suivi thérapeutique. Cette solution a l'avantage d'allier la simplicité d'utilisation au faible coût, tout en garantissant le plus grand confort possible et la sécurité du patient. De plus, la souplesse de cadrage des appareils photo permet l'acquisition d'escarres de tailles très diverses. La tolérance de 5 % d'erreur semble suffisamment large pour que cette approche soit envisageable. En outre, l'utilisation de cet équipement standard largement diffusé ouvre de nouveaux horizons comme le suivi à domicile. Les photographies numériques acquises par l'entourage du patient peuvent être transmises par internet au médecin pour analyse. En utilisant plusieurs images acquises de différents points de vue, il est possible de combiner des informations bidimensionnelles et tridimensionnelles. L'approche 2D permet une analyse colorimétrique des différents tissus composant l'escarre alors que l'aspect 3D rendu possible en mettant en œuvre des méthodes de stéréovision non calibrée conduit au relevé volumétrique. C'est sur ce dernier point que porte le travail développé au cours de cette thèse.

Les techniques de stéréovision non calibrées sont rarement destinées à réaliser des mesures. Elles sont principalement utilisées pour des reconstructions 3D pour une visualisation qualitative sur écran. On peut notamment citer les travaux de l'Institut de Géodésie et de Photogrammétrie à Zurich [Gruen, 2004], qui utilise ces techniques pour reconstruire en trois dimensions les grands bouddhas d'Afghanistan, détruits par les Talibans en mars 2001. Les statues n'existant plus, leurs données numériques proviennent de photographies diffusées sur Internet ou de photos de voyages de touristes. De façon générale, plus le nombre d'images utilisées est important, meilleure sera la qualité de la reconstruction. Partant de ce principe, les séquences vidéo sont fréquemment utilisées en reconstruction 3D non calibrée. Un état de l'art de ces techniques appelées « structure from motion » est présentée dans [Pollefeys, 2002]. Nous n'avons pas retenu l'utilisation de vidéos pour notre projet, car l'espace de stockage numérique serait trop important.

De par la nature de ces applications, il existe peu de travaux traitant de la précision de ces techniques. Une des contributions de cette thèse est d'évaluer la qualité de la reconstruction que l'on peut obtenir par ces techniques non calibrées. L'approche retenue pour traiter cette

problématique, est de réaliser une chaîne de reconstruction globale dont les différents algorithmes sont choisis et optimisés dans le but de réduire l'erreur commise sur le modèle 3D en sortie de chaîne.

Dans ces travaux, nous nous plaçons dans le cas d'une reconstruction tridimensionnelle non calibrée à partir de seulement deux vues. Le choix de se limiter à deux vues est justifié par plusieurs raisons. Premièrement, un nombre de vues limité est l'assurance d'un protocole d'acquisition simple, et d'un espace mémoire de stockage faible. Cela favorise alors des relevés 3D fréquents pour un meilleur suivi thérapeutique. Ensuite, utiliser seulement deux vues pour obtenir une erreur de mesure inférieure à 5 % constitue un véritable challenge.

Ce mémoire est divisé en deux parties. La première partie regroupe les chapitres 1 à 4 et présente la chaîne de reconstruction 3D qui a été mise en place pour cette application ainsi que les apports originaux aux différentes étapes. La seconde partie traite de l'évaluation de la qualité de cette chaîne par des simulations et expérimentations sur des images réelles.

- Le chapitre 1 présente le principe d'une reconstruction tridimensionnelle par les techniques de stéréovision non calibrée. Il dresse un état de l'art de ces méthodes et présente les différentes étapes nécessaires à toute reconstruction 3D. Celles-ci sont détaillées dans les chapitres suivants.
- Le chapitre 2 se focalise sur la problématique de l'estimation de la matrice fondamentale, une étape clé de la chaîne de reconstruction présentée au chapitre 1. Les méthodes existantes pour réaliser cette tâche sont expliquées, puis des tests sont réalisés afin de choisir le meilleur algorithme pour notre application.
- Le chapitre 3 traite du délicat problème de la mise en correspondance des images, étape préalable à toute reconstruction 3D. Un processus itératif original permet d'améliorer la densité de la mise en correspondance.
- Le chapitre 4 termine la description de la chaîne de reconstruction par les étapes d'autocalibrage, de triangulation et d'optimisation du modèle 3D par ajustement de faisceaux.
- Le chapitre 5 ouvre la seconde partie de ce mémoire avec une évaluation de la qualité des reconstructions 3D fournies par notre chaîne.
- Le chapitre 6 permet d'établir le protocole d'acquisition qui assure d'être dans la configuration optimale des prises de vue minimisant les erreurs de reconstruction 3D.
- Enfin, le chapitre 7 termine ce mémoire en évaluant la précision sur le volume calculé à partir des modèles 3D fournis par notre chaîne de reconstruction.

En conclusion, nous récapitulons les principales contributions de notre chaîne de reconstruction, les améliorations à apporter et les perspectives ouvertes par cette thèse.

PREMIÈRE PARTIE

-

LA CHAÎNE DE RECONSTRUCTION 3D

Chapitre 1

Principe de la stéréovision non calibrée

Ce chapitre présente les bases théoriques de la reconstruction tridimensionnelle non calibrée. Les différentes étapes nécessaires pour obtenir le modèle 3D sont passées en revue, définissant ainsi une chaîne de reconstruction 3D. Le cas particulier où seulement deux vues sont disponibles est également abordé.

Sommaire

1.1	Modèle géométrique de formation des images.....	23
1.1.1	Le modèle sténopé.....	25
1.2	Géométrie épipolaire.....	28
1.2.1	Relations géométriques.....	28
1.2.2	Relations analytiques.....	29
1.3	Calibrage des capteurs d'image.....	30
1.3.1	Méthodes de calibrage.....	30
1.3.1.1	Calibrage avec un objet de référence ou mire.....	30
1.3.1.2	Auto-calibrage.....	31
1.3.2	Stratification des reconstructions 3D.....	31
1.3.2.1	Reconstruction projective.....	31
1.3.2.2	Reconstruction affine.....	32
1.3.2.3	Reconstruction métrique.....	32
1.3.2.4	Reconstruction euclidienne.....	34
1.3.3	Définition du problème de l'auto-calibrage et méthodes de résolution.....	34
1.3.3.1	Approches directes.....	35
1.3.3.2	Approches stratifiées.....	35
1.3.4	Cas de deux vues.....	36
1.4	Triangulation.....	37
1.5	Optimisation par ajustement de faisceaux.....	37
1.6	Conclusion.....	38

La vision humaine utilise de façon innée la stéréovision. Lorsque nous regardons un objet réel, l'image captée par l'œil droit est légèrement différente de celle perçue par l'œil gauche. Ce léger décalage entre les deux yeux nous permet de percevoir le relief, même si nous n'avons aucune connaissance a priori sur l'objet (fig. 1.1). Cependant, ce n'est pas la seule technique mise en œuvre par le cerveau car un grand nombre d'informations a priori peut nous renseigner sur le relief et la distance de l'objet. C'est cette connaissance a priori des formes et des tailles des objets qui nous permet d'appréhender le relief des objets sur des photographies, où le processus de triangulation n'est plus possible car la scène est plane.

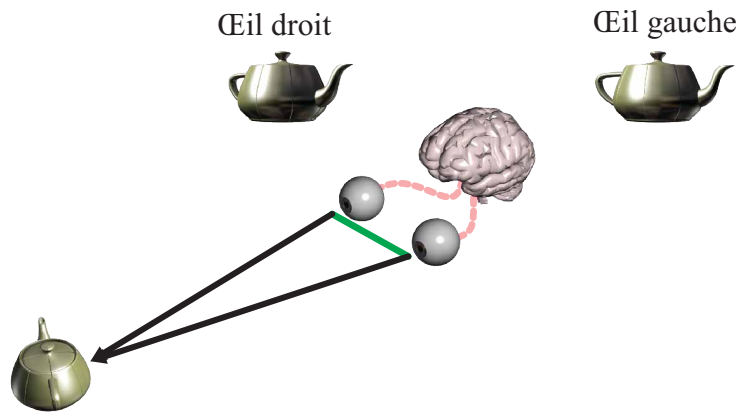


Figure 1.1 : Vision humaine.

Ce chapitre présente les bases théoriques de la stéréovision par ordinateur. Dans un premier temps, le modèle géométrique de formation des images est présenté. Puis les relations géométriques entre deux images sont détaillées. Avant de pouvoir réaliser la triangulation des éléments mis en correspondance entre les deux images, il faut connaître les positions relatives et les caractéristiques des capteurs d'image. C'est l'objet du calibrage, présenté en troisième section de ce chapitre avant le processus de triangulation. Pour finir, une étape d'optimisation du modèle 3D, appelée ajustement de faisceaux, est réalisée.

1.1 Modèle géométrique de formation des images

Un appareil photographique réalise la projection de la scène 3D sur le plan du capteur d'image (fig. 1.2). Les rayons lumineux en provenance de la scène sont focalisés sur le plan image par l'objectif. La modélisation optique la plus simple d'un objectif est le modèle de la lentille mince convergente, définie par sa distance focale f . Les rayons lumineux qui arrivent parallèles sur la lentille convergent tous en un point sur un plan perpendiculaire à l'axe optique, situé à la distance f du centre optique de la lentille (fig. 1.3). Ce plan est appelé plan focal, c'est sur celui-ci que se forme l'image des objets situés à une distance infinie de la lentille. Lorsque l'objet n'est pas situé à l'infini, l'image se forme en arrière du plan focal. Un dispositif de mise au point composé de lentilles permet de replacer l'image sur le plan focal.

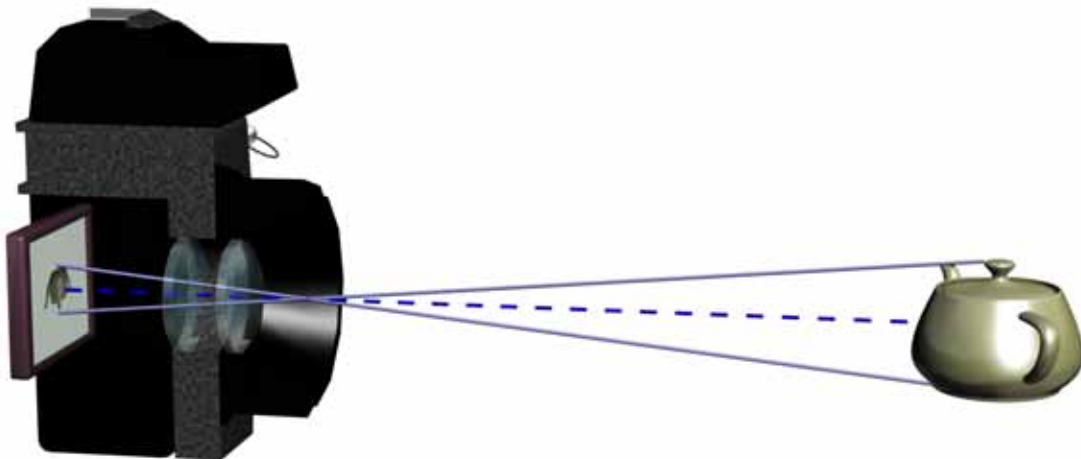


Figure 1.2 : Création d'une image par un appareil photo.

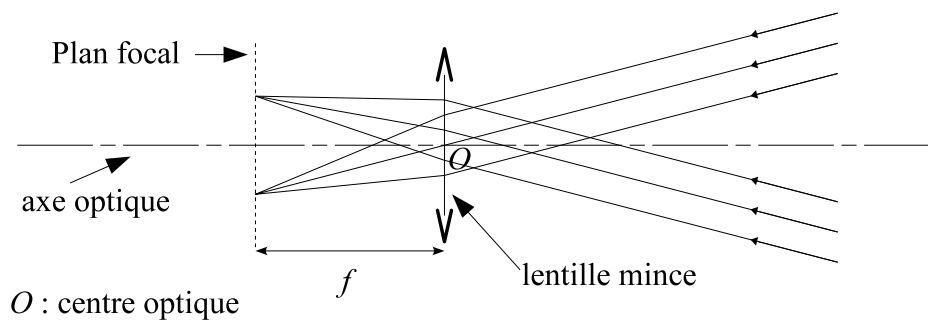


Figure 1.3 : Lentille mince convergente.

Une autre caractéristique des lentilles minces est que les rayons lumineux passant par le centre optique ne sont pas déviés. Tous ces rayons qui peuvent provenir de n'importe quel point de l'objet concourent en un seul (le centre optique) puis forment l'image sur le plan focal. Historiquement, c'est ce principe qui a été mis en œuvre par les premiers appareils photos qui étaient dépourvus d'objectif, et restent utilisés par certains passionnés et artistes. Il s'agit simplement d'une chambre noire percée d'un minuscule trou dans une de ses faces, réalisé par exemple grâce à une aiguille (fig. 1.4). Ce trou, appelé sténopé, a donné son nom au modèle sténopé, très simple, mais suffisant en vision par ordinateur pour caractériser la formation d'une image par un appareil photo. Ce modèle est encore appelé « pinhole » pour « trou d'épingle ». En pratique, la précision des images y reste limité intrinsèquement par le phénomène de diffraction.

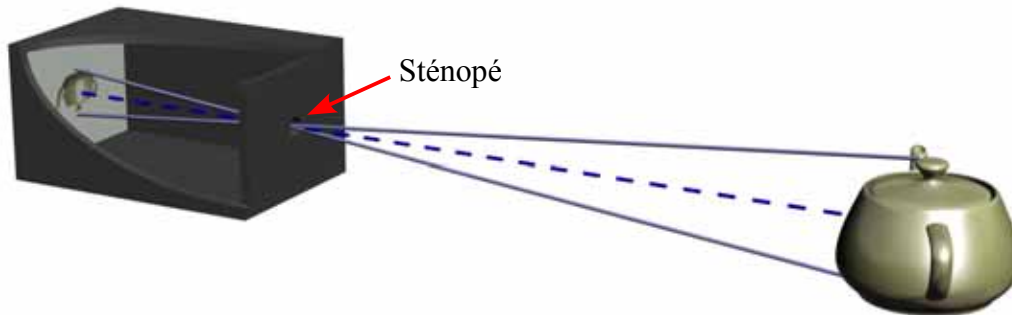


Figure 1.4 : Appareil sténopé.

1.1.1 Le modèle sténopé

Il s'agit de trouver les relations mathématiques entre un point dans l'espace et son image sur le capteur. On désigne par (O, x, y, z) le repère lié au capteur d'image. L'origine du repère est placée sur le centre optique, l'axe z le long de l'axe optique, les z positifs allant vers l'objet. Les axes x et y correspondent respectivement aux directions horizontale et verticale de l'image et forment un trièdre direct avec l'axe z . L'image se forme donc sur le plan perpendiculaire à l'axe optique, situé à la coordonnée $z = -f$. L'image produite à cet endroit est inversée car située derrière le centre optique. La même image non inversée est située sur le plan parallèle à la coordonnée $z = f$. Le dispositif est alors celui représenté par la figure 1.5.

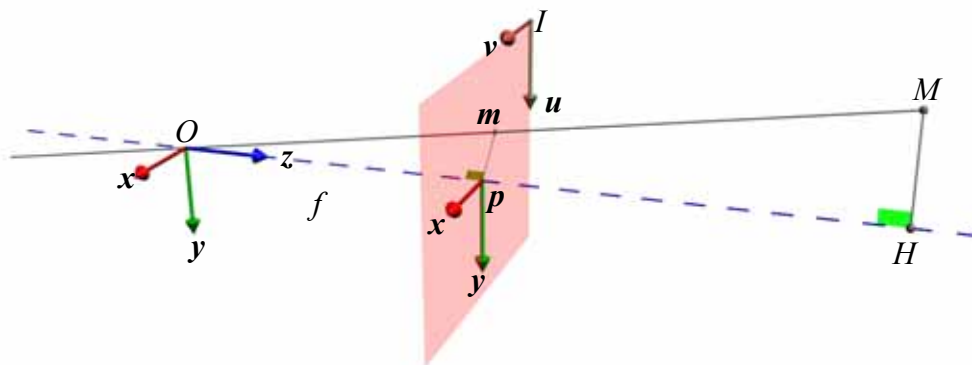


Figure 1.5 : Définition des repères pour le modèle sténopé.

Le point p d'intersection entre le plan image et l'axe optique est appelé point principal. Définissons enfin le point 3D M de coordonnées $(x, y, z)^T$, son image m et sa projection orthogonale H sur l'axe optique. Dans la configuration des triangles semblables Opm et OHM , les coordonnées du point m sont :

$$\mathbf{m}_{/(O, x, y, z)} = \begin{pmatrix} f \frac{x}{z} \\ f \frac{y}{z} \\ f \end{pmatrix} \quad (1.1)$$

En munissant le plan image du repère (p, x, y) , les coordonnées de m dans l'image sont directement déduites de l'équation 1.1 en supprimant la troisième coordonnée. Cependant, ce repère n'est pas idéal. En effet, les coordonnées de m devraient être exprimées en pixels par rapport au coin supérieur gauche I dans le repère image (I, u, v) . Pour cela, il suffit de réaliser un changement de repère :

$$\mathbf{m}_{I(I, u, v)} = \begin{pmatrix} f \frac{x}{z} k_u + u_0 \\ f \frac{y}{z} k_v + v_0 \end{pmatrix} \quad (1.2)$$

où (u_0, v_0) sont les coordonnées en pixels du point p dans le repère (I, u, v) et k_u et k_v sont respectivement les facteurs de conversion horizontal et vertical. Ils permettent par exemple de passer des coordonnées exprimées en millimètres à des coordonnées en pixels. Dans ce cas, l'unité pour ces valeurs serait le pixel/mm.

Toutes ces relations peuvent s'exprimer sous forme matricielle en utilisant les coordonnées homogènes. En coordonnées homogènes, un point dans un espace de dimension N est défini par un vecteur de dimension $N+1$ dont les N premières composantes sont les coordonnées du point définies à un facteur d'échelle près non nul, et la $N+1^{\text{ème}}$ valeur est ce facteur d'échelle. Ainsi, en dimension 3, le point M précédent est aussi bien défini par le vecteur $(x, y, z, 1)^T$ que par le vecteur $(s \cdot x, s \cdot y, s \cdot z, s)^T$ pour $s \neq 0$. Par la suite, le signe \propto signifiera « égal à un facteur près non nul ».

Les coordonnées homogènes permettent donc d'écrire directement la projection du point M sur l'image par l'équation :

$$\mathbf{m} \propto \begin{pmatrix} f k_u & 0 & u_0 & 0 \\ 0 & f k_v & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} M \quad (1.3)$$

La sous-matrice constituée des 3 premières colonnes de la matrice de l'équation 1.3 est appelée matrice des paramètres intrinsèques du capteur d'image. Dans la littérature, elle est parfois rencontrée avec un paramètre supplémentaire qui est l'angle entre les vecteurs u et v . Cet angle n'est pas forcément égal à 90° en fonction des technologies utilisées pour fabriquer le capteur. Ainsi, la définition de cette matrice devient :

$$A = \begin{pmatrix} \alpha_u & -\alpha_u \cot(\phi) & u_0 \\ 0 & \frac{\alpha_v}{\sin(\phi)} & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.4)$$

où $\alpha_u = f k_u$ et $\alpha_v = f k_v$, et ϕ est l'angle entre les axes u et v .

L'équation 1.3 n'est valable que si les coordonnées de M sont exprimées dans le repère attaché au capteur. Si ce n'est pas le cas, il faut opérer un changement de repère par une isométrie de l'espace, c'est-à-dire la combinaison d'une translation et d'une rotation. La figure

1.6 schématise cette transformation. La première étape consiste à translater le monde de sorte que le capteur d'image C vienne sur l'origine O , puis dans un second temps, d'effectuer la rotation qui fera coïncider les axes.

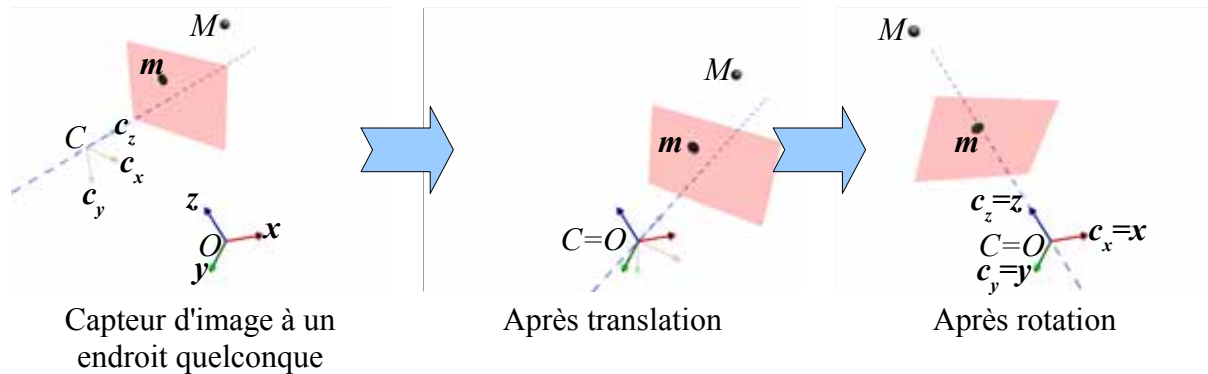


Figure 1.6 : Changement de repère pour placer le capteur d'image sur l'origine.

Mathématiquement, et toujours en coordonnées homogènes, cette transformation s'écrit :

$$\begin{pmatrix} R & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} t \\ M \end{pmatrix} \quad (1.5)$$

où R est la matrice de rotation de taille 3×3 qui décrit l'orientation du capteur et t le vecteur de translation du capteur vers l'origine.

En prenant en compte le changement de repère de l'équation 1.5 et la projection effectuée par l'équation 1.3, les coordonnées du point m sont donc calculées à partir de celles M par l'équation suivante :

$$m \propto A(R|Rt)M \quad (1.6)$$

Cette équation définit la matrice P de projection du capteur d'image :

$$P \propto A(R|Rt) \quad (1.7)$$

où A est la matrice composée des 5 paramètres intrinsèques et $(R|Rt)$ est construite à partir des paramètres extrinsèques du capteur d'image. Bien que la matrice des paramètres extrinsèques soit de taille 3×4 , les paramètres extrinsèques sont au nombre de 6, car la rotation est complètement déterminée par seulement 3 paramètres comme les angles d'Euler. Le modèle sténopé est entièrement caractérisé par cette matrice P et contient donc 11 paramètres.

Les projections de la scène sur des plans images différents introduisent des relations entre ces images.

1.2 Géométrie épipolaire

Les relations géométriques existantes entre deux prises de vue constituent la géométrie épipolaire. Ces liens entre les images peuvent être appréhendés de façon géométrique ou analytique.

1.2.1 Relations géométriques

La figure 1.7 montre un système stéréoscopique composé de deux capteurs d'image disposés en C_1 et C_2 . Un point M dans l'espace se projette respectivement en m_1 et m_2 sur les images 1 et 2. Le point M et les deux centres optiques C_1 et C_2 forment un plan Π , appelé plan épipolaire, qui coupe les deux plans images en deux droites l_1 et l_2 nommées droites épipolaires. Ces deux droites coupent la ligne de base (C_1C_2) respectivement en deux points e_1 et e_2 . Ces points sont fixes quelque soit le point M générateur, comme l'atteste la construction réalisée à partir du point N . Les points e_1 et e_2 sont appelés les épipoles des images. L'épipole e_1 est l'image de C_2 sur le premier capteur d'image, inversement l'épipole e_2 est l'image de C_1 sur le second capteur d'image.

Lorsque la géométrie épipolaire est connue, elle peut aider la mise en correspondance. En ne connaissant que le projeté m_1 , son point 3D associé se trouve quelque part le long du rayon (C_1m_1). Donc le correspondant de m_1 sur l'image 2 se trouve sur la projection de cette droite, c'est-à-dire le long de la droite l_2 . Ainsi l'espace de recherche du correspondant est réduit à une recherche monodimensionnelle, diminuant ainsi les temps de calcul et les risques d'erreur. La relation entre un point et sa droite épipolaire dans la seconde image peut être définie analytiquement.

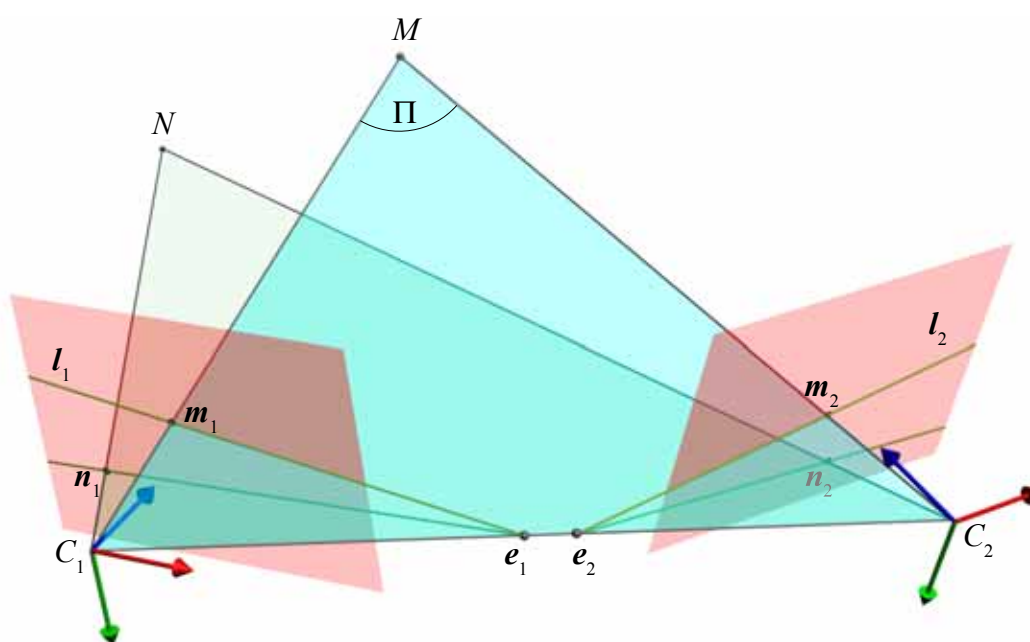


Figure 1.7 : Géométrie épipolaire.

1.2.2 Relations analytiques

Il existe différentes méthodes pour déterminer la relation analytique entre un point \mathbf{m}_1 et sa droite épipolaire associée \mathbf{l}_2 [Hartley, 2004]. Nous détaillons ici une dérivation algébrique à partir des matrices de projection, issue des travaux de Xu et Zhang parus en 1996 [Xu, 1996].

Soit P_1 et P_2 les matrices de projection associées aux capteurs d'image. Le rayon $(C_1\mathbf{m}_1)$ est obtenu algébriquement par la résolution de l'équation $P_1 M \propto \mathbf{m}_1$. La solution n'est évidemment pas unique et donne une famille de solutions représentée par l'ensemble des points de la droite $(C_1\mathbf{m}_1)$. La famille des points M est alors définie par :

$$M(\lambda) \propto P_1^+ \mathbf{m}_1 + \lambda C_1 \quad (1.8)$$

où $P_1^+ = P_1^T (P_1 P_1^T)^{-1}$ est la matrice pseudo inverse de P_1 et λ un scalaire. On peut choisir arbitrairement deux points sur cette droite pour déterminer leurs projections sur la seconde image et ainsi déterminer la droite \mathbf{l}_2 . Deux points particuliers sont $P_1^+ \mathbf{m}_1$ pour $\lambda=0$ et le centre de projection C_1 pour $\lambda=\infty$. Les images de ces deux points par le second capteur sont respectivement $P_2 P_1^+ \mathbf{m}_1$ et $P_2 C_1 \propto \mathbf{e}_2$. La droite épipolaire \mathbf{l}_2 joint ces deux points et est définie par le produit vectoriel $\mathbf{l}_2 \propto \mathbf{e}_2 \times (P_2 P_1^+ \mathbf{m}_1)$. Ce produit vectoriel peut s'écrire de façon matricielle par l'équation :

$$\mathbf{l}_2 \propto [\mathbf{e}_2]_{\times} P_2 P_1^+ \mathbf{m}_1 \propto F \mathbf{m}_1 \quad (1.9)$$

où la notation $[\mathbf{x}]_{\times}$ signifie $\begin{pmatrix} 0 & -x_3 & x_2 \\ x_3 & 0 & -x_1 \\ -x_2 & x_1 & 0 \end{pmatrix}$.

Cette équation permet de définir la matrice $F \propto [\mathbf{e}_2]_{\times} P_2 P_1^+$, appelée matrice fondamentale. Cette matrice représente une application d'un espace projectif à 2 dimensions dans un espace à 1 dimension, elle est donc de rang 2. Comme le point \mathbf{m}_2 appartient à la droite \mathbf{l}_2 , le produit $\mathbf{m}_2^T \mathbf{l}_2$ est nul. Ainsi, on obtient la relation suivante entre deux points mis en correspondance :

$$\mathbf{m}_2^T F \mathbf{m}_1 = 0 \quad (1.10)$$

Etant donné que le choix de l'image de référence est arbitraire, le même raisonnement peut être fait en inversant le rôle des images. Dans ce cas, la matrice fondamentale obtenue aurait été la transposée de celle de l'équation 1.10. Ainsi, de la même façon que $\mathbf{l}_2 \propto F \mathbf{m}_1$, on a $\mathbf{l}_1 \propto F^T \mathbf{m}_2$.

La matrice fondamentale est entièrement définie par l'équation 1.10 à partir d'un ensemble suffisant de correspondances de points entre les deux images. Le chapitre 2 traite des méthodes d'estimation de la matrice fondamentale.

Enfin, cette matrice intègre à la fois les paramètres intrinsèques et extrinsèques des capteurs d'image car elle est construite à partir des deux matrices de projections. Il est d'ailleurs possible de déterminer ces paramètres à partir de la matrice fondamentale.

1.3 Calibrage des capteurs d'image

Le principe du calibrage des caméras est connu depuis des décennies en photogrammétrie [Slama, 1980]. Il consiste à déterminer la relation entre les coordonnées tridimensionnelles d'un point de la scène observée et sa projection dans l'image. Autrement dit, il s'agit de trouver les paramètres intrinsèques et extrinsèques des capteurs d'image. C'est une étape incontournable dès que l'on veut appréhender le monde 3D pour en faire une reconstruction euclidienne, localiser un objet ou guider un robot mobile, par exemple.

Une première famille de méthodes utilise un objet de référence ou mire, de dimensions parfaitement connues. La seconde, au contraire, s'affranchit de cette contrainte ; on parle alors d'auto-calibrage.

1.3.1 Méthodes de calibrage

1.3.1.1 Calibrage avec un objet de référence ou mire

Les premières méthodes de calibrage nécessitent l'utilisation d'une mire tridimensionnelle de grande précision (par exemple un cube), équipement coûteux et fastidieux à mettre en œuvre. En effet, la diversité des applications nécessite de recourir à plusieurs mires adaptées à chaque champ de vue. De plus, la précision de l'estimation des paramètres dépend du soin apporté aux mesures faites dans l'image mais aussi à la fabrication de la mire. Une première amélioration est proposée par [Lavest, 1998] sous la forme d'une optimisation multi-images par ajustement de faisceaux, qui permet de réestimer conjointement les paramètres et la structure de la mire à partir de valeurs initiales assurant la convergence.

Une seconde génération de méthodes propose un calibrage utilisant des mires planes [Tsai, 1986], [Sturm, 1999], [Zhang, 2000]. Beaucoup plus faciles à réaliser, elles représentent aujourd'hui un standard en vision par ordinateur. Plusieurs outils sont disponibles sur internet comme par exemple [Bouguet, 2006]. Ils permettent de réaliser un calibrage à partir d'une série d'images d'une mire prises sous des points de vue différents. Le calibrage multi-images conduit à des résultats plus stables. Enfin, quelques chercheurs se sont attachés à simplifier encore les mires jusqu'à des objets linéaires [Zhang, 2004], [Hammarstedt, 2005].

1.3.1.2 Auto-calibrage

Utiliser une mire peut vite s'avérer contraignant hors d'un environnement de laboratoire. Une solution à ce problème a été initialement apportée par [Faugeras, 1992] et [Hartley, 1992] pour des paramètres fixes. Ils proposent un autocalibrage du capteur d'image qui repose sur les équations de Kruppa déjà connues en photogrammétrie. Le calibrage s'effectue directement à partir d'un ensemble de correspondances entre les images sans requérir la présence d'une mire. Ces résultats ont ensuite été étendus aux cas d'un ou plusieurs paramètres variables, notamment la focale [Pollefeys, 1998].

1.3.2 Stratification des reconstructions 3D

Selon les informations à extraire d'une reconstruction, le calibrage peut varier entre 4 niveaux : projectif puis affine puis métrique et enfin euclidien. Cette notion de hiérarchisation de l'espace 3D a été définie par Faugeras [Faugeras, 1995] et depuis utilisée par de nombreux scientifiques, comme Pollefeys [Pollefeys, 1999]. Le figure 1.8 montre les représentations d'un cube dans la hiérarchie des différentes reconstructions. Le niveau d'une reconstruction définit le type de transformation géométrique liant l'objet et son modèle 3D.



Figure 1.8 : Représentation d'un cube selon les niveaux géométriques de reconstruction 3D.

1.3.2.1 Reconstruction projective

Le premier niveau de cette hiérarchie est la reconstruction projective, où la reconstruction est fidèle à la réalité à une transformation projective près. Une transformation projective est représentée dans l'espace 3D par une matrice de taille 4×4 non singulière. Une telle transformation dispose de 15 degrés de liberté, puisque définie à un facteur d'échelle près. Il est établi qu'il est possible d'obtenir une reconstruction projective à partir d'un ensemble de correspondances entre deux vues non calibrées, grâce à une décomposition de la matrice fondamentale [Hartley, 1992]. Lors d'une telle reconstruction, seul le birapport est conservé.

1.3.2.2 Reconstruction affine

La reconstruction affine est fidèle à la réalité à une transformation affine près. En coordonnées homogènes, une transformation affine est définie par une matrice de la forme :

$$T_{A\infty} \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.11)$$

Ce type de transformation présente 12 degrés de liberté. Pour passer d'une reconstruction projective à une reconstruction affine, il faut déterminer un plan particulier appelé le plan à l'infini. Dans l'espace projectif, ce plan est défini par l'ensemble des points dont le facteur d'échelle des coordonnées homogènes est nul. Ainsi, les coordonnées du plan à l'infini sont $\Pi_\infty = (0 \ 0 \ 0 \ 1)^T$. Ce plan est perçu comme contenant les points de fuite de deux droites parallèles. Dans l'espace projectif, ces points particuliers ne sont pas obligatoirement sur le plan de coordonnées $(0 \ 0 \ 0 \ 1)^T$, autrement dit, le plan à l'infini n'est pas à sa position canonique. Il suffit alors de déterminer les coordonnées du plan contenant les points limites puis d'estimer la transformation projective qui l'appliquera à sa position canonique. Une fois les coordonnées du plan déterminées dans l'espace projectif, la transformation est de la forme :

$$T_{PA\infty} \begin{pmatrix} A_{3 \times 3} & \mathbf{0}_3 \\ \pi_\infty^T & 1 \end{pmatrix} \quad \text{avec } \det A_{3 \times 3} \neq 0 \quad (1.12)$$

où $\mathbf{0}_3$ est le vecteur nul de dimension 3, π_∞^T le vecteur composé des 3 premières coordonnées du plan lorsque la quatrième est rendue égale à 1. Le choix de la matrice $A_{3 \times 3}$ étant arbitraire, on peut choisir la matrice identité. En plus du birapport, une transformation affine conserve les rapports de distances selon une même direction ainsi que le parallélisme. Seuls ces éléments peuvent alors être mesurés sur une reconstruction affine.

1.3.2.3 Reconstruction métrique

La couche directement supérieure concerne la reconstruction métrique. Ce type de reconstruction est fidèle à la scène réelle à une similarité près. Les similarités sont des compositions de rotations et de translations complétées par une mise à l'échelle isotropique. Elles comportent donc 7 degrés de liberté : 3 pour la rotation, 3 pour la translation et 1 pour le facteur d'échelle. Dans l'espace projectif, ces transformations sont définies par :

$$T_M \propto \begin{pmatrix} \sigma r_{11} & \sigma r_{12} & \sigma r_{13} & t_x \\ \sigma r_{21} & \sigma r_{22} & \sigma r_{23} & t_y \\ \sigma r_{31} & \sigma r_{32} & \sigma r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.13)$$

où les éléments r_{ij} constituent une matrice de rotation R vérifiant donc $R^{-1} = R^T$, σ est le facteur d'échelle et $(t_x \ t_y \ t_z)^T$ le vecteur de translation. Lorsque la reconstruction est métrique, les éléments mesurables sont les distances relatives et les angles.

Pour permettre une reconstruction métrique, il faut identifier une conique particulière appelée la conique absolue. Ce concept est plus abstrait que le plan à l'infini. Lorsque la reconstruction est métrique, cette conique, notée Ω , est définie par l'ensemble des points $M = (x \ y \ z \ w)^T$ tels que :

$$x^2 + y^2 + z^2 = 0 \quad \text{et} \quad w = 0 \quad (1.14)$$

Le recours à deux équations pour définir cette conique est peu pratique. L'objet dual de cette conique dans l'espace 3D est alors utilisé pour pallier ce problème. Il s'agit de la quadrique duale absolue, notée Q_∞^* , composée des plans tangents à la conique absolue. Son expression canonique est la suivante :

$$Q_\infty^* \propto \begin{pmatrix} Id_{3 \times 3} & \mathbf{0}_3 \\ \mathbf{0}_3^T & 0 \end{pmatrix} \quad (1.15)$$

Si seul le plan à l'infini est considéré, alors la conique absolue peut se représenter mathématiquement par une matrice. De façon générale, une conique se représente dans le plan par une matrice C de taille 3×3 définissant l'ensemble des points \mathbf{m} par l'équation :

$$\mathbf{m}^T C \mathbf{m} = 0 \quad (1.16)$$

On définit également la conique duale de C , notée C^* , par l'ensemble des droites \mathbf{l} tangentes à C respectant l'équation :

$$\mathbf{l}^T C^* \mathbf{l} = 0 \quad (1.17)$$

où C^* est la matrice adjointe de C . Lorsque la conique n'est pas dégénérée, alors C est de rang 3 et $C^* \propto C^{-1}$.

Dans le cas de la conique absolue, l'expression mathématique de Ω dans le plan à l'infini est :

$$\omega_\infty \propto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (1.18)$$

La conique absolue se projette également en une conique dans les images. Il est plus simple d'utiliser la duale de la conique absolue pour réaliser cette projection, sachant qu'une seule

matrice la définit. Il vient alors que la duale de l'image de la conique absolue est donnée par la relation suivante pour le capteur d'image i :

$$\omega_i^* \propto P_i Q_\infty^* P_i^T \quad (1.19)$$

Lorsque Q_∞^* est à sa position canonique, la relation suivante est vérifiée [Hartley, 2004] :

$$\omega_i^* \propto A_i A_i^T \quad (1.20)$$

où A_i est la matrice des paramètres intrinsèques du capteur définie à l'équation 1.4.

Pour passer d'une reconstruction affine à une reconstruction métrique, il faut déterminer la transformation affine qui placera la conique absolue sur sa position canonique. Cette transformation peut être choisie comme étant [Hartley, 1994] :

$$T_{AM} \propto \begin{pmatrix} A^{-1} & \mathbf{0}_3 \\ \mathbf{0}_3^T & 1 \end{pmatrix} \quad (1.21)$$

1.3.2.4 Reconstruction euclidienne

Ce type de reconstruction diffère du niveau métrique par le fait que l'échelle est fixée. Ainsi, les distances peuvent être mesurées de façon absolue. La reconstruction est connue à une transformation rigide près, c'est-à-dire la composition d'une rotation et d'une translation. Ces transformations n'ont donc que 6 degrés de liberté : 3 pour la rotation et 3 pour la translation, et s'écrivent sous la forme :

$$T_E \propto \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (1.22)$$

où les éléments r_{ij} constituent une matrice de rotation.

Pour passer d'une reconstruction affine à une reconstruction euclidienne, il suffit de réaliser une mise à l'échelle. Le facteur d'échelle ne peut être connu qu'en disposant d'une distance de référence dans la scène.

1.3.3 Définition du problème de l'auto-calibrage et méthodes de résolution

Soient m images de la scène 3D à reconstruire. A partir des matrices fondamentales reliant les différentes images, il est donc possible de déterminer une reconstruction projective de cette scène notée $\{P_i, M_j\}$ où P_i est la matrice de projection de la $i^{\text{ème}}$ image et M_j le $j^{\text{ème}}$

point 3D. Le but de l'auto-calibrage est alors de déterminer la transformation projective T_{PM} de l'espace 3D de sorte que $\{P_i T_{PM}, T_{PM}^{-1} M_j\}$ soit une reconstruction métrique. Dans ce cas, la matrice $P_i T_{PM}$ peut se décomposer selon l'équation 1.7 en $P_i T_{PM} = A_i (R_i | R_i \mathbf{t}_i)$. En fixant des contraintes sur les paramètres des caméras, c'est-à-dire sur A_i et/ou R_i et \mathbf{t}_i , il est possible de déterminer la transformation T_{PM} .

Les méthodes d'auto-calibrage peuvent être classées en deux types : celles qui déterminent directement la transformation T_{PM} et celles qui la déterminent par une approche stratifiée.

1.3.3.1 Approches directes

Parmi ce type d'approches, une méthode consiste à déterminer la quadrique duale absolue dont l'image est la conique duale absolue (équation 1.19). L'équation 1.20 permet de fixer des contraintes sur ω_i^* à partir de celles sur A_i . L'idée est alors de transférer ces contraintes vers Q_∞^* [Triggs, 1997]. Une fois la matrice Q_∞^* connue, il suffit de déterminer la transformation T_{PM} qui replacera la quadrique duale absolue à sa position canonique.

Une méthode analogue récente utilise la quadrique absolue de droites à la place de la quadrique duale absolue [Valdés, 2006]. La quadrique absolue de droites est définie par l'ensemble des droites coupant la conique absolue. L'utilisation de cette nouvelle quadrique débouche sur un algorithme linéaire pour la détermination du cadre métrique de la reconstruction.

Enfin, la méthode qui semble être la plus répandue se base sur les équations de Kruppa. Ces équations ont été appliquées à la vision par ordinateur par Faugeras et al. en 1992 [Faugeras, 1992]. Elles ont fait l'objet de nombreuses recherches depuis. On peut citer par exemple [Zeller, 1996], [Lourakis, 1999], [Lourakis, 2000a], [Ma, 2000] et [Sturm, 2001]. Les équations de Kruppa sont définies pour un couple d'images par l'équation (voir [Hartley, 2004] pour une démonstration) :

$$[\mathbf{e}_2]_\times \omega_2^* [\mathbf{e}_2]_\times = F \omega_1^* F^T \quad (1.23)$$

La connaissance de l'épipole \mathbf{e}_2 dans la seconde image et de la matrice fondamentale F permet de déterminer les images ω_1^* et ω_2^* de la conique duale absolue, et ainsi de retrouver les paramètres intrinsèques par l'équation 1.20, grâce à la factorisation de Cholesky.

1.3.3.2 Approches stratifiées

Les méthodes stratifiées procèdent en deux étapes pour auto-calibrer les capteurs. Dans un premier temps, le plan à l'infini est déterminé, permettant ainsi de se placer dans un cadre affine. Puis une fois dans ce cadre, l'estimation de la matrice A des paramètres intrinsèques est simplifiée car il existe une solution linéaire. La détermination du plan à l'infini peut se faire en

appliquant une méthode appelée contrainte du module, lorsque les paramètres intrinsèques sont fixes entre toutes les vues. Pour plus d'informations sur cette méthode, le lecteur peut se référer à [Pollefeys, 1999].

1.3.4 Cas de deux vues

Le nombre d'images nécessaires pour obtenir une reconstruction métrique est déterminé par les connaissances a priori sur les paramètres intrinsèques des vues. Ces connaissances peuvent être de deux types : soit le paramètre est connu, soit il est inconnu mais fixe entre toutes les images. Pour déterminer la structure de l'objet dans un cadre métrique, il faut bloquer 8 degrés de liberté (15 projectifs moins 3 rotations moins 3 translations moins 1 pour l'échelle). Un paramètre connu apporte une contrainte par vue alors qu'un paramètre inconnu mais fixe en apporte une de moins. Ainsi, le nombre de vues nécessaires peut être déterminé grâce à la relation [Pollefeys, 1998] :

$$n \times (\#connus) + (n - 1) \times (\#fixes) \geq 8 \quad (1.24)$$

où n est le nombre de vues, $\#connus$, le nombre de paramètres intrinsèques connus, et $\#fixes$, le nombre de paramètres fixes. L'angle ϕ entre les axes de l'image est généralement connu et égal à $\pi/2$, ce qui permet d'obtenir une reconstruction métrique à partir de 8 vues. Dans le cas où uniquement deux vues sont disponibles, alors il faut connaître en plus la position du point principal et le rapport d'aspect (α_u/α_v). Le paramètre restant est la distance focale des images. La seule méthode possible permettant l'auto-calibrage à partir de seulement deux vues, est l'utilisation des équations de Kruppa. Sturm propose un algorithme clairement détaillé [Sturm, 2001] et étudié [Sturm, 2005] pour déterminer la distance focale des images dans le cas où ce paramètre est fixe.

Les algorithmes permettant de réaliser l'auto-calibrage présentent tous des singularités empêchant la détermination de la solution dans certaines conditions. Sturm a étudié ces configurations singulières dans un cadre général pour l'auto-calibrage [Sturm, 1997]. Dans le cas de deux vues dont tous les paramètres intrinsèques sont connus sauf la distance focale, l'auto-calibrage est impossible lorsque les axes optiques sont coplanaires [Sturm, 2002]. Ceci est très gênant puisque cette configuration de prise de vue est souvent adoptée par les capteurs stéréoscopiques et instinctive lors de prises de vue libres. Sturm a montré que lorsque la focale est fixe entre les deux prises de vue, alors la singularité ne s'obtient que pour un mouvement orbital (fig. 1.9). La configuration orbitale est respectée lorsque les axes optiques se coupent et que les positions des prises de vue sont équidistantes de cette intersection.

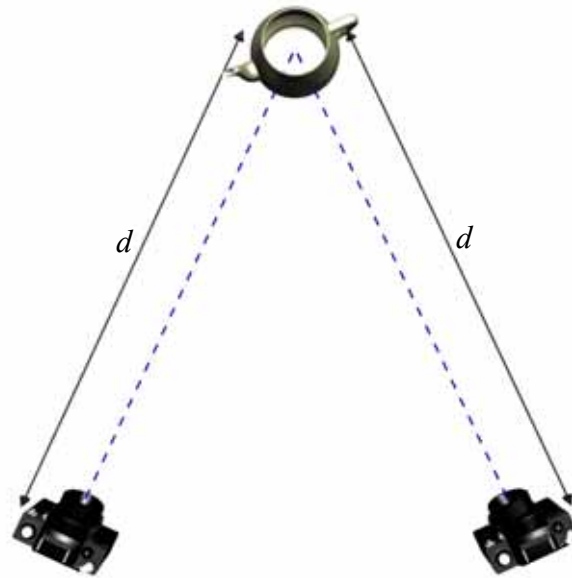


Figure 1.9 : Mouvement orbital.

1.4 Triangulation

La dernière étape d'une reconstruction 3D est la détermination de la structure de la scène. Pour ce faire, on calcule la droite de vue issue du centre optique et passant par chaque projection. Les droites de vue de deux projections homologues se coupent en un point 3D unique. Dans le cas où seulement deux images sont considérées, les deux rayons issus d'un couple de points appariés forment un triangle avec la ligne de base. Les centres optiques correspondant à deux sommets du triangle, la triangulation consiste à déterminer le troisième sommet à partir des directions données par les deux rayons lumineux.

Dans la pratique, les rayons lumineux ne se coupent pas rigoureusement du fait de la présence de bruit dans les images. Il s'agit alors de déterminer le point 3D qui satisfait au mieux la configuration des rayons lumineux selon un certain critère, dépendant du cadre géométrique de la reconstruction. Hartley et Sturm ont comparé les différentes méthodes existantes de triangulation dans [Hartley, 1997a]. Dans le cadre métrique ou euclidien, l'implémentation valable la plus simple est la méthode linéaire par moindres carrés, développée dans le chapitre 4.

1.5 Optimisation par ajustement de faisceaux

Cette dernière étape permet de raffiner le modèle 3D reconstruit ainsi que les paramètres des capteurs d'image. Il s'agit de trouver les matrices de projections P_i et l'ensemble des points 3D M_j qui minimisent l'erreur de reprojection :

$$\min_{P_i, M_j, i, j} \sum d(P_i M_j, \mathbf{m}_{ij}) \quad (1.25)$$

où \mathbf{m}_{ij} est le point détecté sur l'image i correspondant au point 3D M_j et $d(\hat{\mathbf{m}}, \mathbf{m})$ est la distance entre le point \mathbf{m} et son estimation par reprojection $\hat{\mathbf{m}}$.

L'algorithme couramment utilisé pour résoudre ce genre de problème de minimisation est l'algorithme de Levenberg-Marquardt. La descente doit être initialisée par une solution proche de la solution recherchée pour éviter de converger vers un minimum local. L'initialisation se fait grâce aux méthodes présentées aux sections précédentes. Le coût de calcul d'un tel algorithme dépend de la paramétrisation du problème. Le nombre de paramètres doit être minimal pour assurer une convergence rapide. De nombreuses paramétrisations existent et dépendent directement du problème de reconstruction 3D traité. La paramétrisation optimale diffère selon que les paramètres intrinsèques ou extrinsèques sont connus ou fixes et du cadre géométrique choisi (projectif, affine ou métrique) [Bartoli, 2001], [Malis, 2001].

D'un point de vue implémentation, Lourakis propose deux types d'algorithmes. Le premier est l'algorithme de Levenberg-Marquardt général, pouvant être utilisé pour tout problème de minimisation non linéaire [Lourakis, 2005]. Le second est une version optimisée pour le problème de l'ajustement de faisceaux en vision par ordinateur utilisant l'algorithme de Levenberg-Marquardt épars [Lourakis, 2006]. Les détails mathématiques de ce type d'implémentation sont disponibles dans [Lourakis, 2004].

1.6 Conclusion

Nous avons présenté dans ce chapitre les bases théoriques d'une reconstruction 3D à partir d'images en donnant au lecteur des pistes pour approfondir cette problématique. Le processus de reconstruction peut être schématisé par la figure 1.10. Cependant, lorsque l'on dispose de seulement deux images, sans aucune autre connaissance, plusieurs hypothèses sont nécessaires pour permettre l'auto-calibrage. Tout d'abord, les paramètres intrinsèques inconnus des capteurs d'image sont réduits à la distance focale, en fixant le rapport d'aspect des pixels à 1, l'angle entre les axes horizontal et vertical à 90° et le point principal au centre des images. Ensuite, le mouvement présent entre les deux capteurs d'image ne doit ni être orbital, ni une translation pure. Par ailleurs, pour permettre une reconstruction euclidienne, il est nécessaire de connaître une distance dans la scène.

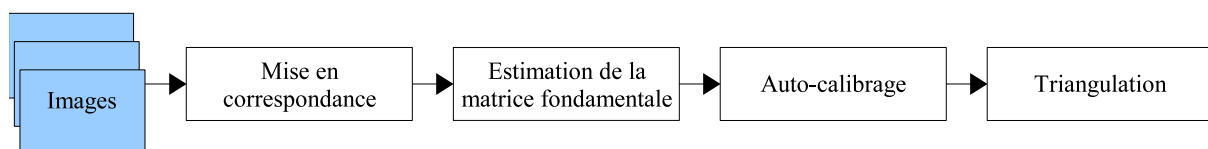


Figure 1.10 : Chaîne de reconstruction 3D à partir d'images non calibrées.

C'est sur ces choix que se base la suite de ce mémoire. L'étape au cœur de ce problème est l'estimation de la matrice fondamentale qui s'appuie sur la mise en correspondance. Le chapitre suivant est consacré entièrement à l'estimation de cette matrice car elle est utilisée dans notre processus de mise en correspondance, détaillé dans le chapitre 3.

Chapitre 2

Estimation de la matrice fondamentale

La matrice fondamentale étant l'élément clé de la stéréovision, son estimation doit être la plus précise possible. Ce chapitre passe en revue les méthodes d'estimation de cette matrice, puis aboutit sur le choix de la méthode la plus adéquate dans le cas d'une reconstruction 3D non calibrée.

Sommaire

2.1 Méthodes d'estimation de la matrice fondamentale.....	41
2.1.1 Caractéristiques communes à tous les algorithmes.....	42
2.1.2 La méthode des moindres carrés.....	43
2.1.3 La technique des moindres carrés orthogonaux.....	43
2.1.4 La méthode de Bookstein.....	44
2.1.5 La méthode itérative des moindres carrés repondérés.....	45
2.1.6 L'algorithme LMedS.....	46
2.1.7 L'algorithme MAPSAC.....	47
2.2 Analyse des estimations de la matrice fondamentale.....	48
2.2.1 La chaîne de reconstruction 3D.....	48
2.2.2 Protocole d'évaluation.....	49
2.2.3 Simulations.....	51
2.2.3.1 Analyse de la robustesse au bruit de localisation des points d'intérêt.....	51
2.2.3.2 Analyse de la robustesse aux faux appariements.....	54
2.3 Discussion sur le choix de la méthode pour la stéréovision non calibrée.....	56
2.4 Conclusion.....	58

À partir des points mis en correspondance entre les deux images, l'équation 1.10 nous montre qu'il est possible de calculer la matrice fondamentale. Dans des conditions réelles, le bruit présent dans les images interdit l'égalité stricte de cette équation. Il s'agit alors d'estimer la matrice fondamentale au mieux pour satisfaire cette équation. Différents critères existent pour cette estimation, ayant engendré autant d'algorithmes de résolution. Armangué et Salvi ont réalisé une étude de 19 algorithmes très répandus [Armangué, 2003]. Ils ont établi un classement en trois catégories des méthodes d'estimation de la matrice fondamentale : les méthodes linéaires, itératives et enfin robustes. Le critère utilisé pour comparer les matrices fondamentales estimées est l'erreur résiduelle définie par la distance moyenne séparant un point et sa droite épipolaire. Nous avons choisi une approche plus globale qui consiste à évaluer les méthodes en bout de chaîne par l'erreur 3D obtenue.

La première partie de ce chapitre présente les méthodes d'estimation de la matrice fondamentale et les algorithmes retenus dans notre cadre d'évaluation. La seconde partie traite ensuite des résultats obtenus avec chacun des algorithmes. Enfin, une troisième partie discute les choix réalisés pour le problème de la reconstruction tridimensionnelle non calibrée.

2.1 Méthodes d'estimation de la matrice fondamentale

L'estimation de la matrice fondamentale se fait par la résolution de l'équation 1.10. Pour simplifier les notations dans ce chapitre, les points sur la première image sont notés $\mathbf{m}_i = (u_i \ v_i \ w_i)^T$ et les points sur la seconde image sont notés $\mathbf{m}_j' = (u_j' \ v_j' \ w_j')^T$. Lorsque $i = j$, les points \mathbf{m}_i et \mathbf{m}_j' sont appariés. Avec ces notations, l'équation 1.10 peut se mettre sous la forme d'un système linéaire :

$$Z \mathbf{f} = \mathbf{0}_9 \quad (2.1)$$

avec

$$\mathbf{f} = (F_{11}, F_{12}, F_{13}, F_{21}, F_{22}, F_{23}, F_{31}, F_{32}, F_{33})^T \quad \text{où } F_{ij} \text{ est l'élément } (i, j) \text{ de } F \quad (2.2)$$

et

$$Z = \begin{pmatrix} u_1' u_1 & u_1' v_1 & u_1' w_1 & v_1' u_1 & v_1' v_1 & v_1' w_1 & w_1' u_1 & w_1' v_1 & w_1' w_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ u_n' u_n & u_n' v_n & u_n' w_n & v_n' u_n & v_n' v_n & v_n' w_n & w_n' u_n & w_n' v_n & w_n' w_n \end{pmatrix} \quad (2.3)$$

La notation $\mathbf{0}_9$ dénote le vecteur nul de dimension 9.

Malgré les 9 inconnues de l'équation 2.1, il faut rappeler que la matrice fondamentale n'a que 7 degrés de liberté car elle est définie à un facteur d'échelle près et de rang 2.

Partant de ce constat, il est possible d'estimer cette matrice à partir de seulement 7 couples de points [Zhang, 1998]. Dans ce cas Z est de taille 7×9 et de rang 7. Par une décomposition en valeurs singulières de Z , on obtient deux vecteurs singuliers \mathbf{f}_1 et \mathbf{f}_2 de Z qui

correspondent respectivement à deux matrices F_1 et F_2 , générant le noyau de Z . De par l'homogénéité des équations, la solution est une famille de matrices de la forme :

$$F = \alpha F_1 + (1 - \alpha) F_2 \quad (2.4)$$

Le paramètre α est fixé grâce à la contrainte du rang de F égal à 2. Ceci s'exprime par l'équation $\det(F) = 0$, aboutissant à un polynôme de degré 3 en α , dont le nombre de solutions réelles est au maximum de 3.

Cette méthode présente deux inconvénients majeurs. Tout d'abord, elle n'accepte pas plus de 7 points, ce qui la rend très sensible au bruit de localisation des points. Deuxièmement, la solution n'est pas toujours unique et son résultat dépend du choix des 7 points parmi l'ensemble des correspondances. Cependant, cette méthode est utilisée dans les algorithmes robustes, par sélection aléatoire successive des 7 correspondances.

Nous présentons dans les sous-sections suivantes les algorithmes sélectionnés pour estimer cette matrice, après avoir détaillé deux caractéristiques communes à toutes les méthodes d'estimation. Nous commençons par les méthodes linéaires puis nous enchaînons par les méthodes itératives avant de finir avec les méthodes robustes. Les méthodes linéaires et itératives permettent de s'affranchir du bruit de localisation des points dans les images, alors que les méthodes robustes sont capables de détecter et d'éliminer les fausses correspondances. Parmi les méthodes linéaires, nous avons retenu les moindres carrés [Longuet-Higgins, 1981], la technique des moindres carrés orthogonaux [Torr, 1997] et l'adaptation de la méthode de Bookstein par Torr et Fitzgibbon [Torr, 2003]. Le groupe des méthodes itératives est représenté par la méthode des moindres carrés itératifs pondérés proposée par Torr et Murray [Torr, 1997]. Enfin, les algorithmes LMedS [Zhang, 1998] et MAPSAC de Torr [Torr, 2002b] ont été choisis pour représenter les méthodes robustes.

2.1.1 Caractéristiques communes à tous les algorithmes

Deux caractéristiques sont communes à toutes les méthodes implémentées. Préalablement au calcul de F , les coordonnées images sont normalisées selon la technique d'Hartley [Hartley, 1997b] qui consiste à effectuer une translation des points de sorte que le barycentre se trouve à l'origine, puis, à mettre à l'échelle toutes les coordonnées de façon à ce que la distance moyenne des points à l'origine soit égale à $\sqrt{2}$. Cette étape préalable permet de réduire les instabilités numériques provoquées par les grandes valeurs des coordonnées en pixels. Pour chaque image, les matrices de transformation T et T' sont calculées à partir des coordonnées des points, de sorte que $\hat{m} = T m$ et $\hat{m}' = T' m'$ où \hat{m} et \hat{m}' sont les points normalisés. Cette normalisation est annulée à la fin de l'estimation de la matrice fondamentale en appliquant la transformation inverse : $F = T'^T \hat{F} T$.

La seconde caractéristique commune à tous les algorithmes est une étape finale permettant d'assurer que F soit de rang 2. Pour cela, si la contrainte du rang n'est pas incluse au sein de l'algorithme, la valeur estimée de F est remplacée par la matrice de rang 2 la plus proche. Ce remplacement est effectué en utilisant une décomposition en valeurs singulières de F [Hartley, 1997b] : si $F = U S V^T$ est la décomposition en valeurs singulières de F avec la

matrice diagonale $S = \text{diag}(r, s, t)$, alors la matrice $F' = U S' V^T$ avec $S' = \text{diag}(r, s, 0)$ est la matrice de rang 2 la plus proche selon la norme de Frobenius.

2.1.2 La méthode des moindres carrés

Il s'agit de la première méthode de calcul de la matrice fondamentale, introduite en 1981 par Longuet-Higgins [Longuet-Higgins, 1981]. Elle nécessite un minimum de 8 points. Lorsque plus de points sont disponibles, la solution est obtenue par une minimisation aux moindres carrés du résidu de l'équation 1.10 :

$$\min_F \sum_i (m_i^T F m_i)^2 \quad (2.5)$$

Sachant que F est définie à un facteur d'échelle près, il est possible de fixer un élément de F à 1, usuellement F_{33} . Avec les notations de l'équation 2.1, la solution est obtenue par :

$$f' = -(Z'^T Z')^{-1} Z'^T z_9 \quad (2.6)$$

où Z' est la matrice composée des 8 premières colonnes de Z , z_9 est la dernière colonne de Z et f' est le vecteur des 8 premiers éléments de f , le neuvième étant égal à 1.

2.1.3 La technique des moindres carrés orthogonaux

Cette technique, connue sous le nom d'algorithme 8 points, est une variante de la précédente. Le problème 2.5 est équivalent à la minimisation de :

$$\min_f f^T Z^T Z f \quad (2.7)$$

La solution est obtenue par une analyse des vecteurs propres. Le vecteur f est le vecteur propre de $Z^T Z$ correspondant à la plus petite valeur propre. Il peut être déterminé par une décomposition en valeurs singulières de la matrice $Z^T Z$. En posant $U S V^T$ la décomposition en valeurs singulières de $Z^T Z$, le vecteur f est la dernière colonne de la matrice V .

La différence entre ces deux techniques réside dans le type d'erreur calculée entre les points et leur droite épipolaire (fig. 2.1). Cette méthode est appelée orthogonale car la distance mesurée entre les points et les droites épipolaires est prise perpendiculairement à la droite et est donc plus rigoureuse.

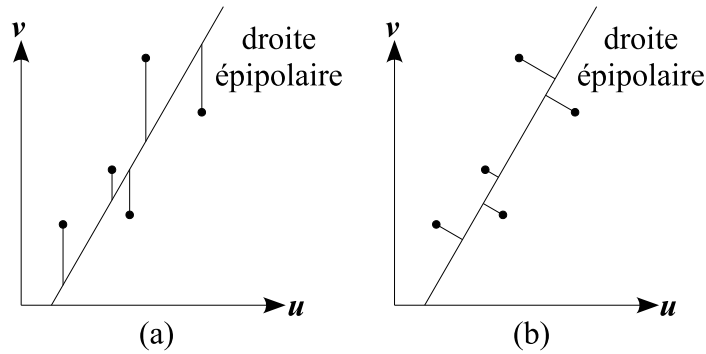


Figure 2.1 : Minimisation aux moindres carrés classiques (a) et orthogonaux (b).

2.1.4 La méthode de Bookstein

Torr et Fitzgibbon ont adapté la méthode d'estimation de coniques de Bookstein [Bookstein, 1979] au problème de la matrice fondamentale [Torr, 2003]. La matrice fondamentale pouvant être vue comme une conique dans l'espace \mathbb{R}^4 , Bookstein recommande de recalculer la conique d'une manière invariante aux transformations euclidiennes. Il faut alors déterminer une norme invariante à ce groupe de transformations ; les auteurs ont montré que le carré de la norme de Frobenius sur la sous-matrice supérieure gauche de taille 2×2 de F est un choix simple et efficace. La détermination de \mathbf{f} se fait par la minimisation de l'équation 2.7 soumise à la contrainte que la norme de \mathbf{f} doit être constante. Cette contrainte s'exprime par :

$$\mathbf{f}^T J \mathbf{f} = F_{11}^2 + F_{12}^2 + F_{21}^2 + F_{22}^2 = k \quad (2.8)$$

où k est une constante et $J = \text{diag}(1, 1, 0, 1, 1, 0, 0, 0, 0)$.

Les auteurs proposent de réarranger \mathbf{f} de sorte que $\mathbf{f}^T = (\mathbf{f}_1^T | \mathbf{f}_2^T)$ avec $\mathbf{f}_1 = (F_{11}, F_{12}, F_{21}, F_{22})^T$ et $\mathbf{f}_2 = (F_{13}, F_{23}, F_{31}, F_{32}, F_{33})^T$ et de réarranger Z en conséquence de sorte que l'équation 2.1 soit toujours vérifiée. Avec ces arrangements, notons $M = Z^T Z$. La matrice M étant symétrique se décompose en 3 sous-matrices M_1 , M_2 et M_3 :

$$M = \begin{pmatrix} M_1 & M_2 \\ M_2^T & M_3 \end{pmatrix} \quad (2.9)$$

de sorte que

$$\mathbf{f}^T M \mathbf{f} = \mathbf{f}_1^T M_1 \mathbf{f}_1 + 2 \mathbf{f}_1^T M_2 \mathbf{f}_2 + \mathbf{f}_2^T M_3 \mathbf{f}_2 \quad (2.10)$$

La contrainte 2.8 s'écrit alors $\mathbf{f}_1^T \mathbf{f}_1 = k$, ce qui implique que le vecteur \mathbf{f}_1 est fixe au cours de la minimisation. Ainsi, le minimum est obtenu pour :

$$\frac{\partial \mathbf{f}^T M \mathbf{f}}{\partial \mathbf{f}_2} = 2 M_2^T \mathbf{f}_1 + 2 M_3 \mathbf{f}_2 = 0 \quad (2.11)$$

d'où :

$$\mathbf{f}_2 = -M_3^{-1} M_2^T \mathbf{f}_1 \quad (2.12)$$

puis en remplaçant \mathbf{f}_2 dans l'équation 2.10 :

$$\mathbf{f}^T M \mathbf{f} = \mathbf{f}_1^T (M_1 - M_2 M_3^{-1} M_2^T) \mathbf{f}_1 = \mathbf{f}_1^T Q \mathbf{f}_1 \quad (2.13)$$

Le vecteur \mathbf{f}_1 est alors obtenu par analyse des vecteurs propres comme dans la sous-section précédente puis le vecteur \mathbf{f}_2 est déterminé par l'équation 2.12.

2.1.5 La méthode itérative des moindres carrés repondérés

Torr et Murray ont montré que les approches linéaires ne sont pas optimales car la distribution des résidus $\mathbf{m}'_i^T F \mathbf{m}_i$ n'est pas gaussienne [Torr, 1997]. En s'appuyant sur les travaux de Sampson [Sampson, 1982], ils ont établi que les résidus doivent être pondérés par leurs variances. La fonction de coût à minimiser s'exprime alors par :

$$\min_F \sum_i (p_i \mathbf{m}'_i^T F \mathbf{m}_i)^2 \quad (2.14)$$

où p_i est le poids optimal. Une estimation au premier ordre de la variance des résidus permet d'écrire :

$$p_i = \frac{1}{\nabla r_i} \quad (2.15)$$

avec

$$\begin{aligned} r_i = & \mathbf{f}_1 u_i' u_i + \mathbf{f}_2 u_i' v_i + \mathbf{f}_3 u_i' w_i + \mathbf{f}_4 v_i' u_i \\ & + \mathbf{f}_5 v_i' v_i + \mathbf{f}_6 v_i' w_i + \mathbf{f}_7 w_i' u_i + \mathbf{f}_8 w_i' v_i + \mathbf{f}_9 w_i' w_i \end{aligned} \quad (2.16)$$

et

$$\nabla r_i = \sqrt{r_{iu}^2 + r_{iv}^2 + r_{iu'}^2 + r_{iv'}^2} \quad (2.17)$$

où r_{iu} , r_{iv} , $r_{iu'}$ et $r_{iv'}$ sont les dérivées partielles du résidu de l'équation 2.16 selon les axes des deux images. Par exemple, $r_{iu} = \mathbf{f}_1 u_i' + \mathbf{f}_4 v_i' + \mathbf{f}_7 w_i'$. Comme le calcul des poids nécessite le calcul des résidus et vice-versa, la minimisation se fait par un processus itératif à partir des p_i initiaux fixés à 1 et d'une matrice fondamentale initialisée par la méthode de Bookstein. À chaque itération, les points présentant une forte variance se verront attribuer un faible poids par l'équation 2.15 et interviendront plus faiblement dans la minimisation de l'équation 2.14.

2.1.6 L'algorithme LMedS

Cette technique robuste permet, en plus d'estimer la matrice fondamentale, de détecter les fausses correspondances de l'ensemble d'appariements, appelées *outliers*. Son but est de résoudre le problème de minimisation non linéaire suivant :

$$\min_F \operatorname{median}_i (r_i(F))^2 \quad (2.18)$$

où $r_i(F)$ est le résidu de la correspondance $(\mathbf{m}_i, \mathbf{m}_i')$, c'est-à-dire la différence entre les valeurs mesurées (les points) et les valeurs estimées (les droites épipolaires).

Le résultat de la minimisation est alors la matrice fondamentale engendrant la plus petite médiane possible des carrés des résidus, d'où le nom anglais de la méthode *Least Median of Squares*. Cette minimisation est réalisée par l'algorithme décrit en figure 2.2.

Soit n correspondances $(\mathbf{m}_i, \mathbf{m}_i'; i=1, \dots, n)$:

Répéter k fois :

- Sélectionner aléatoirement 7 correspondances (rappelons que l'estimation de F nécessite un minimum de 7 correspondances.)
- Calculer la matrice fondamentale F_j à partir des 7 correspondances.
- Calculer la médiane Med_j des carrés des résidus selon l'expression :

$$\operatorname{median}_{i=1, \dots, n} (d^2(\mathbf{m}_i', F_j \mathbf{m}_i) + d^2(\mathbf{m}_i, F_j^T \mathbf{m}_i'))$$
 où d est la distance orthogonale entre un point et sa droite épipolaire.

Conserver la matrice fondamentale correspondante au Med_j minimal.

Figure 2.2 : Algorithme LMedS.

Le nombre d'itérations k est déterminé à partir de la probabilité recherchée P d'avoir au moins un sous-ensemble de 7 correspondances correctes, cela dans le cas où les données peuvent contenir une certaine proportion d'*outliers* ϵ . Cette probabilité s'exprime par :

$$P = 1 - (1 - (1 - \epsilon)^7)^k \quad (2.19)$$

ainsi k est déterminé par :

$$k = \frac{\ln(1 - P)}{\ln(1 - (1 - \epsilon)^7)} \quad (2.20)$$

Dans notre implémentation, nous avons choisi le cas le moins favorable où $\epsilon = 50\%$ et où on veut s'assurer une probabilité de bonne sélection de 99%. La valeur k vaut alors 587.

Les *outliers* sont ensuite détectés parmi l'ensemble des correspondances par un seuillage des résidus des correspondances. Le seuil est fixé à $2,5\hat{\sigma}$ où $\hat{\sigma}$ est l'estimation robuste de l'écart-type donnée par l'expression :

$$\hat{\sigma} = 1,4826 \left(1 + \frac{5}{(n-7)} \right) \sqrt{\text{Med}_j} \quad (2.21)$$

où Med_j est la plus petite médiane estimée par l'algorithme de la figure 2.2. Ainsi une correspondance est considérée comme *outlier* si $r_i^2 > (2,5\hat{\sigma})^2$ où :

$$r_i^2 = d^2(\mathbf{m}_i', F_j \mathbf{m}_i) + d^2(\mathbf{m}_i, F_j^T \mathbf{m}_i') \quad (2.22)$$

2.1.7 L'algorithme MAPSAC

L'algorithme MAPSAC (*Maximum A Posteriori Sample Consensus*) [Torr, 2002b] est semblable à l'algorithme RANSAC (*RANdom SAMple Consensus*) de Fischler et Bolles [Fischler, 1981] dont l'objectif est de minimiser une fonction de coût définie par :

$$C = \sum_i \rho(r_i^2) \quad (2.23)$$

où la fonction ρ est définie par :

$$\rho(r_i^2) = \begin{cases} 0 & r_i^2 < T \\ \text{constante} & r_i^2 \geq T \end{cases} \quad (2.24)$$

La minimisation est réalisée par un processus itératif semblable à LMedS. A chaque itération, une matrice fondamentale est estimée à partir de 7 correspondances choisies aléatoirement, puis les résidus de chaque appariement sont comparés au seuil T afin de calculer la fonction de coût C . La matrice fondamentale correspondante au minimum de la fonction de coût C est alors conservée.

L'inconvénient de la méthode RANSAC est que de nombreux *outliers* ne sont pas détectés si le seuil T est trop élevé. De plus, avec un seuil trop élevé, de nombreuses réalisations de C peuvent être égales rendant ainsi difficile la détermination d'un minimum. Pour remédier à cela, Torr propose dans son algorithme de faire évoluer la fonction ρ :

$$\rho(r_i^2) = \begin{cases} r_i^2 & r_i^2 < T \\ T & r_i^2 \geq T \end{cases} \quad (2.25)$$

Lorsque le résidu r_i est calculé comme dans l'équation 2.22, le seuil T peut être défini comme étant égal à $5,99\sigma^2$, avec σ pouvant être estimé par l'équation 2.21 [Torr, 2002a].

2.2 Analyse des estimations de la matrice fondamentale

Afin de choisir la méthode d'estimation de la matrice fondamentale la plus adéquate pour notre application, nous avons réalisé plusieurs tests sur des données de synthèse en reproduisant les erreurs qu'il est possible de rencontrer. L'idée de base est de déterminer l'algorithme d'estimation de la matrice fondamentale qui induit l'erreur minimale sur les données 3D obtenues en fin de chaîne de reconstruction. Cette section présente ensuite une chaîne de reconstruction typique dans le cas de la stéréovision non calibrée à partir de 2 vues, puis décrit le protocole de test mis en place avant de discuter les résultats.

2.2.1 La chaîne de reconstruction 3D

La chaîne de reconstruction 3D est déduite du chapitre 1 et peut être décomposée en 4 étapes (fig. 2.3). La matrice fondamentale est estimée à partir de points mis en correspondance entre les deux images, puis les paramètres intrinsèques sont extraits de cette matrice par un algorithme d'auto-calibrage. Les paramètres intrinsèques permettent de construire la matrice de calibrage A , utilisée pour calculer la matrice essentielle E . Cette dernière est la version calibrée de la matrice fondamentale F et vérifie l'équation suivante :

$$E \propto A^T F A \propto R [\mathbf{t}]_{\times} \quad (2.26)$$

Les paramètres extrinsèques contenus dans les matrices R et $[\mathbf{t}]_{\times}$, sont alors calculés par une décomposition de la matrice essentielle. Il est important de ne pas oublier que toutes les valeurs sont déterminées à un facteur d'échelle près qui reste inconnu. La reconstruction 3D ne peut être obtenue que dans un cadre métrique. Les algorithmes choisis pour l'estimation des paramètres extrinsèques sont détaillés dans le chapitre 4. Toutefois, il faut savoir que l'algorithme de décomposition de E en R et \mathbf{t} fixe la norme de \mathbf{t} à 1. Donc, si la norme réelle de ce vecteur est connue, la reconstruction 3D peut être mise à l'échelle afin d'obtenir une reconstruction euclidienne. Une fois tous ces paramètres estimés, les points 3D sont déterminés par triangulation.

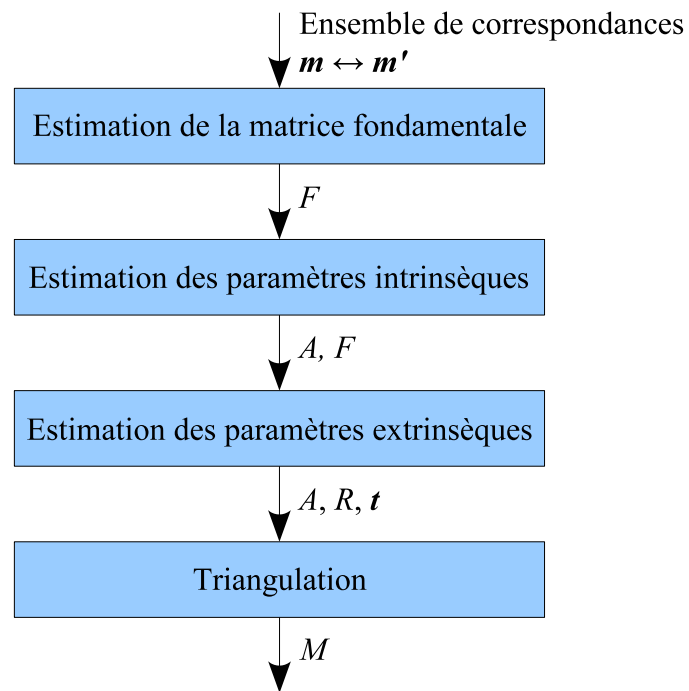


Figure 2.3 : Chaîne de reconstruction 3D.

2.2.2 Protocole d'évaluation

La scène 3D permettant de réaliser les tests de simulation est composée de 100 points 3D X_i ($i = 1, \dots, 100$) générés aléatoirement dans un cube, défini par $-5 < x < 5$, $-5 < y < 5$ et $10 < z < 20$ avec des unités arbitraires. Deux capteurs d'image sont disposés autour de ce cube dont l'un est confondu avec l'origine des coordonnées O alors que l'autre est à la position $O' = (20, 0, 5)^T$ et vise le centre du cube. Les directions verticales des deux capteurs sont confondues ($\mathbf{y} = \mathbf{y}'$). La figure 2.4 représente cette configuration.

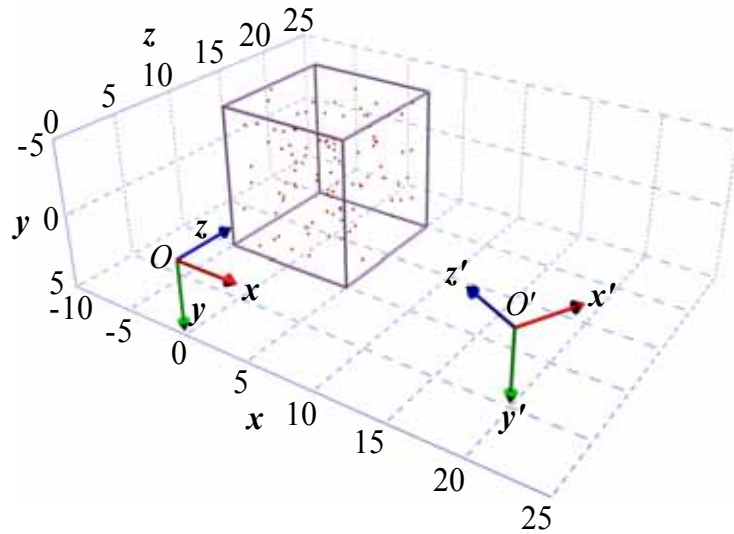


Figure 2.4 : Représentation de la boîte englobante des données.

Les images générées par les capteurs sont de résolution 640×480 et leur distance focale est fixée à $\alpha = 600$ pixels. Cette valeur est proche de la focale standard pour cette résolution qui est égale à 800, la diagonale de l'image. Cette configuration de l'espace et des capteurs est cohérente avec une prise de vue manuelle par un appareil photo numérique. Nous pouvons remarquer également que les positions relatives des capteurs d'image ont été choisies de façon à ce que les points O et O' ne soient pas équidistants du centre du cube, afin d'éviter une configuration singulière gênant l'auto-calibrage. La figure 2.5 représente le cube projeté sur les deux capteurs O et O' ainsi qu'une réalisation aléatoire des 100 points.

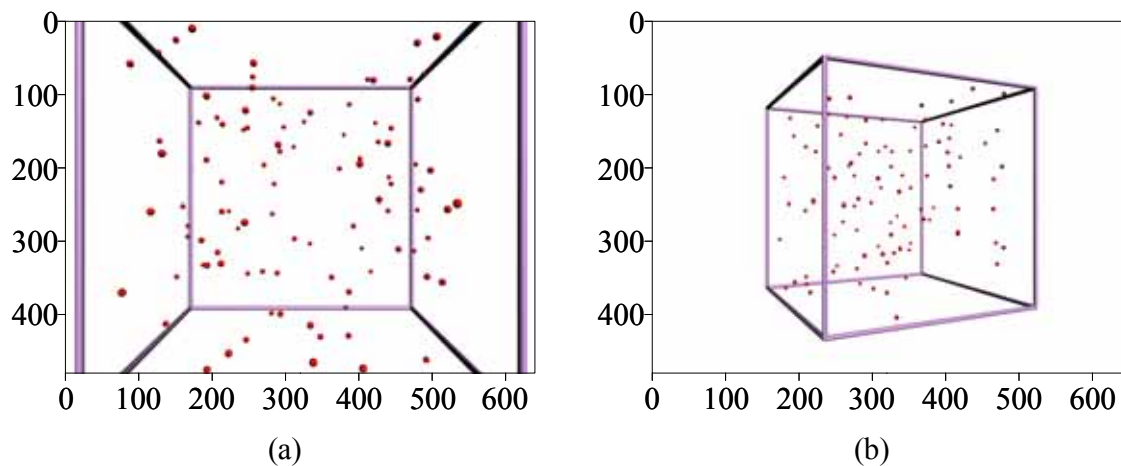


Figure 2.5 : Projection du cube sur les capteurs positionnés en O (a) et en O' (b).

Tous ces choix se traduisent mathématiquement par les matrices suivantes :

$$A = \begin{pmatrix} 600 & 0 & 320 \\ 0 & 600 & 240 \\ 0 & 0 & 1 \end{pmatrix}, \quad R = \begin{pmatrix} 0,44721 & 0 & 0,89443 \\ 0 & 1 & 0 \\ -0,89443 & 0 & 0,44721 \end{pmatrix}, \quad t = \begin{pmatrix} -20 \\ 0 \\ -5 \end{pmatrix} \quad (2.27)$$

où A est la matrice des paramètres intrinsèques des capteurs d'image et R et \mathbf{t} sont respectivement la matrice de rotation et le vecteur de translation du second capteur. Ainsi, la matrice de projection du premier capteur est $P_1 = A(Id_{3 \times 3} | 0)$ et celle du second est $P_2 = A(R | R\mathbf{t})$.

Les points 3D générés dans le cube sont projetés sur les deux capteurs, créant ainsi un ensemble idéal de points 2D en correspondance entre les deux images. Ces correspondances sont le vecteur d'entrée d'une chaîne de reconstruction 3D métrique, représentée sur la figure 2.3. La reconstruction métrique considère la distance entre les deux capteurs égale à 1. Pour passer à la reconstruction euclidienne, le facteur d'échelle est alors déterminé par la distance réelle séparant les capteurs $\|\mathbf{t}\| = 20,62$.

Deux erreurs sont relevées afin de permettre la comparaison des algorithmes. Il s'agit de l'erreur relative e_α sur l'estimation $\hat{\alpha}$ de la distance focale α après l'étape d'auto-calibrage et l'erreur relative e_x sur la distance moyenne entre les estimations \hat{X}_i des points 3D et les points 3D X_i eux-mêmes. L'expression de e_α est donc :

$$e_\alpha = \frac{|\hat{\alpha} - \alpha|}{\alpha} \quad (2.28)$$

L'erreur sur les points 3D est calculée relativement à la plus grande distance de l'objet à reconstruire, c'est-à-dire la diagonale du cube qui vaut $10\sqrt{3}$. L'expression de e_x est alors :

$$e_x = \frac{1}{100} \sum_{i=1}^{100} \frac{\|\hat{X}_i - X_i\|}{10\sqrt{3}} \quad (2.29)$$

2.2.3 Simulations

Lors de l'étape de mise en correspondance préalable à l'estimation de la matrice fondamentale, deux types de bruit peuvent perturber les mesures. Il s'agit des erreurs de localisation des points sur les images et des faux appariements. Le premier type de bruit agit comme une erreur de position, c'est-à-dire qu'un point \mathbf{p} est détecté à la position $\hat{\mathbf{p}} = \mathbf{p} + \mathbf{e}$, où \mathbf{e} est le vecteur d'erreur. Le second type de bruit perturbe l'ensemble des mises en correspondance $(\mathbf{p}_i, \mathbf{p}_i')$ en associant parfois un point \mathbf{p}_i avec un point \mathbf{p}_j' pour $i \neq j$. Ceci se produit lorsque les scores d'appariements des couples $(\mathbf{p}_i, \mathbf{p}_i')$ et $(\mathbf{p}_j, \mathbf{p}_j')$ sont proches, où lorsqu'un des points du couple n'a pas été détecté dans une des images.

Nous avons alors analysé le comportement des six algorithmes retenus précédemment face à ces deux types de bruit. Pour ce faire, nous avons généré ces bruits sur l'ensemble des correspondances idéales obtenues par nos données de synthèse.

2.2.3.1 Analyse de la robustesse au bruit de localisation des points d'intérêt

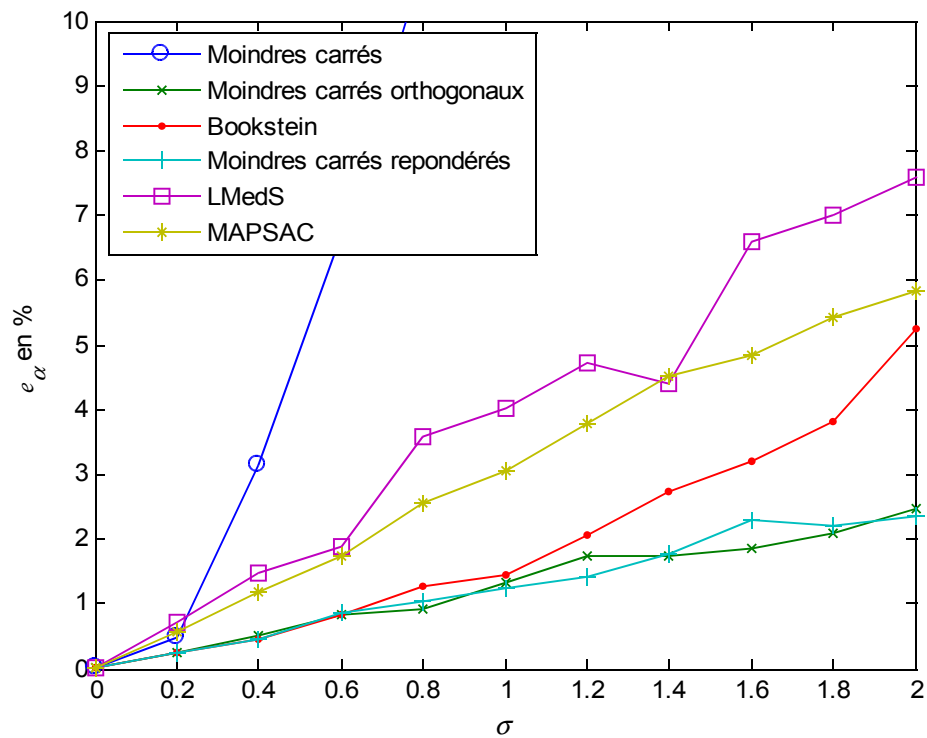
Une fois les points 3D projetés sur les images, un bruit gaussien est ajouté sur leurs positions. L'écart-type de la distribution gaussienne varie entre 0 et 2 pixels, ce qui correspond à un déplacement maximal de 6 pixels pour 98% des cas. Pour stabiliser les résultats, les mesures d'erreur sur la matrice fondamentale sont moyennées sur 100 générations aléatoires des 100 points 3D.

Les résultats de cette expérience sont reportés sur la figure 2.6. Il apparaît clairement que la méthode linéaire des moindres carrés doit absolument être évitée, les erreurs induites par cet algorithme étant très supérieures aux autres dès qu'un peu de bruit est présent ($\sigma \geq 0,4$). Les méthodes robustes (LMedS et MAPSAC) aboutissent à des erreurs globalement deux fois supérieures aux autres méthodes pour ce type de bruit. Les deux algorithmes des moindres carrés orthogonaux et des moindres carrés repondérés présentent des résultats similaires en terme d'erreur, mais celui des moindres carrés orthogonaux offre l'avantage d'être plus rapide, comme l'atteste le tableau 2.1.

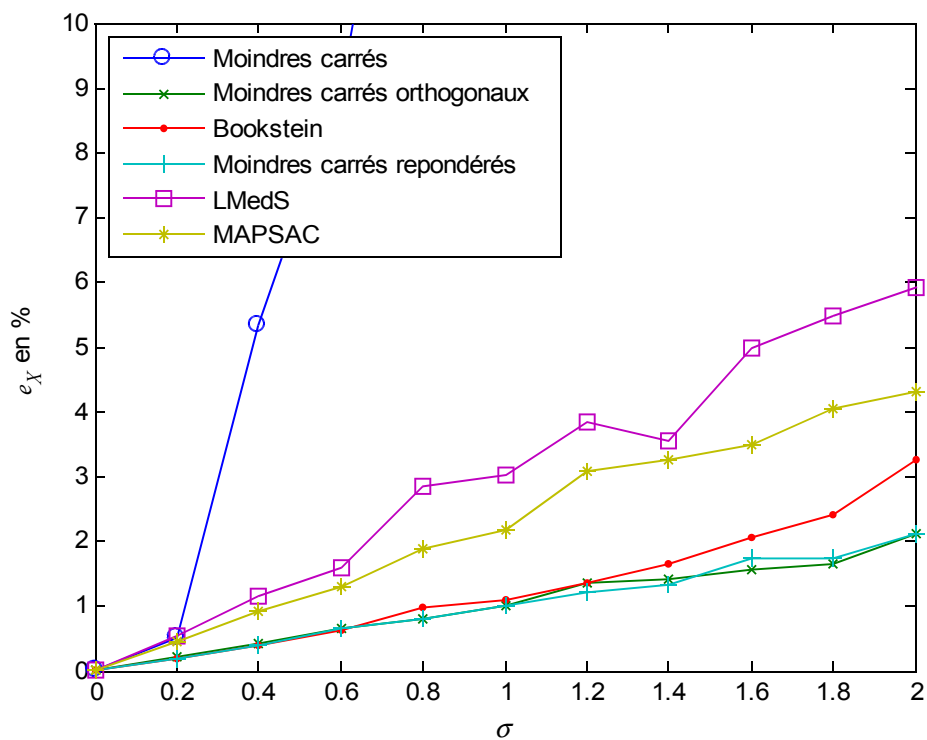
Tableau 2.1 : Temps de calcul de la matrice fondamentale à partir de 100 appariements.

Algorithme	Moindres carrés	Moindres carrés orth.	Bookstein	Moindres carrés rep.	LMedS	MAPSAC
Temps de calcul en ms	1,0	1,3	0,9	10,5	460,8	496,2

En présence de bruit sur la localisation des points d'intérêt, il est donc préférable de choisir la méthode des moindres carrés orthogonaux pour estimer la matrice fondamentale. Cette conclusion est valable uniquement en l'absence totale de faux appariements, ce qui est rarement le cas en pratique. C'est pourquoi l'expérience suivante a été réalisée pour analyser le comportement des algorithmes face aux faux appariements.



(a)



(b)

Figure 2.6 : Analyse de la robustesse des algorithmes d'estimation de la matrice fondamentale face au bruit de localisation. Erreurs sur la distance focale (a) et sur la position des points 3D (b).

2.2.3.2 Analyse de la robustesse aux faux appariements

Pour cette simulation, aucun bruit gaussien n'est ajouté aux positions des points 2D. Les algorithmes sont analysés en fonction du pourcentage de faux appariements présents dans l'ensemble des correspondances. Pour générer n faux appariements à partir des N points idéalement mis en correspondance, les n derniers points sur l'image 2 sont réarrangés par une permutation circulaire comme représentée sur la figure 2.7.

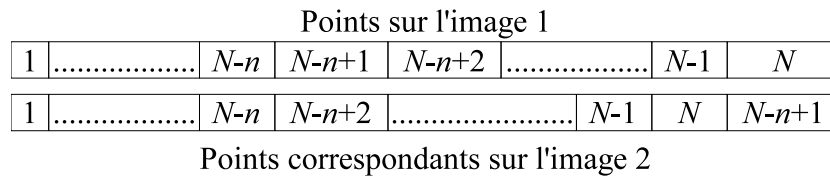
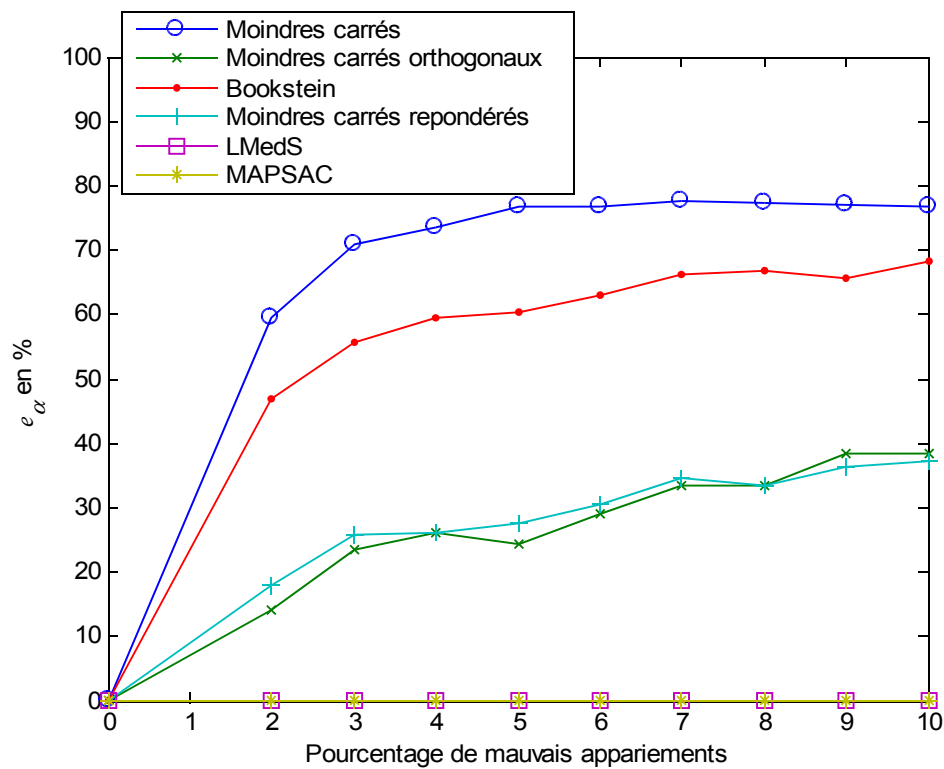


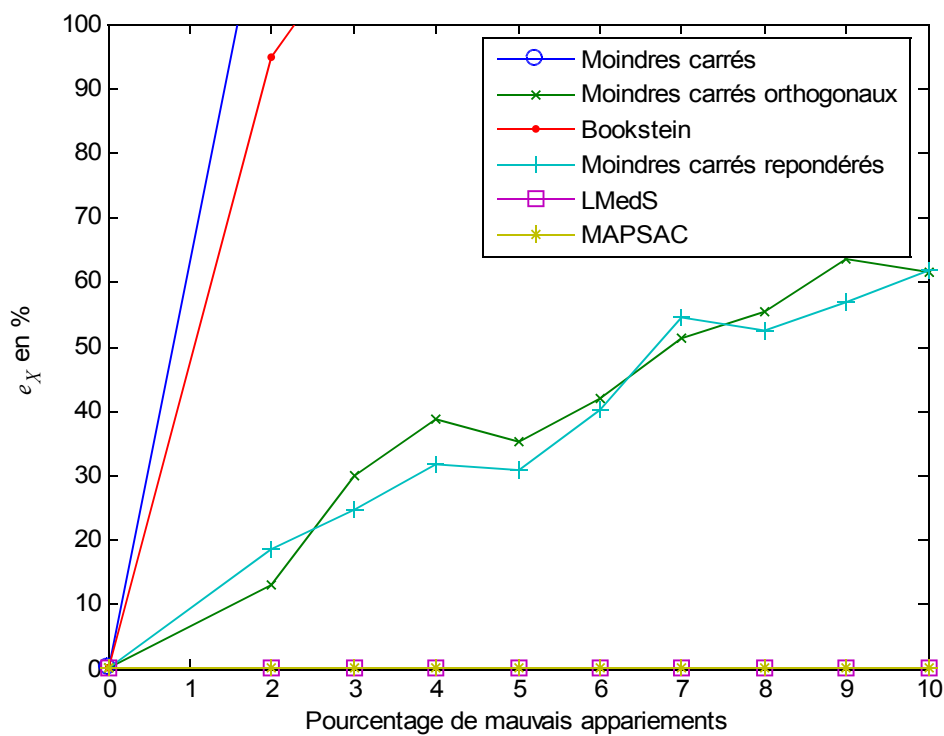
Figure 2.7 : Méthode pour générer n faux appariements.

De la même manière que pour l'analyse du bruit de localisation des points d'intérêt, les erreurs sont moyennées sur 100 réalisations de la génération aléatoire des points 3D, afin d'assurer la stabilité des mesures. Le taux de faux appariements varie entre 0 et 10 %.

Les résultats sont reportés sur la figure 2.8 et prouvent que les méthodes robustes respectent bien leurs objectifs d'être peu sensibles aux faux appariements. Les algorithmes MAPSAC et LMedS présentent toujours une erreur nulle quel que soit le taux de faux appariements. Par ailleurs, nous avons vérifié que tous les faux appariements étaient bien détectés comme tels par ces deux méthodes. Les autres méthodes sont inutilisables lors de la présence de faux appariements au vu des erreurs provoquées dès les faibles taux de faux appariements.



(a)



(b)

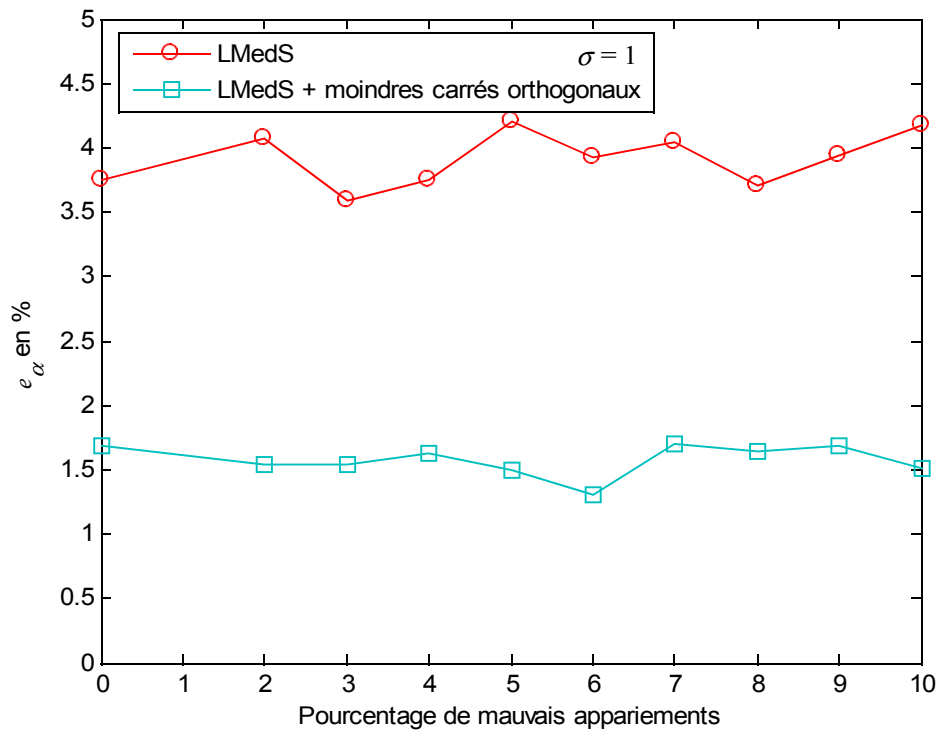
Figure 2.8 : Analyse de la robustesse des algorithmes d'estimation de la matrice fondamentale face aux faux appariements. Erreurs sur la distance focale (a) et sur la position des points 3D (b).

2.3 Discussion sur le choix de la méthode pour la stéréovision non calibrée

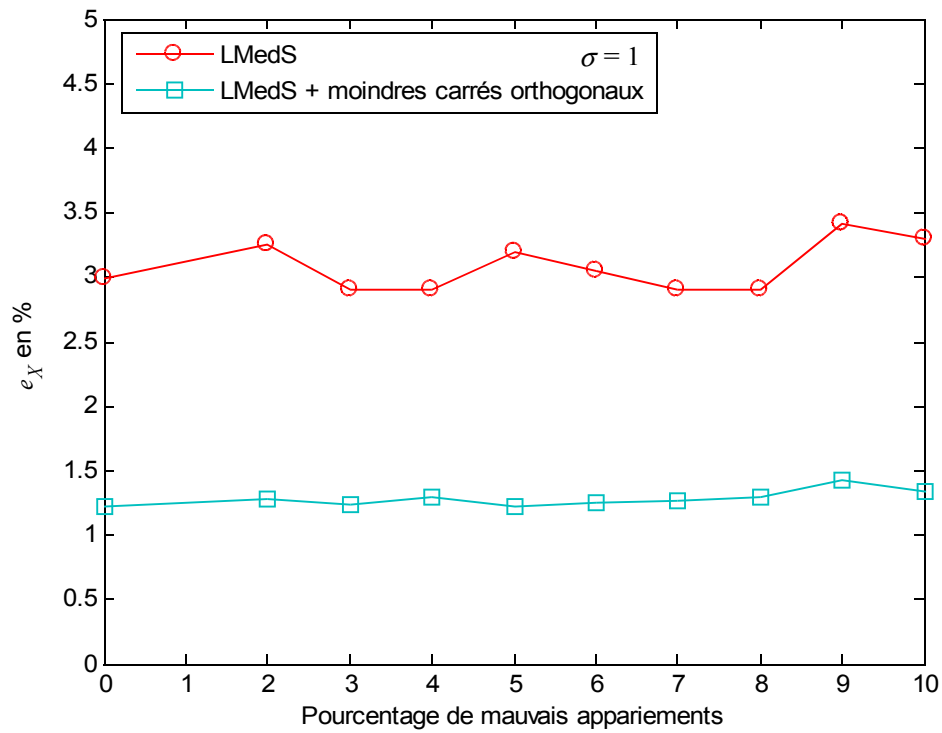
Dans la pratique, les deux types de bruits étudiés dans la section précédente sont présents. Afin de s'affranchir de ces perturbations, on peut imaginer une estimation de la matrice fondamentale en deux étapes. Tout d'abord, une estimation robuste dont l'objectif est de détecter les *outliers* dans l'ensemble de correspondances. Une fois ces faux appariements écartés, il est possible d'estimer la matrice fondamentale sur l'ensemble des correspondances vraies, ne présentant donc plus que du bruit de localisation. Nous avons choisi la méthode LMedS pour détecter les faux appariements car elle est plus rapide que l'algorithme MAPSAC (tab. 2.1). L'estimation de la matrice fondamentale est ensuite réalisée par la méthode des moindres carrés orthogonaux sur l'ensemble des vraies correspondances. Ce choix est justifié par les résultats présentés sur la figure 2.6, où l'algorithme des moindres carrés orthogonaux est la méthode la plus performante face au bruit de localisation.

Afin de quantifier l'amélioration produite par cette technique, nous avons réalisé une troisième simulation combinant les deux types de perturbation. Les 100 appariements sont perturbés par un bruit de localisation d'écart-type $\sigma = 1$ comme lors de la première expérience, et des faux appariements sont générés comme pour la deuxième expérience. Nous avons comparé dans ce cadre la méthode LMedS seule avec la méthode LMedS suivi d'une optimisation par moindres carrés orthogonaux. Les résultats sont présentés dans la figure 2.9.

Les courbes montrent que l'erreur de l'algorithme combinant les méthodes LMedS et des moindres carrés orthogonaux est stable quelque soit le pourcentage de faux appariements. L'optimisation réalisée permet de réduire par un facteur 2 l'erreur engendrée par la méthode LMedS afin d'obtenir des valeurs d'erreur de l'ordre de celles obtenues en absence de fausses correspondances. Grâce à cette combinaison d'algorithmes, nous obtenons une méthode d'estimation de la matrice fondamentale robuste aux deux types de bruit rencontrés en pratique.



(a)



(b)

Figure 2.9 : Comparaison de la méthode LMedS seule avec la méthode LMedS optimisée par moindres carrés orthogonaux. Erreurs sur la distance focale (a) et sur la position des points 3D (b).

2.4 Conclusion

Nous avons passé en revue dans ce chapitre, six méthodes différentes d'estimation de la matrice fondamentale classées en trois catégories : les méthodes linéaires, itératives et robustes. Afin de choisir la méthode la plus appropriée pour notre problématique de stéréovision non calibrée, nous avons analysé les différentes sources d'erreurs possibles. Deux types de bruit sont rencontrés en pratique : le bruit sur la position des points d'intérêt détectés avant l'étape de mise en correspondance et les faux appariements.

Nous avons alors mis en place un cadre de travail permettant d'analyser le comportement des algorithmes d'estimation de la matrice fondamentale face à ces deux types de bruit. Il en résulte que l'algorithme des moindres carrés orthogonaux est le plus adapté pour s'affranchir du bruit de localisation des points d'intérêt alors que les algorithmes robustes permettent d'écartier les faux appariements. L'algorithme LMedS est préféré à son homologue MAPSAC car il est plus rapide.

Les algorithmes LMedS et des moindres carrés orthogonaux sont alors combinés de manière à réduire l'impact des deux types de bruit. L'algorithme LMedS est exécuté préalablement afin de détecter les faux appariements de l'ensemble des correspondances, puis la matrice fondamentale est estimée à partir des vrais appariements par les moindres carrés orthogonaux, réduisant ainsi l'impact du bruit de localisation des points d'intérêt. Cette combinaison des algorithmes est optimale pour réduire l'impact des deux types de bruit sur l'estimation de la matrice fondamentale.

Ces tests des algorithmes d'estimation de la matrice fondamentale analysés dans un cadre global de reconstruction 3D ont fait l'objet d'une communication en 2004 [Albouy, 2004]. Les résultats des simulations réalisées, montrant une erreur sur le modèle 3D inférieure à 1,5 %, sont compatibles avec l'objectif de 5 % d'erreur maximum pour notre application médicale. En outre, les temps de calculs de l'ordre de la demi-seconde sont négligeables dans notre contexte applicatif.

Nous avons établi dans ce chapitre une méthode optimale et robuste d'estimation de la matrice fondamentale en stéréovision non calibrée. Aucune information sur les capteurs d'image n'étant disponible, la matrice fondamentale ne peut être estimée que par les points appariés entre les deux images. L'objet du chapitre suivant concerne cette mise en correspondance dans les images.

Chapitre 3

Mise en correspondance des images

L'estimation de la matrice fondamentale s'appuie sur les points d'intérêt mis en correspondance entre les deux images. La mise en correspondance est donc une étape critique pour la stéréovision non calibrée. Ce chapitre rappelle le contexte des prises de vue de notre application avant de détailler la méthode mise en oeuvre.

Sommaire

3.1	Contexte de prise de vue des escarres.....	61
3.2	Méthodes courantes d'appariement de points d'intérêt.....	63
3.2.1	Détecteur de coins de Harris et Stephens adapté à la couleur.....	63
3.2.2	Caractérisation locale des points d'intérêt.....	64
3.2.3	Méthodes d'appariement.....	65
3.2.3.1	Algorithme « Winner Takes All ».....	65
3.2.3.2	Algorithmes de relaxation.....	66
3.3	Algorithme itératif de mise en correspondance.....	67
3.3.1	Images en entrée de l'algorithme.....	68
3.3.2	Mise en correspondance initiale.....	69
3.3.3	Détection de points d'intérêt.....	70
3.3.4	Raffinement de la mise en correspondance par un processus itératif.....	71
3.3.4.1	Estimation de la matrice fondamentale.....	71
3.3.4.2	Rectification épipolaire.....	71
3.3.4.3	Application de la contrainte épipolaire.....	74
3.3.4.4	Application de la contrainte de voisinage.....	75
3.3.5	Détection et suppression des points 3D aberrants.....	75
3.3.6	Densification par appariement de triangles homologues.....	77
3.4	Évaluation de l'algorithme sur des images réelles.....	79
3.4.1	Protocole d'expérimentation.....	79
3.4.2	Résultats.....	81
3.4.3	Comparaison à une méthode de mise en correspondance dense.....	83
3.5	Conclusion.....	84

La reconstruction 3D obtenue par stéréovision repose entièrement sur la mise en correspondance faite dans les images initiales. Sa qualité conditionne celle du modèle 3D. La question qui se pose alors est de savoir comment juger de la qualité d'une mise en correspondance. Dans le chapitre précédent, nous avons vu que deux types de bruit pouvaient être présents dans une mise en correspondance : d'une part, le bruit de localisation des points d'intérêt engendrant de mauvais appariements, et d'autre part les faux appariements. Les mauvais appariements sont des points présentant un fort résidu par rapport à la matrice fondamentale. Les faux appariements reliant deux projections qui ne correspondent pas au même point physique sur l'objet. Ils sont la plupart du temps éliminés par la contrainte épipolaire mais sont parfois cohérents avec cette dernière. Il faut alors mettre en œuvre d'autres techniques pour les éliminer.

Il est évident que plus ces bruits sont faibles plus la mise en correspondance est de qualité. Cependant, le nombre de points appariés intervient également dans la qualité d'une mise en correspondance ayant pour objectif la reconstruction 3D métrique par auto-calibrage. En effet, Sturm et al. ont montré par expérimentation que plus le nombre de points est important, plus la distance focale estimée par auto-calibrage est précise [Sturm, 2005]. La précision augmente avec le nombre de points en suivant un comportement asymptotique. La convergence vers l'asymptote est plus ou moins rapide selon les positions relatives des capteurs d'image, mais leurs résultats montrent qu'un minimum de 500 points suffit dans la plupart des cas. En outre, plus le nombre de points est important, plus le modèle 3D peut présenter les détails de la scène reconstruite.

L'objectif de ce chapitre est donc de mettre en place une méthode de mise en correspondance offrant le maximum de points appariés pertinents. Le contexte des prises de vue de notre application est rappelé en première partie, justifiant ainsi les choix des méthodes. L'algorithme de mise en correspondance développé est ensuite décrit en détail, puis testé sur des images réelles.

3.1 Contexte de prise de vue des escarres

Pour simplifier les prises de vue, une liberté quasi totale est laissée au photographe amateur. Les images de l'escarre peuvent donc être prises de points de vue très différents. En outre, un grand angle de vergence des axes optiques permet d'avoir une meilleure triangulation. Cette grande liberté à l'acquisition des images se transforme en une difficulté accrue pour les algorithmes de mise en correspondance. Les images pouvant être prises de points de vue très différents sont susceptibles de ne présenter que très peu de parties semblables à cause des effets de perspective. Il faut alors mettre en œuvre des techniques capables de dégager des caractéristiques locales communes aux deux images, afin de les appairer. Ce problème d'appariement d'images prises de points de vue très différents est un problème encore ouvert en vision par ordinateur et fait l'objet de nombreuses recherches. Certaines méthodes s'appuient sur des hypothèses concernant la scène ou les conditions de prises de vue. Par exemple, Elias propose une méthode où les paramètres intrinsèques des capteurs d'image sont connus et où l'image contient un plan de référence [Elias, 2004]. L'appariement de points d'intérêt est ensuite réalisé par une corrélation homographique entre

les plans de référence. Lu et Manduchi proposent une méthode qui nécessite que les épipoles des images soient à l'infini, c'est-à-dire que les plans des capteurs d'image soient confondus [Lu, 2004]. Cette condition n'est jamais vérifiée dans le cas d'images non calibrées. Enfin, l'algorithme de Lourakis et al. nécessite que la scène contienne au minimum deux plans [Lourakis, 2000b]. Dans le cas d'images d'escarres, la scène ne contenant aucun plan et les conditions de prises de vue étant complètement inconnues, la mise en œuvre de telles techniques est impossible.

D'autres recherches ont porté sur l'appariement de primitives plus discriminantes que les points. Bay et al. utilisent les segments de droite comme primitives, appariés grâce aux profils de couleur présents de part et d'autre des segments [Bay, 2005]. Pritchett et Zisserman proposent de grouper les segments appariés de sorte à former des parallélogrammes [Pritchett, 1998]. Une homographie existe entre les parallélogrammes appariés entre les deux images, qui permet de réaliser un appariement de points détectés à l'intérieur des parallélogrammes. Ces travaux ne sont pas transposables à notre application, nos images étant dépourvues de segments de droite.

Enfin, les primitives les plus répandues dans le cadre d'un appariement dans des images de points de vue éloignés, sont les régions d'intérêt appelées *distinguished regions* (DRs) par Matas et al. [Matas, 2002]. Il s'agit de détecter dans les images des régions ayant des caractéristiques discriminantes et invariantes aux changements de points de vue. Matas et al. détectent ces régions par un algorithme de type ligne de séparation des eaux (ou *watershed* [Vincent, 1991]) sur l'intensité des images. Quelle que soit la manière de détecter les régions, elles sont décrites de façon invariante aux transformations affines, rendant ainsi la description robuste aux changements de points de vue. Matas et al. utilisent les moments rotationnels pour cette description [Matas, 2003]. Tuytelaars et Van Gool utilisent d'autres descripteurs appelés moments couleur généralisés [Tuytelaars, 2000]. Enfin, des descripteurs peuvent être calculés à partir des textures des régions, comme le font Schaffalitzky et Zisserman [Schaffalitzky, 2001].

Parmi tous ces descripteurs, le descripteur SIFT (*Scale Invariant Feature Transform*) proposé par Lowe [Lowe, 2004], fait référence dans la communauté de la vision par ordinateur. Il est très tolérant au changement d'échelle, d'éclairage et de point de vue. Une étude comparative menée par Mikolajczyk et Schmid a montré la supériorité de ce descripteur face à ses concurrents, dans le cas de plusieurs transformations [Mikolajczyk, 2005]. Ce descripteur a récemment été utilisé par Delponte et al. dans un algorithme d'appariement d'images basé sur une décomposition en valeurs singulières (SVD) d'une matrice de correspondances, les éléments de cette matrice étant les scores d'appariement [Delponte, 2005].

Mais toutes ces méthodes ne permettent pas de trouver un nombre de correspondances autorisant une reconstruction 3D suffisamment précise. Il est donc nécessaire d'augmenter le nombre d'appariements afin d'améliorer la densité du modèle 3D obtenu en fin de chaîne. Nous avons alors développé un processus itératif permettant de « booster » le nombre de correspondances à partir d'appariements en nombre réduit mais robustes. Ce processus présenté dans une communication [Albouy, 2006], est détaillé dans ce chapitre, après la description des méthodes couramment utilisées en vision par ordinateur pour appairer des points d'intérêt.

3.2 Méthodes courantes d'appariement de points d'intérêt

La mise en correspondance dans les images consiste à appairer des primitives géométriques. Les primitives peuvent être, par exemple, des régions, des courbes, des segments de droite ou des points d'intérêt. En vision par ordinateur, les primitives les plus couramment utilisées sont les points d'intérêt, car les primitives de plus haut niveau sont instables d'une vue à l'autre.

Plusieurs méthodes de détection de points d'intérêt existent dans la littérature, différenciées par la définition des caractéristiques du signal à rechercher. Cependant, le détecteur de coins de Harris et Stephens [Harris, 1988] est le plus couramment utilisé et simple d'implémentation. Ce détecteur est réputé robuste aux changements pouvant intervenir dans les images, tant d'un point de vue luminance que géométrique. Cette réputation est corroborée par les tests comparatifs de détecteurs effectués par Schmid [Schmid, 2000].

3.2.1 Détecteur de coins de Harris et Stephens adapté à la couleur

Le détecteur d'Harris et Stephens utilise la fonction d'auto-corrélation du signal image en niveaux de gris pour détecter les lieux de forte courbure. Une matrice M dépendante de la fonction d'auto-corrélation est calculée pour chaque pixel à partir des dérivées premières de l'image. Dans le cas des images en niveaux de gris, l'expression de cette matrice M est :

$$M = G(\tilde{\sigma}) \otimes \begin{pmatrix} I_x^2(\sigma) & I_x(\sigma)I_y(\sigma) \\ I_x(\sigma)I_y(\sigma) & I_y^2(\sigma) \end{pmatrix} \quad (3.1)$$

où $I_x(\sigma)$ et $I_y(\sigma)$ sont respectivement les dérivées selon les axes x et y de l'image filtrée par une gaussienne d'écart-type σ , et $G(\tilde{\sigma}) \otimes$ dénote le filtrage par une gaussienne d'écart-type $\tilde{\sigma}$. Dans notre implémentation, nous avons fixé les écart-types des gaussiennes à 1. Les dérivées sont calculées par convolution avec des dérivées de gaussiennes dans le but de réduire l'influence des bruits haute fréquence. La réponse du détecteur est alors calculée à partir de cette matrice selon l'équation :

$$R = \text{Det}(M) - k \cdot \text{Trace}^2(M) \quad \text{avec} \quad k = 0.04 \quad (3.2)$$

Les maxima locaux de la réponse du détecteur correspondent aux coins présents dans l'image. Il est possible de ne détecter que les coins les plus riches en information en appliquant un seuil sur la réponse R du détecteur, ou en ne relevant que les N plus grands maxima locaux.

Dans le cas des images couleur, Montesinos et al. proposent d'intégrer les informations couleur à la matrice M de l'équation 3.1 [Montesinos, 1998]. Pour une image représentée dans l'espace de couleurs RVB, la matrice M devient :

$$M = \begin{pmatrix} R_x^2(\sigma) + V_x^2(\sigma) + B_x^2(\sigma) & R_x(\sigma)R_y(\sigma) + V_x(\sigma)V_y(\sigma) + B_x(\sigma)B_y(\sigma) \\ R_x(\sigma)R_y(\sigma) + V_x(\sigma)V_y(\sigma) + B_x(\sigma)B_y(\sigma) & R_y^2(\sigma) + V_y^2(\sigma) + B_y^2(\sigma) \end{pmatrix} \quad (3.3)$$

Une fois des points d'intérêt détectés dans les images, il s'agit de définir des caractéristiques de voisinage, en vue de les appairer entre les images. La méthode la plus simple et la plus répandue est de caractériser les points d'intérêt par les valeurs des pixels présents dans leurs voisinages.

3.2.2 Caractérisation locale des points d'intérêt

Le voisinage d'un point d'intérêt forme une imagerie extraite de l'image qui sera comparée avec les imageries des points d'intérêt des autres images. Cette comparaison est traditionnellement réalisée par une mesure de corrélation entre les motifs. L'inconvénient de ces méthodes est qu'elles ne sont pas robustes aux transformations géométriques de l'image, hormis la translation.

Afin de pallier cet inconvénient, on préfère extraire du voisinage des points des caractéristiques robustes aux transformations géométriques courantes dans les images (translation, rotation, zoom). Ces caractéristiques portent le nom d'invariants. Par exemple, les invariants différentiels sont calculés à partir de combinaisons des dérivées successives de l'image au point considéré. D'autres techniques comme les moments statistiques ou les descriptions fréquentielles sont parfois utilisées pour caractériser l'image au niveau des points d'intérêt. Gouet a réalisé une étude approfondie de ces techniques dans [Gouet, 2000].

Enfin, il existe des techniques basées sur des histogrammes calculés à partir de la distribution des niveaux des pixels entourant le point d'intérêt. Ces techniques sont très discriminantes dans le cas des images couleur et ont l'avantage d'être robustes aux transformations géométriques.

Mikolajczyk et Schmid ont réalisé une étude comparative de descripteurs locaux [Mikolajczyk, 2005]. Neuf descripteurs récents dont SIFT, sont évalués dans l'optique de la mise en correspondance. Les images de test pour la mise en correspondance présentent des transformations affines, des changements d'échelle, des rotations, du flou, de la compression JPEG et des changements d'illumination. Il ressort de cet article que la méthode SIFT est la plus robuste face à toutes ces modifications d'image. Les auteurs ont également mis au point une extension de ce descripteur, baptisée GLOH (*Gradient Location-Orientation Histogram*). Cette méthode se différencie de SIFT par une approche polaire plutôt que cartésienne pour le voisinage des points d'intérêt. De plus, la taille du vecteur descripteur est diminuée grâce à une analyse en composantes principales. Les tests effectués révèlent une légère amélioration des performances par rapport à SIFT.

3.2.3 Méthodes d'appariement

L'appariement entre deux images consiste à trouver les points d'intérêt homologues en se basant sur leurs descriptions. Entre chaque paire de points d'intérêt, un score d'appariement est calculé. La définition de ce score dépend de la caractérisation qui a été faite selon une des méthodes présentées dans la section précédente. De nombreux algorithmes d'appariement existent, nous détaillons ici les plus connus.

3.2.3.1 Algorithme « Winner Takes All »

Cet algorithme est le plus immédiat. Il s'agit d'apparier chaque point de la première image avec celui de la seconde image dont le score d'appariement est le meilleur. En plus d'être simple et rapide, cette méthode garantit le respect de la contrainte d'unicité, selon laquelle un point doit avoir un unique correspondant dans l'autre image. Cependant, elle comporte de forts inconvénients. Tout d'abord, elle n'est pas symétrique, c'est-à-dire qu'elle ne fournirait pas le même résultat en utilisant la seconde image comme image de départ. De plus, l'algorithme « Winner Takes All » présente l'inconvénient d'imposer un correspondant aux points qui sont occultés dans la seconde image.

Pour pallier l'asymétrie de la méthode, il est possible de réaliser un appariement croisé. Le principe est d'appliquer l'algorithme de la première image vers la seconde, puis dans un second temps, de la seconde vers la première. Seuls les appariements identiques entre les deux passes de l'algorithme sont conservés (fig. 3.1).

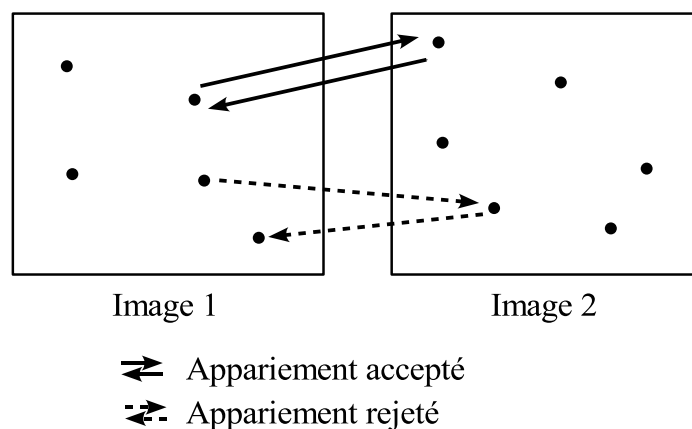


Figure 3.1 : Appariement croisé.

En pratique, il n'est pas nécessaire d'exécuter l'algorithme deux fois. Il suffit de construire une matrice S dont l'élément $S_{i,j}$ est le score d'appariement entre le $i^{\text{ème}}$ point de la première image et le $j^{\text{ème}}$ point de la seconde image. Les maxima selon les lignes et les colonnes indiquent si les couples de points doivent être conservés (fig. 3.2). Si l'élément $S_{i,j}$ est à la fois un maximum pour sa ligne et sa colonne, alors il correspond à un appariement croisé. Inversement, si le maximum de la ligne i et le maximum de la colonne j ne sont pas l'élément $S_{i,j}$, alors l'appariement n'est pas croisé et il est rejeté.

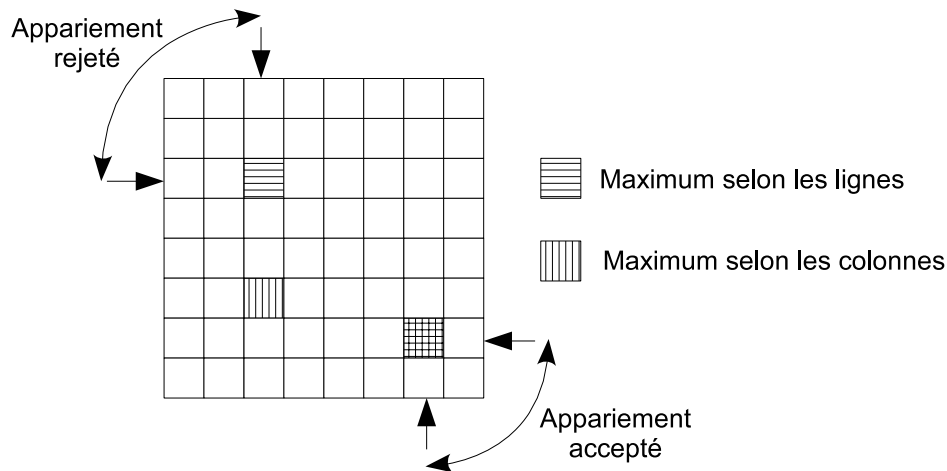


Figure 3.2 : Implémentation de l'appariement croisé.

L'utilisation de l'algorithme « *Winner Takes All* » croisé permet donc de respecter à la fois la symétrie et la contrainte d'unicité. Cependant, dans la pratique, certains scores d'appariements peuvent être très proches, ce qui provoque une ambiguïté qu'il est possible de lever en comparant la différence entre les deux maxima et un seuil. Si l'écart est insuffisant, alors l'appariement est considéré ambigu et donc rejeté. Une autre solution consiste à augmenter la taille des descripteurs pour qu'ils deviennent plus discriminants, mais ceci induit également une augmentation du temps de calcul.

Par ailleurs, certains maxima peuvent ne pas être significatifs car très faibles, mais être tout de même conservés. Ces appariements non significatifs ont un fort risque d'être faux. Pour éviter de sélectionner de tels appariements, on peut procéder à un seuillage des scores. La difficulté réside dans le choix de ce seuil. Un seuil trop faible permettra d'avoir plus d'appariements mais augmentera le risque d'en sélectionner des faux. Inversement, un seuil trop haut risque de ne pas permettre de sélectionner suffisamment d'appariements.

Afin de lever les difficultés précédentes, des algorithmes de relaxation sont fréquemment utilisés en vision par ordinateur. Ces algorithmes ne considèrent plus les couples de points de façon isolée, mais intègrent les points voisins dans la décision d'appariement.

3.2.3.2 Algorithmes de relaxation

Les algorithmes de relaxation permettent de minimiser une fonction d'énergie globale, appelée critère de relaxation. Ce critère est établi en fonction de la cohérence de l'ensemble des appariements. De façon générale, un ensemble d'appariements est cohérent lorsque pour chaque appariement (m_1, m_2) il existe dans le voisinage du point m_1 un nombre important de correspondants des points voisins du second point m_2 . Ce principe, appelé contrainte de voisinage, est illustré par la figure 3.3. Cette contrainte est notamment utilisée par [Zhang, 1994]. Parfois, des contraintes géométriques accompagnent cette contrainte de voisinage. Par exemple, la conservation de l'angle entre les différents points du voisinage est utilisée par [Schmid, 1996].

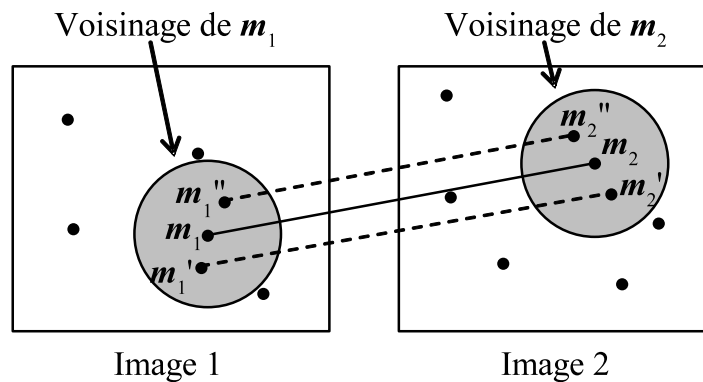


Figure 3.3 : Contrainte de voisinage.

Les détails d'implémentation des algorithmes de relaxation peuvent être trouvés dans [Gouet, 2000]. Ces algorithmes offrent de meilleurs résultats que l'algorithme « Winner Takes All » en terme de nombre de correspondances et de proportion d'appariements vrais. Cependant, l'aspect itératif du processus de relaxation induit un temps de calcul nettement supérieur.

Nous avons testé ces méthodes directement sur nos images en utilisant une mesure de corrélation entre les voisinages des points d'intérêt comme score d'appariement. Malheureusement, dans notre application, les points de vue sont souvent trop écartés pour que ces méthodes offrent une mise en correspondance correcte. Le résultat est alors un ensemble d'appariements peu nombreux ou possédant une majorité de faux appariements. Nous avons donc imaginé un processus itératif de mise en correspondance, initialisé par un ensemble de correspondances peu nombreuses mais robustes. Cet algorithme est détaillé dans la section suivante.

3.3 Algorithme itératif de mise en correspondance

L'objectif de notre algorithme est de réaliser une mise en correspondance suffisamment dense de deux images couleur, prises de deux points de vue assez différents, afin de minimiser l'erreur de la reconstruction 3D. L'algorithme imaginé est itératif. La mise en correspondance s'affine au cours des itérations suivantes. Chaque itération ajoute de nouveaux appariements et supprime les mauvais de l'itération précédente. Notre algorithme est schématisé sur la figure 3.4.

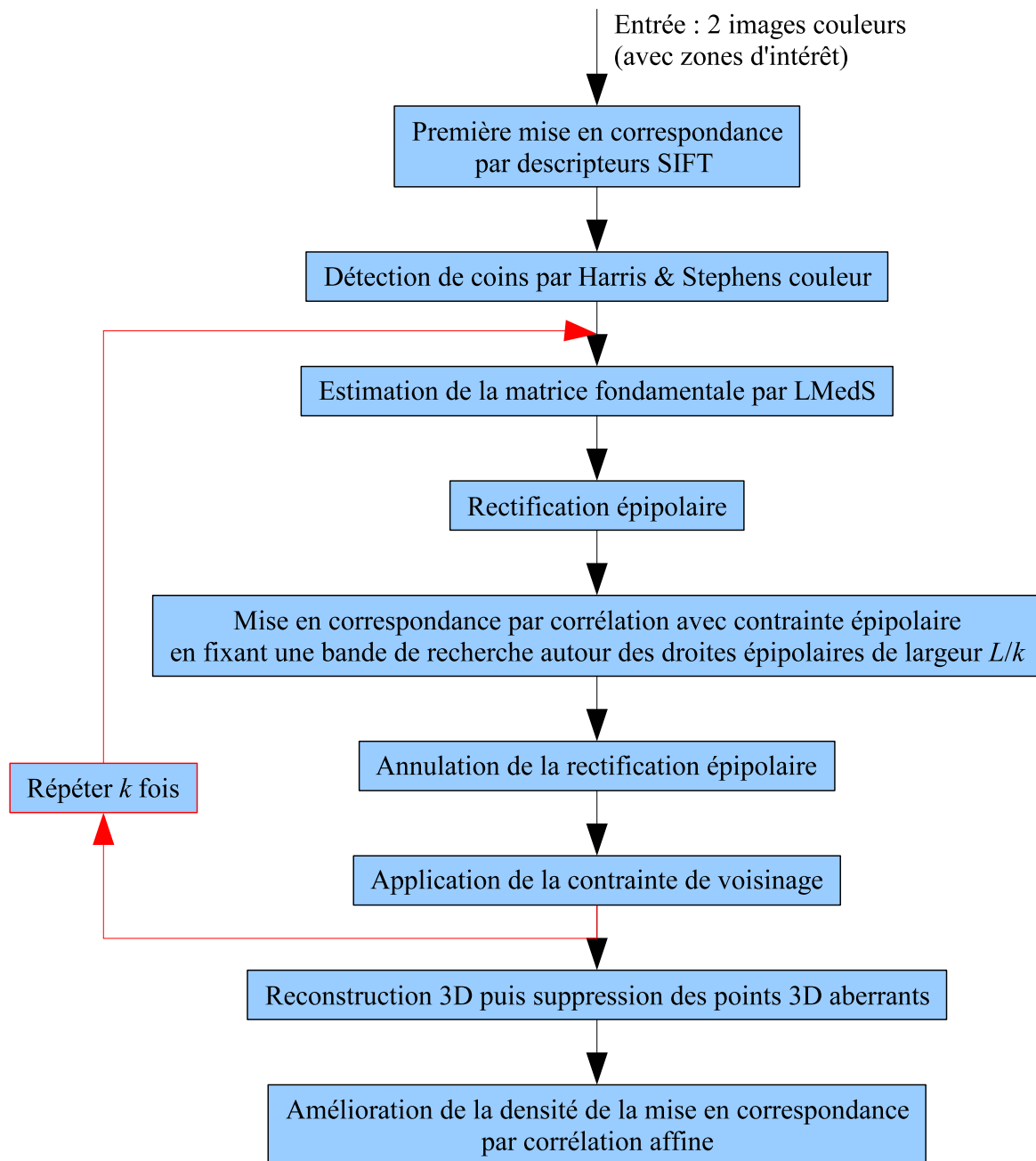


Figure 3.4 : Algorithme de mise en correspondance des images.

Chaque étape de l'algorithme est maintenant détaillée et illustrée dans les sous-sections suivantes.

3.3.1 Images en entrée de l'algorithme

L'algorithme accepte en entrée les deux images à apparier. Sur chaque image, une zone d'intérêt peut être définie par l'utilisateur pour limiter spatialement la zone à traiter. Ceci a le double intérêt d'accélérer les temps de calcul en réduisant les espaces de recherche, et de

fournir en bout de chaîne un modèle 3D qui correspond exactement aux attentes de l'utilisateur. Cette zone définie à main levée n'est donc pas nécessairement rectangulaire. Elle est codée par un masque de l'image où, pour chaque pixel de l'image, une valeur booléenne est définie pour savoir si le pixel est à l'intérieur (1) ou à l'extérieur (0) de la zone. La forme et la topologie de la zone est alors complètement libre. Cette étape d'initialisation manuelle pourra être remplacée par une détection automatique du contour de l'escarre. La figure 3.5 représente les deux images d'entrée et les zones d'intérêt qui serviront à illustrer les autres étapes de l'algorithme. Comme on le voit sur ces images, un repère coloré constitué de deux petites sphères bleue et verte reliées par une tige rouge est disposé au plus proche de l'escarre. Ce repère est utilisé pour fixer le facteur d'échelle de la reconstruction 3D et participe à la mise en correspondance des images en fournissant deux appariements robustes conservés tout au long du processus de mise en correspondance. Les centres des deux sphères sont automatiquement détectés dans les images grâce à leur forme et leurs couleurs caractéristiques.

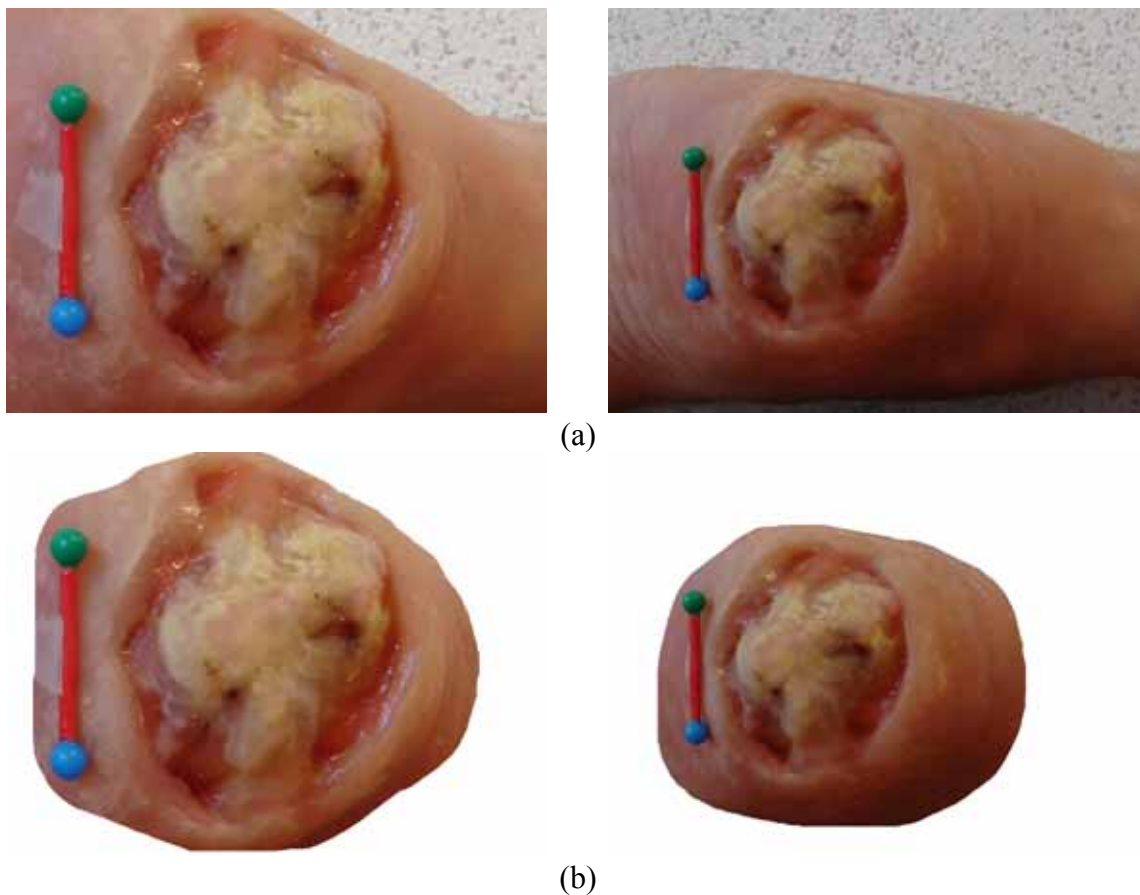


Figure 3.5 : Images en entrée de l'algorithme (a) et leurs zones d'intérêt (b).

3.3.2 Mise en correspondance initiale

Une première mise en correspondance est réalisée en utilisant le descripteur SIFT [Lowe, 2004]. Elle est effectuée par un algorithme de type *winner takes all* sur les scores

d'appariement des descripteurs. Pour chaque point m_i de la première image, deux distances d_1 et d_2 sont calculées, correspondant aux deux distances euclidiennes les plus faibles entre le descripteur de m_i et les descripteurs des points sur la seconde image. Le score d'appariement est alors le rapport d_1/d_2 , qui mesure l'ambiguïté de l'appariement. Un score proche de 0 ($d_1 \ll d_2$) signifie que l'appariement n'est pas ambigu au contraire d'un score proche de 1 ($d_1 \simeq d_2$). Seuls les appariements présentant un score inférieur à 0,6, c'est-à-dire présentant une faible ambiguïté, sont conservés. Le choix de ce seuil est inspiré de l'implémentation de Lowe.

Les étapes suivantes, notamment la rectification épipolaire, nécessitent un minimum de 10 appariements. Dans le cas où ce nombre ne serait pas atteint par l'algorithme de mise en correspondance précédent, les appariements de scores supérieurs à 0,6 sont sélectionnés par ordre croissant afin d'obtenir 20 appariements. Ceci se fait au risque de sélectionner de faux appariements, mais qui sont la plupart du temps éliminés par le calcul robuste de la matrice fondamentale. C'est la raison pour laquelle le nombre de 20 a été choisi, ce qui assure qu'après l'élimination des faux appariements, il reste toujours un minimum de 10 appariements.

La figure 3.6 illustre cette première mise en correspondance où 19 appariements ont été trouvés. Les vecteurs blancs sur l'image sont les vecteurs d'appariement des points entre les deux images, représentés sur la première image.

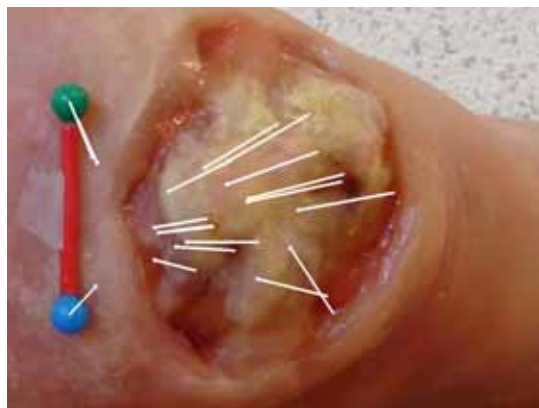


Figure 3.6 : Première mise en correspondance réalisée grâce aux descripteurs SIFT.

3.3.3 Détection de points d'intérêt

Pour servir de base au processus itératif, 2000 coins sont détectés dans les zones d'intérêt des images par l'algorithme d'Harris et Stephens [Harris, 1988] adapté aux images couleur par Gouet [Gouet, 2000]. Un exemple de détection de ces coins est donné en figure 3.7. La détection des coins est réalisée une seule fois en dehors de la boucle de mise en correspondance.



Figure 3.7 : Détection de 2000 coins dans les zones d'intérêt des images.

3.3.4 Raffinement de la mise en correspondance par un processus itératif

Le premier calcul de la matrice fondamentale réalisé à partir des appariements robustes initiaux permet de sélectionner de nouvelles paires de points parmi les 2000 détectés dans les images en appliquant la contrainte épipolaire. Le raffinement de la mise en correspondance s'effectue de façon itérative en diminuant à chaque itération la bande permise de recherche des correspondants le long des droites épipolaires. Ainsi, à chaque itération, la matrice fondamentale et la mise en correspondance s'affinent mutuellement. Chaque itération comprend 5 étapes détaillées dans les paragraphes suivants. Le nombre d'itérations a été arbitrairement fixé à 7 dans un premier temps.

3.3.4.1 Estimation de la matrice fondamentale

La première étape consiste à estimer la matrice fondamentale régissant la géométrie épipolaire entre les deux vues. Cette estimation est réalisée selon la méthode LMedS, suivie de l'algorithme des moindres carrés orthogonaux exécuté sur l'ensemble des *inliers* détectés par LMedS, conformément à l'étude menée au chapitre précédent. Une fois cette matrice obtenue, il est possible d'appliquer la contrainte épipolaire pour rechercher les appariements. Cette contrainte s'applique grâce aux deux étapes suivantes.

3.3.4.2 Rectification épipolaire

Le but de la rectification épipolaire est de déformer les images de sorte que les droites épipolaires se retrouvent parallèles et horizontales. Ainsi la première ligne de pixels de la première image devient une droite épipolaire, et son image dans la seconde image via la matrice fondamentale est également la première ligne de pixels. Cette transformation a pour

but de faciliter la recherche des correspondants dans les images par l'application de la contrainte épipolaire. La recherche est restreinte à une bande horizontale.

Il existe deux méthodes pour réaliser ces transformations. La première consiste en une homographie des images dont le but est de placer les épipoles à l'infini (fig. 3.8b). Loops et Zhang ont établi un algorithme performant pour réaliser cette tâche [Loops, 1999]. L'inconvénient majeur de cette méthode réside dans le fait que les images peuvent devenir très grandes lorsque les épipoles sont proches des images. En outre, cette méthode devient impossible si un des épipoles est à l'intérieur de l'image.

La seconde méthode, appelée rectification générale, consiste à « dérouler » les images autour des épipoles. Le système de coordonnées des images rectifiées devient polaire, chaque ligne de pixels correspondant à une valeur d'angle de la droite épipolaire et chaque colonne à une distance par rapport à l'épipole (fig. 3.8c). Cette méthode fonctionne dans tous les cas de figures, y compris lorsque les épipoles sont inclus dans les images. Nous avons donc préféré cette méthode et avons utilisé l'implémentation proposée par Oram [Oram, 2001].

A titre d'illustration, la figure 3.9 présente les différentes rectifications réalisées au cours des 7 itérations du processus de raffinement. On peut noter que les rectifications sont très semblables à partir de la quatrième itération, prouvant que la géométrie épipolaire est stabilisée sur les itérations suivantes.

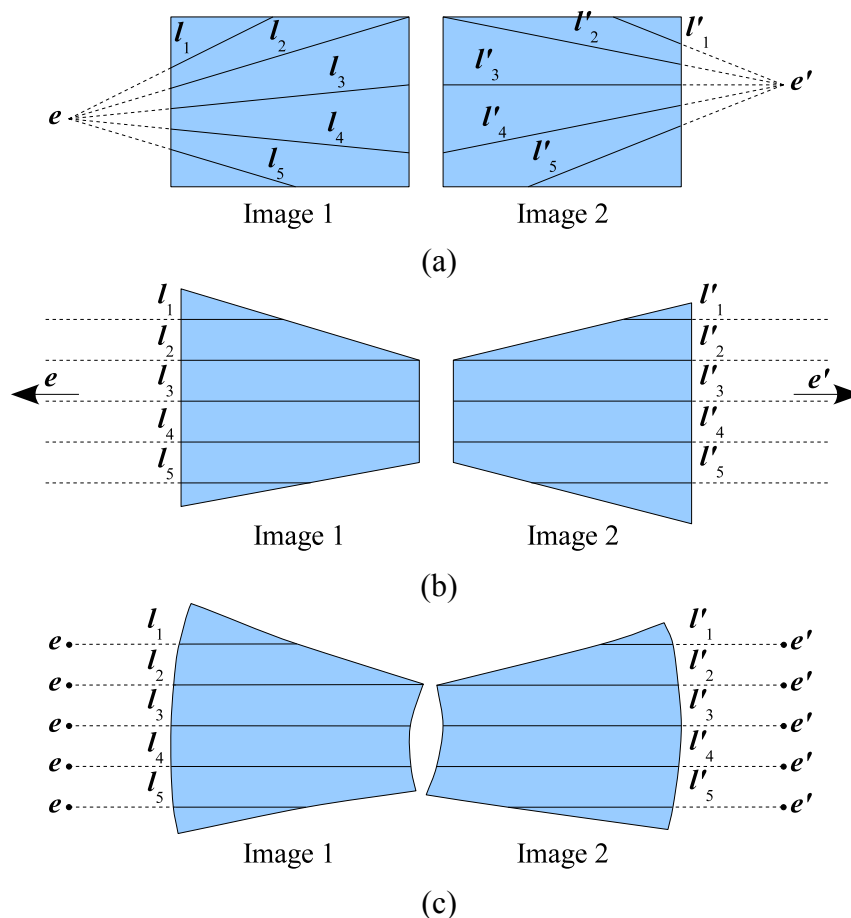


Figure 3.8 : Rectification épipolaire d'images. Images non rectifiées (a), rectification par homographie (b) et rectification générale (c).

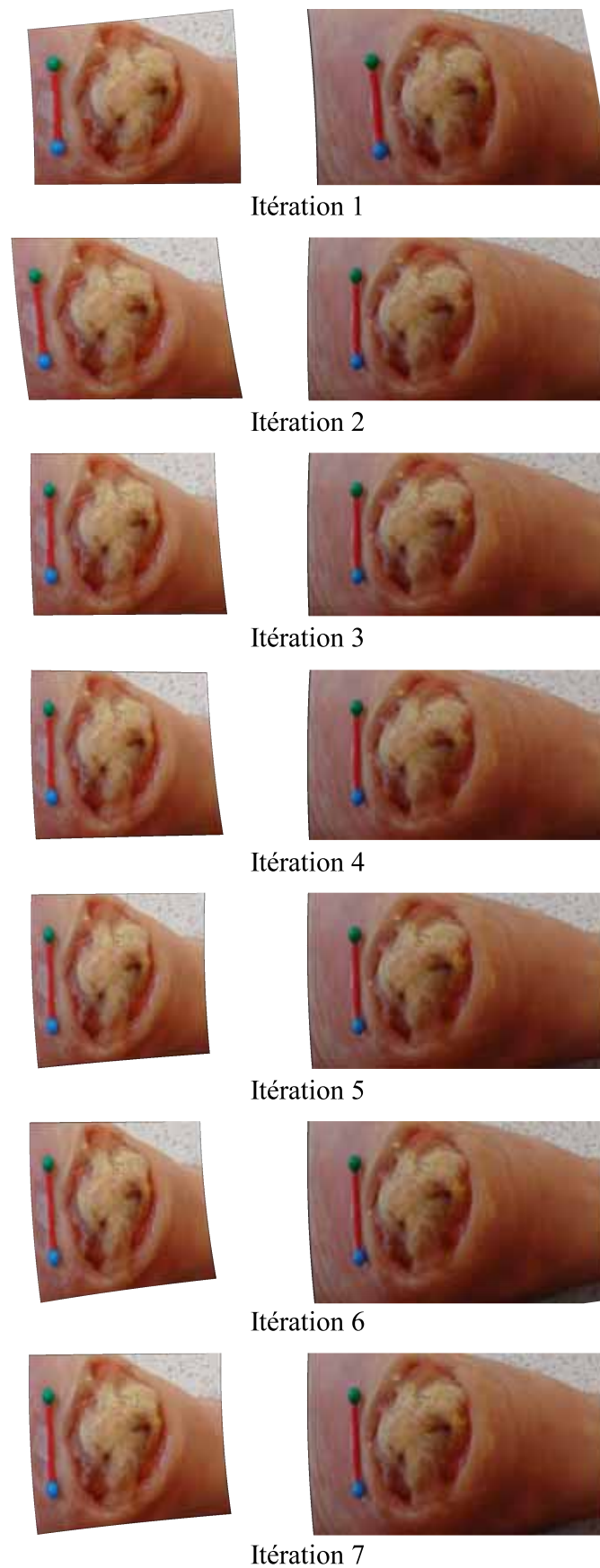


Figure 3.9 : Rectification des images au cours des 7 itérations du raffinement.

3.3.4.3 Application de la contrainte épipolaire

La rectification des images simplifie l'application de la contrainte épipolaire pour la mise en correspondance. La contrainte est généralement appliquée en limitant la recherche des correspondants dans une bande de largeur fixe autour des lignes épipolaires, afin de tolérer d'éventuelles erreurs sur l'estimation de la matrice fondamentale. L'intérêt de notre processus itératif est d'appliquer une largeur de la bande de recherche L/k de plus en plus fine au fil des itérations (fig. 3.10). Cette décroissance implique une contrainte épipolaire de plus en plus forte pour une estimation de plus en plus robuste. En effet, aux premières itérations, la matrice fondamentale est calculée à partir de très peu de points (parfois moins d'une vingtaine), rendant instable l'estimation de la géométrie épipolaire. Plus le nombre d'itérations augmente, plus le nombre de points appariés doit augmenter, rendant ainsi l'estimation de la matrice fondamentale plus fiable. La bande de tolérance autour des droites épipolaires peut alors être réduite afin de limiter le nombre de scores d'appariement ambigus. Dans notre exemple, L est fixé à 35 pixels. La bande de recherche est progressivement réduite jusqu'à plus ou moins 5 pixels à la dernière itération.

Dans chaque bande, la recherche des correspondants est réalisée par l'algorithme *winner takes all* croisé avec pour score de correspondance la mesure ZSAD (*Zero-mean Sum of Absolute Difference*) calculée sur des voisinages de 31×31 pixels autour des points d'intérêt. La mesure ZSAD est donnée par l'expression :

$$\sum_{p=-P}^P \sum_{q=-Q}^Q \left| [I(i+p, j+q) - \bar{I}(i, j)] - [I'(i'+p, j'+q) - \bar{I}'(i', j')] \right| \quad (3.4)$$

où P et Q sont respectivement les demi tailles des voisinages ($P = Q = 15$), I et I' sont les fonctions d'intensité des images rectifiées, (i, j) et (i', j') sont les coordonnées des points d'intérêt sur les images, et \bar{I} et \bar{I}' sont les intensités moyennes des images sur le voisinage de taille $(2P+1) \times (2Q+1)$ centré sur les points d'intérêt. Cette mesure locale est strictement positive, une valeur proche de 0 reflète une grande ressemblance des voisinages.

Par ailleurs, la recherche des points homologues est restreinte à un segment de W pixels (W est fixé à 100 dans cet exemple), ce qui correspond à un seuil de disparité.

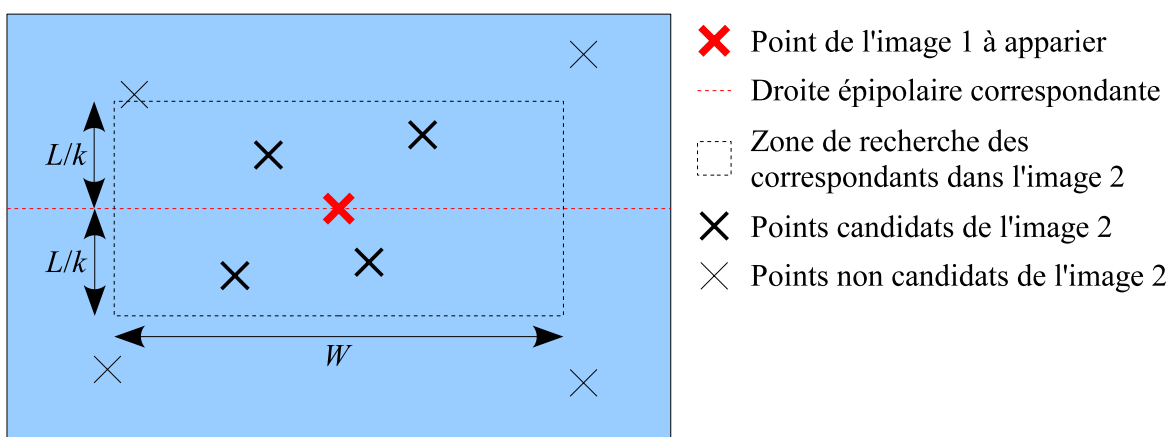


Figure 3.10 : Application de la contrainte épipolaire.

Une fois cette contrainte appliquée, la rectification des images est annulée afin de replacer les images dans un cadre cartésien.

3.3.4.4 Application de la contrainte de voisinage

Cette contrainte est utilisée dans ce processus pour vérifier la cohérence locale des appariements. Si les voisins d'un point d'intérêt sont tous appariés dans une même zone de l'autre image, alors ce point a de fortes chances d'être apparié dans la même zone. L'application de cette contrainte peut-être effectuée par une analyse des vecteurs d'appariements entre la première et la seconde image. Pour chaque vecteur, la moyenne de ses vecteurs voisins est calculée. Sont déclarés « voisins » tous les vecteurs dont le point de départ est situé à moins de 50 pixels du point de départ du vecteur considéré. Si la norme et l'argument du vecteur considéré diffèrent trop du vecteur moyen, alors l'appariement est rejeté. Les seuils de tolérance sont fixés à 25 % pour la norme et $\pi/10$ pour l'argument.

La figure 3.11 illustre l'application de cette contrainte à chaque itération. Les vecteurs représentés en noir correspondent aux appariements jugés aberrants par cette contrainte, et supprimés du lot d'appariements. Leur nombre est indiqué en haut à droite des images. Les vecteurs en blanc sont conservés et leur nombre est indiqué en bas à droite des images.

3.3.5 Détection et suppression des points 3D aberrants

L'ensemble d'appariements obtenu au bout du processus itératif permet déjà d'obtenir un modèle 3D. Cependant, la densité de ce modèle est faible (565 points dans notre exemple) et nécessite d'être augmentée pour obtenir un modèle plus fidèle à la réalité, permettant de voir plus de détails. La méthode mise en place pour « booster » le nombre d'appariements sera détaillée au paragraphe suivant. Comme elle ne tolère pas de faux appariements, nous avons mis en place une étape supplémentaire pour filtrer les correspondances conduisant à une reconstruction 3D erronée. En effet, malgré l'application des trois contraintes précédentes (épipolaire, de disparité et de voisinage), l'expérience montre que certains faux appariements peuvent subsister. Plutôt que de jouer sur les seuils de tolérance de la contrainte de voisinage au risque de rejeter de vrais appariements, nous avons mis en place une technique de filtrage du modèle 3D lui-même.

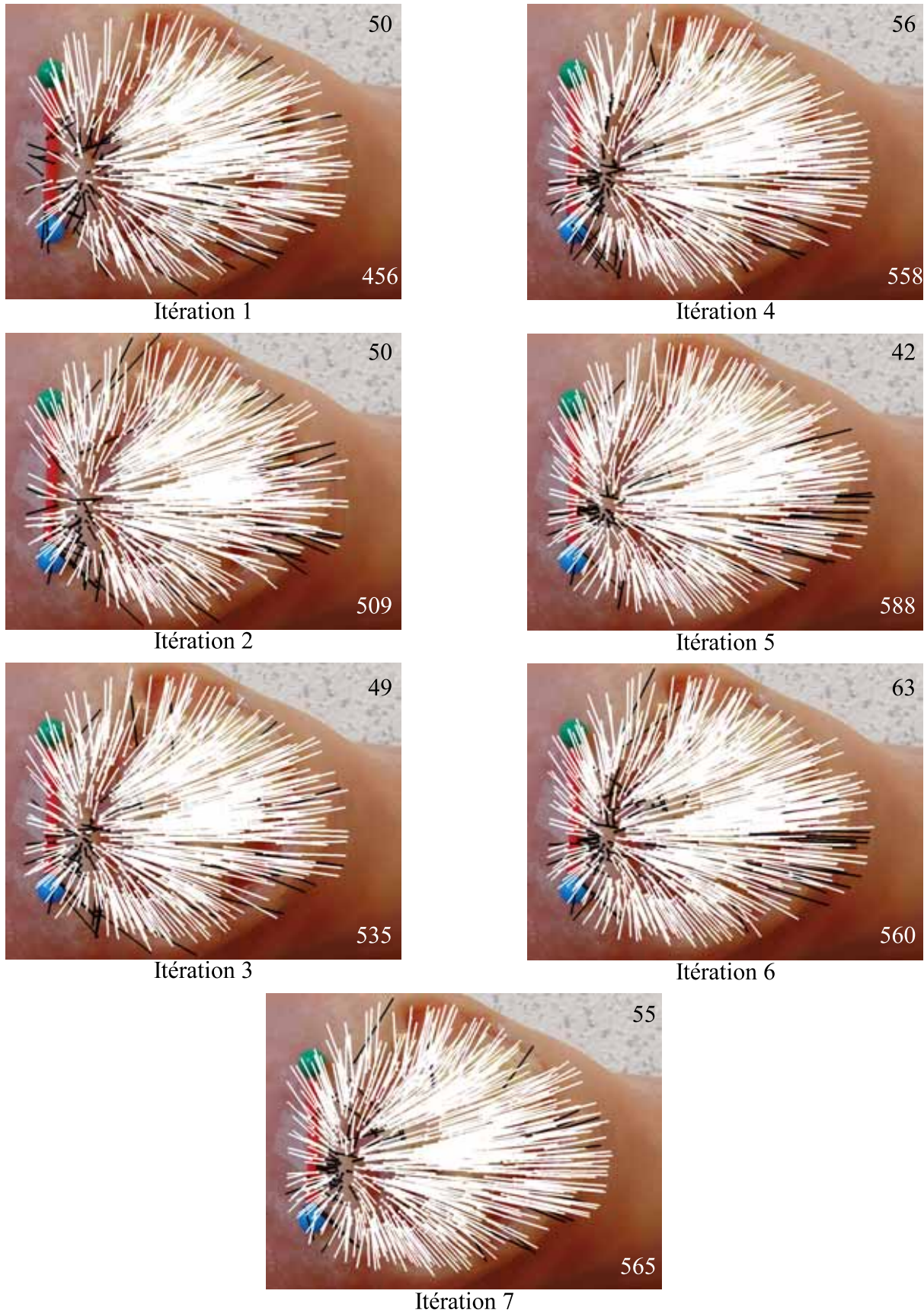


Figure 3.11 : Application de la contrainte de voisinage.

Pour cela, on réalise une reconstruction de surface 3D à partir des appariements obtenus en fin du processus itératif. Les faux appariements provoquent l'apparition de pics et de crevasses sur le modèle 3D, pouvant être assimilés à une sorte de bruit impulsionnel sur le relief de la surface 3D. Ce type de bruit peut être aisément détecté à l'aide d'un filtrage médian de la surface. Dans notre implémentation, le filtre est appliqué sur la distance du point 3D à l'origine, c'est-à-dire le premier capteur. Pour chaque point 3D, la distance médiane est calculée à partir des points 3D directement voisins sur le maillage, c'est-à-dire connectés par le biais d'une arête de triangle de la surface. Si l'écart entre la distance du point et la médiane est trop important, alors le point est considéré comme aberrant, et l'appariement associé au point 3D est supprimé de l'ensemble des correspondances. L'écart des distances est donné par la relation :

$$\frac{\left| a_i - \text{median}_j(a_j) \right|}{\max_j(a_j) - \min_j(a_j)} \quad (3.5)$$

où a_j sont les distances à l'origine des points voisins au point 3D considéré, et a_i est la distance à l'origine du point considéré. Le dénominateur permet de s'affranchir de la distance entre la surface et le capteur d'image. Dans notre expérimentation, la forme des escarres étant localement assez lisse, le seuil sur l'écart de distance a été fixé à 10%.

3.3.6 Densification par appariement de triangles homologues

Cette dernière étape permet de rendre la mise en correspondance plus dense. Le principe est de réaliser une triangulation de Delaunay à partir des points déjà mis en correspondance sur la première image, de reporter cette triangulation sur la seconde image, puis d'établir une mise en correspondance triangle à triangle.

Sous l'hypothèse que la portion triangulaire de l'objet 3D considérée soit plane, la transformation entre la surface de l'objet et son image est alors une homographie. Il existe donc une homographie H_1 entre la surface de l'objet et sa projection sur l'image 1, et une homographie H_2 entre l'objet et l'image 2. La composition de deux homographies étant une homographie, il existe donc une homographie H entre le triangle sur l'image 1 et le triangle sur l'image 2.

Dans la réalité, la portion d'objet incluse dans un triangle n'est pas rigoureusement plane. Mais cette hypothèse permet par l'homographie H de guider la mise en correspondance des points inclus dans le triangle. Le correspondant d'un point se trouvera non loin de son image par H , la différence étant liée au défaut de planéité. Il reste à déterminer cette homographie. Nous pourrions pour cela nous appuyer sur la reconstruction 3D obtenue à l'étape précédente. Cependant, pour rendre cette méthode plus générale et utilisable lorsque la reconstruction 3D n'est pas disponible, nous avons préféré estimer cette homographie par la transformation affine la plus proche. La transformation affine est calculée directement à partir des 3 sommets des triangles sur chacune des images. Le remplacement de l'homographie par une transformation affine ne provoque que peu d'erreurs, d'autant moins que l'inclinaison du plan du triangle 3D est faible par rapport aux plans des capteurs d'image.

Deux triangles homologues subissent donc une transformation affine, dans le but de les placer dans un cadre géométrique identique. Puis, des points caractéristiques supplémentaires sont détectés par l'algorithme d'Harris et Stephens adapté à la couleur. Ces points sont ensuite mis en correspondance en utilisant la corrélation ZSAD avec un algorithme de relaxation. Le score ZSAD d'un couple de points est pondéré par une fonction gaussienne de la distance entre les points d'écart-type 5 pixels pour avantager les points proches. Ceci a pour effet de discriminer les points éloignés du plan moyen.

La figure 3.12 illustre la transformation affine effectuée pour recaler les triangles dans un même référentiel et la figure 3.13 fournit un exemple de mise en correspondance utilisant cette technique.

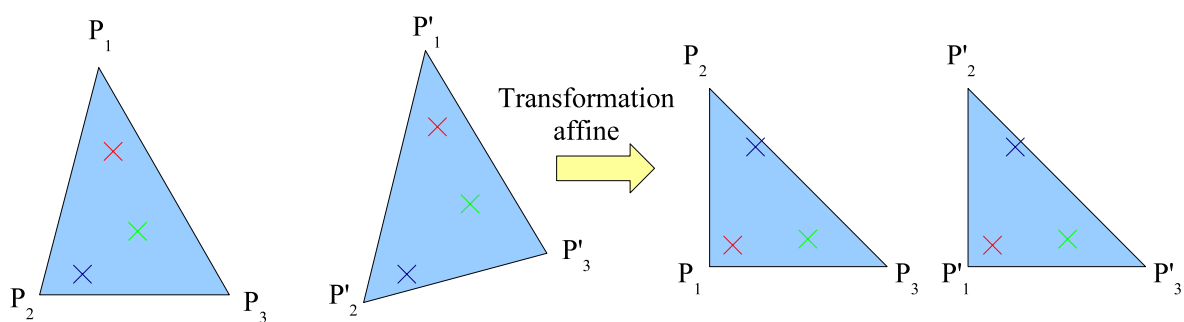


Figure 3.12 : Recalage de triangles homologues dans un même référentiel géométrique.

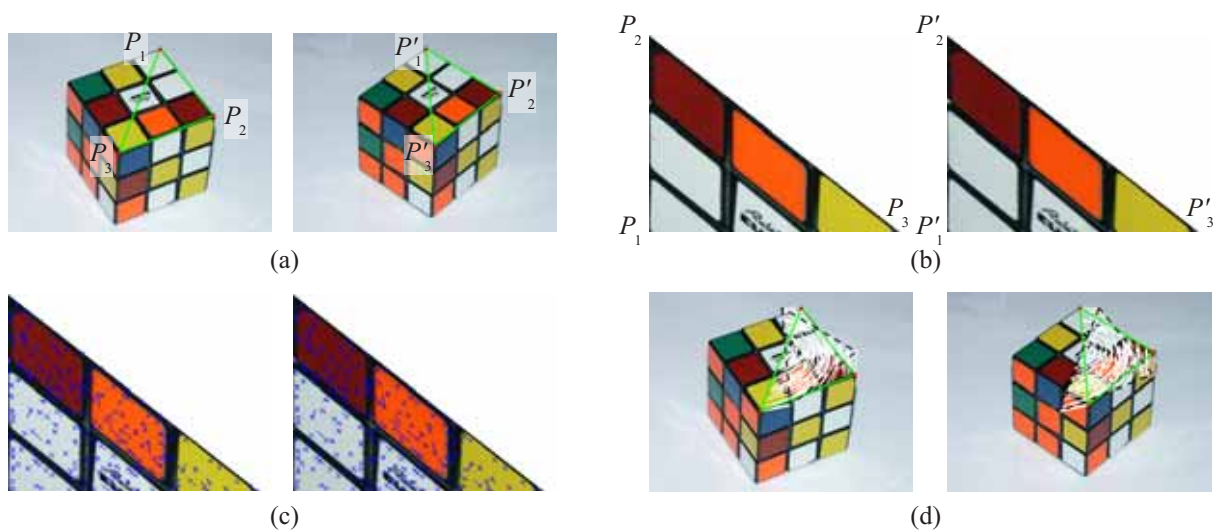


Figure 3.13 : Exemple de mise en correspondance de triangles homologues. Trois paires de sommets de triangles homologues (a). Transformation affine des triangles (b). Mise en correspondance à l'intérieur des triangles (c). Report de la mise en correspondance sur les images originales (d).

Dans l'exemple utilisé pour illustrer le processus itératif, cette dernière étape permet de passer de 565 points appariés à 4501 appariements, correspondant à un gain de 8 sur le nombre de correspondances. La figure 3.14 montre les appariements finaux obtenus par notre algorithme sur la paire d'images de la figure 3.5.

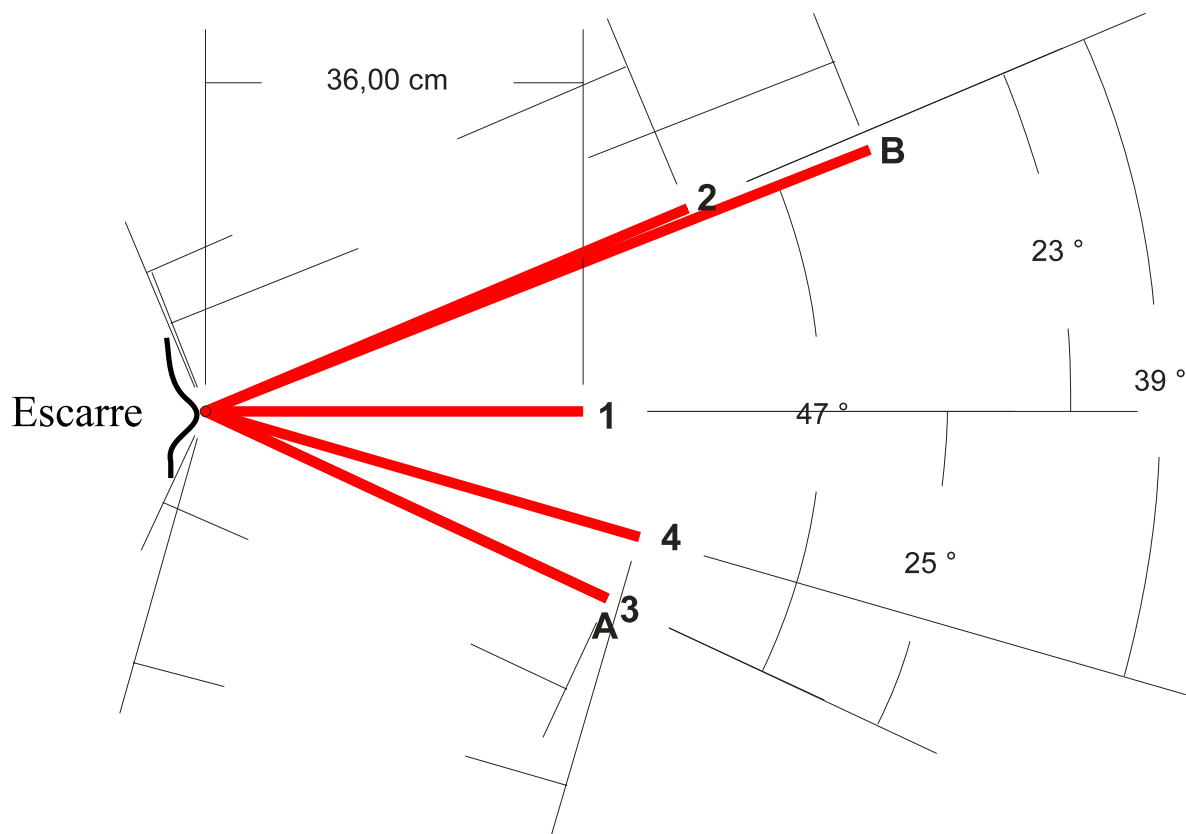


Figure 3.14 : Mise en correspondance finale (4501 appariements).

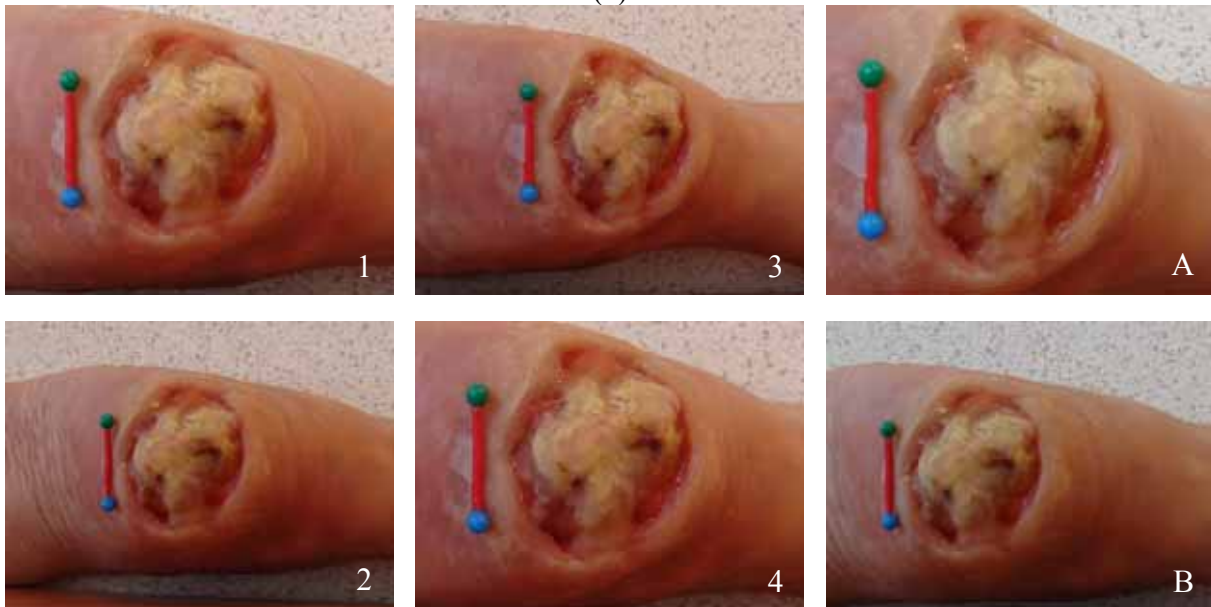
3.4 Évaluation de l'algorithme sur des images réelles

3.4.1 Protocole d'expérimentation

Nous avons analysé cet algorithme de mise en correspondance sur quatre couples d'images réelles. Les images sont des photographies d'une escarre artificielle, prises sous des points de vue différents. La figure 3.15 représente les différentes images ainsi que les positions de leurs prises de vue. Les couples d'images utilisés sont constitués des images A-B, 1-2, 1-3 et 2-4. Les images A et B ont été prises avec un niveau de zoom différent des photos 1 à 4 et ne peuvent donc pas être utilisées avec ces dernières pour réaliser une reconstruction 3D par notre méthode.



(a)



(b)

Figure 3.15 : Points de vue (a) et images (b) utilisés pour l'analyse de la mise en correspondance.

Les données analysées pour caractériser les mises en correspondance sont de trois types. Tout d'abord, la distance moyenne des points à leur droite épipolaire est calculée pour chaque itération. Cette mesure est bornée par la largeur de la bande épipolaire à chaque itération et doit être la plus faible possible. Cette erreur résiduelle est un indicateur de qualité d'estimation

de la matrice fondamentale, sa valeur doit diminuer au fur et à mesure de l'affinement de la mise en correspondance.

Ensuite, le nombre de correspondances obtenues est relevé à chaque itération. L'objectif de cette méthode étant d'obtenir la meilleure mise en correspondance possible, ce nombre doit croître à chaque passe de la boucle. Le nombre de correspondances est majoré par le nombre de points détectés dans les images, c'est-à-dire 2000.

Enfin, le pourcentage de correspondances rejetées à chaque itération par la méthode d'estimation de la matrice fondamentale est observé. On s'attend à ce que ce pourcentage diminue au fil des itérations du fait de l'affinement de la mise en correspondance et du resserrement de la bande de recherche.

3.4.2 Résultats

Les figures suivantes illustrent les résultats des expérimentations précédentes. Les distances moyennes entre les points et les droites épipolaires au cours des itérations sont reportées sur la figure 3.16. L'évolution du nombre de correspondances est représenté sur la figure 3.17 et le pourcentage d'*outliers* est relevé sur la figure 3.18. Enfin, le tableau 3.1 indique le nombre d'appariements obtenus au fil des itérations et après l'étape de raffinement pour chaque couple d'images.

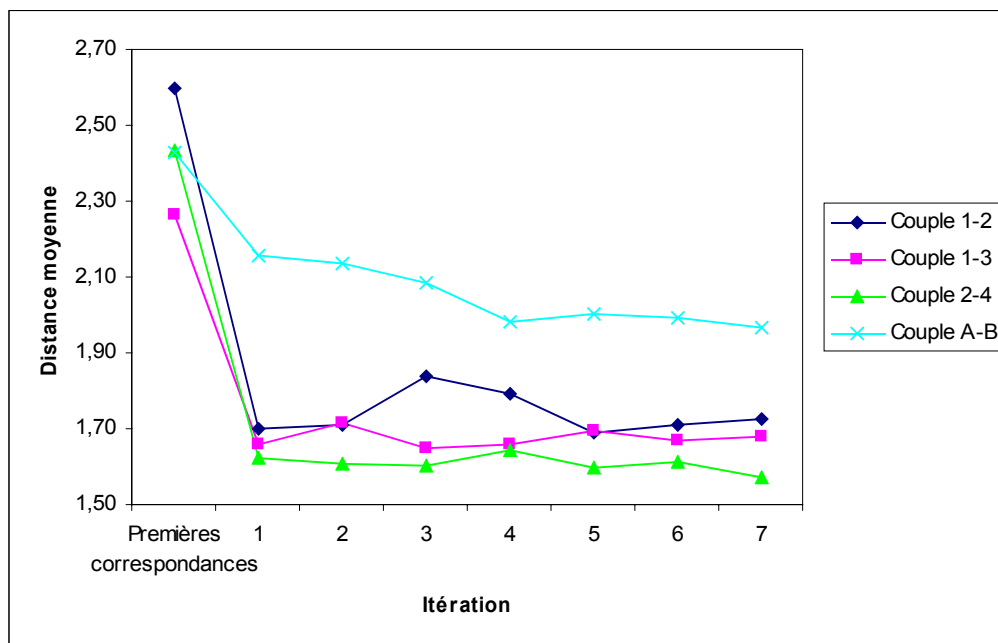


Figure 3.16 : Distance moyenne des points aux droites épipolaires au cours des itérations.

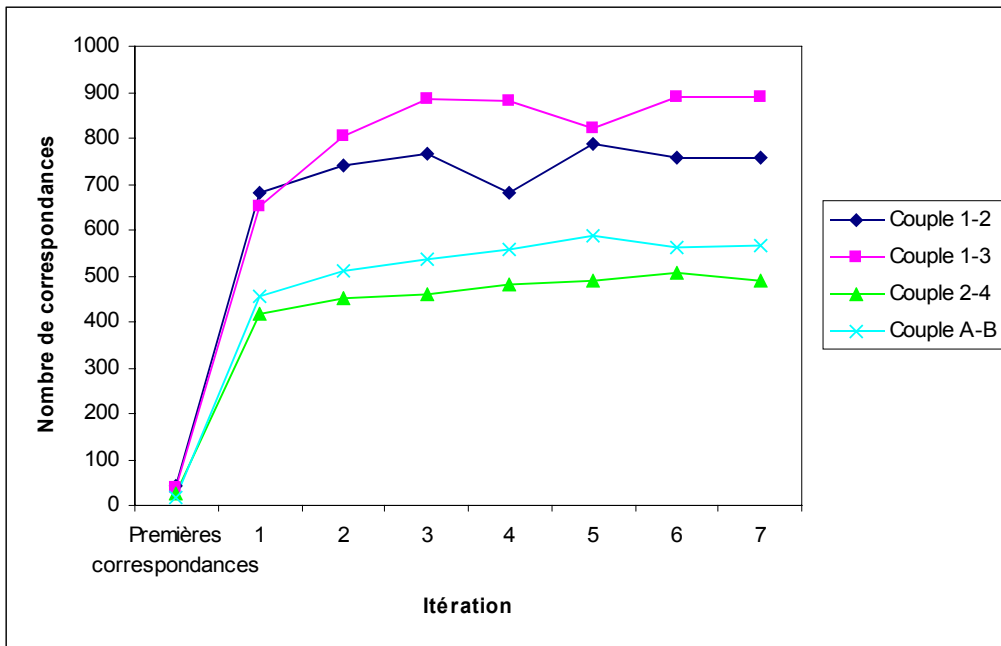


Figure 3.17 : Évolution du nombre de correspondances au cours des itérations.

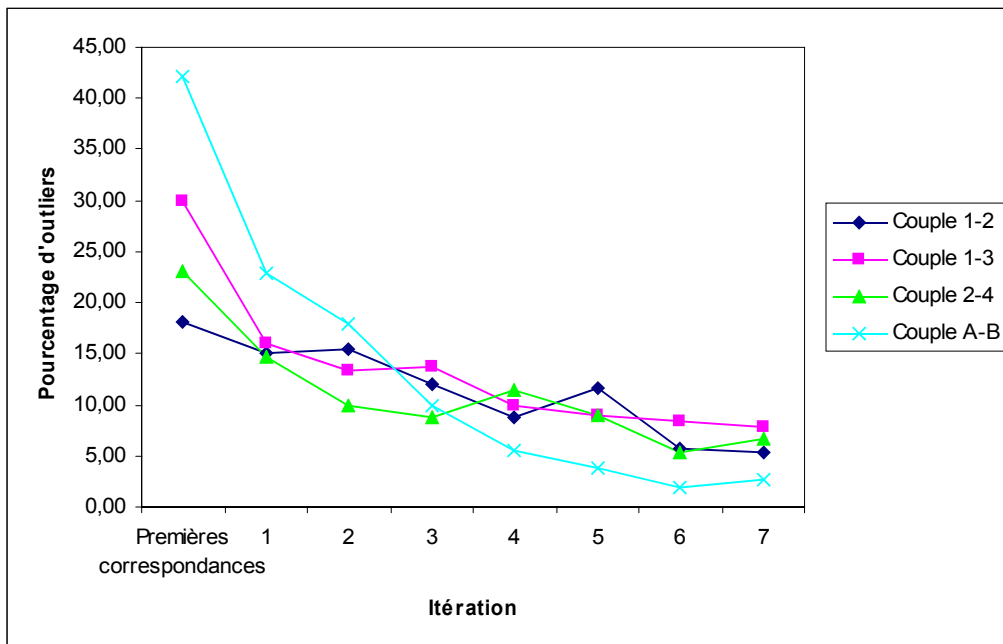


Figure 3.18 : Pourcentage d'outliers détectés par l'estimation robuste de la matrice fondamentale au cours des itérations.

Tableau 3.1 : Nombre d'appariements obtenus par l'algorithme de mise en correspondance.

Nombre d'appariements	Couple 1-2	Couple 1-3	Couple 2-4	Couple A-B
initiaux	44	40	26	19
itération 1	679	653	416	456
itération 2	742	806	451	509
itération 3	768	884	459	535
itération 4	679	880	482	558
itération 5	788	823	489	588
itération 6	756	888	505	560
itération 7	759	889	490	565
finaux	2962	3031	1733	4501
gain	×3,9	×3,4	×3,5	×8,0

Ces résultats confirment l'utilité de cette méthode d'appariement itérative. En effet, la figure 3.17 montre qu'à partir d'un nombre de correspondances très faible (environ une vingtaine), la première itération permet d'obtenir déjà un ensemble de correspondances plus dense d'un facteur entre 15 et 24. Les itérations suivantes permettent d'augmenter légèrement le nombre de correspondances mais surtout d'en améliorer la qualité, comme en témoignent les figures 3.16 et 3.18. Le pourcentage d'*outliers* diminue au fil des itérations et la distance moyenne des points à leurs droites épipolaires diminue ou reste stable selon les images. Enfin, la dernière étape de l'algorithme permet de « booster » le nombre de correspondances en passant de 565 à 4501 (gain de 8) appariements pour le couple A-B, le plus favorable, ou de 889 à 3031 (gain de 3,4) pour le couple 1-3, le moins favorable.

L'algorithme a été implémenté et testé sur un ordinateur équipé d'un processeur Pentium IV cadencé à 3 GHz. Les temps de calcul varient de 2 à 3 minutes, selon les images et le nombre de correspondances obtenues. La résolution des images est de 1024×768 pixels. Ces temps de calcul sont tout à fait compatibles avec l'utilisation prévue dans notre application, d'autant plus que l'algorithme est facilement parallélisable, ce qui permettrait de tirer profit des nouvelles technologies de multi-threading des processeurs récents. Les temps de calcul peuvent ainsi être réduits de moitié.

3.4.3 Comparaison à une méthode de mise en correspondance dense

Nous avons testé sur nos images plusieurs méthodes éprouvées de mise en correspondance dense. Ces méthodes nécessitent que les images soient rectifiées selon la géométrie épipolaire. Nous avons utilisé le logiciel StereoMatcher librement distribué par Microsoft Research [Microsoft, 2001]. Les algorithmes inclus dans ce logiciel ont fait l'objet d'une publication par leurs auteurs Scharstein et Szeliski [Scharstein, 2002]. Les résultats obtenus n'ont pas été à la hauteur de nos espérances (fig. 3.19). Sur la carte de disparité, les zones de transitions brutales entre le blanc et le noir correspondraient à une rupture du modèle 3D (comme une marche d'escalier). Ce genre de cassure n'étant pas présente sur l'escarre

artificielle, il est clair que la carte de disparité obtenue est fautive. Nous pensons que ces erreurs sont liées à la rectification épipolaire. En effet, les algorithmes de mise en correspondance dense ne tolèrent aucune erreur sur la rectification épipolaire. Le correspondant d'un pixel n'est pas détecté s'il se situe en réalité sur la droite épipolaire directement supérieure ou inférieure. Nous avons donc décidé de ne pas utiliser ces techniques. Ces méthodes sont donc réservées à des systèmes calibrés, où la matrice fondamentale est connue avec précision et non pas estimée.

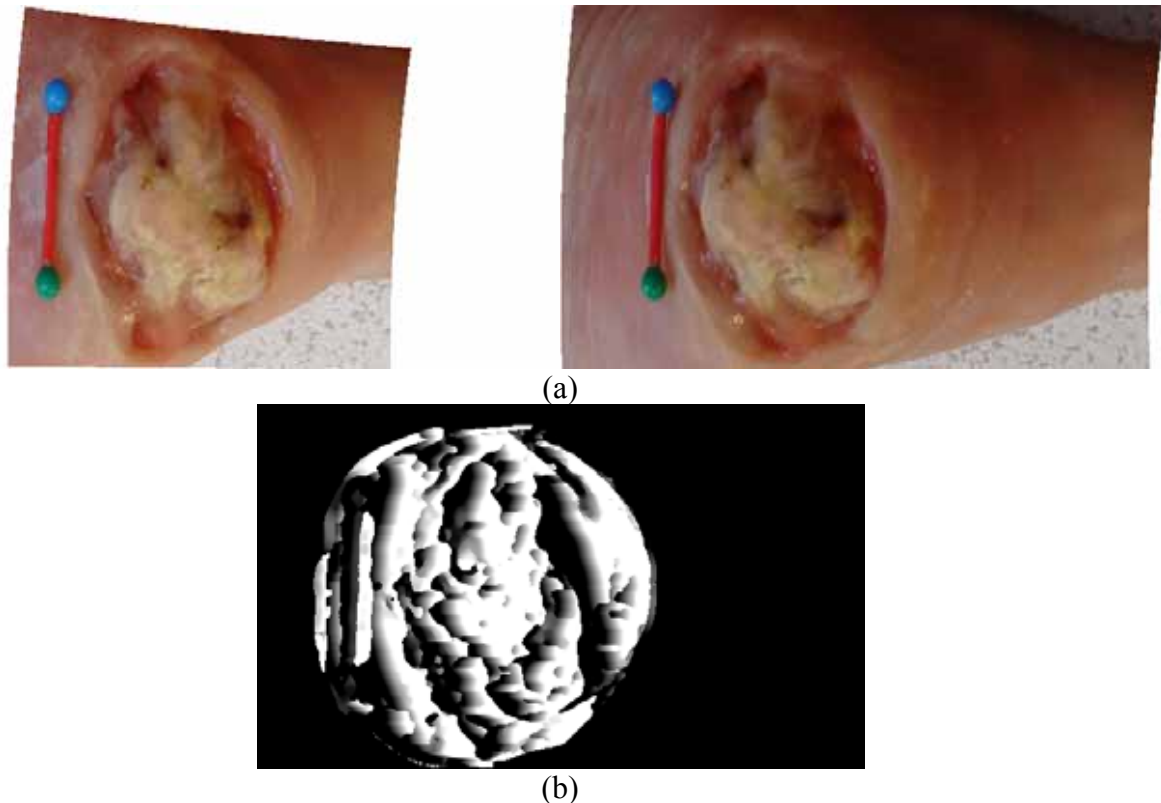


Figure 3.19 : Images rectifiées (a) et carte de disparité associée (b). Les pixels blancs correspondent aux points proches, les pixels noirs aux points éloignés.

La distance entre deux pixels appariés par notre méthode est d'environ entre 5 et 10 pixels, alors qu'une mise en correspondance dense apparie tous les pixels. Cette mise en correspondance peut être qualifiée de semi-dense.

3.5 Conclusion

Après avoir rappelé le contexte applicatif de prise de vue de nos images, nous avons développé dans ce chapitre une méthode d'appariement semi-dense entre deux images prises de points de vue éloignés. L'algorithme mis en oeuvre est itératif et permet d'augmenter la densité d'une première mise en correspondance robuste. Dans notre implémentation, nous avons choisi d'utiliser les descripteurs d'image SIFT, appariés par *winner takes all*, comme initialisation de l'algorithme. La matrice fondamentale estimée à chaque itération permet

d'appliquer la contrainte épipolaire aux appariements après rectification. Cette contrainte est resserrée au fil des itérations, afin de lever l'ambiguïté de certains appariements et d'affiner la géométrie épipolaire. En fin de chaque itération, une contrainte de voisinage est appliquée aux appariements dans le but de détecter les faux appariements non dépistés par la contrainte épipolaire. À la fin du processus itératif, une dernière étape de raffinement de la mise en correspondance supprime les appariements provoquant des points 3D aberrants sur la surface reconstruite par un filtrage médian. À partir de cet ensemble cohérent d'appariements, de nouveaux appariements sont générés par corrélation affine entre des triangles homologues.

Les tests menés sur cet algorithme ont donné des résultats très convaincants. À partir d'une vingtaine d'appariements robustes, alors qu'un algorithme classique aurait donné environ 500 appariements, notre processus itératif permet d'augmenter le nombre de correspondances mais surtout d'améliorer la sélection des plus justes. La dernière étape de densification permet d'obtenir un gain de 4 à 8 sur le nombre de couples en fin de chaîne. Le nombre de correspondances obtenues atteint 1733 à 4501 selon les couples d'images, pour un temps de calcul raisonnable d'environ 2 minutes. Cette méthode d'appariement est utilisée pour la reconstruction 3D des escarres dans les prochains chapitres.

Chapitre 4

Reconstruction euclidienne de la structure 3D

Les méthodes présentées dans les chapitres précédents nous permettent d'obtenir un appariement semi-dense de points d'intérêt entre les deux images, ainsi que la matrice fondamentale régissant la géométrie épipolaire entre les vues. Nous présentons ici la technique utilisée pour déduire la reconstruction 3D euclidienne de la scène à partir des seules informations précédentes.

Sommaire

4.1	Auto-calibrage.....	89
4.1.1	Détermination des paramètres intrinsèques.....	90
4.1.2	Détermination des paramètres extrinsèques.....	92
4.1.3	Validation expérimentale de l'auto-calibrage.....	93
4.1.3.1	Position du point principal.....	95
4.1.3.2	Distance focale.....	97
4.1.3.3	Distorsions optiques.....	99
4.2	Triangulation.....	100
4.3	Optimisation par ajustement de faisceaux.....	101
4.4	Détermination euclidienne du modèle 3D.....	105
4.5	Conclusion.....	105

Comme nous l'avons vu au chapitre 1, la détermination de la structure 3D est possible à partir des points d'intérêt appariés entre les deux images. Mais pour obtenir un modèle métrique, un auto-calibrage des capteurs d'image est nécessaire. Sans cette étape, la structure 3D ne peut être connue qu'à une transformation projective près. Cependant, le cadre métrique n'est pas suffisant pour notre application. En effet, il permet seulement de réaliser des mesures relatives, la structure n'étant déterminée qu'à un facteur d'échelle près. Une distance, au moins, doit alors être connue dans la scène réelle et reportée sur la structure 3D pour fixer ce facteur d'échelle, permettant ainsi la mesure de distances absolues entre les points 3D. On parle alors de reconstruction euclidienne.

Nous détaillons dans ce chapitre tout le processus pour obtenir la structure 3D à partir d'un ensemble de points appariés entre deux images et de la matrice fondamentale estimée à partir de ces points. Le processus est constitué de 3 étapes. Premièrement, les paramètres intrinsèques puis extrinsèques des capteurs d'image sont déterminés par une méthode d'auto-calibrage. Puis, les coordonnées des points 3D sont calculées par triangulation à partir des positions des capteurs d'image. Finalement, la structure est affinée par un ajustement de faisceaux. Ces étapes sont détaillées dans ce chapitre selon cet ordre.

4.1 Auto-calibrage

L'auto-calibrage permet de déterminer les paramètres intrinsèques et extrinsèques des capteurs d'image à partir des seules données images sans utiliser de mire. Comme nous l'avons vu dans le premier chapitre, lorsque seulement deux images sont disponibles, il est nécessaire de faire quelques hypothèses sur les capteurs d'image. Tout d'abord, l'angle ϕ entre les axes horizontal et vertical de l'image est fixé à $\pi/2$, ensuite la position du point principal est supposée au centre de l'image, et enfin le rapport d'aspect des pixels α_u/α_v est considéré égal à 1. Ces suppositions sont valides pour des appareils photo numériques standards, comme nous le validons dans la section 4.1.3. Leur construction assure de disposer de pixels carrés vérifiant les hypothèses sur ϕ et α_u/α_v , et la qualité de leurs optiques place le point principal près du centre des images. Le seul paramètre restant alors à déterminer est la distance focale α exprimée en pixels. La seule méthode alors possible pour l'auto-calibrage repose sur les équations de Kruppa et nous avons utilisé l'algorithme de Sturm pour la détermination de ce paramètre [Sturm, 2005]. Cette méthode impose que la distance focale soit identique entre les deux images, diminuant ainsi les configurations de prises de vue singulières. Les deux capteurs d'image ont alors la même matrice de calibrage, dont l'expression est la suivante :

$$A = \begin{pmatrix} \alpha & 0 & u_0 \\ 0 & \alpha & v_0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4.1)$$

où (u_0, v_0) sont les coordonnées image du point principal, c'est-à-dire le centre des images. Il faut noter que les capteurs d'image partagent la même matrice de calibrage seulement si les images sont de même résolution, c'est-à-dire que les coordonnées du point principal sont identiques dans les deux images. Une fois cette matrice déterminée, on peut calculer la

matrice essentielle E qui est la version calibrée de la matrice fondamentale. Comme nous l'avons vu dans le premier chapitre, l'expression suivante est vérifiée :

$$E \propto A^T F A \propto R[\mathbf{t}]_{\times} \quad (4.2)$$

Ainsi, la matrice essentielle est la composition de la matrice de rotation R et du vecteur \mathbf{t} séparant les deux capteurs d'image. Hartley a mis en place une méthode pour décomposer la matrice E en un produit d'une matrice de rotation et d'une matrice antisymétrique, permettant de retrouver les paramètres extrinsèques [Hartley, 1992].

Les sous-sections suivantes détaillent les méthodes choisies pour l'estimation des paramètres intrinsèques et extrinsèques.

4.1.1 Détermination des paramètres intrinsèques

Avec les hypothèses précédentes, le problème se réduit à déterminer le paramètre α de la matrice de calibrage A des capteurs d'image, définie par l'équation 4.1. Nous détaillons ici l'implémentation de la méthode de Sturm [Sturm, 2005].

Avec pour unique inconnue la distance focale, il est possible de calculer une matrice intermédiaire entre la matrice essentielle et la matrice fondamentale. Cette matrice fondamentale semi-calibrée, notée G , est donnée par l'équation :

$$G \propto \begin{pmatrix} \alpha_0 & 0 & 0 \\ 0 & \alpha_0 & 0 \\ u_0 & v_0 & 1 \end{pmatrix} F \begin{pmatrix} \alpha_0 & 0 & u_0 \\ 0 & \alpha_0 & v_0 \\ 0 & 0 & 1 \end{pmatrix} \propto \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha/\alpha_0 \end{pmatrix} E \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \alpha/\alpha_0 \end{pmatrix} \quad (4.3)$$

où α_0 est un facteur d'échelle utilisé pour la stabilité numérique de l'algorithme.

Dérivées à partir de l'égalité $F^T \mathbf{e}'_F = 0$, les équations de Kruppa s'écrivent [Lourakis, 1999] :

$$F A A^T F^T \propto [\mathbf{e}'_F]_{\times} A A^T [\mathbf{e}'_F]_{\times}^T \quad (4.4)$$

où \mathbf{e}'_F représente les coordonnées de l'épipole de la seconde image dans l'espace non calibré correspondant à la matrice fondamentale. Dans l'espace semi-calibré, cette équation devient :

$$G \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} G^T \propto [\mathbf{e}'_G]_{\times} \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} [\mathbf{e}'_G]_{\times}^T \quad (4.5)$$

avec \mathbf{e}'_G les coordonnées de l'épipole de la seconde image dans l'espace semi-calibré correspondant à G . Cette expression peut être simplifiée en utilisant la décomposition en valeurs singulières (SVD) de la matrice G donnée par :

$$G = U S V^T \quad (4.6)$$

où $S = \text{diag}(a, b, 0)$ est la matrice diagonale composée des valeurs singulières de G avec $a > b > 0$, et U et V sont des matrices orthogonales. Notons \mathbf{u}_i et \mathbf{v}_i respectivement les $i^{\text{èmes}}$ colonnes de U et V . Dans ce cas, on a $\mathbf{e}'_G \propto \mathbf{u}_3$. Ainsi, l'équation 4.5 peut s'écrire :

$$U S V^T \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} V S U^T \propto [\mathbf{u}_3]_{\times} \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} [\mathbf{u}_3]_{\times}^T \quad (4.7)$$

en multipliant l'équation à gauche par U^T et à droite par U , l'orthogonalité de U permet d'écrire :

$$\begin{aligned} S V^T \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} V S &\propto \begin{pmatrix} \mathbf{u}_1^T \\ \mathbf{u}_2^T \\ \mathbf{u}_3^T \end{pmatrix} [\mathbf{u}_3]_{\times} \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} [\mathbf{u}_3]_{\times}^T (\mathbf{u}_1 \ \mathbf{u}_2 \ \mathbf{u}_3) \\ &\propto \begin{pmatrix} \mathbf{u}_2^T \\ -\mathbf{u}_1^T \\ \mathbf{0}^T \end{pmatrix} \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} (\mathbf{u}_2 \ -\mathbf{u}_1 \ \mathbf{0}) \end{aligned} \quad (4.8)$$

or, comme $[\mathbf{u}_3]_{\times}^T \mathbf{u}_3 = \mathbf{0}$, la dernière colonne et la dernière ligne de la matrice calculée par l'équation 4.8 sont des vecteurs nuls. Ainsi, seule la partie 2×2 supérieure gauche est utile et s'écrit :

$$\begin{pmatrix} a \mathbf{v}_1^T \\ b \mathbf{v}_2^T \end{pmatrix} \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} (a \mathbf{v}_1 \ b \mathbf{v}_2) \propto \begin{pmatrix} \mathbf{u}_2^T \\ -\mathbf{u}_1^T \end{pmatrix} \begin{pmatrix} (\alpha/\alpha_0)^2 & 0 & 0 \\ 0 & (\alpha/\alpha_0)^2 & 0 \\ 0 & 0 & 1 \end{pmatrix} (\mathbf{u}_2 \ -\mathbf{u}_1) \quad (4.9)$$

En utilisant le fait que $\mathbf{v}_1, \mathbf{v}_2, \mathbf{u}_1$ et \mathbf{u}_2 sont de norme unitaire, l'équation se simplifie en :

$$\begin{aligned} &\begin{pmatrix} a^2 \left((\alpha/\alpha_0)^2 + V_{31}^2 (1 - (\alpha/\alpha_0)^2) \right) & a b V_{31} V_{32} (1 - (\alpha/\alpha_0)^2) \\ a b V_{31} V_{32} (1 - (\alpha/\alpha_0)^2) & b^2 \left((\alpha/\alpha_0)^2 + V_{32}^2 (1 - (\alpha/\alpha_0)^2) \right) \end{pmatrix} \\ &\propto \begin{pmatrix} (\alpha/\alpha_0)^2 + U_{32}^2 (1 - (\alpha/\alpha_0)^2) & -U_{31} U_{32} (1 - (\alpha/\alpha_0)^2) \\ -U_{31} U_{32} (1 - (\alpha/\alpha_0)^2) & (\alpha/\alpha_0)^2 + U_{31}^2 (1 - (\alpha/\alpha_0)^2) \end{pmatrix} \end{aligned} \quad (4.10)$$

où U_{ij} et V_{ij} sont respectivement les éléments de la $i^{\text{ème}}$ ligne et $j^{\text{ème}}$ colonne des matrices U et V .

Cette égalité à un facteur près permet d'exprimer trois polynômes du second degré en $(\alpha/\alpha_0)^2$ en posant le produit vectoriel des vecteurs contenant les trois éléments différents de

chaque matrice. Les vecteurs étant colinéaires, le produit vectoriel est nul. Deux des équations ainsi formées ont une solution triviale pour $(\alpha/\alpha_0)^2 = 1$ et deviennent linéaires en $(\alpha/\alpha_0)^2$, une fois la solution triviale factorisée. Sturm recommande que ce facteur soit nettement supérieur à la distance focale maximale possible. La focale standard d'un appareil photo est égale à la diagonale de l'image et correspond aux indications de zoom « 1× » d'un appareil photo numérique. Ceux-ci disposant couramment de zoom « 3× » ou « 4× », nous avons choisi d'initialiser la valeur α_0 à 5 fois la distance focale correspondant à un zoom « 4× », c'est-à-dire 20 fois la focale standard. La valeur α_0 est alors donnée par l'équation :

$$\alpha_0 = 20 \sqrt{(2u_0)^2 + (2v_0)^2} \quad (4.11)$$

Sturm recommande d'utiliser la troisième équation, du second degré, pour trouver la solution $(\alpha/\alpha_0)^2$, car elle présente moins de singularités que les équations linéaires. Cette équation est la suivante :

$$\begin{aligned} & \left[a^2(1-U_{31}^2)(1-V_{31}^2) - b^2(1-U_{32}^2)(1-V_{32}^2) \right] (\alpha/\alpha_0)^4 \\ & + \left[a^2(U_{31}^2 + V_{31}^2 - 2U_{31}^2 V_{31}^2) - b^2(U_{32}^2 + V_{32}^2 - 2U_{32}^2 V_{32}^2) \right] (\alpha/\alpha_0)^2 \\ & + a^2 U_{31}^2 V_{31}^2 - b^2 U_{32}^2 V_{32}^2 = 0 \end{aligned} \quad (4.12)$$

Cette équation peut avoir deux solutions, mais en pratique une des solutions est aberrante et très éloignée d'une valeur raisonnable. La solution conservée est la valeur de α la plus proche de la focale standard ($\alpha_0/20$). L'algorithme pour estimer le paramètre α est résumé dans la figure 4.1.

- Calculer la valeur α_0 grâce à l'équation 4.11, ainsi que la matrice G à partir de la matrice fondamentale selon l'équation 4.3.
- Décomposer en valeurs singulières (SVD) la matrice G et récupérer les éléments U_{31} , U_{32} , V_{31} et V_{32} ainsi que les valeurs singulières a et b (équation 4.6).
- Résoudre l'équation 4.12 possédant au plus deux solutions pour α .
- Conserver la solution la plus proche de $\alpha_0/20$.

Figure 4.1 : Algorithme d'auto-calibrage pour la distance focale.

4.1.2 Détermination des paramètres extrinsèques

Une fois les paramètres intrinsèques déterminés, il est possible de calculer la matrice essentielle grâce à l'équation 4.2. Hartley propose une méthode pour décomposer cette matrice en un produit d'une matrice de rotation et d'une matrice antisymétrique [Hartley, 1992]. Nous détaillons maintenant cette méthode.

Tout d'abord, la décomposition en valeurs singulières de la matrice essentielle doit être calculée : $E = U S V^T$ où S est la matrice diagonale composée des valeurs singulières de E , par

ordre décroissant. Puis, les matrices R et $[\mathbf{t}]_{\times}$ sont construites selon les équations suivantes, fournissant chacune deux solutions :

$$R = U B V^T \text{ ou } R = U B^T V^T \quad (4.13)$$

$$[\mathbf{t}]_{\times} = V C V^T \text{ ou } [\mathbf{t}]_{\times} = -V C V^T \quad (4.14)$$

où :

$$B = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix} \text{ et } C = \begin{pmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad (4.15)$$

Cette décomposition n'est pas unique puisqu'il existe quatre possibilités différentes. Comme les matrices R et $[\mathbf{t}]_{\times}$ permettent de construire la matrice de projection du second capteur $P_2 = A(R|R\mathbf{t})$, il existe donc quatre versions de cette matrice selon les choix des matrices R et $[\mathbf{t}]_{\times}$. Les trois versions non valides sont identifiées par la présence de points 3D aberrants, car reconstruits derrière l'un ou l'autre des capteurs. Il faut donc réaliser quatre fois la triangulation de la structure 3D pour choisir P_2 .

4.1.3 Validation expérimentale de l'auto-calibrage

La méthode précédente fait l'hypothèse forte que le point principal est situé au centre des images. Il est établi que la position précise de ce point n'a que peu d'impact sur la qualité de la reconstruction [Urbanek, 2001], [Lourakis, 2000a]. Nous avons voulu vérifier que les défauts de centrage introduits par les appareils photo numériques standards sont négligeables. Pour cela, nous avons réalisé le calibrage complet de 10 appareils photos numériques dont les modèles et principales caractéristiques sont listés dans le tableau . Le calibrage mis en œuvre permet de déterminer les 3 paramètres intrinsèques du modèle sténopé simplifié de l'équation 4.1. Nous avons analysé la position du point principal ainsi que la distance focale en fonction du niveau de zoom des appareils photo. Par ailleurs, le calibrage permet d'estimer 4 paramètres supplémentaires qui modélisent les distorsions optiques des lentilles des objectifs.

Pour réaliser ce calibrage, nous avons utilisé la boîte à outils Matlab développée par Bouguet [Bouguet, 2006]. Le processus de calibrage consiste à prendre une série de 15 photos, sous différents angles, d'une mire constituée d'un damier de dimensions connues (fig. 4.2). Les coins des carreaux sont détectés automatiquement dans les images après avoir initialisé manuellement les quatre coins de la mire. À partir des points détectés, les paramètres du modèle de projection sont déterminés en deux temps. Tout d'abord, les paramètres intrinsèques et extrinsèques sont initialisés en utilisant la technique de Zhang [Zhang, 1999], basée sur la détermination des homographies entre le plan de la mire et les plans image. Ensuite, le modèle complet, incluant les distorsions optiques jusqu'à l'ordre 2, est affiné par la minimisation de l'erreur de reprojection, c'est-à-dire la distance moyenne entre les points détectés dans les images et les points reprojétés de la mire suivant le modèle obtenu.

Tableau 4.1 : Modèles et principales caractéristiques des appareils photos numériques analysés.

Modèle	Diagonale du capteur en pouces	Résolution des images utilisées	Niveaux de zoom optique
Canon EOS 350D	1,05	1728×1152	2×
Canon IXUS 400	0,56	1600×1200	1× - 3×
Canon PowerShot A95	0,56	1600×1200	1× - 3×
Canon PowerShot S2 IS	0,40	1600×1200	1× - 12×
Kyocera Finecam S5R	0,56	1600×1200	1× - 3×
Nikon Coolpix S1	0,40	2048×1536	1× - 3×
Panasonic NV-MX 500 EG	inconnue	640×480	1× - 12×
Sony DSC-H1	0,40	2048×1536	1× - 12×
Sony DSC-P52	0,37	1632×1224	1× - 2×
Sony DSC-P92	0,56	2048×1536	1× - 3×

La modélisation utilisée pour les distorsions optiques est celle de Heikkilä et Silvén [Heikkilä, 1997]. Elle corrige le modèle sténopé, qui n'est qu'une approximation de la projection réelle, en incorporant les déformations radiales et tangentielles des objectifs. L'expression de ces distorsions se fait dans le plan image en coordonnées normalisées, c'est-à-dire sans utiliser les paramètres intrinsèques. Un point M de coordonnées $(X, Y, Z)^T$ exprimées dans le repère lié au capteur d'image, se projette en \mathbf{m} de coordonnées normalisées :

$$\mathbf{m} = \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} X/Z \\ Y/Z \end{pmatrix} \quad (4.16)$$

En définissant le rayon $r = \sqrt{x^2 + y^2}$, les distorsions s'expriment par :

$$\mathbf{m}_d = d_r \mathbf{m} + d_t(\mathbf{m}) \quad (4.17)$$

avec d_r la composante radiale et $d_t(\mathbf{m})$ la composante tangentielle :

$$d_r = 1 + k_1 r^2 + k_2 r^4 + k_3 r^6 + \dots \quad (4.18)$$

$$d_t(\mathbf{m}) = \begin{pmatrix} 2 p_1 x y + p_2 (r^2 + 2 x^2) \\ p_1 (r^2 + 2 y^2) + 2 p_2 x y \end{pmatrix} \quad (4.19)$$

Ainsi, le calibrage à l'ordre 2 que nous avons réalisé détermine les valeurs k_1 , k_2 , p_1 et p_2 .

Les résultats complets de ce calibrage sont disponibles en annexe A.

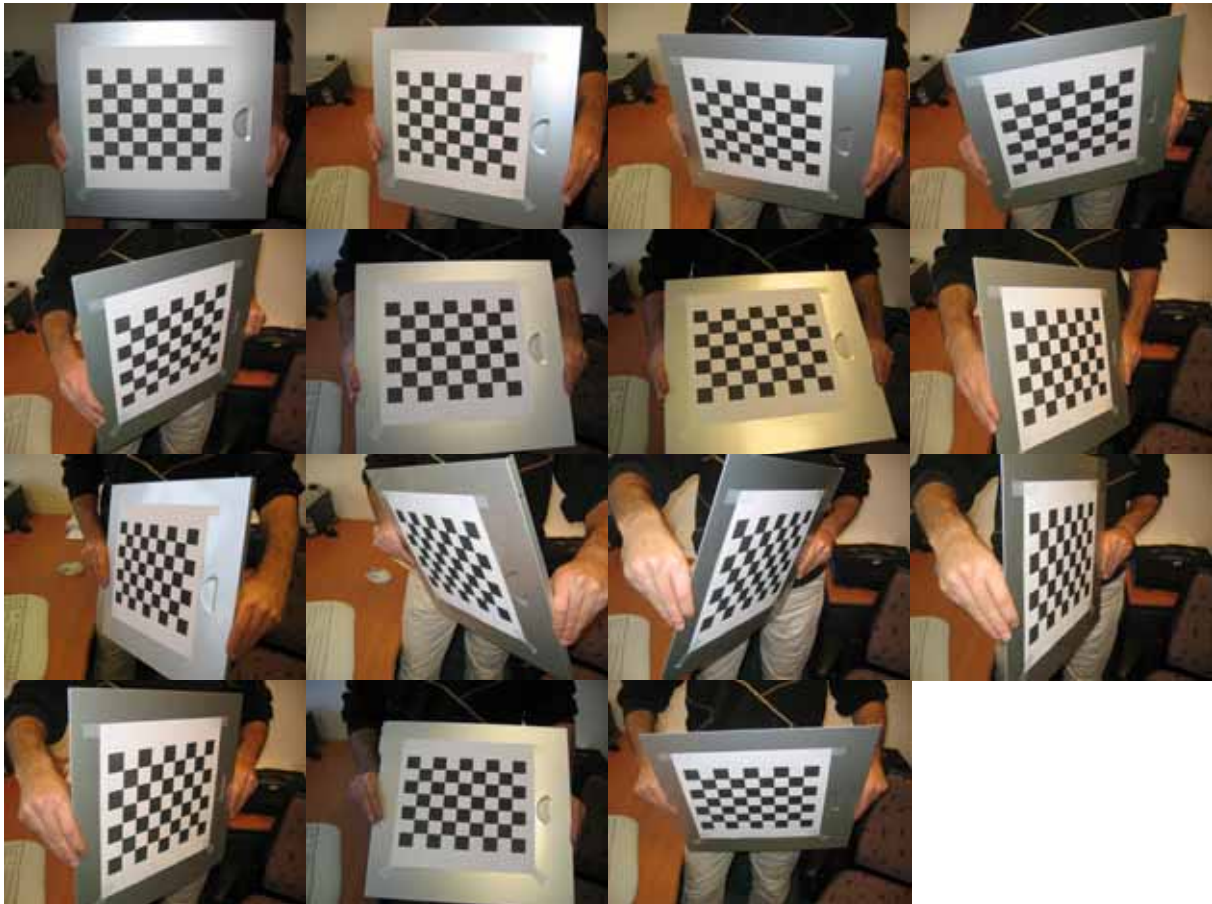


Figure 4.2 : Représentation des 15 images utilisées pour le calibrage du Canon IXUS 400 au zoom 1×.

4.1.3.1 Position du point principal

À titre d'exemple, la figure 4.3 montre la dérive du point principal pour l'appareil Canon PowerShot A95 pour des variations de zoom. Ces résultats sont typiques des appareils photos dont le zoom varie de 1× à 3×. Lorsque le niveau de zoom augmente, le point principal a tendance à s'éloigner du centre de l'image, comme en attestent les résultats du Sony DSC-H1 (fig. 4.4). La partie grisée représente la zone de l'image. Afin de s'assurer que l'erreur sur la position du point principal reste faible, il est donc important que les appareils photos ne soient pas utilisés aux niveaux de zoom supérieurs à 3×.

Les figures ne représentent pas les incertitudes sur les estimations de la position du point principal car elles sont très faibles. Les écart-types sont de l'ordre de 2 pixels pour les niveaux de zoom inférieurs à 3×. Nous avons cependant noté une augmentation de ces incertitudes jusqu'à des valeurs de 100 pixels pour les plus forts niveaux de zoom. Nous expliquons cette baisse de la stabilité de l'estimation par la limite du modèle sténopé. En effet, plus la focale augmente, plus la convergence des faisceaux lumineux devient faible et plus le modèle sténopé ressemble à un modèle de projection parallèle.

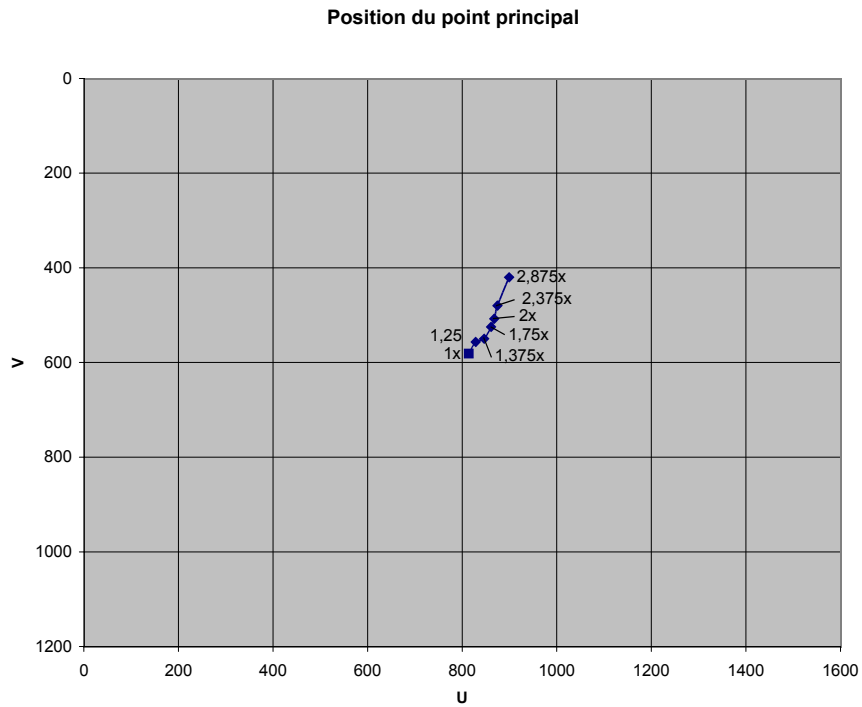


Figure 4.3 : Position du point principal en fonction du niveau de zoom pour l'appareil Canon PowerShot A95.

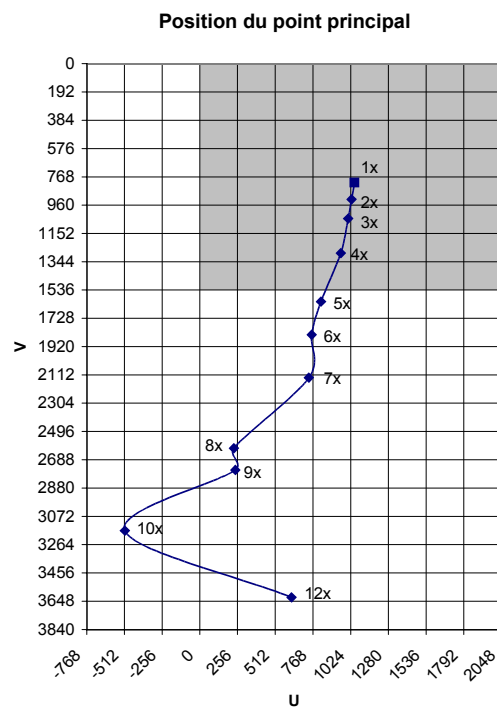


Figure 4.4 : Position du point principal en fonction du niveau de zoom pour l'appareil Sony DSC-H1.

4.1.3.2 Distance focale

Les constructeurs d'appareil photo numérique ont mis en place un standard de format de fichier, permettant d'enregistrer avec l'image des méta-informations. Stockées au format EXIF [JEITA, 2002], elles incluent notamment la distance focale exprimée en millimètres. En connaissant la taille du capteur de l'appareil photo, il est possible de convertir cette valeur en pixels par la formule suivante :

$$\hat{\alpha} = \frac{f}{d} \sqrt{w^2 + h^2} \quad (4.20)$$

où f est la distance focale en millimètres enregistrée dans le fichier image, d est la diagonale du capteur de l'appareil photo, w et h sont respectivement la largeur et la hauteur de l'image en pixels. La valeur d est généralement fournie par le constructeur de l'appareil photo et exprimée en pouces. La deuxième colonne du tableau donne cette valeur pour chaque appareil testé. Seule la valeur pour le Panasonic NV-MX 500 EG n'a pas pu être trouvée, mais contrairement aux autres appareils, celui-ci est en réalité un caméscope numérique.

Nous avons voulu vérifier la concordance de la valeur $\hat{\alpha}$ estimée avec la valeur α déterminée par le calibrage. Pour cela nous avons calculé les facteurs \hat{c} et c de conversion entre la focale exprimée en millimètres et la focale exprimée en pixels. La valeur de c est obtenue par régression linéaire sur plusieurs calibrages à différentes valeurs de focales, alors que le facteur \hat{c} est obtenu à partir de l'équation 4.20 :

$$\hat{c} = \frac{\sqrt{w^2 + h^2}}{d} \quad (4.21)$$

La figure 4.5 présente des résultats typiques obtenus pour l'appareil photo Canon IXUS 400. Des barres d'erreurs de $\pm 0,5$ mm sont indiquées sur les valeurs de focales obtenues par les fichiers images car cette valeur est stockée sous forme entière. Tous les résultats sont résumés dans le tableau 4.2. Une différence relative d'environ 30% entre les valeurs c et \hat{c} est toujours observée. Cette différence s'explique par le fait que la diagonale du capteur fournie par les constructeurs, inclut un certain nombre de pixels non utilisés pour former l'image finale. De ce fait, l'estimation de la focale en pixels par ce moyen est très grossière et ne doit pas être utilisée. L'auto-calibrage est donc le seul moyen pour déterminer la distance focale.

Par ailleurs, certains appareils photos sont équipés d'un mode macro, permettant de faire la mise au point à des distances très faibles de l'objectif. Afin de nous assurer que ce mode de mise au point ne perturberait pas l'auto-calibrage, nous avons calibré les appareils munis de ce mode pour le niveau de zoom le plus faible en faisant varier la distance de mise au point sur la plage des distances permises par ce mode. Les résultats nous ont montré que la distance focale était stable quelle que soit la distance de mise au point. Le mode macro peut donc être utilisé sans crainte de perturber l'auto-calibrage, permettant alors de réaliser la reconstruction 3D de petits objets.

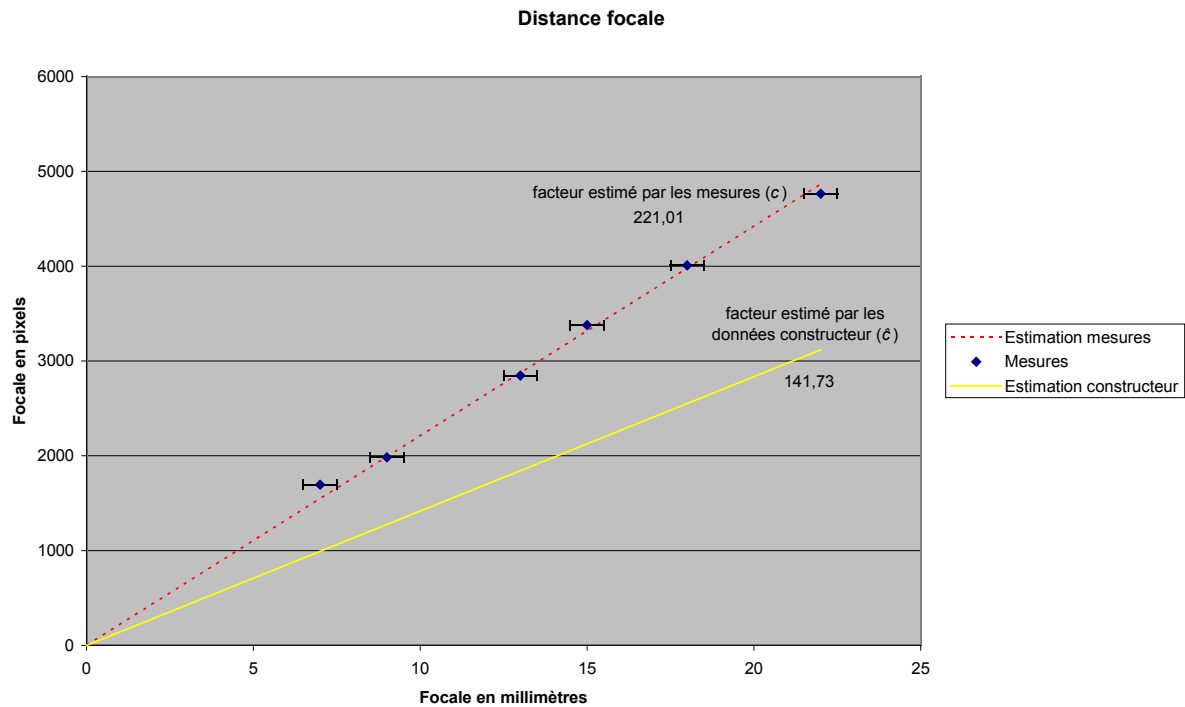


Figure 4.5 : Etude du facteur de conversion entre la focale exprimée en millimètres et en pixels. Résultats obtenus pour le Canon IXUS 400.

Tableau 4.2 : Valeur des facteurs de conversion pour tous les appareils photos testés.

Modèle	Valeur \hat{c}	Valeur c	Différence relative ($\hat{c} - c$)/ \hat{c}
Canon EOS 350D	121,89	77,87	36,11 %
Canon IXUS 400	221,01	141,73	35,87 %
Canon PowerShot A95	215,03	141,73	34,09 %
Canon PowerShot S2 IS	267,85	196,85	26,51 %
Kyocera Finecam S5R	223,27	141,73	36,52 %
Nikon Coolpix S1	352,24	251,97	28,47 %
Panasonic NV-MX 500 EG	776,42	non disponible	
Sony DSC-H1	360,24	251,97	30,06 %
Sony DSC-P52	305,43	216,85	29,00 %
Sony DSC-P92	267,50	181,42	32,18 %

4.1.3.3 Distorsions optiques

Le dernier aspect du calibrage est l'estimation des distorsions optiques. Les distorsions étant des défauts de convergence des lentilles, elles sont maximales lorsque l'angle de champ est le plus grand, c'est-à-dire pour les faibles valeurs de focale. La figure 4.6 représente des résultats typiques correspondant au modèle de distorsion du Kyocera Finecam S5R. Le champ de vecteurs représente les déplacements des pixels causés par les distorsions optiques. Les courbes sont les lignes de niveaux des normes des vecteurs. La distorsion maximale du champ de vision utile de l'image est d'environ 9 pixels dans cet exemple, si l'on considère que le champ de vision utile est constitué du disque maximal inscrit et centré dans l'image. Afin de comparer les résultats des différents appareils photos avec les expérimentations du chapitre 2 sur la matrice fondamentale, la valeur de distorsion maximale est rapportée à des images de taille 640×480 dans le tableau 4.1.

Ces distorsions ne peuvent pas être corrigées dans le cadre de notre application pour privilégier la souplesse d'utilisation. En effet, nous refusons d'obliger les utilisateurs finaux non spécialistes de la vision par ordinateur à opérer un calibrage des appareils, parfois délicat. Les distorsions restant inférieures à 5 pixels dans des images de taille 640×480 , et en s'appuyant sur les expérimentations réalisées pour l'étude de la matrice fondamentale au chapitre 2, on peut penser que l'erreur induite sur le modèle 3D reste inférieure à l'incertitude maximale de 5 % requise pour notre application.

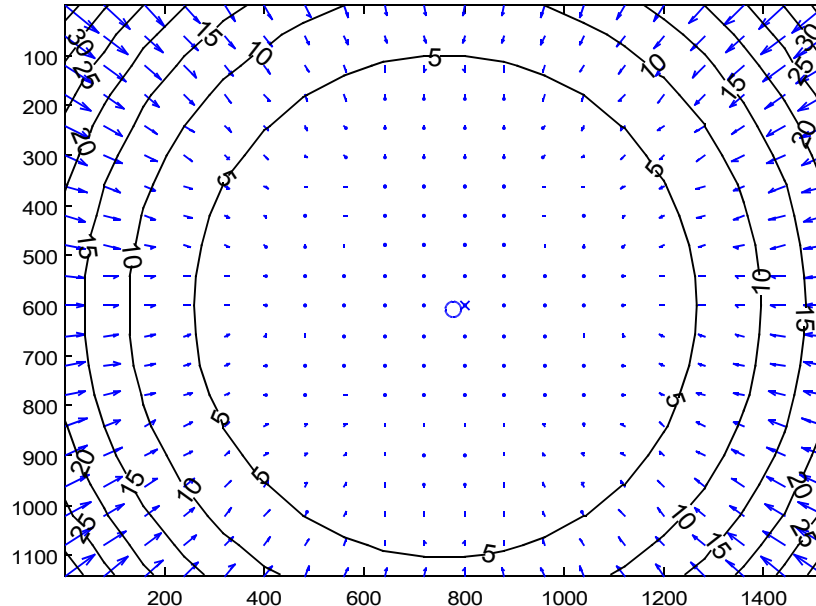


Figure 4.6 : Modèle de distorsion typique des appareils photos numériques standards. Mesures réalisées sur un Kyocera Finecam S5R.

Tableau 4.3 : Distorsions maximales des différents appareils photos analysés, rapportées à des images de taille 640×480 .

Modèle	Distorsion maximale en pixels rapportée à des images de taille 640×480
Canon EOS 350D	1,9
Canon IXUS 400	4,0
Canon PowerShot A95	2,8
Canon PowerShot S2 IS	4,0
Kyocera Finecam S5R	3,6
Nikon Coolpix S1	4,7
Panasonic NV-MX 500 EG	4,0
Sony DSC-H1	4,7
Sony DSC-P52	3,5
Sony DSC-P92	3,1

4.2 Triangulation

Une fois les matrices de projection déterminées, le modèle 3D peut être obtenu par triangulation. La matrice de projection du premier capteur d'image est $P_1 = A \begin{pmatrix} Id_{3 \times 3} \\ \mathbf{0} \end{pmatrix}$ et celle du second est $P_2 = A(R|R\mathbf{t})$. Notons $\mathbf{m}_1 \propto (u_1 \ v_1 \ 1)^T$ et $\mathbf{m}_2 \propto (u_2 \ v_2 \ 1)^T$ les projections images du point $M \propto (x \ y \ z \ 1)^T$ recherché. Les équations suivantes sont alors vérifiées :

$$s_1 \begin{pmatrix} u_1 \\ v_1 \\ 1 \end{pmatrix} = P_1 \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad \text{et} \quad s_2 \begin{pmatrix} u_2 \\ v_2 \\ 1 \end{pmatrix} = P_2 \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (4.22)$$

où s_1 et s_2 sont deux facteurs d'échelle non nuls. En posant P_{11} , P_{12} et P_{13} égaux aux trois lignes de P_1 et P_{21} , P_{22} et P_{23} les trois lignes de P_2 , on obtient $s_1 = P_{13}M$ et $s_2 = P_{23}M$. Il vient alors le système suivant :

$$\begin{aligned} u_1 P_{13} M &= P_{11} M \\ v_1 P_{13} M &= P_{12} M \\ u_2 P_{23} M &= P_{21} M \\ v_2 P_{23} M &= P_{22} M \end{aligned} \quad (4.23)$$

ce qui revient à écrire :

$$\begin{pmatrix} u_1 P_{13} - P_{11} \\ v_1 P_{13} - P_{12} \\ u_2 P_{23} - P_{21} \\ v_2 P_{23} - P_{22} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad (4.24)$$

La matrice de la partie gauche de l'équation 4.24 est de taille 4×4 . En notant $B_{4 \times 3}$ la sous-matrice composée des 3 premières colonnes et $C_{4 \times 1}$ la quatrième colonne, l'équation 4.24 devient :

$$B_{4 \times 3} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = -C_{4 \times 1} \quad (4.25)$$

Ainsi, il suffit de résoudre ce système sur-contraint de 4 équations à 3 inconnues par minimisation aux moindres carrés afin de trouver la position du point M dans l'espace. Il a été montré par Hartley et Sturm que ce type de triangulation est la plus adéquate lorsque la reconstruction 3D est métrique [Hartley, 1997a].

4.3 Optimisation par ajustement de faisceaux

Comme nous l'avons vu au chapitre 1, l'ajustement de faisceaux permet d'affiner tous les paramètres ainsi que la structure 3D, en minimisant l'erreur de reprojection.

Les variables ajustées sont donc les matrices de projection P_i et les points 3D reconstruits M_j . Rappelons que les matrices P_i n'ont que 11 degrés de liberté (les 5 paramètres intrinsèques et les 6 paramètres extrinsèques). Avec deux images, le nombre de paramètres pour les matrices P_i serait donc de $2 \times (5+6) = 22$. Nous montrons ci-après que ce nombre peut être réduit à 6.

Au niveau des paramètres extrinsèques, les positions des capteurs d'image sont définies relativement au premier capteur, placé à l'origine. Ce postulat permet de supprimer donc 6 des 12 paramètres extrinsèques. Ensuite, la reconstruction est déterminée à un facteur d'échelle près, et la distance entre les deux prises de vue est fixée arbitrairement à 1. Le nombre de paramètres extrinsèques passe alors de 6 à 5. L'orientation entre les deux vues peut être définie par un vecteur de dimension 3, colinéaire à l'axe de rotation et dont la norme donne l'angle de rotation. La translation entre les deux capteurs est décrite par 2 paramètres qui sont les deux angles des coordonnées sphériques du vecteur déplacement entre les deux caméras.

Concernant les paramètres intrinsèques, nous avons choisi de ne paramétrer les capteurs d'image que par leur distance focale, le point principal étant fixé au centre des images et les pixels étant carrés. Ainsi le seul paramètre intrinsèque utilisé pour paramétrer une prise de vue est la distance focale. De plus, la distance focale est imposée constante entre les deux prises de vue. Il ne reste donc plus qu'un seul paramètre intrinsèque pour les deux prises de vue.

Nous avons donc 5 paramètres extrinsèques plus 1 paramètre intrinsèque pour caractériser les deux capteurs d'image. Le problème d'ajustement de faisceaux peut alors s'écrire :

$$\min_{\mathbf{a}, \mathbf{b}_j} \sum_{i=1}^2 \sum_{j=1}^n d(Q_i(\mathbf{a}, \mathbf{b}_j), \mathbf{m}_{ij}) \quad (4.26)$$

où n est le nombre de points 3D, \mathbf{b}_j est le vecteur des coordonnées non homogènes du $j^{\text{ème}}$ point 3D, \mathbf{a} est le vecteur des paramètres des capteurs d'image et $Q_i(\mathbf{a}, \mathbf{b}_j)$ est la fonction de projection du point 3D défini par \mathbf{b}_j sur la $i^{\text{ème}}$ image. Le vecteur \mathbf{a} est défini de la façon suivante :

$$\mathbf{a} = (\alpha, r_1, r_2, r_3, \theta_t, \phi_t)^T \quad (4.27)$$

où α est la distance focale exprimée en pixels, $(r_1, r_2, r_3)^T$ est le vecteur des paramètres de la matrice R de rotation entre les capteurs d'image, et $(\theta_t, \phi_t)^T$ est le vecteur de définition de la direction du vecteur unitaire \mathbf{t} de translation entre les deux capteurs. Le passage entre le vecteur $(r_1, r_2, r_3)^T$ et la matrice R , et entre le vecteur $(\theta_t, \phi_t)^T$ et le vecteur \mathbf{t} est détaillé en annexe B.

À ce stade de la chaîne de traitements, la reconstruction 3D est métrique. Nous avons appliqué nos algorithmes à des objets autres que des escarres afin de nous détacher de notre application. L'intérêt de cette expérimentation est aussi de vérifier qualitativement que la géométrie des objets est bien restituée, ce qui est difficile sur des escarres dont les formes ne sont pas usuelles. Les figures 4.7 et 4.8 montrent deux exemples de reconstructions 3D réalisées par nos algorithmes sur des objets différents.



Figure 4.7 : Reconstruction d'une carte géographique en relief. En haut, paire d'images initiales, en bas, trois vues en perspective du modèle 3D obtenu par notre méthode.



Figure 4.8 : Reconstruction d'une statue. En haut, paire d'images initiales, en bas, deux vues en perspective du modèle 3D obtenu par notre méthode.

4.4 Détermination euclidienne du modèle 3D

À ce stade, la structure 3D n'est connue que dans un cadre métrique. La distance entre les deux capteurs d'image est fixée à 1 par l'auto-calibrage. Afin de passer dans un cadre euclidien, il est nécessaire de réaliser une mise à l'échelle du modèle 3D. Pour ce faire, nous plaçons un repère d'échelle dans le champ des images. Ce repère dispose de deux extrémités clairement et précisément identifiables dans les images, et dont l'écartement est connu. Ces deux points sont mis en correspondance et reconstruits par le même processus décrit précédemment. Il suffit alors de mettre à l'échelle le modèle 3D de sorte que la distance entre ces deux points respecte l'écartement réel.

Dans la pratique, nous avons utilisé un repère constitué de deux têtes d'épingles colorées en vert et en bleu, séparées par une tige rouge (fig. 4.9a). Ces couleurs sont facilement identifiables dans nos images, les escarres étant dépourvues de parties vertes ou bleues. Au cours du développement du projet, la présence d'une mire de couleur (fig. 4.9b) est devenue indispensable dans les images, afin de réaliser une correction colorimétrique des images avant leur classification tissulaire. Cette thèse se limitant à l'évaluation géométrique de la plaie, l'exploitation de cette mire couleur ne sera pas développée dans ce mémoire. Cependant, afin de simplifier l'utilisation du système complet, nous avons combiné le repère d'échelle avec la mire couleur. Des points verts et bleus sont disposés aux quatre coins de la mire, rendant ainsi automatique la détection de la mire dans les images. Les dimensions de la mire étant fixes et connues, il est possible d'utiliser plus de points, rendant ainsi plus robuste la détermination du facteur d'échelle.

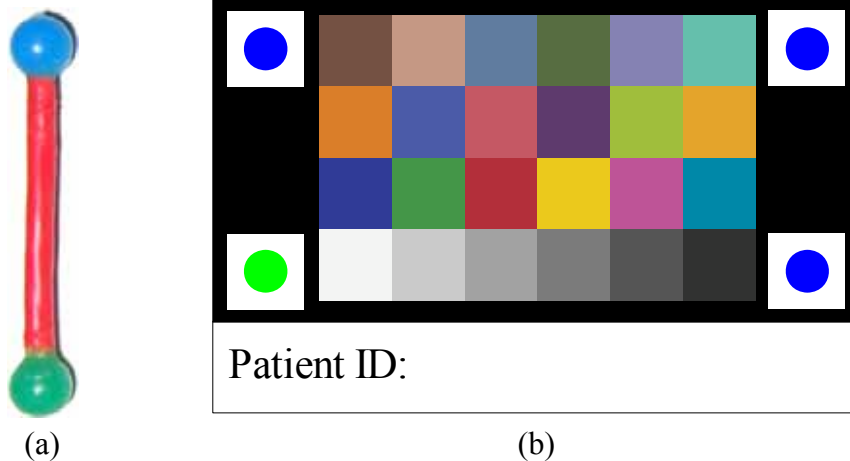


Figure 4.9 : Repère coloré (a) ou mire couleur (b) inséré dans le champ de vue des images.

4.5 Conclusion

Nous avons présenté dans ce chapitre les méthodes permettant d'obtenir un modèle 3D à partir de points mis en correspondance entre deux vues non calibrées. Tout d'abord, un auto-calibrage permet de retrouver les paramètres des capteurs d'image. Cette étape nécessite

néanmoins que certains paramètres soient connus. C'est le cas de l'angle entre les axes des coordonnées images supposé égal à $\pi/2$, du rapport d'échelle des pixels égal à 1 et du point principal positionné au centre des images. La méthode d'auto-calibrage permet alors de retrouver la distance focale ainsi que les positions relatives des capteurs.

Ces hypothèses simplificatrices ont été validées par des expérimentations sur une dizaines d'appareils photos numériques. Les expérimentations ont permis d'établir trois constats importants. Tout d'abord, le niveau de zoom doit être inférieur à $3\times$ pour garantir que le point principal soit proche du centre des images. Ensuite, l'information de focale enregistrée dans les fichiers images par certains appareils photos est peu fiable car elle n'est pas exprimée par rapport à la taille de l'image, mais par rapport à la taille globale du capteur physique. Enfin, les distorsions optiques causées par les objectifs des appareils photos sont suffisamment faibles pour espérer obtenir une erreur sur le modèle 3D inférieure à 5%.

Une fois les paramètres des capteurs d'image déterminés, la scène est reconstruite par triangulation, puis optimisée par ajustement de faisceaux. Ces deux étapes fournissent une structure 3D dans un cadre métrique. La structure 3D euclidienne peut alors être déterminée par une simple mise à l'échelle, à partir d'une distance connue dans la scène.

Ce chapitre conclut la première partie de ce mémoire dédiée à la description de la chaîne de reconstruction 3D. La seconde partie va maintenant s'attacher à évaluer la performance de cette chaîne, c'est-à-dire sa précision métrologique. Après les nombreuses considérations théoriques précédentes, nous allons donc développer les aspects expérimentaux de ces travaux.

SECONDE PARTIE

-

ÉVALUATION DES PERFORMANCES

Chapitre 5

Évaluation de la qualité de la reconstruction

Le modèle 3D d'un objet peut être généré en appliquant les méthodes précédemment décrites à partir de seulement deux vues. Le but de ce chapitre est d'évaluer la qualité de la reconstruction obtenue dans un cas réel.

Sommaire

5.1 Les systèmes de numérisation 3D.....	112
5.1.1 Le système d'Eotech.....	112
5.1.2 Le système de Kréon.....	113
5.1.3 Le système de Vidéometric.....	114
5.2 Évaluation de la qualité d'un modèle 3D.....	114
5.2.1 Méthodes bidimensionnelles.....	114
5.2.1.1 Erreur résiduelle sur le calcul de la matrice fondamentale.....	114
5.2.1.2 Erreur de reprojection.....	115
5.2.1.3 Relation entre les erreurs bidimensionnelles et le modèle 3D.....	115
5.2.2 Méthodes tridimensionnelles.....	116
5.2.2.1 Comparaison de formes.....	116
5.2.2.2 Calcul de distance.....	117
5.3 La méthode de comparaison.....	117
5.4 Résultats.....	119
5.4.1 Choix de la vérité terrain.....	119
5.4.2 Évaluation de notre chaîne de reconstruction.....	120
5.5 Conclusion.....	125

La chaîne de reconstruction 3D décrite dans la première partie de ce mémoire a été mise en place. Les méthodes et algorithmes la constituant ont été évalués de façon modulaire. Pour évaluer la chaîne dans sa globalité, il faut mettre en place un protocole expérimental permettant un contrôle de qualité quantitatif sur la reconstruction finale obtenue, dans des conditions réelles ou aussi proches que possible.

A cette fin, nous avons demandé à Dominique Colladant, expert en effets spéciaux pour le cinéma, de réaliser plusieurs escarres artificielles particulièrement réalistes (fig. 5.1). Les reconstructions 3D de ces pièces par notre méthode ont été comparées aux modèles 3D issus de différents systèmes industriels de numérisation de grande précision. Ce travail de comparaison a été fait en collaboration avec Emilie Koenig [Koenig, 2006] au cours de son stage de master recherche. Ce chapitre livre les détails de cette expérimentation selon le plan suivant. Dans un premier temps, les différents systèmes de reconstruction 3D sont présentés. Puis, la méthode de comparaison des reconstructions est expliquée avant d'analyser les différences entre les reconstructions.



Figure 5.1 : Fausses escarres utilisées pour les numérisations.

5.1 Les systèmes de numérisation 3D

Nous avons eu accès à trois systèmes différents de numérisation 3D. Ce sont les systèmes des sociétés Eotech, Kréon et Vidéometric. Ces systèmes font appel à trois techniques différentes s'appuyant sur un calibrage préalable.

5.1.1 Le système d'Eotech

La société Eotech (<http://www.eotech.fr/>) utilise une méthode de triangulation associée à un éclairage structuré pour reconstruire un objet en 3 dimensions. L'éclairage utilise la Technique de Projection Miniature (MPT) selon un code Gray avec déphasage (fig. 5.2). Cette technique repose sur la conception d'une lentille spéciale, sur laquelle les franges à projeter ont été gravées avec une très grande précision. La première partie de la lentille sert à la projection d'un code Gray. Le nombre de franges est successivement doublé à chaque projection. La somme de ces projections élémentaires compose une série de raies dont les intensités obéissent à un code Gray. La seconde partie de la lentille correspond à un déphasage réalisé par la projection de sinusoides sur l'objet en décalant soit les raies soit l'objet. La configuration standard utilise neuf plans pour le code Gray et quatre pour le déphasage. Cette technique permet de déterminer une information de contraste local et une information de phase sur chaque pixel. Les images permettent ainsi de calculer la position 3D de chaque point de l'image de façon automatique par triangulation, en tenant compte des paramètres de calibrage du capteur utilisé. Plusieurs prises de vue peuvent être recollées pour couvrir un objet de grandes dimensions. La numérisation d'une escarpe artificielle est réalisée en 5 minutes environ et génère un nuage de points dont le nombre est compris entre 12 000 et 140 000 selon la précision demandée. La position des points est certifiée à environ 20 μm .

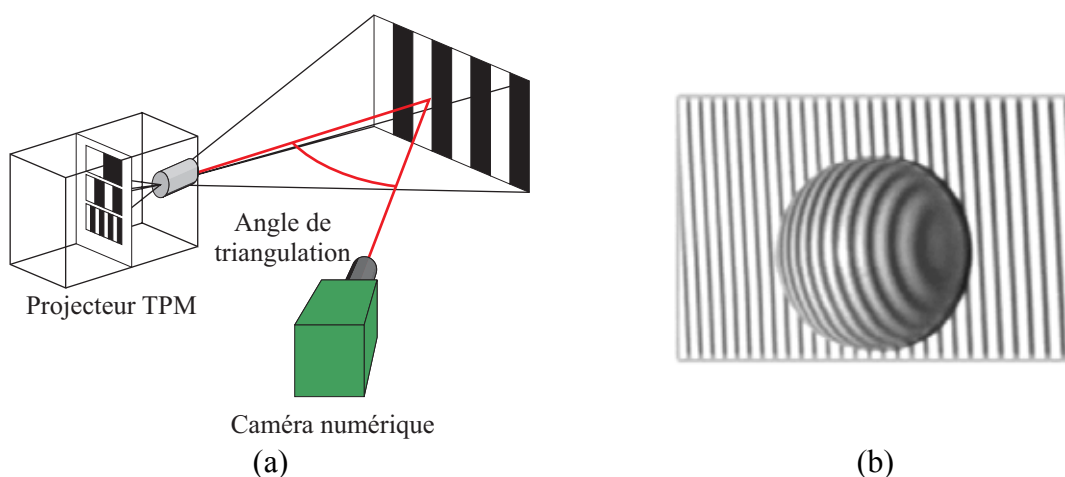


Figure 5.2 : Système de vision 3D d'Eotech (a). Exemple de projection de franges (b).

5.1.2 Le système de Kréon

La société Kréon (<http://www.kreon3d.com/>) numérise des pièces grâce à un système de triangulation laser utilisant une caméra et un dispositif de déplacement contrôlé (bras articulé ou axes de commandes numériques). Le principe est le suivant (fig. 5.3) :

- Projection d'un plan laser sur l'objet à numériser
- Numérisation en temps réel de la scène filmée par une caméra
- Extraction du profil laser dans les images
- Transformation des profils en un ensemble de points 3D
- Déplacement du système autour de la pièce pour obtenir un modèle 3D complet de l'objet.

La numérisation permet d'extraire de la vidéo un ensemble de points 2D dans un système de coordonnées (ligne, colonne) en précision subpixellique. Puis le calibrage détermine la relation entre le système de coordonnées caméra et un système de coordonnées cartésien. Après triangulation, les points du profil laser sont localisés dans l'espace. Le résultat final est un ensemble de courbes de points 3D. Pour numériser les pieds artificiels, un balayage manuel profil par profil utilisant le processus ci-dessus a été utilisé. Les profils sont recalés automatiquement dans le référentiel fixe du bras supportant la tête de mesure. La densité de numérisation est d'environ 500 points par profil scanné. La numérisation d'une escarre artificielle prend environ 25 minutes pour obtenir un nuage de 70 000 à 80 000 points. La précision de ces points est certifiée par la société Kréon à 40 μm .

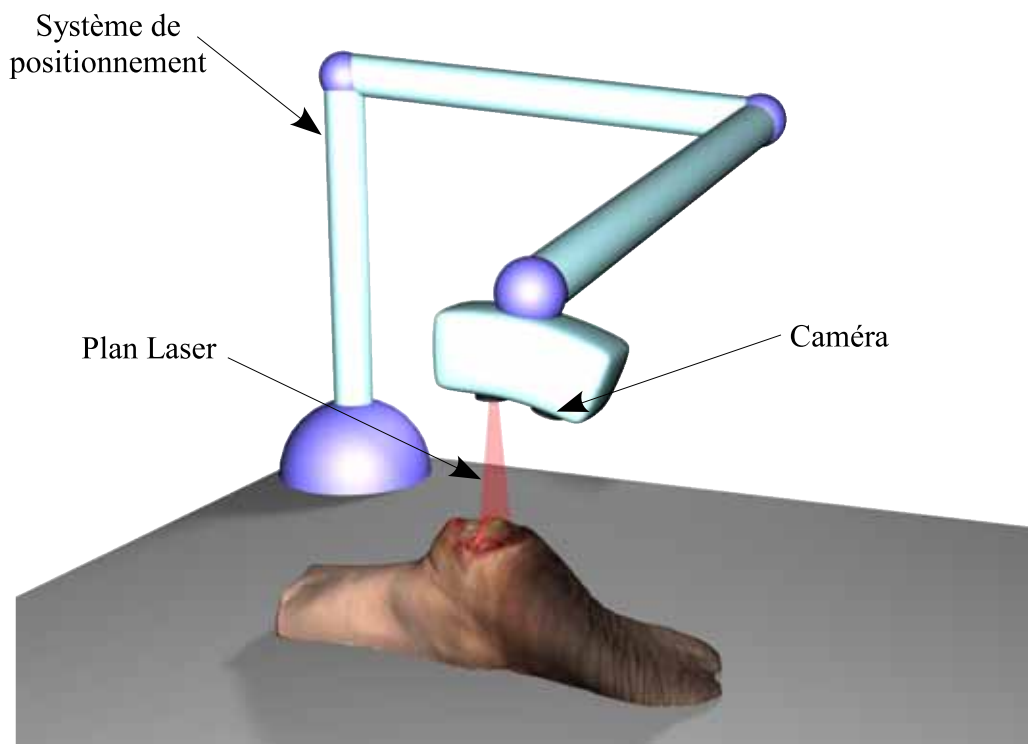


Figure 5.3 : Schéma de principe du système de Kréon.

5.1.3 Le système de Vidéometric

Le système de Vidéometric (<http://www.videometric.com/>) utilise la stéréovision calibrée. Deux caméras rigidement liées dans un tube en carbone prennent des images de l'objet. Pour améliorer la mise en correspondance entre les deux images, des motifs texturés respectant le code Gray sont projetés sur la surface de l'objet au moyen d'un vidéo projecteur. Pour numériser la pièce en entier, 13 acquisitions ont été effectuées et recalées. Le recalage entre les différents nuages est initialisé manuellement, puis optimisé en utilisant l'algorithme Iterative Closest Point (ICP) [Besl, 1992]. Le nuage final comporte environ 660 000 points. Une simplification est apportée en éliminant tous les triangles coplanaires de la surface reconstruite, portant le nombre de points à 85 000 environ. De plus un maillage non uniforme est appliqué pour apporter une simplification supplémentaire au modèle. Le processus global prend environ 2 secondes par acquisition, soit 1 minute pour les 13 acquisitions (en comptant le temps pris pour déplacer le système). L'obtention du nuage final prend de 5 à 10 minutes et les mesures 3D sont données avec une précision de 50 μm .

5.2 Évaluation de la qualité d'un modèle 3D

L'évaluation de la qualité d'un modèle 3D peut être envisagée de deux façons, selon qu'une vérité terrain est connue ou non. Dans ce dernier cas, l'évaluation de la qualité se fait par différents critères qui dépendent de la technique exploitée pour la numérisation 3D. Les critères couramment utilisés en stéréovision non calibrée sont l'erreur résiduelle sur le calcul de la matrice fondamentale et l'erreur de reprojection. Ces deux erreurs sont calculées dans les plans des images et peuvent par conséquent être considérées comme bidimensionnelles. Dans le cas où la vérité terrain est connue, les techniques d'évaluation sont indépendantes de la modalité d'acquisition du modèle 3D. Il s'agit alors de comparer deux nuages de points dans l'espace en utilisant des critères tridimensionnels. Nous présentons dans cette section ces deux types de méthodes.

5.2.1 Méthodes bidimensionnelles

5.2.1.1 Erreur résiduelle sur le calcul de la matrice fondamentale

Cette erreur calculée dans les plans des images est la distance moyenne entre les points appariés et leurs droites épipolaires. Son expression est donnée par l'équation suivante :

$$e_F = \frac{1}{2N} \sum_{i=1}^N \left(d(\mathbf{x}_i, F^T \mathbf{x}_i') + d(\mathbf{x}_i', F \mathbf{x}_i) \right) \quad (5.1)$$

où \mathbf{x}_i et \mathbf{x}'_i sont deux points appariés entre les deux images, F est la matrice fondamentale, N le nombre d'appariements et $d(\mathbf{x}, l)$ est la distance orthogonale entre le point \mathbf{x} et la droite l .

Cette erreur, exprimée en pixels, est un facteur de qualité de la mise en correspondance qui est l'origine de la reconstruction 3D. L'erreur résiduelle sur la matrice fondamentale est alors par conséquent un facteur de qualité du modèle 3D.

5.2.1.2 Erreur de reprojection

On appelle reprojection la projection du modèle 3D reconstruit dans un des plans des images. Les points ainsi reprojétés ne sont pas identiques aux points initialement appariés car la triangulation introduit un décalage par la résolution au sens des moindres carrés du système d'équations des projections. L'erreur de reprojection est alors la distance moyenne entre les points mis en correspondance et leurs reprojétés respectifs. Elle s'exprime en pixel par l'équation suivante :

$$e_{reproj} = \frac{1}{2N} \sum_{i=1}^N (d(\mathbf{x}_i, P X_i) + d(\mathbf{x}'_i, P' X_i)) \quad (5.2)$$

où \mathbf{x}_i et \mathbf{x}'_i sont deux points appariés entre les deux images correspondant au point 3D X_i , P et P' sont les matrices de projection des deux images, N le nombre d'appariements et $d(\mathbf{x}, \hat{\mathbf{x}})$ est la distance entre les points \mathbf{x} et $\hat{\mathbf{x}}$.

5.2.1.3 Relation entre les erreurs bidimensionnelles et le modèle 3D

Il est clair que plus ce type d'erreur est grand, plus le modèle 3D est de mauvaise qualité. Une forte erreur résiduelle sur F pour un appariement signifie que les points appariés n'appartiennent pas strictement au même plan épipolaire. Par conséquent, la reconstruction 3D de cet appariement par une triangulation stricte est impossible. Le point reconstruit correspond alors à une estimation de l'appariement cohérente avec la géométrie épipolaire, introduisant ainsi une erreur de reprojection. Les deux erreurs présentées précédemment sont fortement corrélées comme en atteste la figure 5.4. Les données de cette figure ont été obtenues pour 25 reconstructions 3D réalisées par notre méthode, où nous avons relevé l'erreur résiduelle sur F et l'erreur de reprojection. Les données montrent un coefficient de corrélation de 0,99 pour une régression linéaire.

Cependant, ces erreurs ne peuvent donner qu'un indice sur la qualité de la reconstruction 3D. En effet, pour une erreur résiduelle sur F nulle, il n'est pas garanti que le point reconstruit soit correct sur le modèle 3D, à cause de l'indétermination la position des correspondants le long des droites épipolaires.

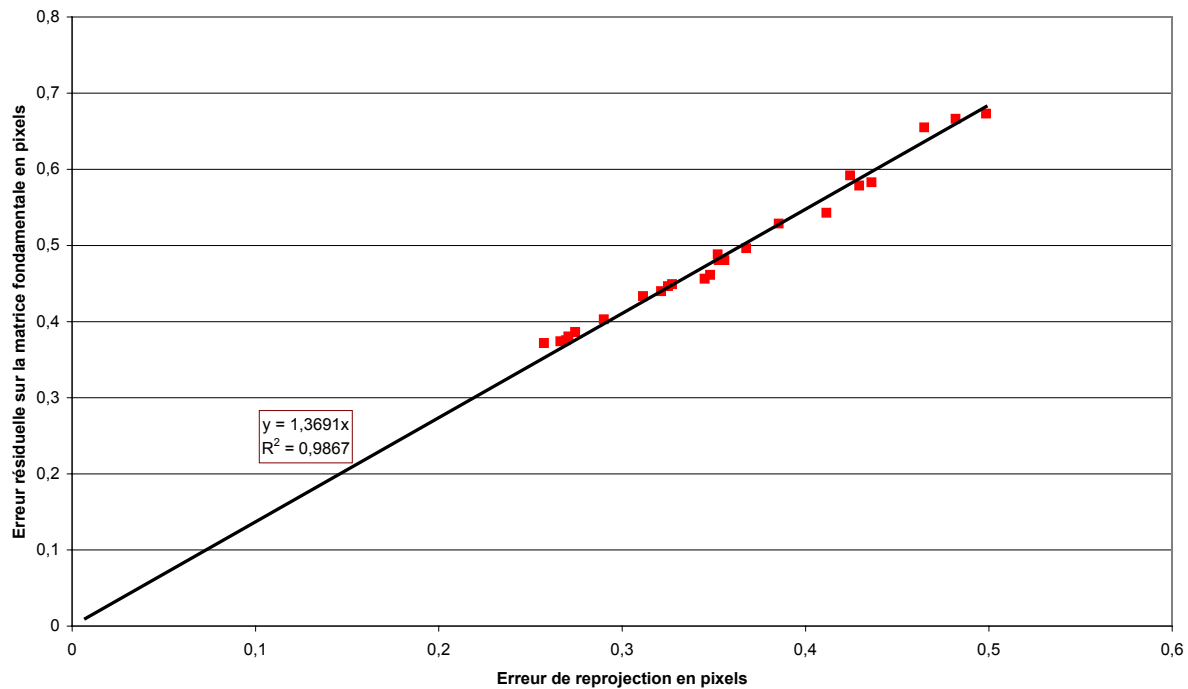


Figure 5.4 : Corrélation entre l'erreur résiduelle sur la matrice fondamentale et l'erreur de reprojection.

5.2.2 Méthodes tridimensionnelles

Lorsqu'une vérité terrain est disponible, il est possible de calculer des facteurs de qualité directement sur le modèle 3D. Ces différents facteurs peuvent être classés en deux familles : ceux qui comparent les formes de l'objet reconstruit vis à vis de la vérité terrain et ceux qui calculent une distance entre les deux nuages de points.

5.2.2.1 Comparaison de formes

Parmi ces méthodes, nombreuses sont celles qui font appel à des descripteurs de formes invariants aux transformations rigides, tels la courbure, les moments statistiques ou les primitives géométriques. L'*Angular Radial Transform* [Ricard, 2005] et la transformée de Fourier 3D [Vranić, 2005] sont des descripteurs répandus. D'autres approches sont plus géométriques telle que celle de Novotni [Novotni, 2001] qui mesure la ressemblance entre deux objets 3D en comparant leur volume de différence. L'objet de référence est dilaté graduellement et le nombre de points de l'objet à comparer présent à l'intérieur de l'objet de référence est mémorisé pour chaque valeur de dilatation. Une deuxième passe est réalisée en permutant les rôles des objets. Les deux courbes ainsi obtenues sont analysées. Une courbe strictement horizontale est gage de bonne ressemblance des objets, à l'inverse d'une courbe

qui croit lentement. Enfin, citons un dernier critère proposé par Farag et Eid [Farag, 2004]. Le principe de leur comparaison est de découper l'espace 3D en voxels et de calculer un facteur de qualité pour chacun d'eux. Ce facteur de qualité local est une distance calculée entre les distributions des deux nuages de points inclus dans le voxel. Un facteur de qualité global peut être donné par l'analyse de l'histogramme des facteurs locaux.

L'inconvénient des ces méthodes est qu'elles ne proposent qu'une mesure relative de la qualité de la reconstruction. Elle permettent de dire seulement qu'une reconstruction est meilleure ou moins bonne qu'une autre vis-à-vis de la vérité terrain. Pour quantifier la qualité de nos reconstructions, nous avons besoin d'une mesure absolue.

5.2.2.2 Calcul de distance

Le critère le plus simple à évaluer est la distance résiduelle moyenne après recalage entre un point et son plus proche voisin dans le nuage de référence. Cette distance est souvent le critère minimisé pour recalcr les deux nuages de points dans un même référentiel. La valeur calculée peut être la distance euclidienne entre les points, la distance de Mahalanobis, ou autre. Pour un calcul optimal de cette distance, les nombres de points des nuages comparés doivent être du même ordre. Lorsque les nuages de points n'ont pas la même densité, le nuage de référence est maillé afin de former une surface, et la distance est calculée entre les points du nuage de l'objet à comparer et leur projection orthogonale sur la surface de référence. C'est cette technique que nous avons utilisée pour évaluer nos comparaisons car nous disposons de nuages de points de densités très différentes. Les reconstructions obtenues par notre méthode comptent environ 3 000 points alors que les numérisations par systèmes calibrés en dénombrent jusqu'à 140 000. Par ailleurs, la mesure de distance euclidienne résiduelle entre les reconstructions nous permet d'avoir une mesure métrologique absolue pour quantifier la qualité de notre méthode. Nous détaillons cette méthode dans la section suivante.

5.3 La méthode de comparaison

Les données fournies par les différents systèmes de numérisation 3D sont des nuages de points 3D. Un maillage est calculé à partir de ces points afin de former la surface de l'objet numérisé. Les différentes surfaces fournies par les systèmes d'acquisition ne sont pas exprimées dans le même référentiel. Il est donc indispensable de réaliser un recalage entre les différents nuages de points avant de pouvoir les comparer, c'est-à-dire de trouver la transformation rigide qui replace les données dans un même référentiel. Le recalage s'effectue en deux étapes. Tout d'abord, un recalage grossier est initialisé manuellement entre les deux nuages de points à comparer, puis, le recalage est affiné en utilisant une technique de minimisation d'erreur entre les deux nuages. L'algorithme le plus répandu pour cette tâche est l'ICP qui minimise la distance euclidienne moyenne entre les points des deux nuages. Afin de nous assurer que l'étape de recalage manuel n'influence pas les résultats, nous avons réalisé une vingtaine de fois cette étape pour obtenir autant de comparaisons entre un modèle 3D et

sa vérité terrain. Les résultats ont montré une grande stabilité, nous permettant de conclure que le recalage manuel n'a que peu d'influence sur le résultat final.

Pour réaliser en pratique les recalages entre les différentes reconstructions, nous avons utilisé le logiciel 3DReshaper de la société Technodigit. Le recalage grossier est réalisé en appariant manuellement au moins 3 points entre les deux surfaces à recaler, puis un recalage basé sur l'algorithme ICP est appliqué. Le logiciel permet également de calculer et de visualiser les erreurs entre les deux surfaces à comparer. Les erreurs relevées par le logiciel sont les distances euclidiennes entre chaque point de la première surface et sa projection orthogonale sur la surface de référence. Cette distance est signée selon que le point analysé est devant (signe positif) ou derrière (signe négatif) la surface de référence. Le côté de la surface est défini par la direction du vecteur normal à la surface comme représenté sur la figure 5.5.

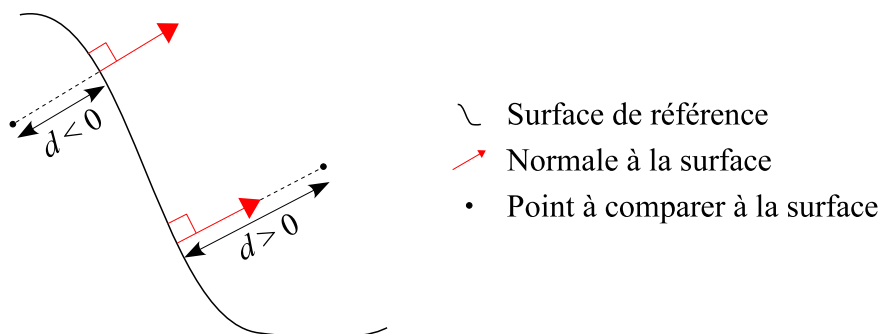


Figure 5.5 : Mesure d'erreur entre un nuage de points et une surface de référence.

La visualisation de ces erreurs se fait par l'intermédiaire de cartes d'erreurs sur la surface de référence. La figure 5.6 montre un exemple de carte d'erreur pour un nuage de points obtenu par notre algorithme de stéréovision et une surface de référence fournie par la société Kréon. La distance entre chaque point du nuage et la surface de référence est codée par une couleur sur la surface de référence. La différence du nombre de points entre la surface de référence fournie par Kréon (50 000 points) et notre modèle 3D (3 000 points) est clairement visible sur la figure. La distribution des distances exprimées en millimètres est affichée sur la gauche de la figure.

Pour des questions de disponibilité des systèmes de numérisation industriels, seule la troisième escarre de la figure 5.1 a été utilisée pour les expérimentations.

Dans un premier temps, nous avons comparé les trois numérisations obtenues par les systèmes industriels entre elles, afin de nous assurer de disposer d'une vérité terrain de qualité. Nous notons pour chaque comparaison l'erreur minimale, maximale, moyenne et l'écart-type de l'erreur. Nous utilisons également la valeur *RMS* de l'erreur résiduelle donnée par l'équation :

$$RMS = \sqrt{\frac{1}{N} \sum_{i=1}^N d_i^2} \quad (5.3)$$

où N est le nombre de points du nuage à comparer et d_i est la distance signée du $i^{\text{ème}}$ point à la surface de référence.

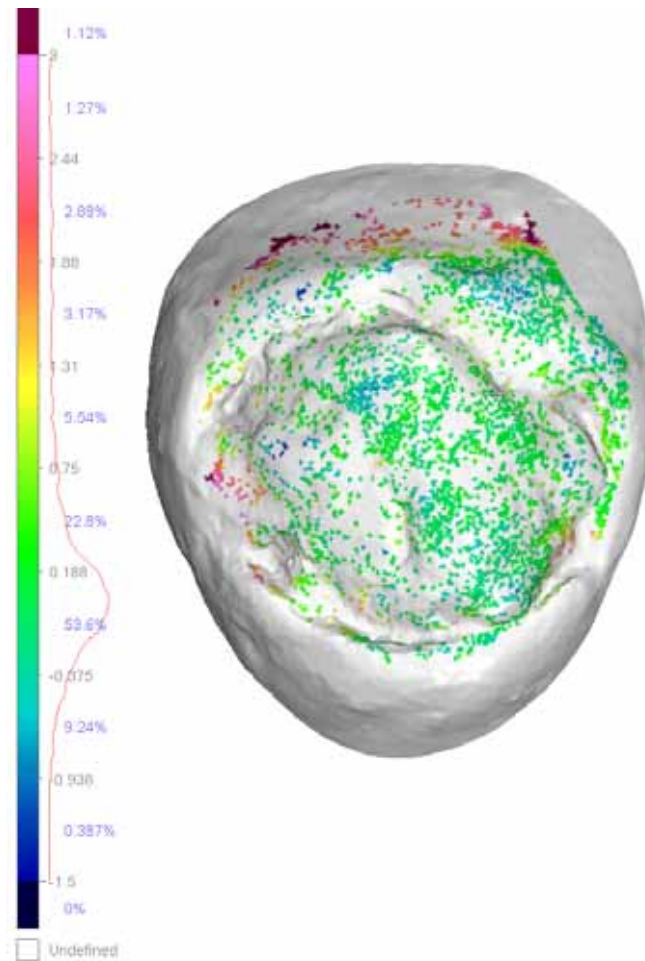


Figure 5.6 : Exemple de carte d'erreur fournie par le logiciel 3DReshaper.

5.4 Résultats

5.4.1 Choix de la vérité terrain

Les trois systèmes industriels ont été comparés deux à deux. La première comparaison est celle de Kréon par rapport à Vidéometric, la seconde entre Eotech et Vidéometric et la troisième entre Kréon et Eotech. Les résultats sont affichés sous forme de cartes d'erreurs sur la figure 5.7 et résumés dans le tableau 5.1.

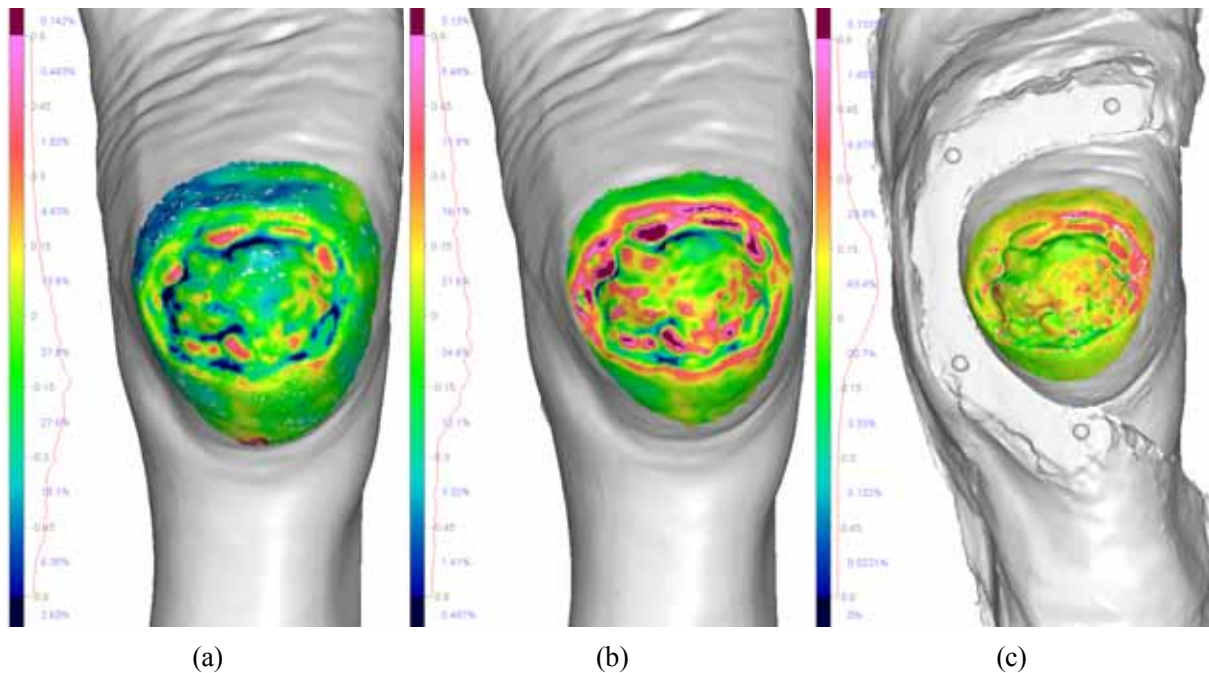


Figure 5.7 : Cartes d'erreurs des comparaisons Kréon / Vidéometric (a), Eotech / Vidéometric (b) et Kréon / Eotech (c).

Tableau 5.1 : Distances résiduelles entre les numérisations industrielles.

Surface de référence	Nuage de points	Distance résiduelle (en mm)				
		Minimum	Maximum	Moyenne	Écart-type	RMS
Vidéometric	Kréon	-0,913	0,985	0,0736	0,2607	0,2709
Vidéometric	Eotech	-1,123	1,264	-0,1718	0,2521	0,3051
Eotech	Kréon	-0,642	0,733	0,0880	0,1301	0,1570

Ces résultats montrent que les comparaisons avec des données issues de Vidéometric présentent des écarts deux fois plus importants que pour les deux autres systèmes. La distance RMS est de l'ordre de 0,3 mm (deux premières lignes du tableau), alors que la distance RMS entre Kréon et Eotech est d'environ 0,15 mm. En analysant les cartes d'erreurs a et b de la figure 5.7, on observe que les écarts entre Vidéometric et les autres sont principalement situés sur les zones à fort relief (les creux en bleu et les bosses en rouge). Cette observation laisse à penser que les algorithmes mis en œuvre par Vidéometric pour simplifier le nuage de points ont tendance à lisser la surface reconstruite. Nous avons alors choisi une des deux autres numérisations, celle de Kréon, comme vérité terrain pour valider nos données.

5.4.2 Évaluation de notre chaîne de reconstruction

La vérité terrain dont nous disposons nous permet d'évaluer notre algorithme de reconstruction 3D. Pour cela, nous avons réalisé 20 reconstructions 3D de l'escarre artificielle, divisées en deux lots. Dix reconstructions 3D ont été réalisées à partir de la même paire

d'images afin de quantifier la stabilité de la reconstruction. En effet, l'algorithme étant basé sur une méthode de type RANSAC pour l'estimation de la matrice fondamentale, le processus aléatoire n'assure pas la reproductibilité exacte des résultats. Le premier lot a été réalisé à partir de deux prises de vue obtenues par l'appareil photo numérique Sony DSC-H1, pour des points de vue distants de 42 et 68 cm et de vergence de 47° , avec un niveau de zoom d'environ $3\times$. Ce niveau de zoom a été utilisé pour avoir les meilleurs cadrages et mises au point possibles sans avoir à utiliser le mode macro de l'appareil. Le second lot de données a été obtenu pour des prises de vue en mode macro au niveau de zoom $1\times$, distantes d'environ 10 et 12 cm de l'escarre, la vergence étant toujours d'environ 45° . Les photographies ayant servi à ces reconstructions sont reproduites en figure 5.8.

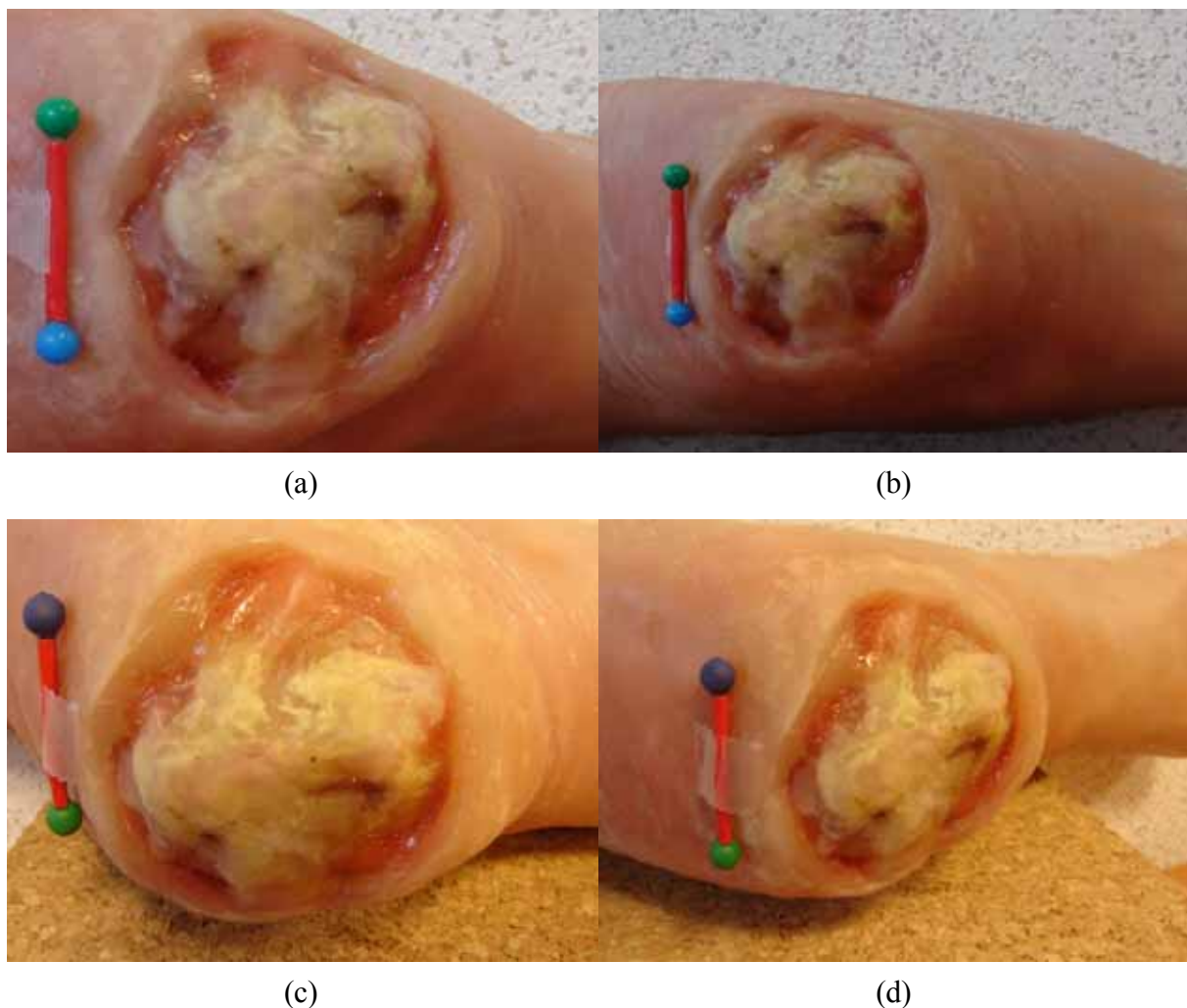


Figure 5.8 : Paires d'images utilisées pour les deux lots de reconstructions 3D. Premier lot de données en mode normal (a) et (b), second lot de données en mode macro (c) et (d).

Le tableau 5.2 donne les caractéristiques des 20 reconstructions 3D, c'est-à-dire la distance focale obtenue par auto-calibrage et le nombre de points 3D pour chaque reconstruction. On observe que l'estimation de la focale est plus stable en mode macro qu'en mode normal. L'écart-type des focales rapporté à la focale estimée moyenne est de 20 % pour le mode normal et de 13 % en macro. De même, le nombre moyen de points reconstruits en mode macro (2992) est légèrement supérieur à celui obtenu en mode normal (2520).

Tableau 5.2 : Caractéristiques des reconstructions 3D réalisées par notre algorithme.

	Numéro de la reconstruction	Focale exprimée en pixel	Nombre de points 3D
Premier lot de reconstructions (mode normal)	1	3956	2615
	2	4707	2362
	3	5160	3161
	4	5545	2552
	5	6080	2585
	6	6291	3009
	7	6830	1980
	8	6831	2287
	9	7372	2322
	10	7906	2329
Second lot de reconstructions (mode macro)	1	1748	3061
	2	1537	3052
	3	1529	2335
	4	1523	3065
	5	1509	2960
	6	1464	3133
	7	1647	3334
	8	2018	2785
	9	1832	3205
	10	1301	2995

Les résultats des comparaisons de ces 20 reconstructions par rapport à la surface de référence de Kréon sont fournis dans le tableau 5.3. On observe que globalement, le mode macro permet une reconstruction plus précise, avec une erreur RMS moyenne de 1 mm contre 1,5 mm pour le mode normal. Par ailleurs, des erreurs de mise en correspondance sur les parties peu texturées des images génèrent de fortes erreurs de reconstruction en dehors de la zone d'intérêt de l'escarre, comme l'atteste la figure 5.9. En supprimant manuellement ces zones manifestement erronées, les erreurs de reconstruction sont nettement diminuées. Le tableau 5.4 reporte ces résultats une fois les zones de mauvaises correspondances supprimées. L'erreur RMS moyenne en mode normal passe alors à 0,72 mm et en mode macro à 0,59 mm. Le mode macro présente toujours une erreur plus faible que l'utilisation du mode normal. On peut donc conclure de ces expérimentations que les prises de vue doivent être effectuées le plus près possible de l'objet à numériser, et que dans ce cas notre méthode a une précision de 0,59 mm sur un objet de taille 5×5 cm. Enfin, la série des dix reconstructions 3D présentent des résultats similaires, vérifiant la stabilité de notre méthode. La figure 5.10 montre une carte d'erreur typique des reconstructions 3D obtenues pour la reconstruction n°4 en mode macro.

Tableau 5.3 : Erreurs résiduelles entre les reconstructions 3D et la surface de référence Kréon.

	Recon- struction	Erreur résiduelle (en mm)				
		Minimum	Maximum	Moyenne	Écart-type	RMS
Premier lot de reconstructions (mode normal)	1	-2,129	9,518	0,6898	1,4149	1,5741
	2	-2,413	5,103	0,3883	1,2402	1,2996
	3	-2,775	11,924	0,7733	2,1117	2,2489
	4	-3,101	3,866	0,0228	0,6853	0,6856
	5	-2,313	12,434	0,7127	2,2901	2,3985
	6	-15,944	19,001	1,4189	2,8497	3,1835
	7	-1,847	3,332	0,1856	0,6288	0,6556
	8	-2,816	8,698	0,2253	1,0309	1,0552
	9	-1,875	5,328	0,3542	1,0269	1,0863
	10	-2,533	4,646	-0,0362	1,0337	1,0343
Second lot de reconstructions (mode macro)	1	-2,258	7,074	0,5931	1,2413	1,3758
	2	-1,758	4,158	0,2782	0,7862	0,8340
	3	-1,778	4,04	0,2204	0,6958	0,7299
	4	-2,37	5,541	0,2460	0,8604	0,8949
	5	-1,181	6,603	0,3547	0,9225	0,9884
	6	-1,589	5,182	0,2850	0,7972	0,8466
	7	-1,667	7,793	0,4700	1,0602	1,1597
	8	-2,813	7,524	0,2487	1,1968	1,2224
	9	-1,513	12,521	0,7063	1,6841	1,8263
	10	-1,01	6,982	0,2610	0,7134	0,7596

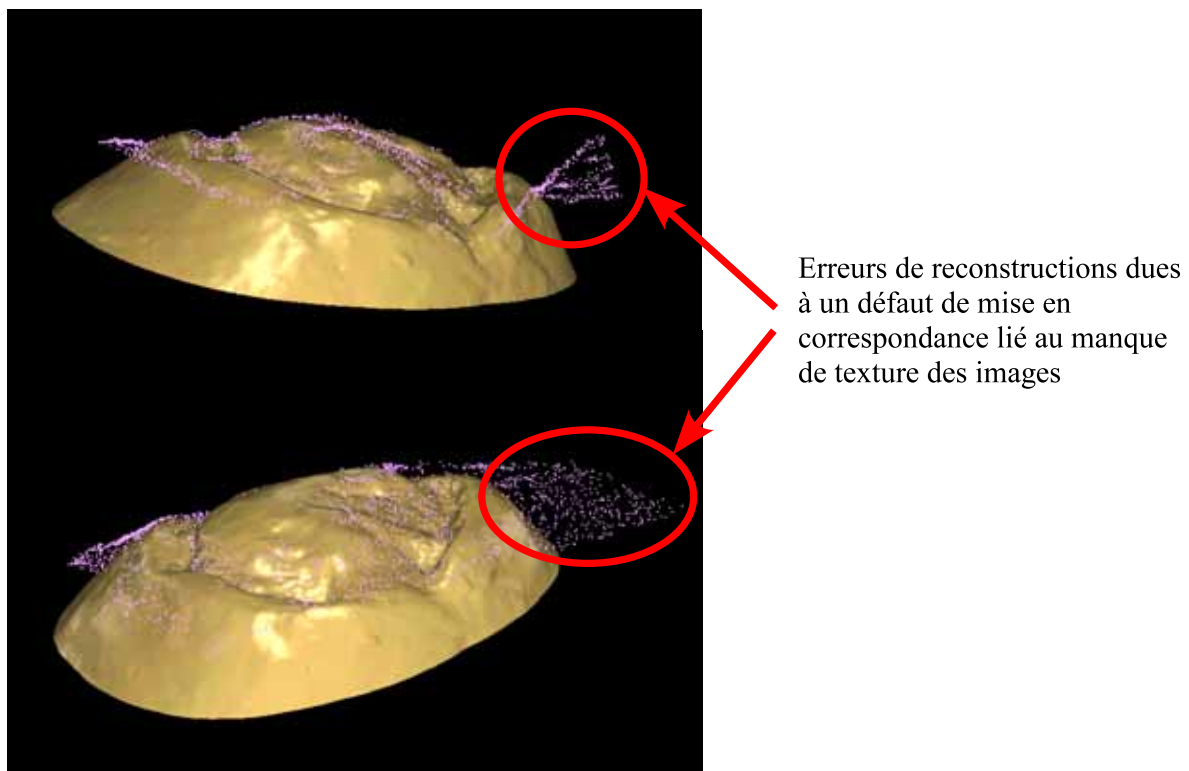


Figure 5.9 : Visualisation des erreurs de reconstructions liées aux problèmes d'appariements.

Tableau 5.4 : Erreurs résiduelles entre les reconstructions 3D et la surface de référence Kréon, après suppression manuelle des points 3D aberrants.

	Recon- struction	Erreur résiduelle (en mm)				
		Minimum	Maximum	Moyenne	Écart-type	RMS
Premier lot de reconstructions (mode normal)	1	-2,129	2,8	0,4455	1,0115	1,1053
	2	-2,413	3,049	0,3516	1,2386	1,2876
	3	-2,775	2,259	-0,181	0,5144	0,5147
	4	-3,101	1,454	-0,0814	0,5432	0,5492
	5	-2,313	1,652	-0,0886	0,4921	0,5003
	6	-1,748	2,968	0,3035	0,6925	0,7561
	7	-1,847	1,925	0,0852	0,4777	0,4853
	8	-2,816	2,109	-0,0322	0,4903	0,4913
	9	-1,875	3,536	0,1041	0,6869	0,6948
	10	-2,533	4,646	-0,0362	1,0337	0,8057
Second lot de reconstructions (mode macro)	1	-2,053	3,516	0,2435	0,6545	0,6983
	2	-1,855	2,884	0,0803	0,4608	0,4678
	3	-1,8	2,25	0,1860	0,5533	0,5838
	4	-2,42	3,107	0,1223	0,6235	0,6354
	5	-1,177	2,991	0,1698	0,5303	0,5568
	6	-1,583	2,08	0,0758	0,4411	0,4183
	7	-1,661	3,7	0,2225	0,6692	0,7052
	8	-1,162	2,856	0,2557	0,6509	0,6993
	9	-1,362	2,68	0,2116	0,5524	0,5916
	10	-1,013	3,747	0,1698	0,5627	0,5878

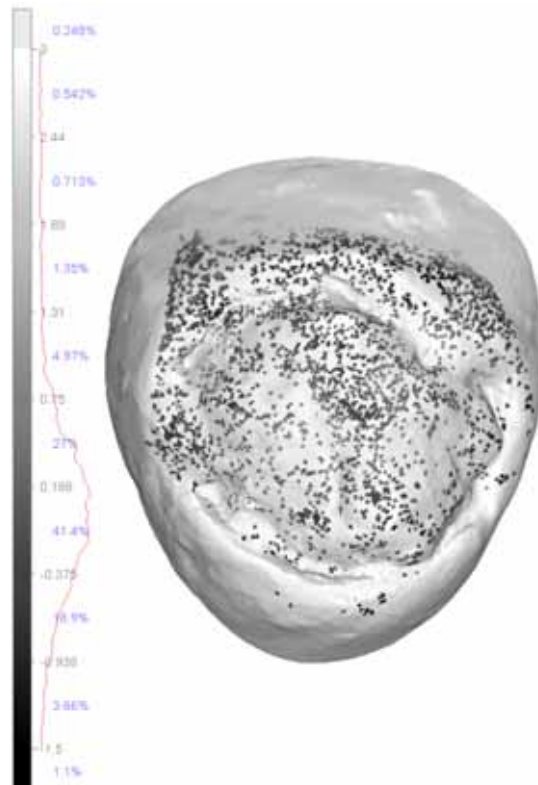


Figure 5.10 : Carte d'erreur typique des reconstructions 3D obtenues par notre méthode.

5.5 Conclusion

Ce premier chapitre concernant la validation de notre chaîne de reconstruction 3D, a permis d'évaluer la qualité des données 3D reconstruites. Dans un premier temps, une vérité terrain a été fournie par des scanners 3D industriels. Trois numérisations industrielles d'une même escarre artificielle étaient disponibles, dont une se démarquait des autres par une surface plus lisse. Cette dernière a été écartée et la vérité terrain a été choisie parmi les deux autres numérisations.

À partir de cette vérité terrain, il a été possible de quantifier la qualité de nos reconstructions 3D de l'escarre artificielle réalisées à partir de deux couples d'images différents, se différenciant par les distances de prises de vue. Le premier couple a été pris à environ 55 cm de l'escarre avec un niveau de zoom d'environ $3\times$, alors que le second couple a été acquis à l'aide du mode macro de l'appareil photo à des distances de l'ordre de 10 cm. Il en résulte que le second couple offre une distance résiduelle plus faible par rapport à la vérité terrain. La distance moyenne entre les points 3D issus de nos reconstructions et la surface de référence est d'environ 0,7 mm pour le premier couple d'images et de 0,6 mm pour le second. Ces erreurs sont à rapporter à la taille de l'objet numérisé qui peut être assimilé à un cube de 5 cm de côté. Par rapport à la dimension de ce cube, les erreurs de localisation des points 3D sont de 1,4 et 1,2 % selon les couples d'images.

Ces résultats sont très encourageants et laissent présager le respect effectif de la limite de 5 % de précision sur la mesure du volume d'escarres réelles sur des patients. Par ailleurs, la méthodologie de comparaison des surfaces 3D mise en place dans ce chapitre va contribuer à établir dans le prochain chapitre une configuration de prise de vue optimale, assurant la meilleure reconstruction possible.

Chapitre 6

Choix d'un protocole de prise de vue

Le précédent chapitre a permis d'évaluer expérimentalement la qualité des reconstructions 3D. Il s'agit maintenant d'analyser l'influence du positionnement de l'appareil photo sur la qualité de cette reconstruction, avec au final la détermination de la configuration de prise de vue optimale.

Sommaire

6.1 Protocole expérimental.....	129
6.2 Résultats.....	131
6.3 Conclusion.....	137

La qualité de la reconstruction 3D dépend directement de la position des prises de vue. En effet, des prises de vue très rapprochées simplifient la mise en correspondance, car les images sont alors très semblables. Cependant, le cas des points de vue proches est une source d'imprécision pour la triangulation, car les rayons issus de chaque plan image et pointant un même point 3D sont presque parallèles. À l'inverse, des points de vue très différents améliorent le résultat de la triangulation, l'idéal étant que l'orientation relative entre les deux points de vue soit de 90° . Mais dans ce cas, la mise en correspondance devient ardue du fait de la grande différence d'aspect des images. Des phénomènes d'occultation peuvent également survenir lors de points de vue très écartés, rendant alors impossible la mise en correspondance dans les régions occultées.

Il s'agit donc de trouver un compromis entre la facilité de mise en correspondance et la qualité de la triangulation. Nous avons pour cela mis en place une expérience où la qualité de la reconstruction 3D est analysée en fonction de la position relative des deux prises de vue. Nous présentons dans un premier temps les données et le protocole expérimental, puis, nous aborderons les résultats avant de conclure sur la configuration optimale de prise de vue.

6.1 Protocole expérimental

Nous avons fait usiner en prototypage rapide un objet de référence constitué de deux portions de sphères collées sur un plan (fig. 6.1). L'utilisation de portions de sphères permet d'acquérir des données sous toutes les orientations possibles, tout en minimisant les risques d'occultation sans toutefois les annuler. La petite sphère sert de détrompeur pour initialiser le recalage de la reconstruction avec le modèle numérique de prototypage rapide de l'objet. En effet, l'étape d'initialisation du recalage impose de pointer manuellement 3 points homologues sur les deux surfaces, ce qui est loin d'être évident sur un maillage dépourvu de texture (le capteur de Kréon ne la fournit pas).

Un vidéo projecteur nous permet de projeter une texture sur l'objet pour la mise en correspondance. En effet, le matériau dans lequel l'objet a été usiné est d'aspect mat et de couleur uniforme. La position de l'appareil photo est également ajustable à l'aide d'un bras articulé, permettant de connaître précisément la position de l'appareil. Ce dispositif est représenté sur la figure 6.2. Nous avons pris 26 photos numériques dont les positions sont repérées sur la figure 6.3. Les couples d'images utilisés pour la stéréovision sont constitués de la vue 1 associée avec l'une des 25 autres vues. Ainsi, 25 reconstructions 3D, numérotées de 2 à 26, sont générées, avec un angle inter-vues et une distance relative des deux prises de vue variables.

La position du projecteur dans ce dispositif ainsi que la texture projetée doivent être judicieusement choisies. La texture utilisée est une génération aléatoire de la couleur de chaque pixel. Cette projection permet de détecter un maximum de coins dans les images. La position du projecteur a été choisie de façon à ne pas gêner le bras métallique portant l'appareil photo, tout en garantissant une projection la plus orthogonale possible à la surface de l'objet. Elle permet aussi d'avoir une texture de bonne qualité sur tout l'objet. Toutefois, un

léger étalement de la texture due à la sphéricité de la pièce est visible sur la partie droite de l'objet (fig 6.1b). Ce problème d'étalement n'a pas dégradé les reconstructions 3D.

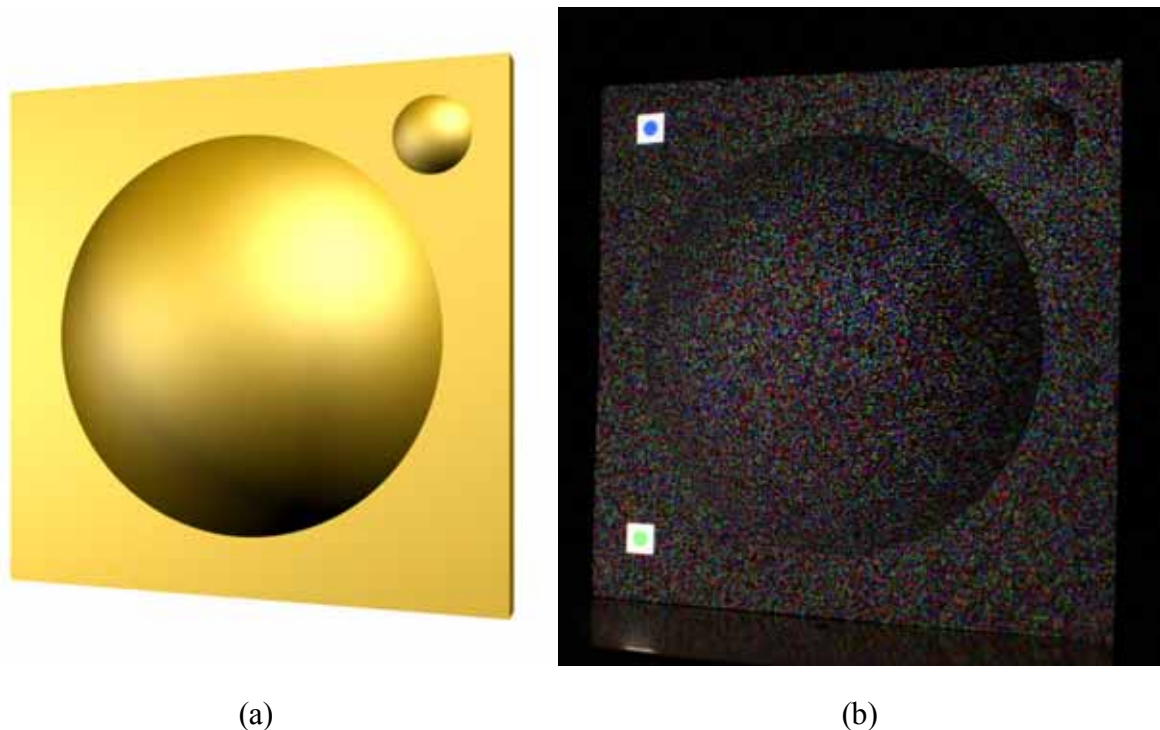


Figure 6.1 : Modèle numérique de référence de l'expérimentation (a). Objet réel avec texture projetée (b). Les points vert et bleu permettent de fixer le facteur d'échelle de la reconstruction 3D.

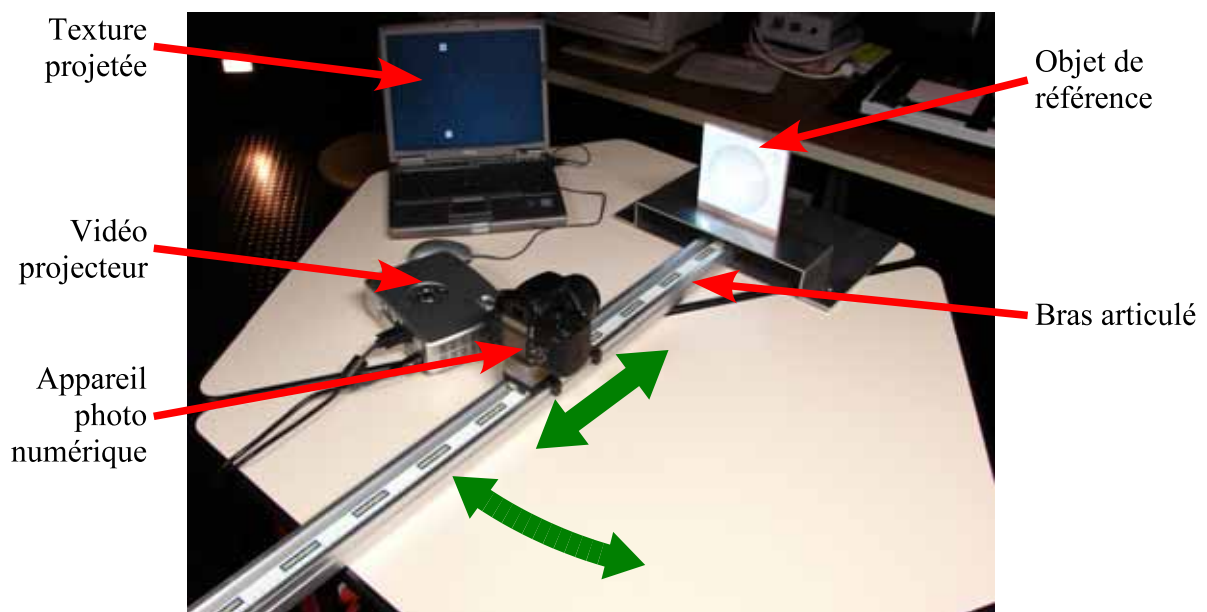


Figure 6.2 : Dispositif de projection de texture et de contrôle de prises de vue.

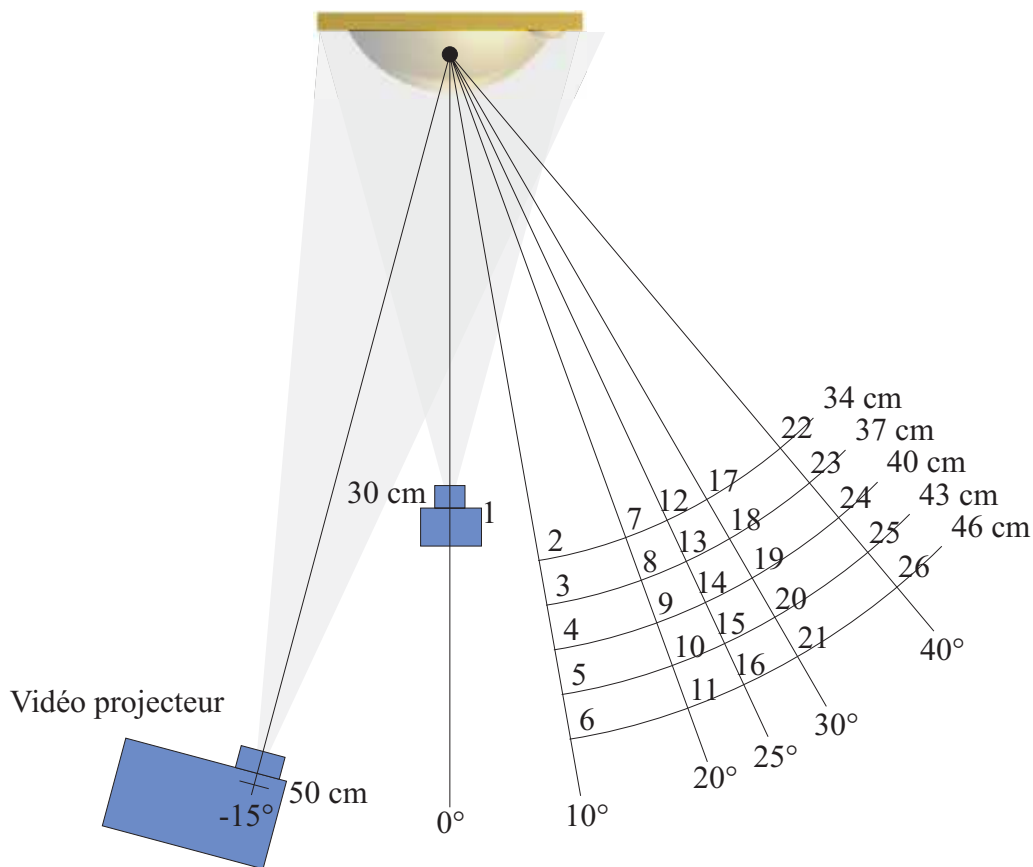


Figure 6.3 : Positions de l'appareil photo pour la série de prises de vue (1 à 26).

6.2 Résultats

L'évaluation de la précision des reconstructions est réalisée par rapport au modèle numérique ayant servi à la construction de l'objet, la conformité de l'usinage étant certifiée par un contrôle par palpeur sur machine à mesurer tridimensionnelle. Le tableau 6.1 donne les résultats obtenus pour les 25 reconstructions. L'erreur RMS reste faible et stable, aux alentours de 1 mm jusqu'à la position 13, c'est-à-dire pour des angles de vergence inférieurs à 25° et pour un rapport de distances de points de vue inférieur à 1,23 lorsque l'angle vaut 25°. Au-delà, l'erreur augmente énormément. En réalité, l'erreur augmente à cause de l'apparition d'occultations dans les images. La partie gauche de la grosse sphère est masquée dans les vues dont l'angle est supérieur à 25°. La reconstruction 3D est alors impossible pour cette partie de la sphère, formant un méplat sur la sphère (fig. 6.4). Cette malformation génère localement de grosses erreurs qui modifient de façon significative son écart-type.

Tableau 6.1 : Erreur résiduelle pour les 25 reconstructions de l'objet de référence.

Seconde vue	Erreur résiduelle (en mm)				
	Minimum	Maximum	Moyenne	Écart-type	RMS
2	-4,627	3,063	0,2938	0,9672	1,0108
3	-4,353	3,764	0,1045	0,7852	0,7921
4	-5,066	3,757	0,3077	1,0131	1,0588
5	-5,528	4,292	0,2336	1,0001	1,0270
6	-4,549	4,608	0,1297	1,0040	1,0123
7	-3,634	3,371	0,3139	0,9453	0,9961
8	-4,763	3,812	0,0900	0,8975	0,9020
9	-3,706	3,365	0,2507	0,9392	0,9709
10	-4,965	4,422	0,1376	0,9145	0,9248
11	-3,62	4,797	0,2403	0,9559	0,9857
12	-6,596	4,498	0,0915	1,0390	1,0431
13	-6,898	3,957	0,0744	0,9236	0,9266
14	-12,026	2,298	-0,1303	1,2963	1,3029
15	-9,051	3,133	-0,2847	1,5811	1,6065
16	-9,947	4,509	-0,1118	1,1963	1,2015
17	-7,093	3,774	0,2552	1,3343	1,3585
18	-8,029	4,085	-0,0435	1,2669	1,2677
19	-7,979	6,302	-0,2156	1,4917	1,5072
20	-8,319	2,867	-0,2542	1,3698	1,3932
21	-10,017	2,614	-0,2952	1,6976	1,7231
22	-8,238	4,548	-0,2594	1,5490	1,5705
23	-10,368	2,939	-0,4814	2,2101	2,2620
24	-11,04	6,929	-0,3737	2,0449	2,0788
25	-10,519	4,031	-0,6852	1,9832	2,0983
26	-11,914	3,629	-0,8439	2,1878	2,3450



Figure 6.4 : Défaut de reconstruction lié aux occultations. Reconstruction obtenue par les images 1-19.

Afin de nous affranchir des erreurs liées aux occultations sur la partie gauche de l'objet, nous avons mesuré l'erreur résiduelle sur la seule moitié droite de l'objet. Les résultats sont donnés par le tableau 6.2. Le graphique de la figure 6.5 représente ces résultats en fonction de la position des vues. L'erreur est stable quelles que soient les positions relatives des vues. Ces mesures ne présentant pas de minimum marqué, il n'est alors pas possible de statuer sur les meilleures conditions de prises de vue pour la reconstruction 3D.

Tableau 6.2 : Erreur résiduelle pour les 25 reconstructions de l'objet de référence, après suppression des zones occultées.

Seconde vue	Erreur résiduelle (en mm)				
	Minimum	Maximum	Moyenne	Écart-type	RMS
2	-3,4	2,799	0,0156	0,7444	0,3645
3	-4,596	8,734	0,1031	0,8463	0,3458
4	-5,256	2,954	0,0584	0,7885	0,7906
5	-5,928	3,145	-0,0030	0,8619	0,8620
6	-4,605	4,181	0,0543	0,7485	0,7505
7	-3,101	2,711	-0,0175	0,6821	0,6823
8	-4,455	3,739	0,0120	0,7559	0,7560
9	-3,854	2,625	0,0710	0,7880	0,7912
10	-4,834	6,069	-0,0148	0,7333	0,7334
11	-2,722	3,329	0,0551	0,7845	0,7864
12	-4,304	4,689	-0,0496	0,7496	0,7164
13	-4,155	3,428	0,0137	0,6907	0,6908
14	-4,046	2,367	-0,0190	0,7559	0,7562
15	-4,43	2,811	0,0279	0,6931	0,6937
16	-5,073	4,367	0,0398	0,6883	0,6894
17	-4,178	3,218	0,1254	0,9095	0,9181
18	-3,304	4,066	0,0337	0,8501	0,8507
19	-5,059	2,37	-0,0141	0,6926	0,6927
20	-4,865	2,852	0,0351	0,6664	0,6673
21	-5,222	2,406	-0,0567	0,7019	0,7042
22	-4,658	3,806	-0,2307	0,9911	1,0176
23	-7,415	2,741	-0,2247	1,0510	1,0747
24	-8,237	7,193	0,0297	1,1568	1,1572
25	-4,011	3,533	-0,0969	0,7615	0,7677
26	-6,927	3,512	-0,0997	0,8461	0,8519

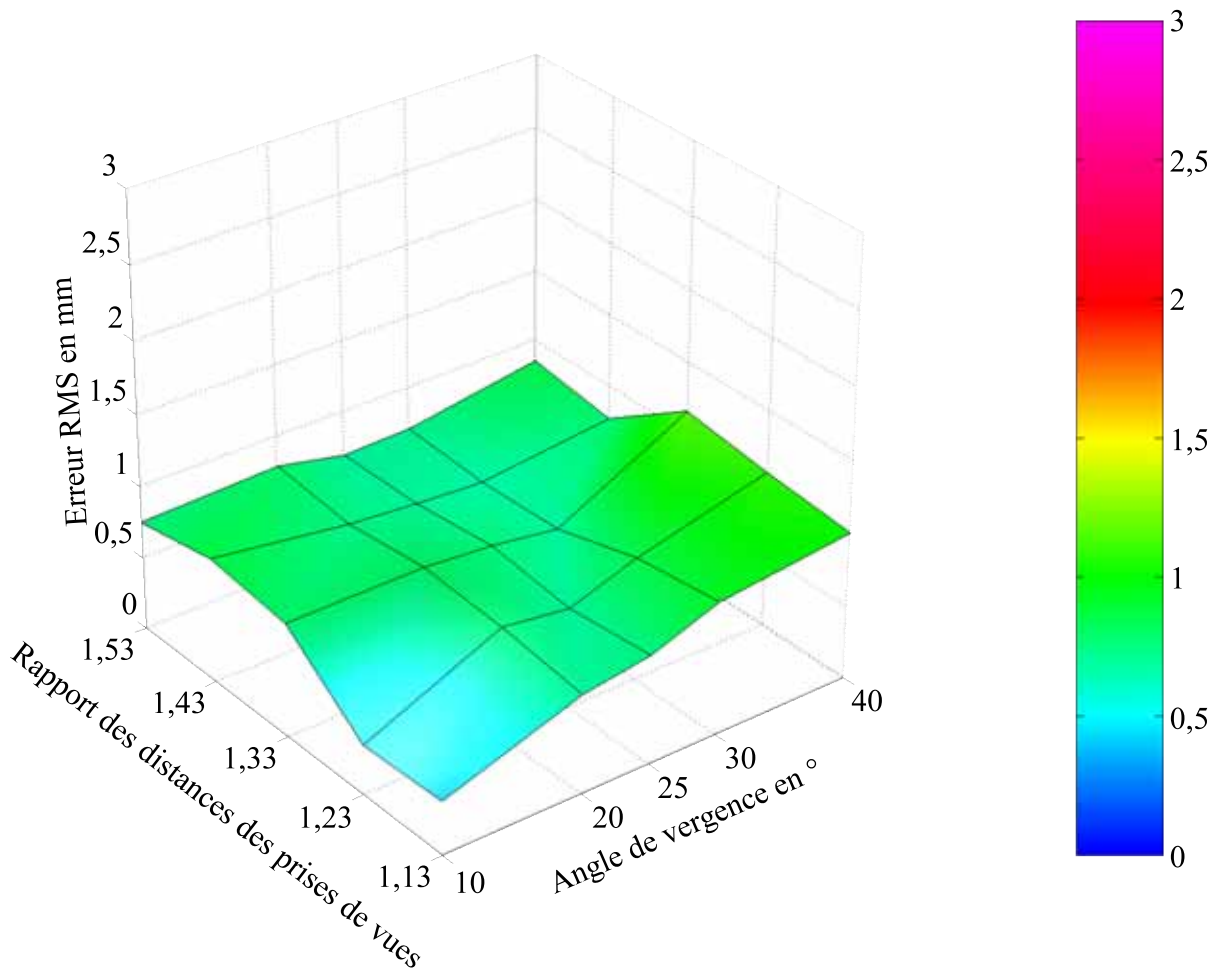
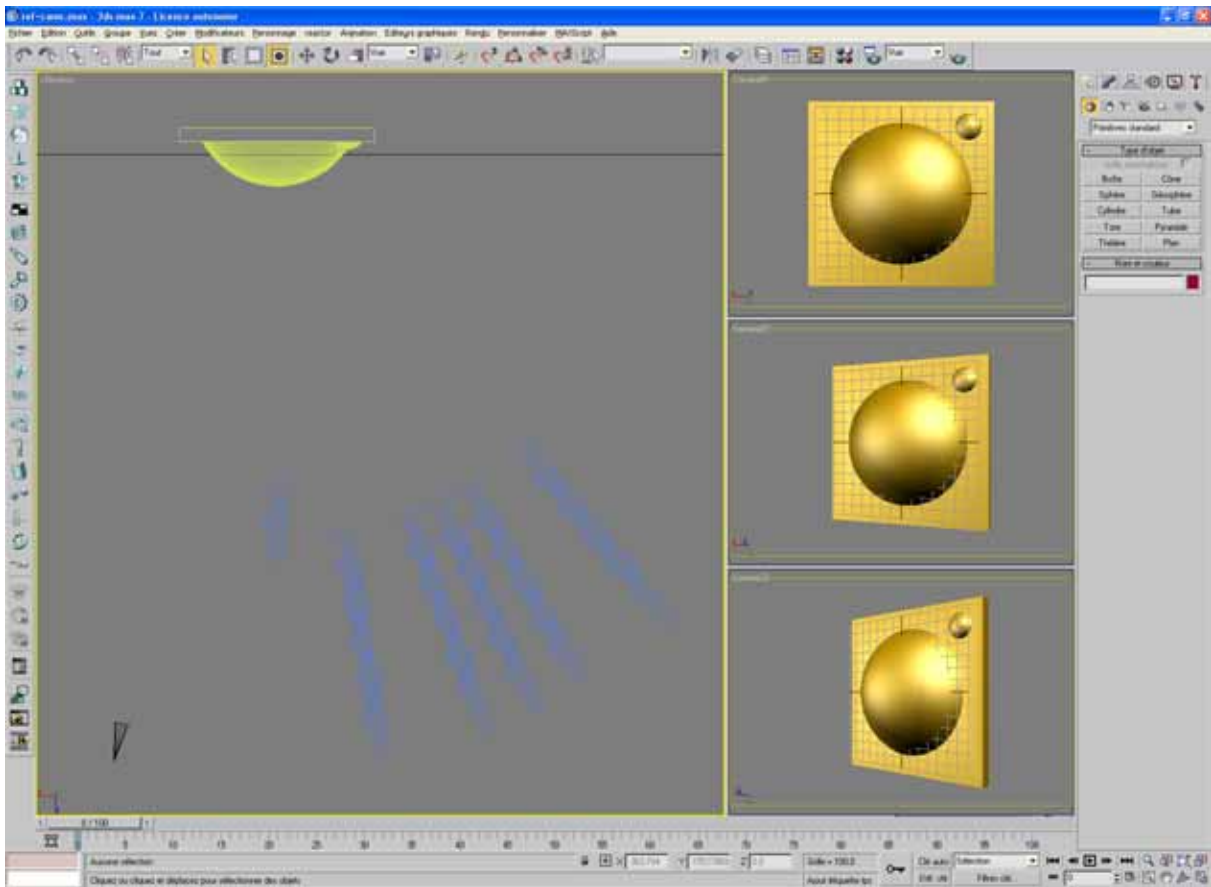
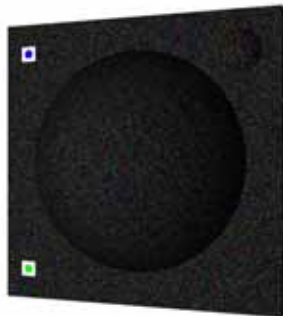


Figure 6.5 : Erreur RMS en fonction de la position des vues pour des images réelles.

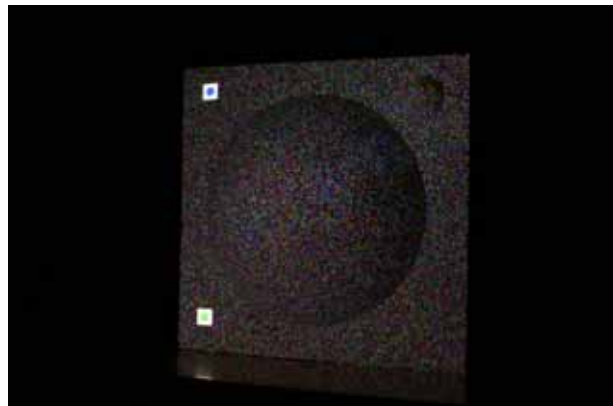
Nous avons alors conçu une nouvelle expérience. L'idée est de simuler complètement toute la chaîne de reconstruction 3D, depuis la capture d'image jusqu'au modèle final. Nous rappelons que dans le chapitre 2 nous avons réalisé une première simulation mais elle se limitait à la génération aléatoire de points dans un cube virtuel, excluant toute la partie de mise en correspondance incluse ici. Le même protocole expérimental que précédemment est utilisé, à la différence près que les images ne sont plus acquises par un appareil photo, mais synthétisées à l'aide d'un logiciel d'infographie 3D. Toute l'expérience précédente a ainsi été simulée grâce au logiciel 3DStudio Max (fig. 6.6). Le placage de texture est simulé par un projecteur virtuel et les vues sont réalisées en plaçant des caméras virtuelles sur les 26 positions. L'avantage de cette simulation est de s'affranchir du bruit généré par les conditions réelles (défaut de mise au point, distorsions optiques, bruit électronique, etc...). L'erreur RMS résiduelle après suppression des problèmes d'occultation est présentée par le tableau 6.3 et représentée en fonction de la position des vues en figure 6.7. La reconstruction pour la position 26 n'a pas pu être réalisée, la mise en correspondance ayant échoué. On observe alors un minimum faiblement marqué pour les positions 12-13, c'est-à-dire pour un angle de vergence de 25° et un rapport de distances des points de vue à l'objet compris entre 1,13 et 1,23.



(a)



(b)



(c)

Figure 6.6 : Simulation des prises de vue grâce à 3D Studio Max. Capture d'écran du logiciel (a), simulation de la vue 7 (b), vue 7 réelle (c).

Tableau 6.3 : Erreur résiduelle pour les 24 reconstructions de l'objet de référence par simulation, après suppression des zones occultées.

Seconde vue	Erreur résiduelle (en mm)				
	Minimum	Maximum	Moyenne	Écart-type	RMS
2	-3,051	3,195	0,1105	0,7216	0,7300
3	-4,596	8,734	0,1031	0,8463	0,8525
4	-5,206	3,593	-0,0106	0,8798	0,8799
5	-3,791	5,85	0,0505	0,7378	0,7395
6	-3,934	3,833	0,0405	0,6704	0,6716
7	-4,537	2,922	-0,0021	0,5967	0,5968
8	-2,921	4,411	0,1133	0,6899	0,6992
9	-2,898	4,056	0,0877	0,6303	0,6364
10	-4,433	3,742	0,0312	0,6009	0,6017
11	-6,07	3,887	0,0070	0,5630	0,5631
12	-2,33	2,489	0,0481	0,4995	0,5018
13	-2,466	4,436	0,0491	0,5265	0,5288
14	-5,543	5,867	-0,0101	0,6725	0,6726
15	-3,53	5,715	-0,0034	0,6366	0,6366
16	-7,167	3,69	0,1703	0,9121	0,9279
17	-3,701	3,92	0,1221	0,5653	0,5654
18	-6,211	3,267	-0,0308	0,8228	0,8233
19	-9,694	3,201	-0,2461	1,3778	1,3996
20	-10,421	3,251	-0,4355	1,4692	1,5324
21	-9,719	3,419	-0,2917	1,3605	1,3914
22	-8,672	4,413	-0,7585	1,7638	1,9201
23	-7,724	2,559	-0,5556	1,6165	1,7094
24	-8,488	4,934	-0,8161	2,0049	2,1648
25	-9,475	3,92	-0,9646	2,0896	2,3017

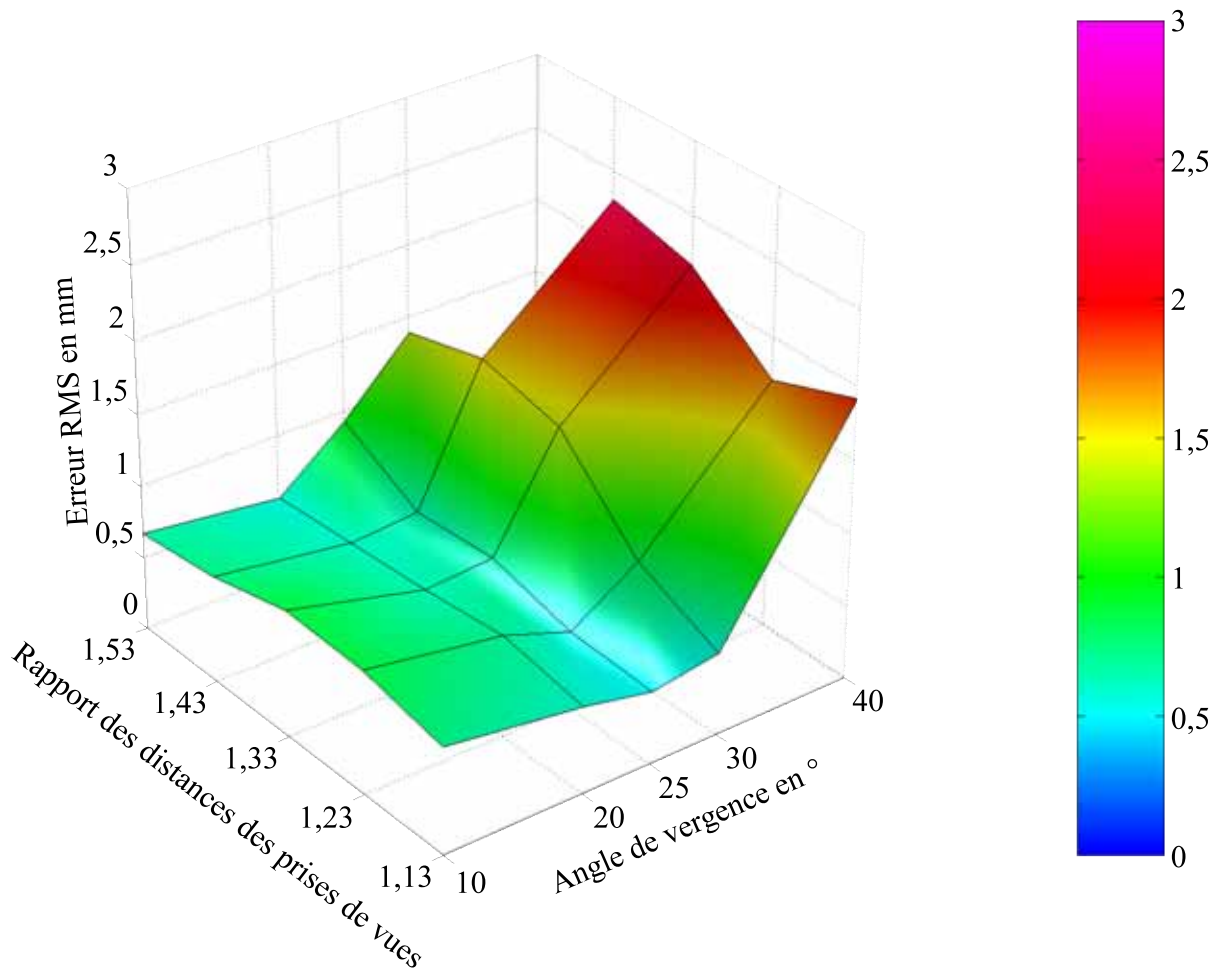


Figure 6.7 : Erreur RMS en fonction de la position des vues pour des images de synthèse.

6.3 Conclusion

Nous avons présenté dans ce chapitre deux expériences réalisées dans le but de déterminer la meilleure configuration de prise de vue pour notre méthode de reconstruction 3D. Vingt-cinq configurations ont été testées, faisant varier à la fois l'angle de vergence et le rapport des distances des points de vue à l'objet acquis. Les tests menés en simulation complète de toute la chaîne de stéréovision montrent que l'erreur de reconstruction est minimisée pour un angle de vergence de 25° et un rapport de distances des points de vue d'environ 1,2. Cependant, le minimum de l'erreur n'est pas fortement marqué, les prises de vue acquises avec une vergence plus faible présentant des erreurs du même ordre de grandeur. De plus, lorsque la chaîne de stéréovision est analysée avec des images réelles, ce minimum n'est plus visible, masqué par les erreurs induites par le bruit inhérent aux conditions réelles.

La configuration de prise de vue qui permet une reconstruction 3D optimale peut être déduite de ces expériences. L'utilisateur doit prendre deux photos de l'objet avec un angle de vergence d'environ 25° et en reculant légèrement l'appareil photo pour la seconde prise de

vue. Le protocole n'a pas besoin d'être plus précis car l'erreur provoquée par un déplacement de la position idéale est négligeable. Ainsi, une grande liberté est assurée à l'utilisateur au moment de la prise de vue, ce qui est un atout de la méthode pour une utilisation routinière en milieu hospitalier. L'utilisateur peut alors se concentrer sur des points plus critiques comme la mise au point, l'éclairage ou le cadrage. Ce protocole de prise de vue est résumé par la figure 6.8.

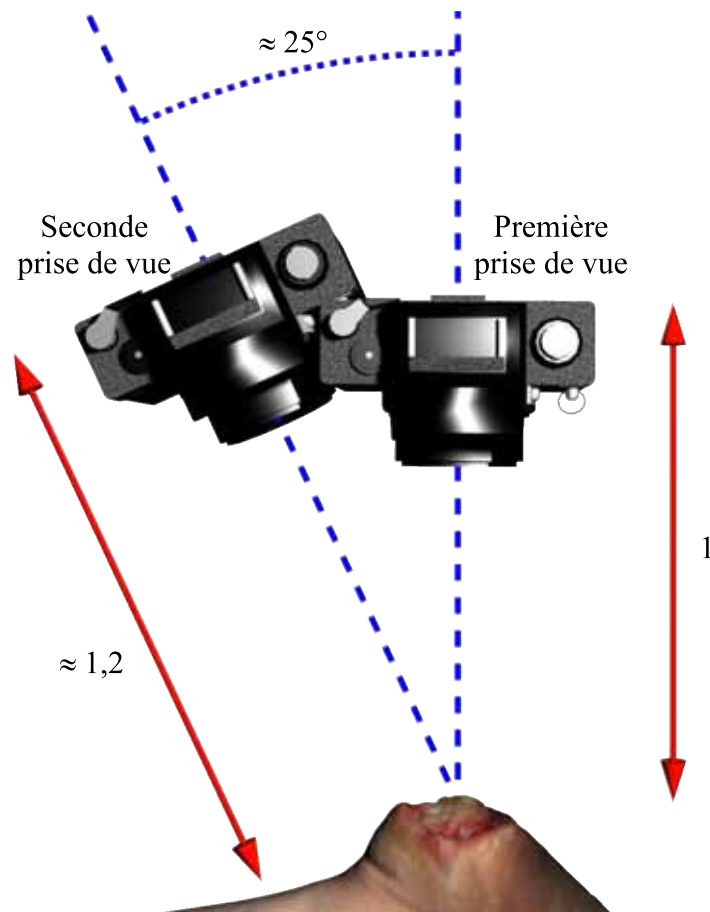


Figure 6.8 : Configuration de prise de vue optimale.

Chapitre 7

Évaluation de la précision volumétrique

L'objectif visé dans le cadre de notre application médicale est le relevé volumétrique d'escarres. En effet, le volume d'une plaie est un indice pertinent pour juger de l'évolution de celle-ci. Les tests ont porté jusqu'à présent sur la conformité de la surface reconstruite avec une référence, mais l'évaluation du volume n'a pas été faite. Ce chapitre présente les calculs et l'expérimentation menés pour réaliser cette évaluation.

Sommaire

7.1 Stratégie.....	140
7.2 Calcul du volume sous une surface.....	142
7.2.1 Calcul d'un volume élémentaire.....	143
7.2.1.1 Cas où le triangle est entièrement au dessus du plan de référence.....	144
7.2.1.2 Cas où un sommet du triangle est sous le plan de référence.....	144
7.2.1.3 Cas où deux sommets du triangle sont sous le plan de référence.....	145
7.2.2 Transformation du plan de référence.....	146
7.2.2.1 Écriture analytique.....	146
7.2.2.2 Algorithme.....	149
7.2.3 Calcul des normales aux triangles.....	149
7.3 Expérimentation sur des images réelles.....	151
7.3.1 Protocole de test.....	151
7.3.2 Résultats.....	153
7.4 Conclusion.....	155

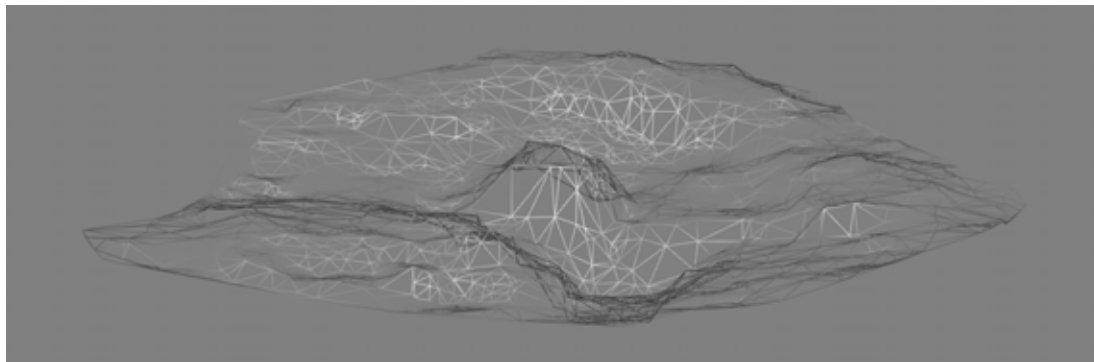
Ces travaux sont guidés par le cadre applicatif du suivi thérapeutique des escarres. Les mesures utiles au médecin (longueur, largeur, profondeur, volume de la plaie...) peuvent être déduites du modèle 3D, moyennant quelques traitements supplémentaires. En particulier, le volume est une valeur très importante pour le diagnostic médical, et nous avons cherché à connaître la précision des mesures volumétriques effectuées sur les modèles 3D issus de notre méthode de reconstruction. Ce chapitre commence par livrer la stratégie mise en œuvre, puis il décrit la méthode utilisée pour estimer le volume. Les tests effectués sur des images réelles sont ensuite présentés. Les résultats obtenus permettent de valider globalement notre chaîne de reconstruction.

7.1 Stratégie

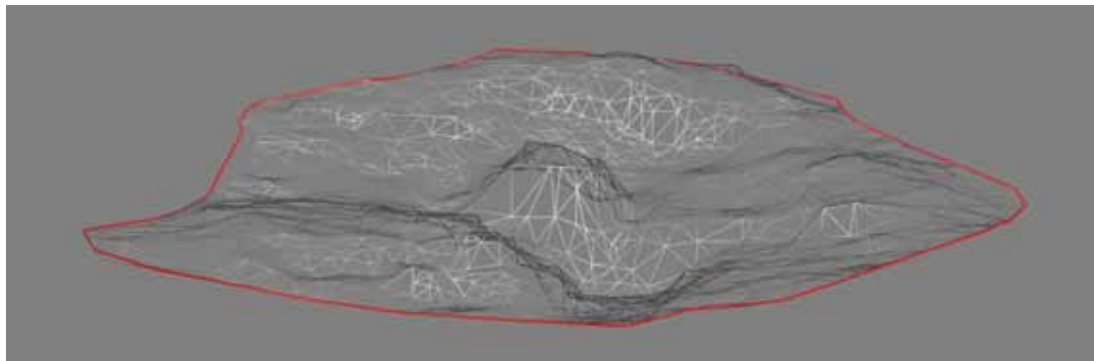
En sortie de la chaîne de reconstruction, nous disposons d'un modèle tridimensionnel de surface sous la forme d'un maillage de triangles. Les triangles sont obtenus par la triangulation de Delaunay appliquée à l'ensemble des points 2D de la première image, puis reportés sur les points 3D. La surface de l'objet est ainsi modélisée par un ensemble de facettes triangulaires (fig. 7.1a).

Le calcul d'un volume requiert en principe une surface fermée, qui n'est pas disponible sur ces données brutes. Dans le cas du suivi thérapeutique d'escarres, le volume intéressant est la cavité creusée par la plaie par rapport au niveau initial de la peau saine. Son estimation nécessite donc l'extrapolation d'une surface de « bouchage » virtuelle de la peau saine à partir des bords de la plaie. Pour arriver à ce résultat, trois étapes doivent être mises en œuvre. La première est l'extraction du contour de la plaie sur les vues stéréoscopiques, afin de reporter ce contour sur le modèle 3D (fig. 7.1b). La seconde étape est la construction de la surface de « bouchage » (fig. 7.1c et d). L'ultime étape concerne le calcul du volume compris entre les deux surfaces : celle de l'escarre reconstruite et celle de la peau saine extrapolée.

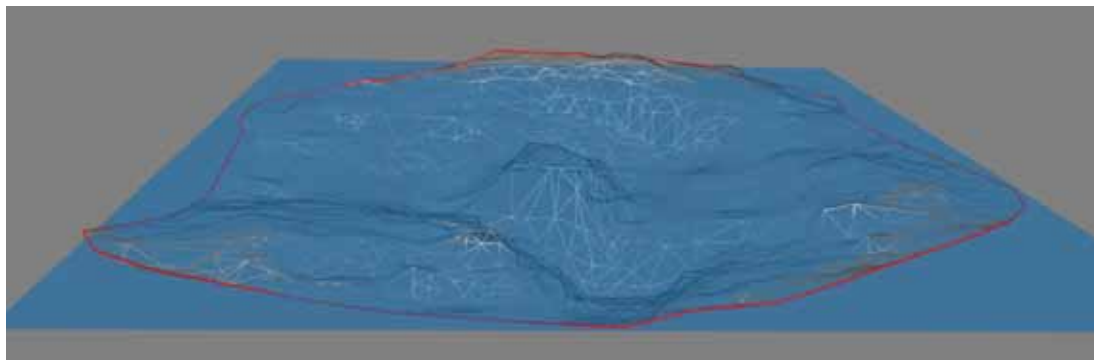
Afin de pouvoir évaluer notre méthode de calcul volumétrique, nous avons choisi en première approximation d'utiliser une surface de bouchage plane. Ce modèle simpliste pourra dans le futur être amélioré pour tenir compte de la courbure du corps. La définition du contour de la plaie reste un problème ouvert. Sa résolution s'appuiera sur les travaux de segmentation couleur et de classification des tissus en cours de développement dans une thèse conjointe. L'extraction des deux courbes appariées dans les images stéréoscopiques pourra être fortement contrainte par la connaissance des positions géométriques des vues issue de nos travaux. En attendant ces outils, nous avons délibérément choisi de simplifier ces deux premières étapes préalables au calcul du volume. Le plan est estimé par une minimisation aux moindres carrés effectuée sur les points du contour. Cette stratégie a permis de valider globalement notre chaîne de reconstruction en testant la méthode sur un objet de volume connu possédant une face plane.



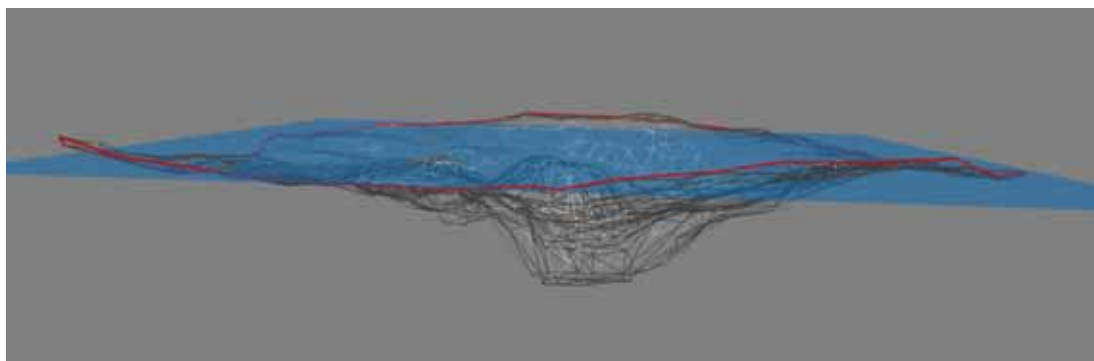
(a)



(b)



(c)



(d)

Figure 7.1 : Maillage triangulaire de la surface reconstruite sur un cas réel d'escarre située sur le sacrum (a). Contour de la plaie reporté sur le modèle 3D (b). Plan de bouchage ajusté sur la surface maillée (c) et (d). Le calcul de volume correspond à la sommation des volumes élémentaires sous chaque facette entre la surface maillée et le plan de référence.

Il est parfois possible, pour certaines escarres, que le tissu cicatriciel dépasse de la surface de la peau saine. On parle alors d'hyperbourgeonnement (fig. 7.2). Le volume de ces zones de l'escarre est également un indice important pour le diagnostic médical et ne doit pas être pris en compte dans le volume de l'escarre. Ainsi, il faut considérer deux volumes distinct, celui sous la surface interpolée de la peau saine, et celui au-dessus. Le principe du calcul reste le même, il s'agit de négliger tantôt les parties de l'escarre au dessus de la surface de la peau et tantôt en dessous.

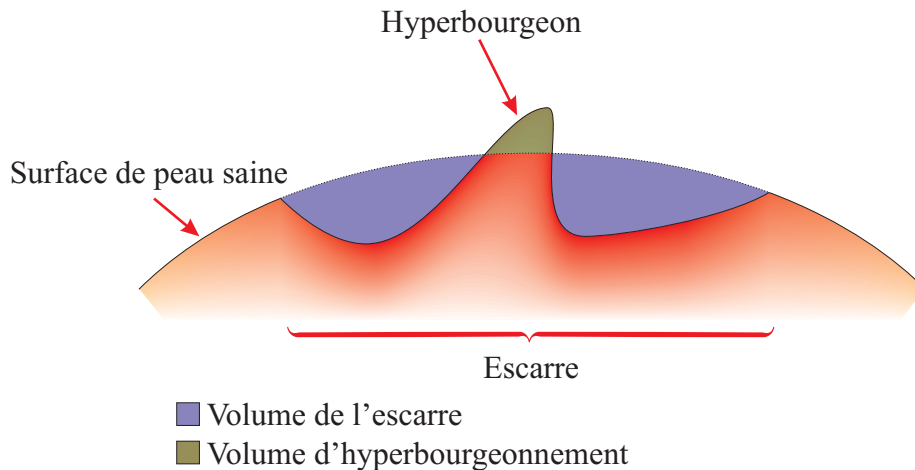


Figure 7.2 : Volume de l'escarre et volume d'hyperbourgeonnement.

7.2 Calcul du volume sous une surface

La méthode développée ci-après permet de calculer le volume compris entre une surface et un plan de référence. Le calcul du volume correspond à la sommation des volumes élémentaires sous chaque facette entre la surface maillée et le plan de référence (fig. 7.3). Le calcul d'un volume élémentaire est rappelé dans la section 7.2.1.

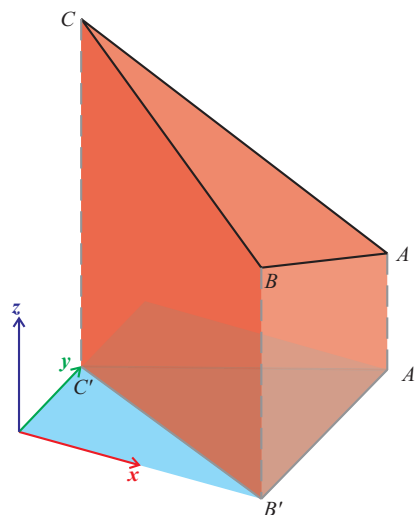


Figure 7.3 : Prisme correspondant au volume élémentaire entre une facette triangulaire (A,B,C) et sa projection orthogonale (A',B',C') sur le plan de référence ($z = 0$).

Sans nuire à la généralité du calcul, le plan de bouchage est défini par l'équation $z = 0$ et sert de plan de référence. Nous expliquerons dans la section 7.2.2 comment ramener n'importe quel plan à cette situation. Dans le cas d'une surface repliée sur elle-même, une projection sur le plan peut correspondre à plusieurs points de la surface. Sur l'exemple donné en figure 7.4, il faut ajouter les volumes élémentaires des triangles supérieurs et soustraire les volumes élémentaires des triangles inférieurs pour déterminer le volume de l'octaèdre. Le signe d'un volume élémentaire est défini par la normale, dont le calcul est présenté dans la section 7.2.3.

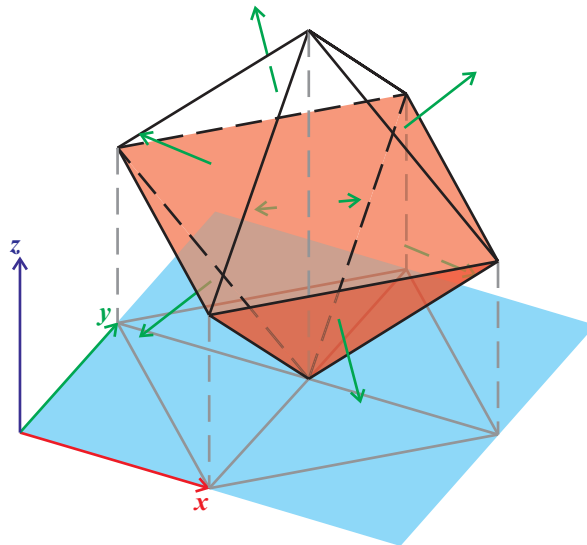


Figure 7.4 : Dans le cas d'une surface repliée sur elle-même, il faut utiliser les normales pour définir le signe des volumes élémentaires.

7.2.1 Calcul d'un volume élémentaire

Quatre cas de figure sont possibles pour le calcul du volume élémentaire sous un triangle :

- le triangle est entièrement au-dessus du plan de référence ;
- un sommet du triangle est sous le plan de référence ;
- deux sommets du triangle sont sous le plan de référence ;
- le triangle est entièrement au-dessous du plan de référence.

Le dernier cas ne mérite pas de développement puisqu'il correspond à un volume élémentaire nul. Les calculs des trois premiers cas sont détaillés ci-après en adoptant les notations suivantes : les sommets $A = (x_A, y_A, z_A)^T$, $B = (x_B, y_B, z_B)^T$ et $C = (x_C, y_C, z_C)^T$ du triangle sont définis de sorte que $z_A \leq z_B \leq z_C$. On définit également leurs projections respectives sur le plan $z = 0$ par $A' = (x_A, y_A, 0)^T$, $B' = (x_B, y_B, 0)^T$ et $C' = (x_C, y_C, 0)^T$.

7.2.1.1 Cas où le triangle est entièrement au dessus du plan de référence

Ce cas est représenté sur la figure 7.3. Dans ce cas trivial, le volume élémentaire est égal à l'aire du triangle projeté sur le plan $z=0$ multipliée par la hauteur moyenne du prisme. Ainsi, le volume est calculé par l'équation suivante :

$$v = \frac{1}{2} \left| x_C(y_B - y_A) + x_B(y_A - y_C) + x_A(y_C - y_B) \right| \frac{(z_A + z_B + z_C)}{3} \quad (7.1)$$

7.2.1.2 Cas où un sommet du triangle est sous le plan de référence

Ce cas est représenté sur la figure 7.5.

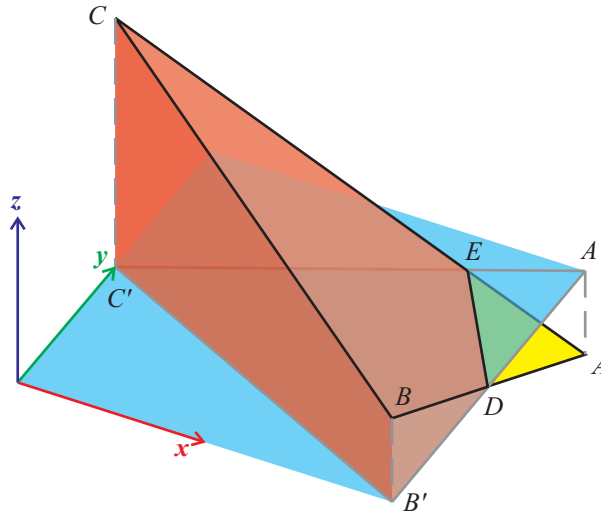


Figure 7.5 : Cas où un sommet du triangle est sous le plan de référence.

Dans ce cas, il ne faut pas tenir compte de la partie du triangle qui est sous le plan de référence. Si l'arête AC (respectivement AB) coupe le plan de référence en E (respectivement D), alors le volume élémentaire est égal au volume sous le quadrilatère $CBDE$. Pour calculer le volume, on se rapporte au cas de la sous-section précédente en coupant le quadrilatère en deux triangles CBE et EBD et en ajoutant les volumes élémentaires correspondants à chaque triangle.

Les points D et E sont définis par les relations suivantes :

$$\overrightarrow{A'E} = \frac{z_A}{z_C} \overrightarrow{EC'} \Rightarrow \overrightarrow{OE} = \overrightarrow{OA'} + \frac{z_A}{z_A - z_C} \overrightarrow{A'C'} \quad (7.2)$$

$$\overrightarrow{A'D} = \frac{z_A}{z_B} \overrightarrow{DB'} \Rightarrow \overrightarrow{OD} = \overrightarrow{OA'} + \frac{z_A}{z_A - z_B} \overrightarrow{A'B'} \quad (7.3)$$

Le volume se calcule alors par l'équation simplifiée suivante :

$$v = \frac{\alpha}{6} \left(\frac{z_C(z_C + z_B)}{z_C - z_A} - \frac{z_A z_B^2}{(z_A - z_B)(z_A - z_C)} \right) \quad (7.4)$$

avec $\alpha = |x_A(y_B - y_C) + x_B(y_C - y_A) + x_C(y_A - y_B)|$

7.2.1.3 Cas où deux sommets du triangle sont sous le plan de référence

Ce cas est représenté sur la figure 7.6.

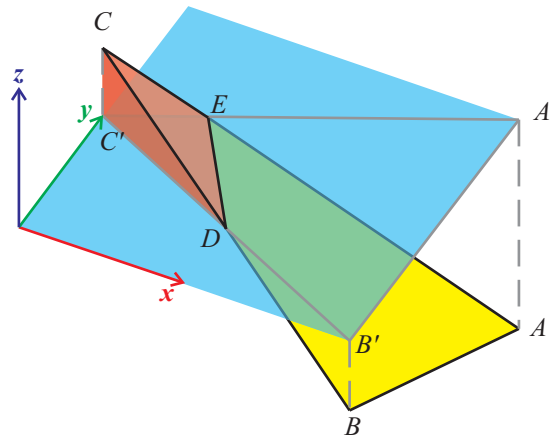


Figure 7.6 : Cas où deux sommets du triangle sont sous le plan de référence.

Comme dans le cas précédent, il ne faut pas tenir compte de la partie du triangle qui est sous le plan de référence. Ainsi, si la droite AC (respectivement BC) coupe le plan de référence en E (respectivement D), alors le volume élémentaire est égal au volume sous le triangle CDE. Les points D et E sont définis par les relations suivantes :

$$\overrightarrow{C'E} = \left| \frac{z_C}{z_A} \right| \overrightarrow{EA'} \Rightarrow \overrightarrow{OE} = \overrightarrow{OC'} + \frac{z_C}{z_C - z_A} \overrightarrow{C'A'} \quad (7.5)$$

$$\overrightarrow{C'D} = \left| \frac{z_C}{z_B} \right| \overrightarrow{DB'} \Rightarrow \overrightarrow{OD} = \overrightarrow{OC'} + \frac{z_C}{z_C - z_B} \overrightarrow{C'B'} \quad (7.6)$$

Le volume se calcule alors par l'équation simplifiée suivante :

$$v = \frac{\alpha z_C^3}{6(z_C - z_B)(z_C - z_A)} \quad (7.7)$$

avec $\alpha = |x_A(y_B - y_C) + x_B(y_C - y_A) + x_C(y_A - y_B)|$

7.2.2 Transformation du plan de référence

Les calculs de volumes élémentaires présentés dans la section précédente se réalisent entre les triangles et le plan $z = 0$. Dans le cas où la surface maillée de triangles est « bouchée » par un plan quelconque, il faut effectuer un changement de repère pour faire coïncider le plan de référence avec le plan $z = 0$. La transformation recherchée est rigide. Il est possible de déterminer cette transformation en décomposant le mouvement à réaliser.

Dans un premier temps le plan de référence peut être placé parallèlement au plan $z = 0$. Pour cela, il suffit d'appliquer une rotation selon l'axe x de sorte à rendre le plan parallèle à l'axe y , puis une rotation selon l'axe y pour le rendre parallèle à l'axe x . Le plan étant alors parallèle aux deux axes x et y , il est parallèle au plan $z = 0$. Ces deux rotations doivent être réalisées de sorte que le volume à calculer se trouve « au dessus » du plan de référence, c'est-à-dire vers les z positifs. On peut noter que la troisième rotation classique autour de l'axe z n'a pas d'importance, on la choisit nulle.

Dans un second temps, il faut appliquer une translation selon l'axe z pour placer le plan transformé exactement sur le plan $z = 0$. De même que la rotation autour de l'axe z , les translations selon x et y n'ont pas d'influence et on les choisit alors nulles. En définitive, la transformation se résume à l'application successive de deux rotations R_x puis R_y et une translation T_z .

7.2.2.1 Écriture analytique

En utilisant les coordonnées homogènes, cette transformation se définit par :

Soit $P = (a \ b \ c \ d)^T$ le plan de bouchage,

soit $P_{ref} = (0 \ 0 \ 1 \ 0)^T$ le plan de référence $z = 0$,

soient R_x et R_y respectivement les rotations d'angles θ et ϕ autour des axes x et y définies par :

$$R_x = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & \cos(\theta) & -\sin(\theta) & 0 \\ 0 & \sin(\theta) & \cos(\theta) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \text{ et } R_y = \begin{pmatrix} \cos(\phi) & 0 & \sin(\phi) & 0 \\ 0 & 1 & 0 & 0 \\ -\sin(\phi) & 0 & \cos(\phi) & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

soit T_z , la translation selon l'axe z définie par :

$$T_z = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

notons H la transformation rigide complète des points définie par :

$$H = T_z R_y R_x \quad (7.8)$$

Par définition, les équations suivantes sont respectées :

$$H^{-T} P \propto P_{ref} \quad (7.9)$$

$$H^T P_{ref} \propto P \quad (7.10)$$

Des équations 7.8 et 7.10, on déduit l'équation suivante :

$$\begin{pmatrix} -\sin(\phi) \\ \cos(\phi) \sin(\theta) \\ \cos(\phi) \cos(\theta) \\ t_z \end{pmatrix} = k \begin{pmatrix} a \\ b \\ c \\ d \end{pmatrix} \quad (7.11)$$

Ce système comporte 4 équations et 4 inconnues (θ , ϕ , t_z et k un réel non nul) dont les solutions sont :

$$\begin{aligned} t_z &= k d \\ \phi &= -\arcsin(k a) \\ \theta &= \arcsin\left(\frac{k b}{\sqrt{1-k^2 a^2}}\right) \\ k &= \pm \frac{1}{\sqrt{a^2 + b^2 + c^2}} \end{aligned} \quad (7.12)$$

Le signe de k est indéterminé et choisi arbitrairement positif. Recalculons le produit $H^T P_{ref} / k$ avec les valeurs de θ , ϕ , t_z et k calculées par la résolution du système, afin de connaître l'influence de ce choix arbitraire :

$$\frac{1}{k} H^T P_{ref} = \begin{pmatrix} a \\ b \\ |c| \\ d \end{pmatrix} \quad (7.13)$$

Donc pour que l'équation 7.10 soit respectée avec le choix $k > 0$, il faut que c soit positif. Cette condition ne pose pas de problème grâce aux coordonnées homogènes. Il est possible de multiplier par -1 le vecteur P de définition du plan de bouchage sans en modifier la signification géométrique. Dans ce cas, avec la condition $c > 0$, nous obtenons les paramètres recherchés de la transformation rigide :

$$\begin{aligned}
 t_z &= \frac{d}{\sqrt{a^2+b^2+c^2}} \\
 \phi &= -\arcsin\left(\frac{a}{\sqrt{a^2+b^2+c^2}}\right) \\
 \theta &= \arcsin\left(\frac{b}{\sqrt{b^2+c^2}}\right)
 \end{aligned} \tag{7.14}$$

Après simplification il vient :

$$H = \frac{1}{\sqrt{a^2+b^2+c^2}} \begin{pmatrix} \sqrt{b^2+c^2} & -\frac{ab}{\sqrt{b^2+c^2}} & -\frac{ac}{\sqrt{b^2+c^2}} & 0 \\ 0 & \frac{c\sqrt{a^2+b^2+c^2}}{\sqrt{b^2+c^2}} & -\frac{b\sqrt{a^2+b^2+c^2}}{\sqrt{b^2+c^2}} & 0 \\ a & b & c & d \\ 0 & 0 & 0 & \sqrt{a^2+b^2+c^2} \end{pmatrix} \tag{7.15}$$

On remarque alors que cette matrice est indéterminée lorsque $b=c=0$, c'est-à-dire lorsque le plan de bouchage est d'équation $x = \text{constante}$. Dans ce cas, il suffit de fixer $\theta=0$ et $\phi=\pi/2$ pour ramener le plan de bouchage au plan $z=0$. Ainsi, H a la forme suivante :

$$H = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{7.16}$$

Dans ce cas, l'équation 7.11 équivaut à :

$$t_z = kd \quad \text{et} \quad k = \frac{-1}{a} \tag{7.17}$$

Donc en définitive, H s'écrit sous la forme suivante :

$$H = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 0 & \frac{-d}{a} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{7.18}$$

Cependant, les équations 7.15 et 7.18 ne permettent pas d'assurer que la surface se trouve du côté des z positifs par rapport au plan de référence. Or, le calcul des volumes élémentaires selon la méthode exposée à la section 7.2.1 repose sur cette hypothèse. Pour assurer cette condition, on teste la transformation H sur un point de l'espace dont on sait que l'image après transformation se situe du côté des z positifs (ou négatifs). Si la transformation par H de ce point donne une composante z négative (respectivement positive) alors il faut « retourner » le

plan de référence. Ce retournement peut s'effectuer en multipliant à gauche H par une matrice de rotation d'angle π autour de l'axe x , donnée par :

$$R_{x,\pi} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (7.19)$$

7.2.2.2 Algorithme

Les calculs précédents ont été implémentés en utilisant le pseudo code donné par l'algorithme de la figure 7.7. Une fois la matrice H calculée, la transformation est appliquée à tous les points 3D.

• **Entrées de l'algorithme**

- $P = (a \ b \ c \ d)^T$: Plan de bouchage en coordonnées homogènes,
- $M = (m_1 \ m_2 \ m_3 \ m_4)^T$: Coordonnées homogènes du point définissant le côté positif du plan.

• **Début**

Si $c < 0$ alors

Affecter $-P$ à P

Fin si

Si $b = c = 0$ alors

Affecter équation 7.18 à H

Sinon

Affecter équation 7.15 à H

Fin si

Calculer $N = (n_1 \ n_2 \ n_3 \ n_4)^T = HM$

Si $\text{signe}(n_3) \neq \text{signe}(n_4)$ alors

Affecter $R_{x,\pi}H$ à H

Fin si

• **Fin**

Figure 7.7 : Algorithme de calcul de la matrice H .

7.2.3 Calcul des normales aux triangles

La normale d'une surface en un point est le vecteur unitaire orthogonal à la surface en ce point et dont le sens est dirigé vers l'extérieur de l'objet défini par la surface. Dans le cas d'un objet défini par une surface maillée de triangles, on a donc autant de normales que de triangles.

La direction d'une normale à un plan est calculable par le produit vectoriel de deux vecteurs du plan non colinéaires. Ainsi, dans le cas d'un triangle ABC , la normale au triangle peut être donnée par l'expression suivante :

$$\vec{n} = \pm \frac{\vec{AB} \times \vec{AC}}{\|\vec{AB} \times \vec{AC}\|} \quad (7.20)$$

où \vec{n} est la normale au triangle et \times désigne le produit vectoriel. Le sens de \vec{n} est défini de façon géométrique, de sorte que le vecteur soit dirigé vers l'extérieur de l'objet. Dans le cas du calcul du volume de l'objet, seul le sens et la direction de la normale sont utiles. Le calcul est allégé en ne normalisant pas le vecteur \vec{n} . La difficulté réside alors dans la détermination du sens du vecteur pointant vers le côté extérieur de l'objet.

Nous avons expliqué précédemment que le maillage de la surface 3D est basé sur une triangulation de Delaunay effectuée sur les projections dans la première image. La définition des normales des triangles s'appuie également sur ce maillage 2D.

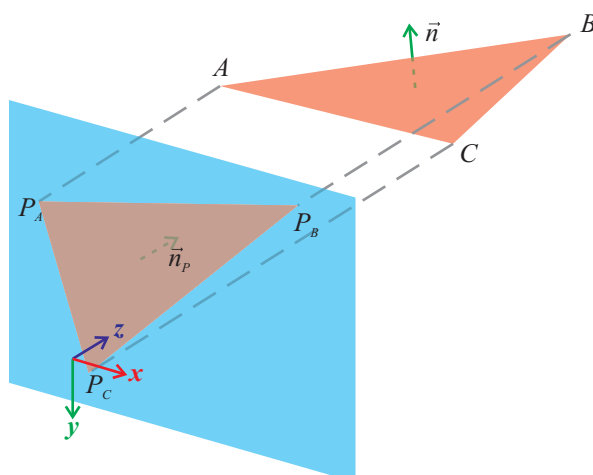


Figure 7.8 : Les normales aux facettes sont définies en s'appuyant sur la triangulation initiale effectuée dans la première image.

Considérons un triangle $P_B P_C$ sur le plan de la première image et son correspondant 3D ABC (figure 7.8). Le repère étant lié à la première image, on sait que toutes les normales des triangles 2D doivent être selon les $z > 0$, c'est-à-dire vers l'extérieur du volume à analyser. Les normales sont donc dirigées vers l'intérieur du corps du patient. La normale \vec{n}_p au triangle $P_A P_B P_C$ est donc calculée par l'équation suivante :

$$\vec{n}_p = \delta \overrightarrow{P_A P_B} \times \overrightarrow{P_A P_C} \quad (7.21)$$

où $\delta = \pm 1$ de sorte que $\vec{n}_p = k \vec{z}$ avec $k > 0$.

Une fois la valeur de δ déterminée, le calcul est reporté sur les points 3D A , B et C correspondants respectifs des points P_A , P_B et P_C . Ainsi, la normale \vec{n} au triangle ABC est calculée par l'équation :

$$\vec{n} = \delta \overrightarrow{AB} \times \overrightarrow{AC} \quad (7.22)$$

7.3 Expérimentation sur des images réelles

7.3.1 Protocole de test

Afin de quantifier la précision sur le calcul du volume obtenu par notre chaîne de reconstruction 3D, nous avons effectué des mesures sur un objet de volume étalonné, possédant une face plane. Il s'agit d'un bloc de liège dans lequel une cavité a été grossièrement creusée (fig. 7.9). Le volume correspondant à la quantité de liège manquante est estimé par une double pesée de l'objet, à vide puis, après un remplissage soigneux de la cavité en eau. Vingt doubles pesées ont été réalisées à l'aide d'une balance précise au centième de gramme. Les résultats de ces pesées sont donnés dans le tableau 7.1. La moyenne de ces pesées nous donne une très bonne estimation du volume de référence, après conversion par la masse volumique de l'eau, soit $39,12 \pm 1,03 \text{ cm}^3$.



Figure 7.9 : Bloc de liège utilisé pour évaluer la précision volumétrique.

Nous avons ensuite réalisé plusieurs mesures de volume par notre méthode de reconstruction 3D avec différents couples d'images. Le plan de bouchage, correspondant à la face plane du bloc de liège, est estimé à partir des points du périmètre de la surface reconstruite, c'est-à-dire les points périphériques du maillage (fig. 7.10). Afin de minimiser l'influence du bruit dans les données, l'estimation du plan se fait par minimisation aux moindres carrés du problème suivant :

$$\sum_i (M_i^T P)^2 \quad (7.23)$$

où les points M_i sont les points du périmètre de la cavité, et P les coordonnées du plan de bouchage.

L'algorithme d'estimation du volume nécessite également un point connu pour lever l'ambiguïté sur le côté du plan de référence qui doit être dirigé vers les z positifs. Ce point est alors choisi comme étant l'isobarycentre de tous les points du maillage.

Tableau 7.1 : Estimation du volume de référence par 20 pesées successives de la masse d'eau contenue dans la cavité.

	Poids à vide en g	Poids plein en g	Volume correspondant en cm^3
	74,65	114,08	39,43
	74,76	113,20	38,44
	74,80	114,48	39,68
	74,89	113,90	39,01
	75,07	113,95	38,88
	74,93	115,25	40,32
	75,10	116,08	40,98
	75,22	116,24	41,02
	75,16	114,63	39,47
	75,30	114,97	39,67
	75,38	112,95	37,57
	76,05	114,59	38,54
	76,48	114,02	37,54
	76,31	113,61	37,30
	76,78	115,73	38,95
	76,79	116,99	40,20
	76,89	115,37	38,48
	77,07	116,02	38,95
	77,22	116,01	38,79
	77,61	116,81	39,20
Moyenne	75,82	114,94	39,12
Écart-type	0,97	1,19	1,03

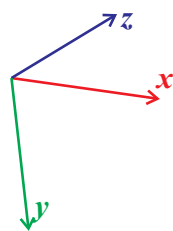
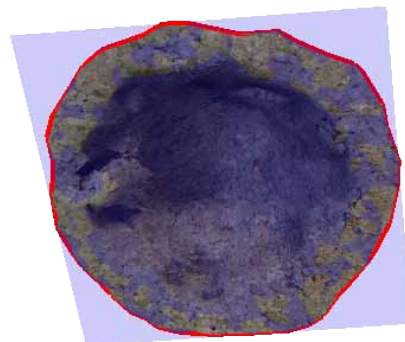


Figure 7.10 : Estimation du plan de référence.

7.3.2 Résultats

Pour cette expérimentation, 16 couples d'images ont été utilisés pour les reconstructions 3D. Des exemples de ces couples d'images sont montrés sur la figure 7.11. Les images ont été acquises à l'aide d'un appareil photo Canon IXUS 400, au niveau de zoom le plus faible et en utilisant le mode macro. Leur résolution est de 1024×768 pixels et les prises de vue ont été réalisées à main levée en respectant la configuration optimale définie au chapitre 6. Pour chaque couple, 40 reconstructions 3D sont réalisées afin d'obtenir une valeur de volume moyenne, s'affranchissant de l'aspect stochastique propre à l'algorithme d'estimation de la matrice fondamentale. Ainsi, un total de 640 reconstructions 3D sont exploitées pour cette expérience. Le tableau 7.2 et la figure 7.12 montrent les résultats obtenus. L'erreur relative exprimée en pourcentage est définie par la formule suivante :

$$e_r = \frac{v_{ref} - v}{v_{ref}} \quad (7.24)$$

où v_{ref} et v sont respectivement le volume de référence et le volume moyen mesuré.

Tableau 7.2 : Moyennes et écarts-types des volumes mesurés sur 40 reconstructions pour chaque paire d'images.

Couple d'images	Volume moyen en cm ³	Écart-type en cm ³	Erreur relative par rapport à la vérité terrain
1	39,09	0,41	0,09 %
2	39,69	0,32	-1,46 %
3	39,11	0,58	0,03 %
4	36,97	0,48	5,51 %
5	38,96	0,41	0,40 %
6	38,45	0,76	1,73 %
7	41,17	0,21	-5,24 %
8	38,80	0,36	0,81 %
9	40,11	0,26	-2,54 %
10	41,24	0,90	-5,43 %
11	38,16	0,58	2,45 %
12	38,15	0,65	2,48 %
13	38,07	0,17	2,68 %
14	40,02	1,25	-2,29 %
15	38,05	0,57	2,73 %
16	37,19	0,49	4,93 %
Moyenne	38,95	0,52	0,43 %
Écart-type	1,25	0,28	3,19 %

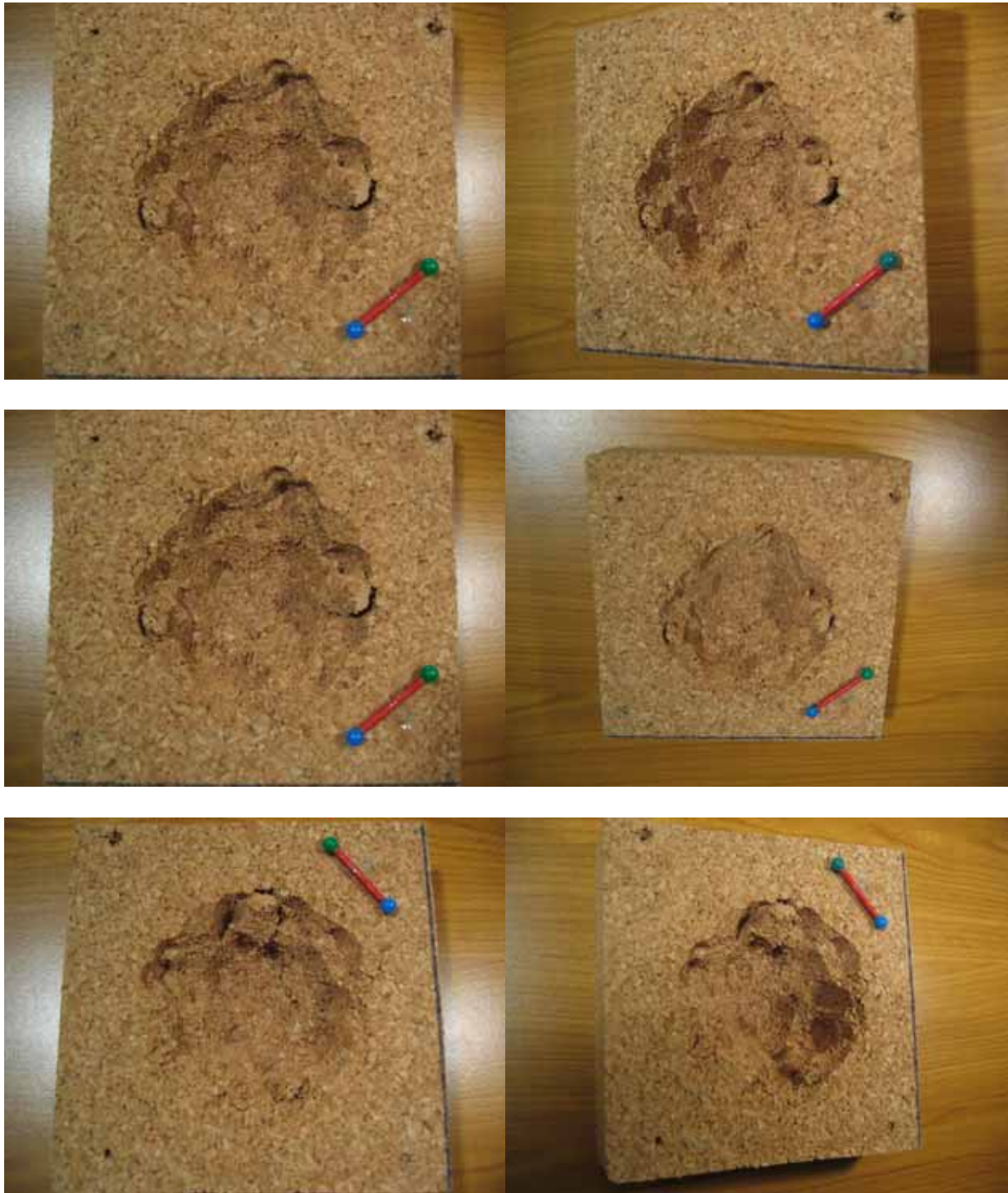


Figure 7.11 : Représentation de 3 des 16 couples d'images ayant servi à l'expérimentation.

Ces résultats montrent que l'écart à la référence sur l'estimation du volume est en moyenne assez faible, de l'ordre de 0,43 %. De plus, la mesure du volume est précise, l'écart-type des mesures sur les 16 couples d'images étant de $1,25 \text{ cm}^3$, soit 3,19 % de la moyenne. Ces résultats présentent donc de nettes améliorations par rapport aux résultats préliminaires que nous avons publié en 2005 [Albouy, 2005]. Nous obtenions alors une précision de l'ordre de 13 % et un écart à la référence de 3,5 %. Cette amélioration est due aux progrès réalisés sur la mise en correspondance des images et l'estimation de la distance focale. En effet, lors de

ces précédents travaux, la mise en correspondance n'était réalisée que par des méthodes simples de corrélation, ne permettant pas de grands changements de points de vue, indispensables pour une bonne triangulation. Concernant la distance focale, son estimation dépend de la matrice fondamentale qui est estimée par une méthode de type RANSAC. Pour le même ensemble de correspondances, l'aspect stochastique de l'algorithme d'estimation n'assure pas d'obtenir exactement le même résultat. Afin de stabiliser tout le processus d'auto-calibrage de l'estimation de la matrice fondamentale à la détermination de la distance focale, le calcul est réitéré 10 fois et la distance focale moyenne est conservée.

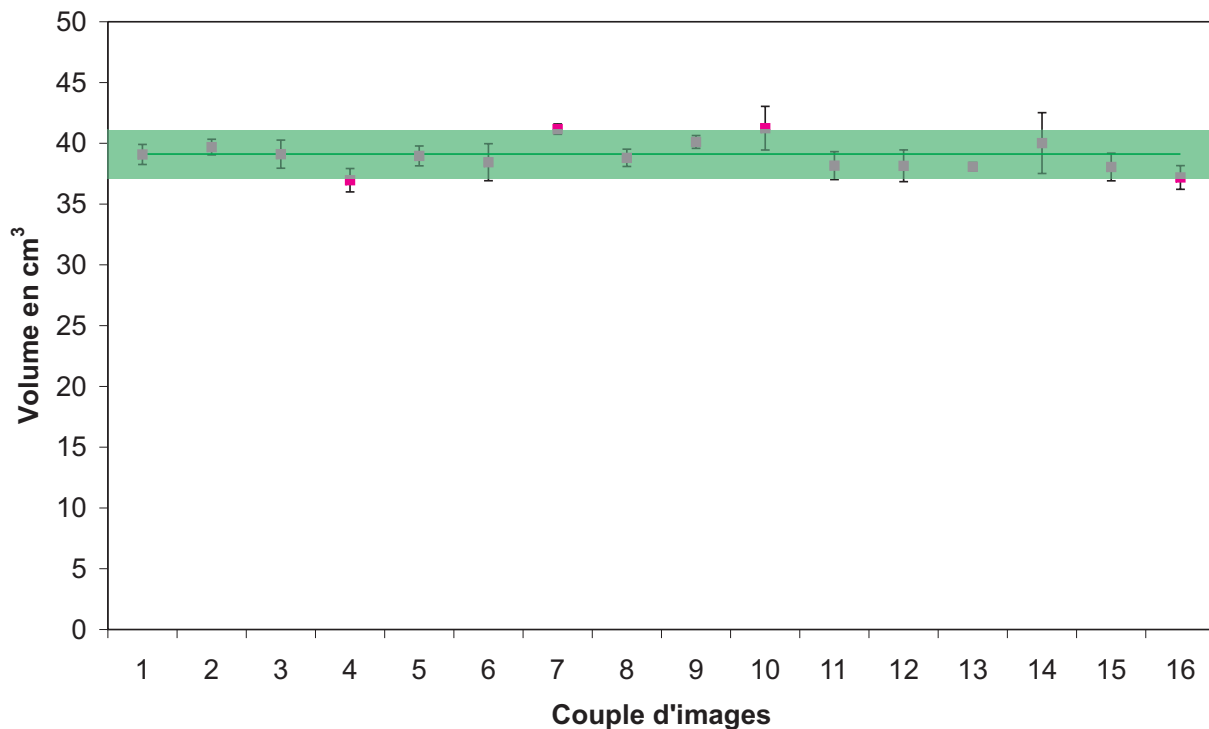


Figure 7.12 : Intervalles de confiance à 95 % sur les 40 mesures volumétriques effectuées pour chacune des 16 paires d'images. La bande colorée correspond au même intervalle sur le volume de référence obtenu par les 20 pesées.

7.4 Conclusion

Après avoir présenté la méthode de calcul du volume d'une cavité, nous avons mené une expérimentation dans le but de valider les mesures réalisables avec notre chaîne de reconstruction 3D. Seize modèles ont été reconstruit à partir de 16 couples de photographies d'une cavité creusée dans un bloc de liège. Le volume de cette cavité est calculé par notre méthode sur les modèles et comparé à la valeur étalonnée. Les résultats obtenus montrent un biais moyen de 0,43 % et une stabilité exprimée par l'écart-type des mesures de 3,19 % par rapport au volume moyen.

Ces résultats sont donc compatibles avec le cahier des charges de notre application qui nous impose une erreur inférieure à 5 %. Cependant, le volume réel d'une escarre ne peut pas

être calculé en bouchant l'escarre avec un plan. En effet, le volume d'une escarre doit être estimé par rapport à la surface de la peau saine, qui ne forme pas obligatoirement un plan, notamment sur des régions anatomiques galbées comme le talon. Il faut alors refermer le volume de l'escarre par une surface respectant la continuité des tissus. Dans un second temps, cette surface de raccordement devrait aussi respecter l'intégrité physique du membre avant lésion. Des modèles de surface plus réalistes devraient donc être appliqués pour le bouchage. Sachant que pour le médecin, c'est l'évolution temporelle du volume qui prime plutôt que le volume lui-même, un bouchage approximatif est suffisant dans la mesure où le biais de mesure reste constant durant la cicatrisation, qui va contracter le contour de l'escarre.

Conclusion et perspectives

Cette thèse est consacrée au développement d'une chaîne complète de reconstruction 3D à partir de deux vues non calibrées d'un objet, avec pour application le relevé volumétrique d'escarres. La stéréovision non calibrée est une approche inédite dans l'aide au suivi thérapeutique des escarres. De plus, elle n'est classiquement pas abordée sous un aspect métrologique avec une approche d'évaluation quantitative de la qualité de la reconstruction. Nos travaux apportent également une solution plus générale de mise en correspondance semi-dense entre deux images de point de vue éloignés. Nous synthétisons ci-après les principales contributions apportées.

Bilan : une chaîne complète pour la reconstruction 3D

Nous résumons ici les principaux points que nous avons abordés dans ce mémoire ainsi que les contributions apportées.

Une approche d'évaluation quantitative et globale

Tout au long de nos travaux, les différentes méthodes ont été étudiées et testées dans le but de minimiser l'erreur finale sur la reconstruction 3D. Tout d'abord, l'étape au cœur de la chaîne de traitements, l'estimation de la matrice fondamentale, a fait l'objet de simulations dans le chapitre 2, visant à déterminer l'algorithme d'estimation optimal au sens de l'erreur 3D. Habituellement, ces algorithmes sont validés par l'erreur résiduelle sur la matrice fondamentale, une information 2D qui ne reflète pas l'erreur 3D finale, comme nous l'avons montré au chapitre 5. Il en résulte que l'algorithme optimal dans notre contexte est la méthode des *Least Median of Squares* (LMedS) suivi d'un raffinement par les moindres carrés orthogonaux. Sur les données de synthèse, la combinaison de ces deux méthodes permet d'obtenir une erreur 3D inférieure à 1,5 % pour des niveaux réalistes de bruit dans les images. Les résultats de ces simulations ont été corroborés par les comparaisons des modèles 3D obtenus à partir d'images réelles avec une vérité terrain issue de scanners industriels, décrites au chapitre 5. L'erreur de localisation des points 3D est d'environ 1,3 % grâce notamment à

l'implémentation d'une méthode de mise en correspondance semi-dense de vues éloignées. Enfin, nous avons recherché la configuration optimale des deux vues qui minimise l'erreur 3D. Le protocole d'acquisition des images est simple : s'assurer d'un angle de vergence de 25° avec un rapport de distances des prises de vue d'environ 1,2. L'optimum de la configuration n'est pas fortement marqué, ce qui laisse une liberté relative pour les prises de vue. Au final, le chapitre 7 montre que l'erreur sur le calcul du volume est de 0,43 % en moyenne avec un écart-type de 3,19 %, prouvant que la méthode est suffisamment précise et stable pour notre application.

Une mise en correspondance semi-dense de vues éloignées

Un autre apport de cette thèse a été la mise en place d'un processus de mise en correspondance semi-dense à partir de vues éloignées. La méthode est itérative, initialisée par un ensemble d'appariements robustes mais en nombre limité. Chaque itération permet d'affiner un peu plus cet ensemble d'appariements en appliquant une contrainte épipolaire de plus en plus serrée. La géométrie épipolaire initiale estimée à partir de quelques appariements initiaux est approximative. Une largeur de bande de recherche importante est alors nécessaire autour des droites épipolaires pour sélectionner de nouveaux appariements. Ceux-ci vont permettre de réestimer la matrice fondamentale de façon plus robuste à l'itération suivante, lui donnant ainsi une estimation plus rigoureuse. Ainsi, au fil des itérations, la géométrie épipolaire s'affine et la largeur de la bande de recherche peut être diminuée, permettant de lever l'ambiguïté de certains appariements nouvellement sélectionnés. Ceci a pour effet d'augmenter le nombre d'appariements au fur et à mesure des itérations. À la fin du processus, une mise en correspondance par corrélation affine à l'intérieur de triangles homologues permet de « booster » le nombre d'appariements. Ainsi, en 2 à 3 minutes sur un ordinateur cadencé à 3 GHz, environ 3 000 points sont appariés sur des images de résolution 1024×768 pixels.

Une nouvelle approche dans l'examen clinique des escarres

Cette thèse ouvre la voie d'une nouvelle technique pour le suivi thérapeutique des escarres et la mise au point de nouveaux traitements lors d'études cliniques. Elle offre de nombreux avantages par rapport aux méthodes existantes. Tout d'abord, l'absence de calibrage préalable du système rend la méthode d'acquisition très simple, ce qui est un atout majeur comparé aux autres applications de la vision par ordinateur aux escarres. Contrairement aux systèmes calibrés, la grande flexibilité de prise des vues permet d'observer des plaies de taille très différente et d'accéder plus facilement à des zones anatomiques souvent délicates en préservant le confort du patient. Ensuite, l'utilisation d'appareils photos numériques standards permet une large diffusion grâce à un coût modique, améliorant la qualité des soins en centre hospitalier et rendant envisageable les soins à domicile. Les photographies peuvent être simplement envoyées à une équipe médicale par internet. Les premiers résultats obtenus par le biais de cette thèse confirment que notre approche est parfaitement ciblée pour des examens cliniques lors du suivi thérapeutique d'escarres.

Perspectives

Nous voyons de nombreuses perspectives à ces travaux.

Extension à 3 vues et plus

L'utilisation de seulement deux vues peut parfois être problématique lorsque le relief de l'escarre est particulièrement prononcé. Des problèmes d'occultation sont alors possibles, formant des zones fortement erronées sur le modèle 3D. L'utilisation de vues supplémentaires permettrait d'accéder à ces zones.

Dans un premier temps, les vues supplémentaires peuvent être utilisées deux à deux en mettant en œuvre les mêmes méthodes que celles développées au cours de cette thèse. Chaque nouvelle paire de vues offrant un nuage de points 3D qui viendrait compléter le nuage déjà reconstruit. Pour calculer le volume de l'escarre, il serait nécessaire de faire évoluer la technique de maillage proposée dans ce mémoire limitée à seulement deux vues.

Dans un second temps, la série de vues pourrait être traitée trois par trois en utilisant le tenseur trifocal, équivalent de la matrice fondamentale dans le cas de trois vues. La contrainte apportée par le tenseur trifocal est plus forte que la contrainte épipolaire car deux points appariés dans deux images définissent leur correspondant dans la troisième vue. Cela permettrait d'avoir une mise en correspondance plus robuste encore.

Vers un calcul de volume plus réaliste

Dans l'état actuel des travaux, le volume lacunaire de l'escarre est calculé avec une surface de bouchage plane. Or la plupart des zones anatomiques couramment atteintes d'escarres ont des courbures parfois prononcées. Il serait alors intéressant de boucher l'escarre en interpolant la surface de la peau saine à partir des bords de l'escarre. Cette interpolation doit se faire de façon à conserver la continuité de la direction de la peau saine en limite de l'escarre. En outre, il est envisageable de la réaliser en s'appuyant sur des modèles anatomiques en fonction de la partie du corps lésée. En effet, un talon n'a pas la même forme qu'un sacrum.

Vers une interaction 2D – 3D

Le projet ESCALE mène de front deux aspects de l'examen clinique des escarres. L'aspect 3D qui a été développé dans cette thèse pour mesurer les dimensions d'une escarre, et l'aspect 2D dont l'objectif est de classifier automatiquement les différents tissus composants l'escarre par la couleur. Ces deux approches peuvent mutuellement s'aider. Les informations

tridimensionnelles peuvent lever les ambiguïtés de classification grâce à l'utilisation de plusieurs vues. A l'inverse, la segmentation 2D des images peut fournir de nombreux points caractéristiques le long des contours des tissus facilement appariables, renforçant ainsi l'étape de mise en correspondance et par conséquent le modèle 3D.

ANNEXES

Annexe A

Résultats du calibrage des appareils photos

Cette annexe regroupe les résultats complets du calibrage de dix appareils photos numériques, présenté au chapitre 4.

Sommaire

A.1 Position du point principal.....	165
A.2 Distance focale.....	171
A.3 Distorsions optiques.....	175

Le calibrage présenté au chapitre 4 porte sur la détermination des paramètres intrinsèques des appareils photos, ainsi que sur quatre paramètres caractérisant les distorsions optiques des objectifs. Tous ces paramètres ont été présentés au chapitre 4. Cette annexe regroupe les résultats de tous les appareils photos analysés. Le tableau A.1 rappelle les modèles et les principales caractéristiques des appareils photos testés.

Tableau A.1 : Modèles et principales caractéristiques des appareils photos numériques analysés

Modèle	Diagonale du capteur en pouces	Résolution des images utilisées	Niveaux de zoom optique
Canon EOS 350D	1,05	1728×1152	2×
Canon IXUS 400	0,56	1600×1200	1× - 3×
Canon PowerShot A95	0,56	1600×1200	1× - 3×
Canon PowerShot S2 IS	0,40	1600×1200	1× - 12×
Kyocera Finecam S5R	0,56	1600×1200	1× - 3×
Nikon Coolpix S1	0,40	2048×1536	1× - 3×
Panasonic NV-MX 500 EG	inconnue	640×480	1× - 12×
Sony DSC-H1	0,40	2048×1536	1× - 12×
Sony DSC-P52	0,37	1632×1224	1× - 2×
Sony DSC-P92	0,56	2048×1536	1× - 3×

Trois types d'informations ont été relevés grâce au calibrage. Premièrement, nous avons suivi la position du point principal en fonction du niveau de zoom de l'appareil photo. Deuxièmement, la distance focale exprimée en millimètres fournie par les appareils photos et la distance focale exprimée en pixels obtenue par calibrage, ont été analysées dans l'espoir de trouver une relation simple entre les deux valeurs. Troisièmement, les distorsions optiques ont été observées afin de s'assurer que les perturbations qu'elles causent à la reconstruction 3D, sont inférieures à la limite de 5% d'erreur tolérée par notre application. Les trois prochaines sections présentent ces trois types de résultats pour chaque appareil photo.

A.1 Position du point principal

Les figures A.1 à A.10, présentent la position du point principal de chaque appareil photo en fonction du niveau de zoom de l'objectif.

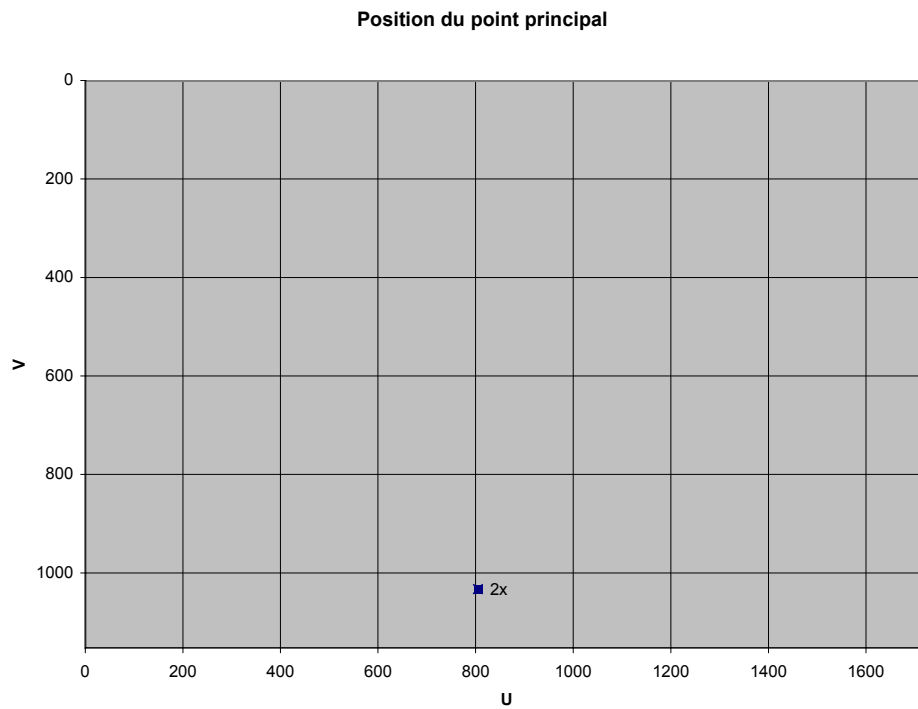


Figure A.1 : Position du point principal en fonction du niveau de zoom pour le Canon EOS 350D.

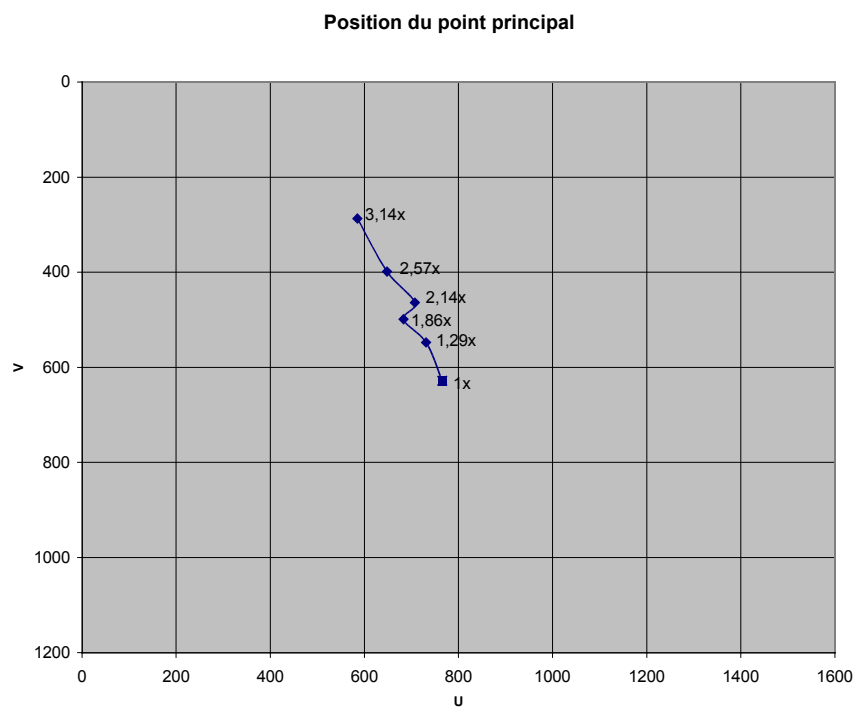


Figure A.2 : Position du point principal en fonction du niveau de zoom pour le Canon IXUS 400.

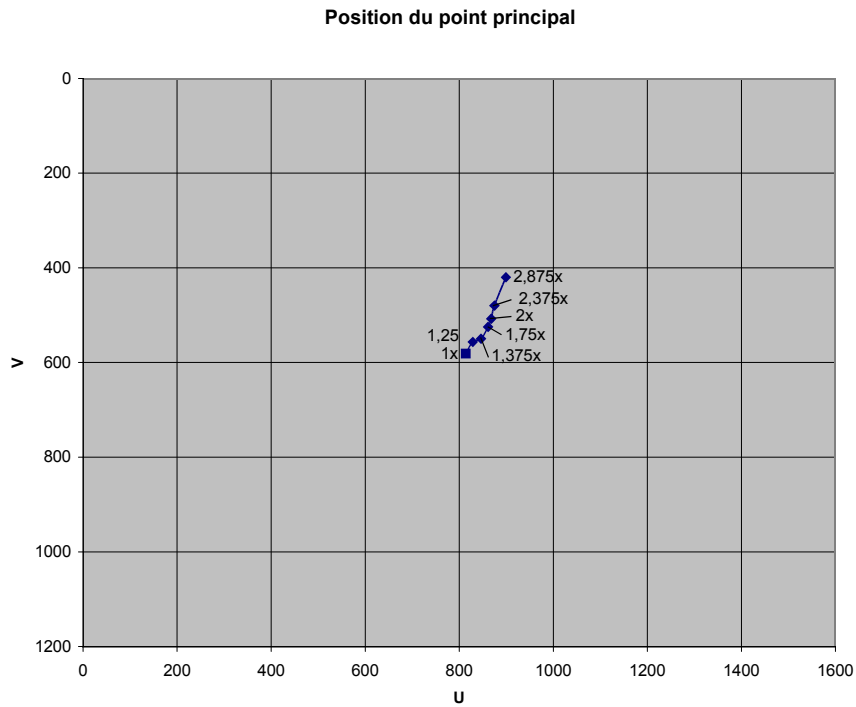


Figure A.3 : Position du point principal en fonction du niveau de zoom pour le Canon PowerShot A95.

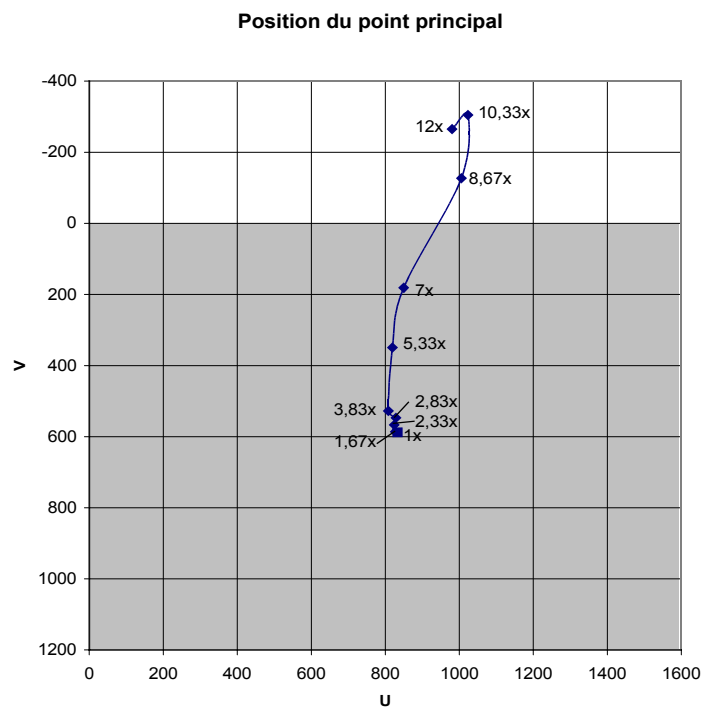


Figure A.4 : Position du point principal en fonction du niveau de zoom pour le Canon PowerShot S2 IS.

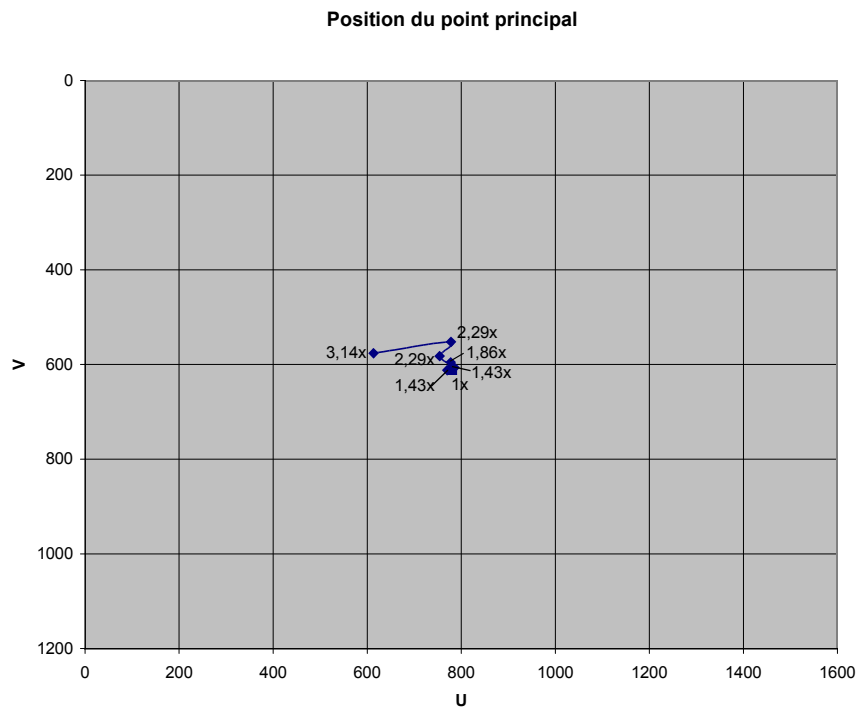


Figure A.5 : Position du point principal en fonction du niveau de zoom pour le Kyocera Finecam S5R.

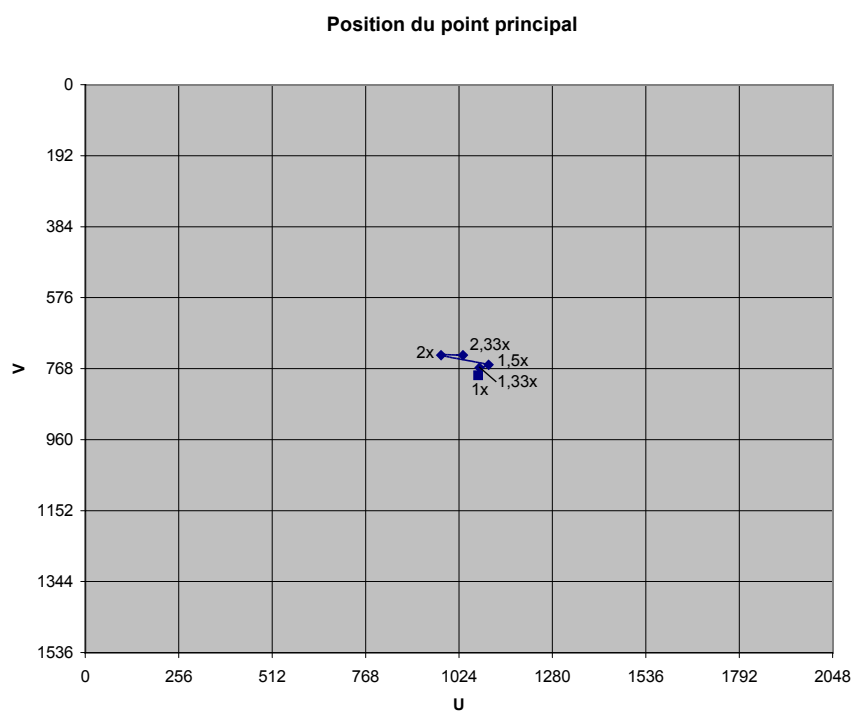


Figure A.6 : Position du point principal en fonction du niveau de zoom pour le Nikon Coolpix S1.

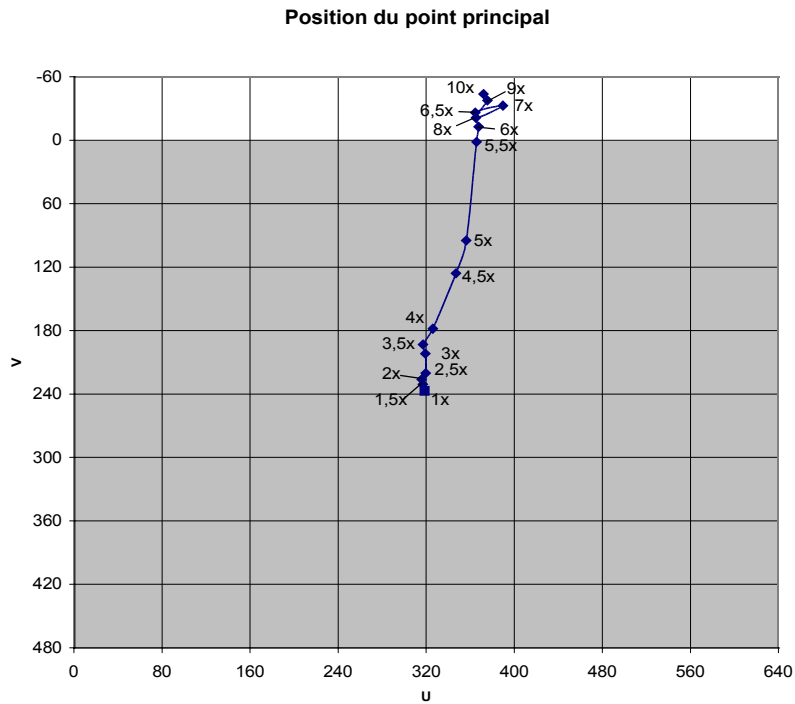


Figure A.7 : Position du point principal en fonction du niveau de zoom pour le Panasonic NV-MX 500 EG.

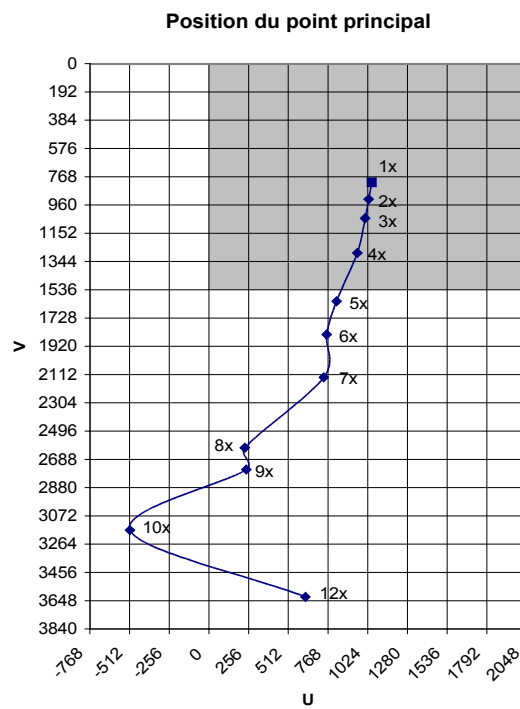


Figure A.8 : Position du point principal en fonction du niveau de zoom pour le Sony DSC-H1.

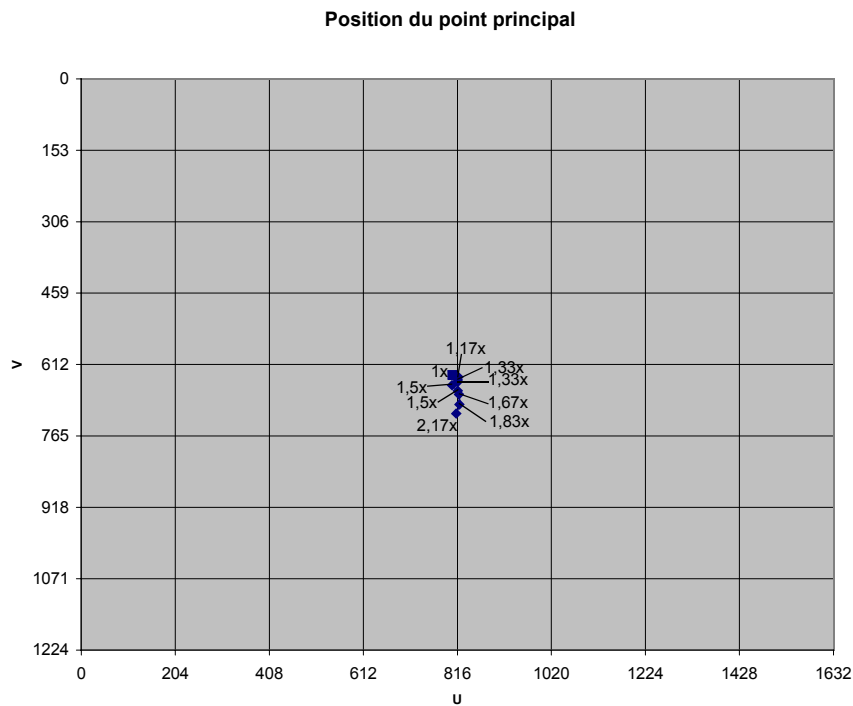


Figure A.9 : Position du point principal en fonction du niveau de zoom pour le Sony DSC-P52.

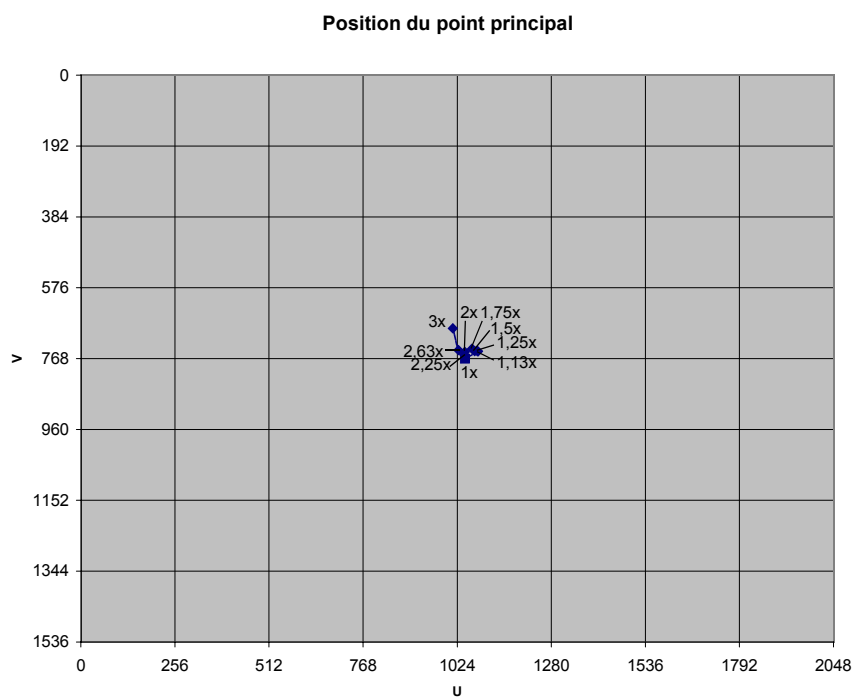


Figure A.10 : Position du point principal en fonction du niveau de zoom pour le Sony DSC-P92.

A.2 Distance focale

Les figures A.11 à A.19, montrent les relations entre la distance focale exprimée en millimètres fournie par les fichiers images et la distance focale exprimée en pixels obtenue par calibrage. Sur chaque figure, la droite en pointillés représente la régression linéaire sur les mesures obtenues par calibrage, alors que la pente de la droite en trait plein est le facteur de conversion \hat{c} calculé grâce aux données constructeur (cf. chapitre 4). Aucun résultat pour le Panasonic NV-MX 500 EG n'a pu être obtenu car cet appareil n'enregistre pas l'information de focale dans les fichiers image.

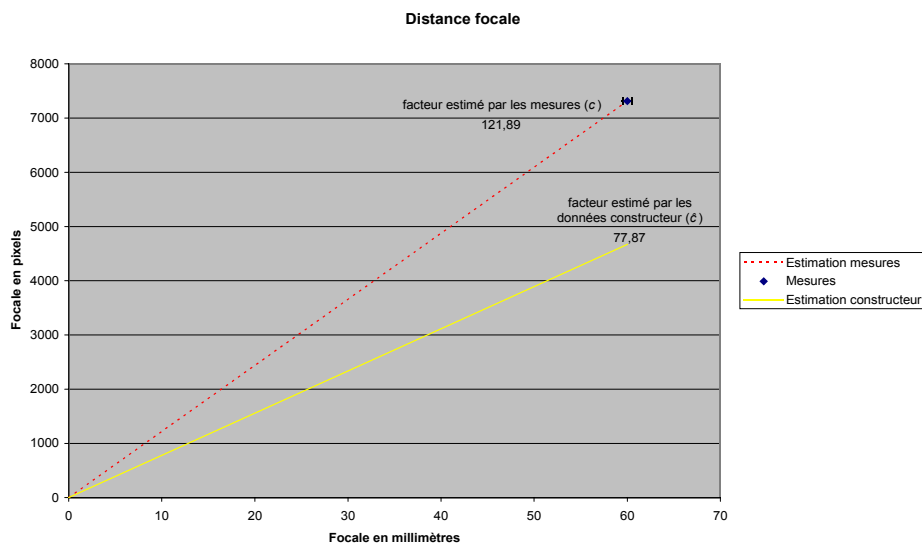


Figure A.11 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Canon EOS 350D.

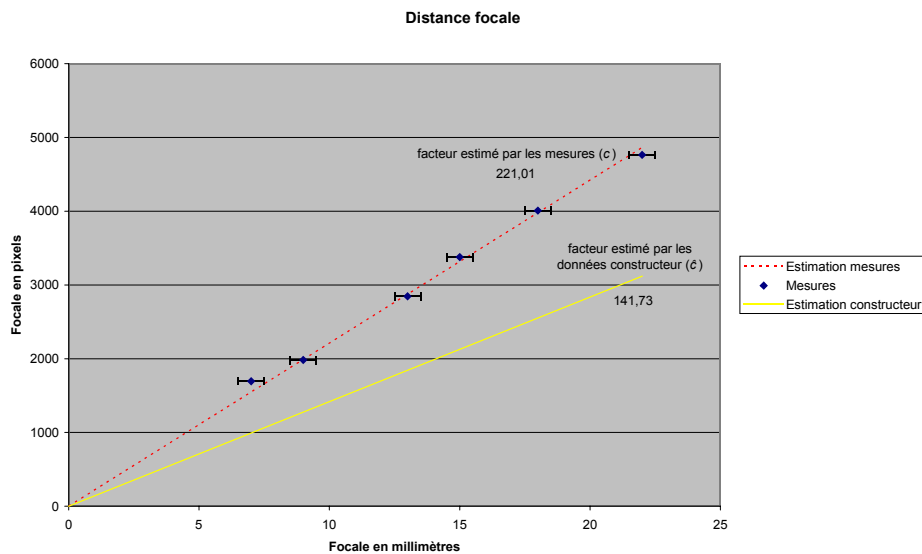


Figure A.12 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Canon IXUS 400.

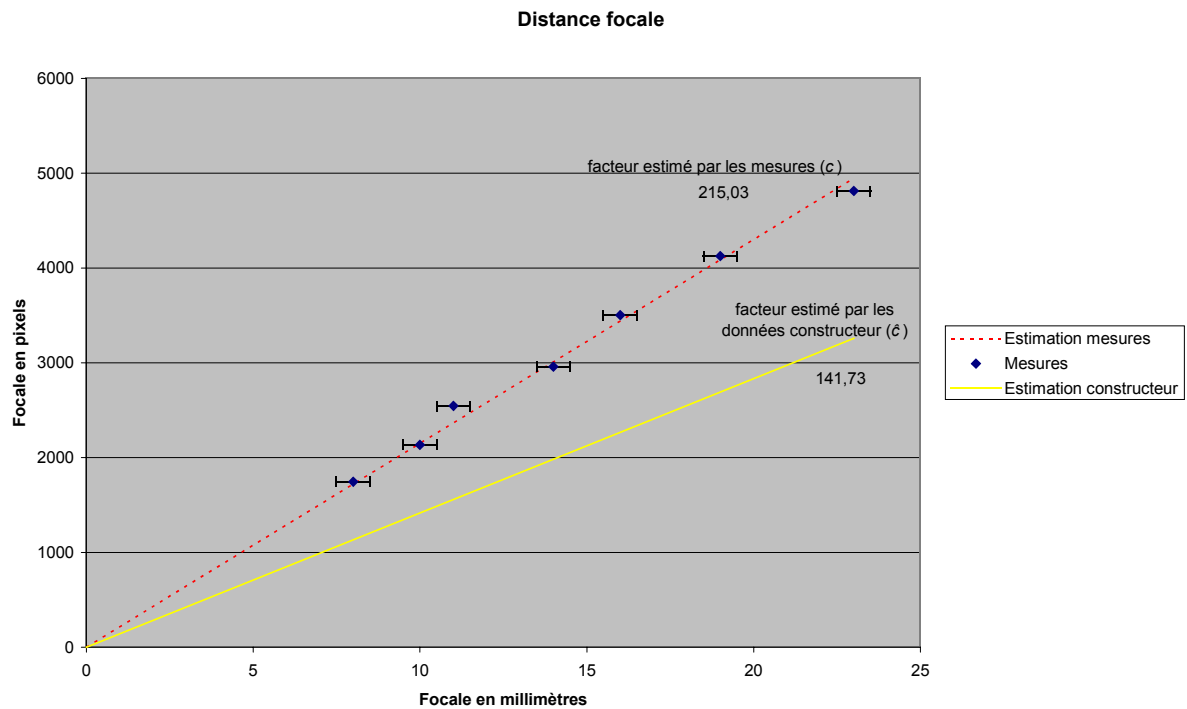


Figure A.13 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Canon PowerShot A95.

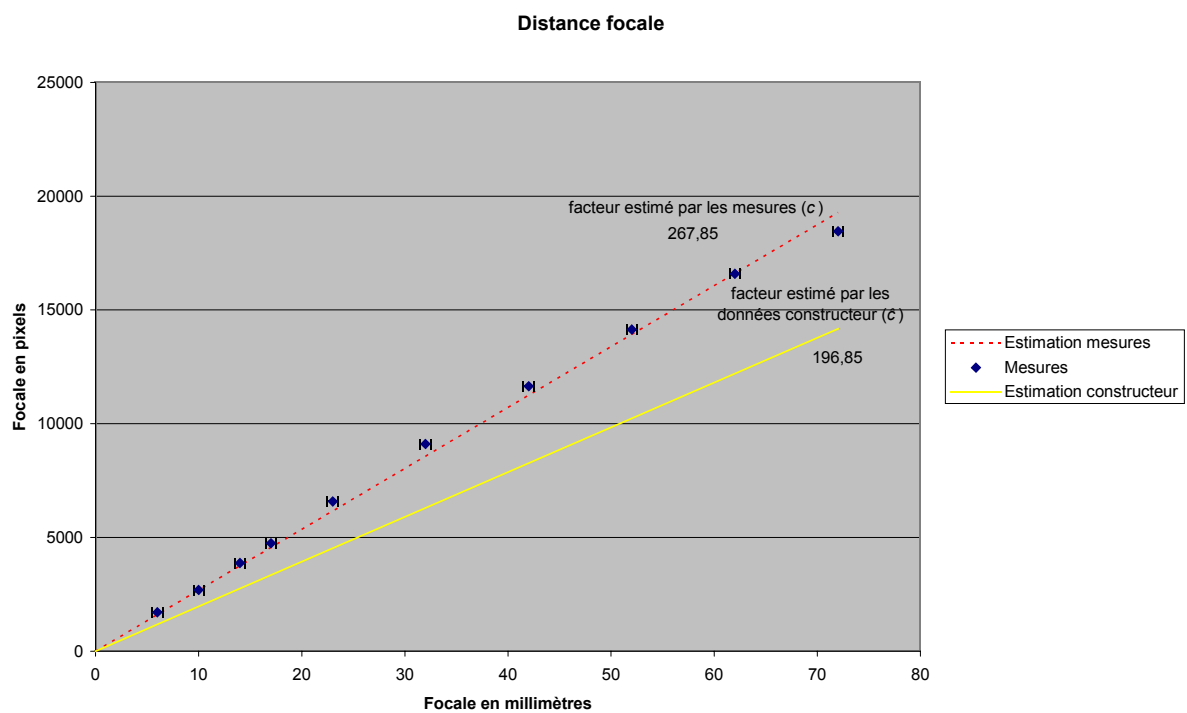


Figure A.14 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Canon PowerShot S2 IS.

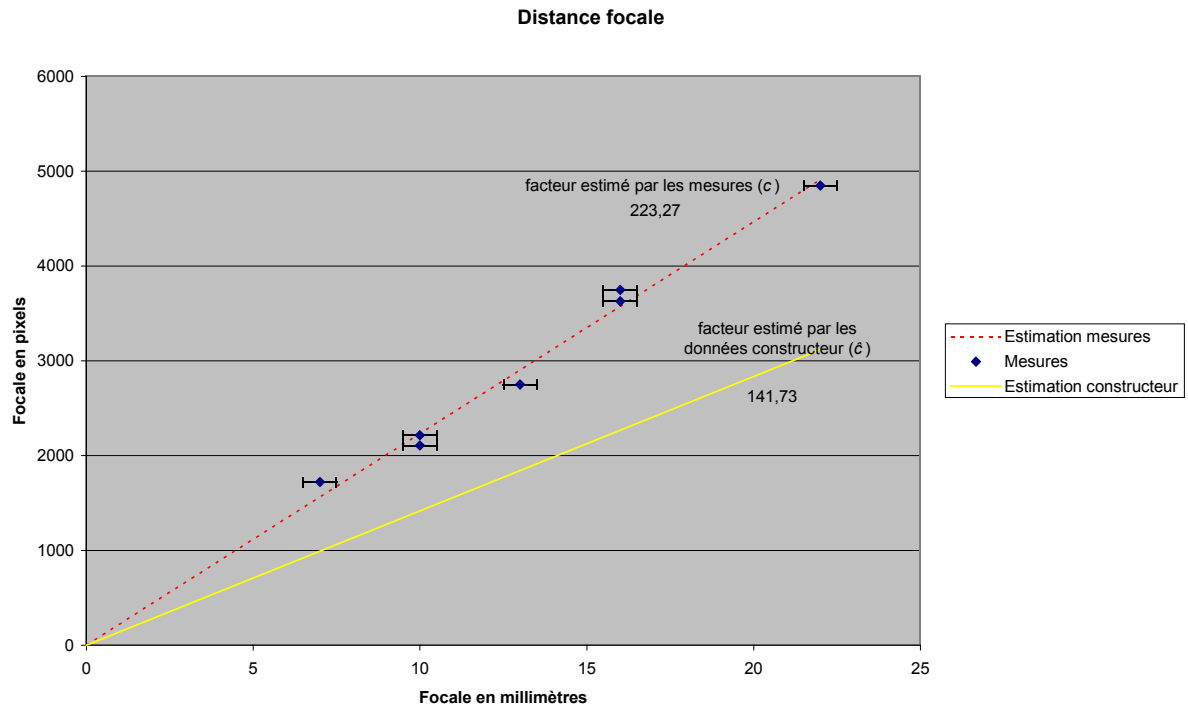


Figure A.15 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Kyocera Finecam S5R.

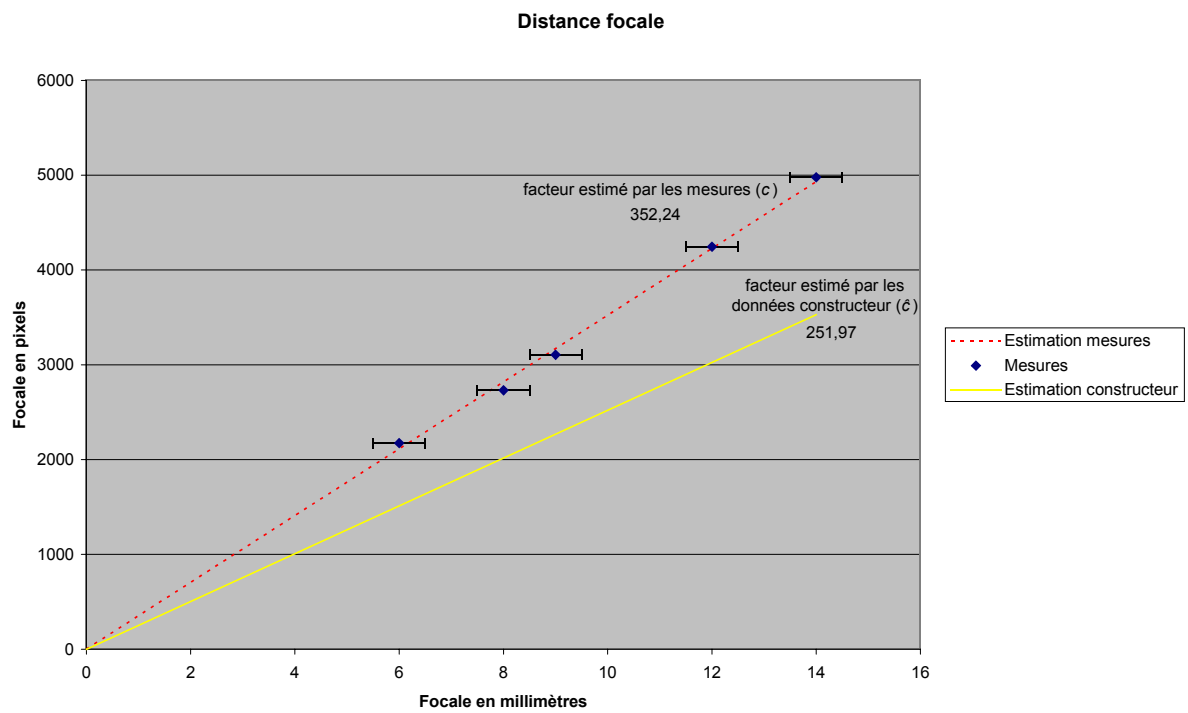


Figure A.16 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Nikon Coolpix S1.

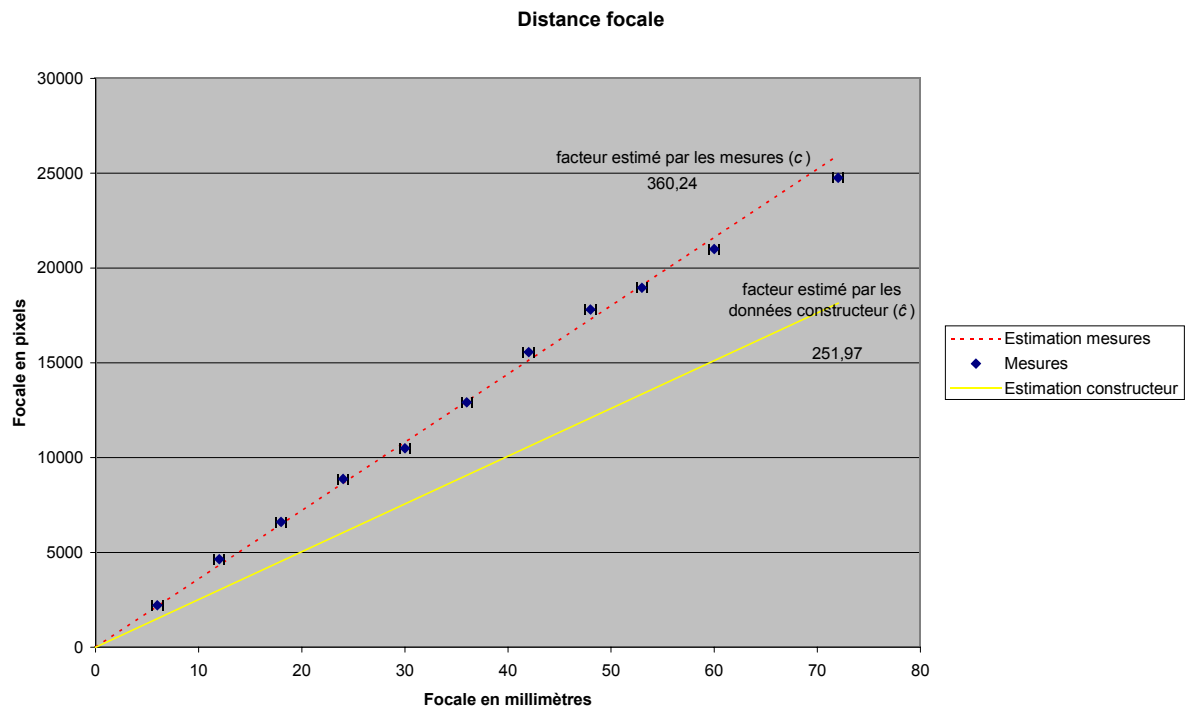


Figure A.17 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Sony DSC-H1.

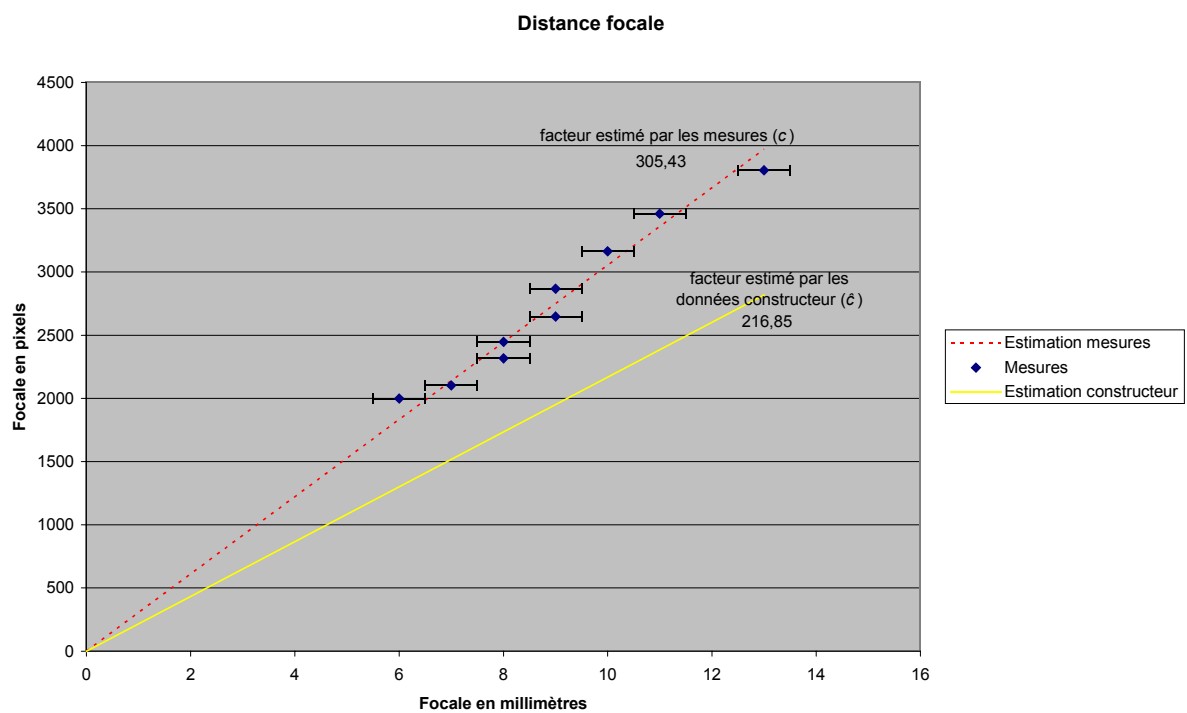


Figure A.18 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Sony DSC-P52.

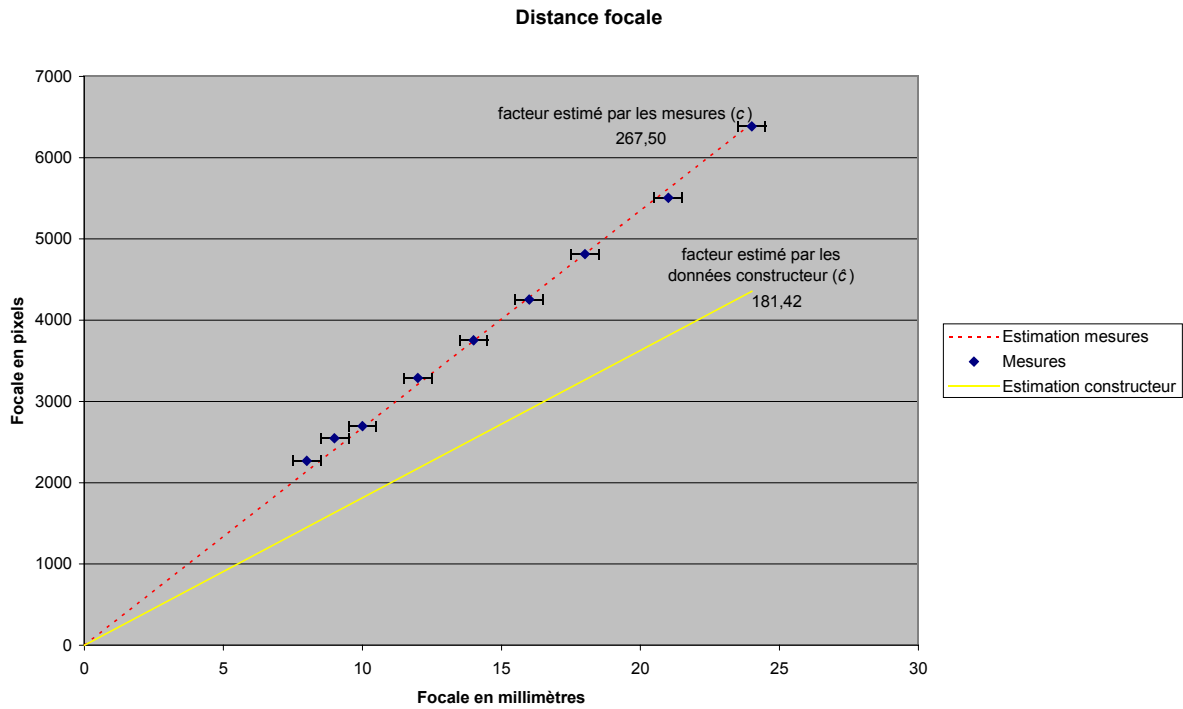


Figure A.19 : Relation entre la distance focale exprimée en millimètres et en pixels pour le Sony DSC-P92.

A.3 Distorsions optiques

Les figures A.20 à A.29, représentent les distorsions optiques causées par les objectifs des appareils photos. Les distorsions représentées ont été obtenues pour les focales minimales, causant le plus de distorsion. Les vecteurs des figures indiquent la déviation locale des points image. Les normes de ces vecteurs sont indiquées à l'aide de courbes de niveaux superposées au champ de vecteurs.

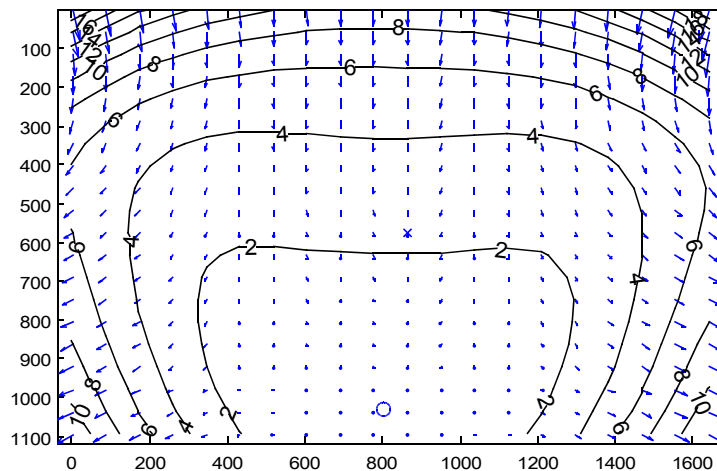


Figure A.20 : Distorsions optiques de l'objectif du Canon EOS 350D.

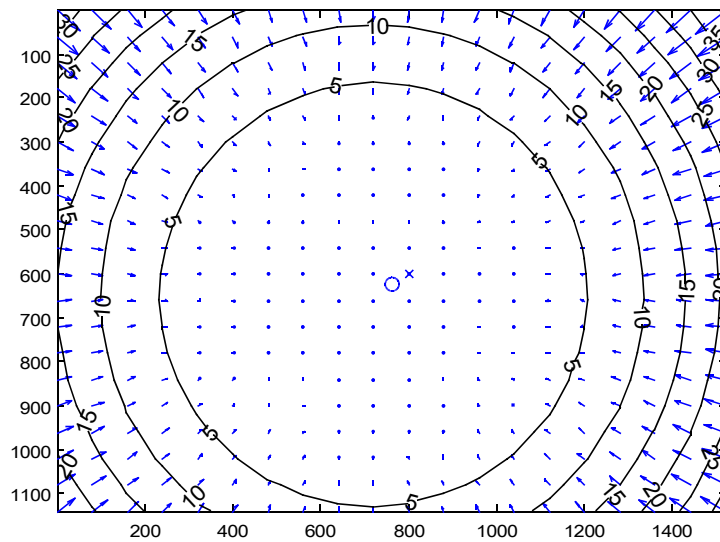


Figure A.21 : Distorsions optiques de l'objectif du Canon IXUS 400.

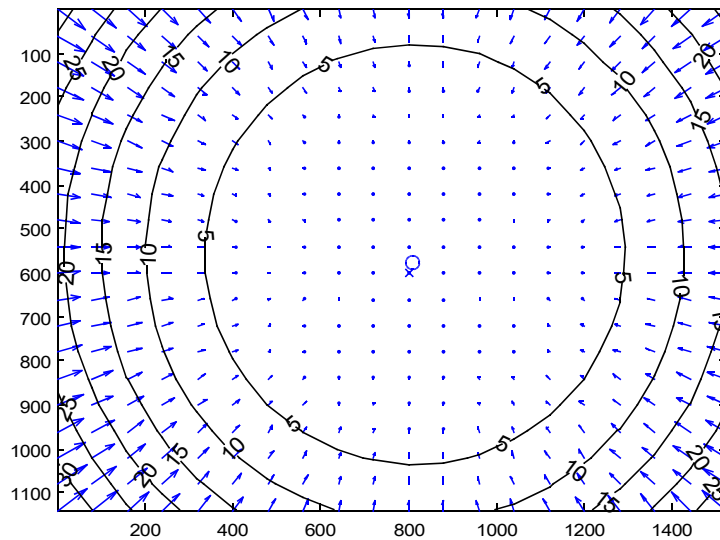


Figure A.22 : Distorsions optiques de l'objectif du Canon PowerShot A95.

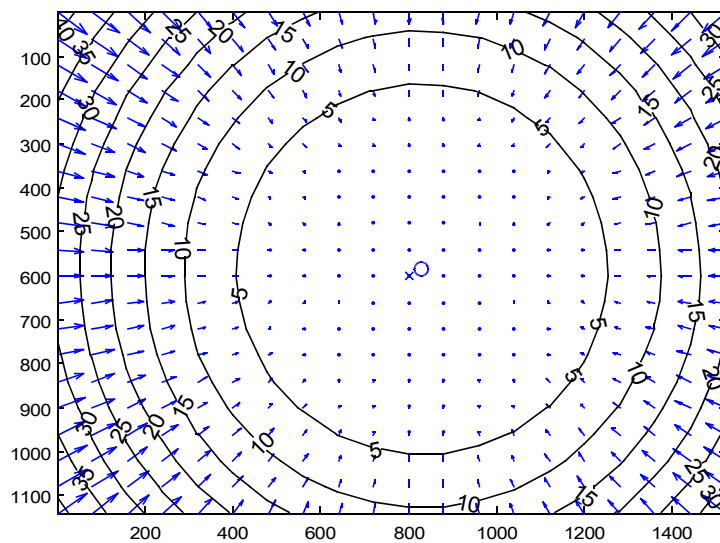


Figure A.23 : Distorsions optiques de l'objectif du Canon PowerShot S2 IS.

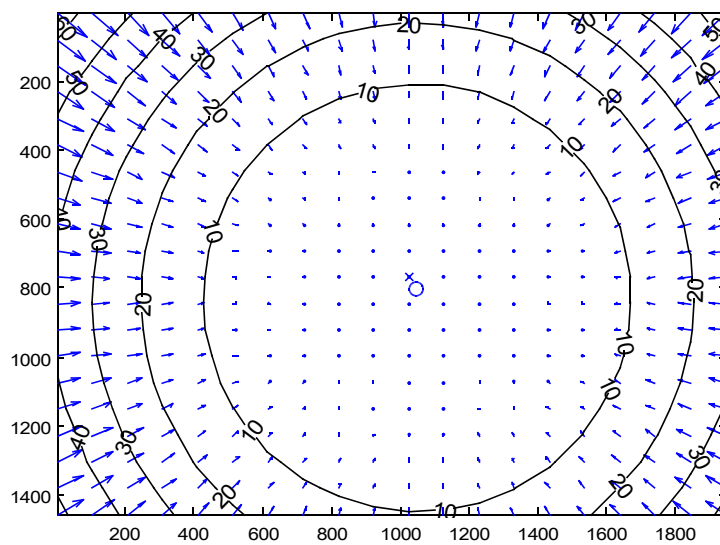


Figure A.27 : Distorsions optiques de l'objectif du Sony DSC-H1.

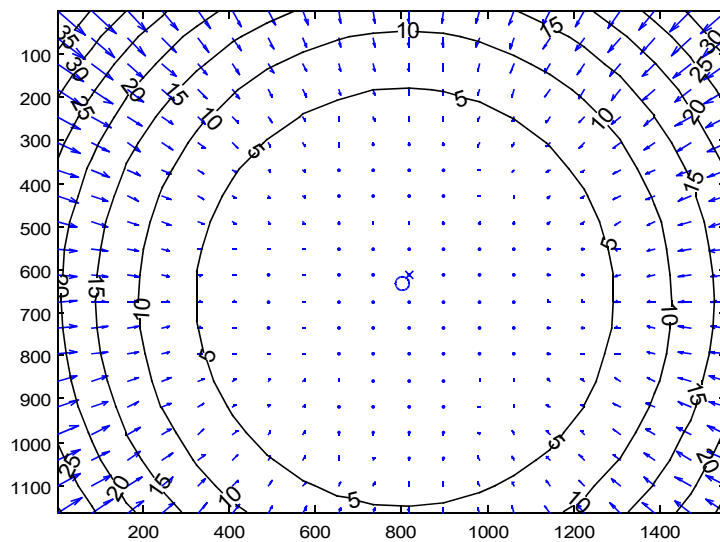


Figure A.28 : Distorsions optiques de l'objectif du Sony DSC-P52.

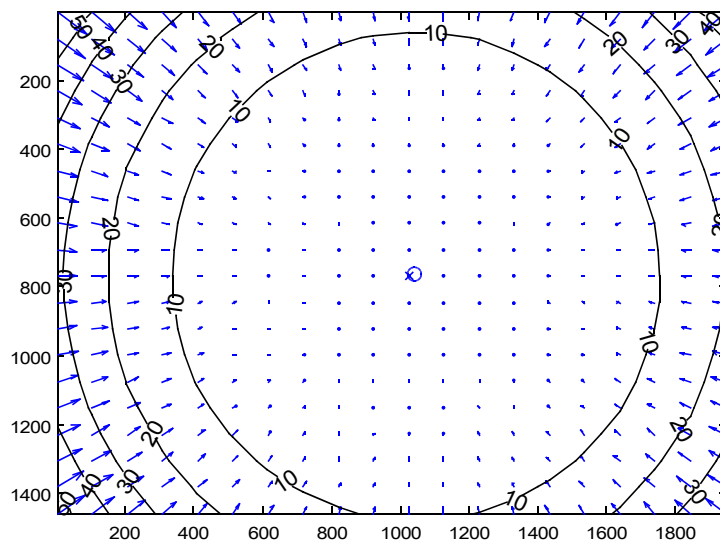


Figure A.29 : Distorsions optiques de l'objectif du Sony DSC-P92.

Annexe B

Paramétrage de l'ajustement de faisceaux

Les paramètres de la minimisation effectuée à l'étape d'ajustement de faisceaux permettent de caractériser les capteurs d'images. Cette annexe décrit les relations mathématiques entre les paramètres de la minimisation et les matrices des capteurs d'images.

Sommaire

B.1 Paramétrage de la rotation.....	181
B.1.1 De R vers r.....	181
B.1.2 De r vers R.....	181
B.2 Paramétrage de la translation.....	182
B.2.1 De t vers θ et ϕ	182
B.2.2 De θ et ϕ vers t.....	182

Le vecteur des paramètres de la minimisation réalisée à l'étape de l'ajustement de faisceaux regroupe les informations de position et d'orientation du second capteur d'image. La matrice de rotation R est paramétrée par un vecteur \mathbf{r} de dimension 3 dont la direction désigne l'axe de la rotation, et la norme la valeur de l'angle. Le vecteur \mathbf{t} , défini à un facteur d'échelle près, est lui paramétré par 2 valeurs θ et ϕ , qui sont les angles de la direction du vecteur en coordonnées sphériques. Cette annexe décrit le passage entre le paramétrage pour l'ajustement de faisceaux et la matrice R et le vecteur \mathbf{t} .

B.1 Paramétrage de la rotation

B.1.1 De R vers \mathbf{r} .

La première étape consiste à déterminer l'axe de la rotation. Le vecteur unitaire de l'axe de rotation, noté \mathbf{v} , est déterminé comme étant le vecteur propre correspondant à la valeur propre unitaire de R . Autrement dit, \mathbf{v} est obtenu en résolvant l'équation $(R - Id)\mathbf{v} = \mathbf{0}$. Ensuite, la valeur de l'angle de la rotation est obtenue grâce aux équations suivantes :

$$\begin{aligned} 2 \cos(\psi) &= \text{trace}(R - 1) \\ 2 \sin(\psi) &= \mathbf{v}^T (R_{32} - R_{23}, R_{13} - R_{31}, R_{21} - R_{12})^T \end{aligned} \quad (\text{B.1})$$

La valeur ψ est alors donnée par une fonction arc-tangente à deux arguments, à partir des valeurs du sinus et du cosinus de l'angle, comme la fonction `atan2(y, x)` du langage C/C++. Enfin, rappelons que le vecteur \mathbf{r} est calculé par $\mathbf{r} = \psi \mathbf{v}$.

B.1.2 De \mathbf{r} vers R .

En écrivant $\mathbf{r} = \psi \mathbf{v}$ où $\psi = \|\mathbf{r}\|$, alors la matrice R s'écrit :

$$R = Id + \sin(\psi)[\mathbf{v}]_{\times} + (1 - \cos(\psi))[\mathbf{v}]_{\times}^2 \quad (\text{B.2})$$

B.2 Paramétrage de la translation

B.2.1 De t vers θ et ϕ .

La transformation est simple : il s'agit de passer d'un système de coordonnées cartésiennes à un système sphérique. Sachant que le vecteur t est de norme unitaire, ce passage s'exprime par :

$$\begin{aligned}\phi &= \arccos\left(\frac{t_3}{\sqrt{t_1^2+t_2^2+t_3^2}}\right) = \arccos(t_3) \\ \theta &= \arctan\left(\frac{t_2}{t_1}\right) + \pi u_0(-t_1)\text{signe}(t_2) = \text{atan2}(t_2, t_1)\end{aligned}\tag{B.3}$$

où $u_0(x)$ est la fonction de Heaviside définie par :

$$u_0(x) = \begin{cases} 0 & \text{si } x \leq 0, \\ 1 & \text{sinon} \end{cases}\tag{B.4}$$

L'utilisation d'une fonction arc-tangente à deux arguments simplifie le calcul de θ .

B.2.2 De θ et ϕ vers t .

Il s'agit de la transformation inverse :

$$\begin{aligned}t_1 &= \sin(\phi)\cos(\theta) \\ t_2 &= \sin(\phi)\sin(\theta) \\ t_3 &= \cos(\phi)\end{aligned}\tag{B.5}$$

BIBLIOGRAPHIE

Bibliographie

[Albouy, 2004] : B. Albouy, S. Treuillet and Y. Lucas, *Fundamental Matrix Estimation Revisited Through a Global 3D Reconstruction Framework*, Advanced Concepts for Intelligent Vision Systems, Bruxelles, Belgique, p. 185-192, 31 août - 3 septembre 2004

[Albouy, 2005] : B. Albouy, S. Treuillet et Y. Lucas, *Mesure Volumétrique d'Escarres à partir de Vues Stéréoscopiques non Calibrées*, GRETSI, Louvain-la-Neuve, Belgique, vol. 1, p. 385-388, 6-9 septembre 2005

[Albouy, 2006] : B. Albouy, S. Treuillet and Y. Lucas, *Accurate 3D Structure Measurements from Two Uncalibrated Views*, Advanced Concepts for Intelligent Vision Systems, Anvers, Belgique, p. 1111-1121, 18-21 septembre 2006

[Armangué, 2003] : X. Armangué and J. Salvi, *Overall view regarding fundamental matrix estimation*, Image and Vision Computing, vol. 21, n° 2, p. 205-220, février 2003

[Bartoli, 2001] : A. Bartoli and P. Sturm, *Three New Algorithms for Projective Bundle Adjustment with Minimum Parameters*, Rapport de Recherche INRIA, n° 4236, août 2001

[Bay, 2005] : H. Bay, V. Ferrari and L. Van Gool, *Wide-Baseline Stereo Matching with Line Segments*, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, vol. 1, p. 329-336, 20-25 juin 2005

[Besl, 1992] : P.J. Besl and N.D. McKay, *A Method for Registration of 3D Shapes*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, n° 2, p. 239-256, février 1992

[Bookstein, 1979] : F. Bookstein, *Fitting Conic Sections to Scattered Data*, Computer Vision, Graphics, and Image Processing, vol. 9, n° 1, p. 56-71, janvier 1979

[Bouguet, 2006] : J.Y. Bouguet, *Camera Calibration Toolbox for Matlab*, http://www.vision.caltech.edu/bouguetj/calib_doc/, 2006

[Delponte, 2005] : E. Delponte, F. Isgro, F. Odono and A. Verri, *SVD-Matching using SIFT Features*, Vision, Video, and Graphics (2005), Edinburg, Royaume-Uni, p. 125-132, 7-8 juillet 2005

- [Elias, 2004] : R. Elias, *Wide Baseline Matching Through Homographic Transformation*, 17th International Conference on Pattern Recognition (ICPR'04), Cambridge, Royaume-Unis, vol. 4, p. 130-133, 23-26 août 2004
- [escarre.fr] : Association PERSE, Société Asymptote, *Le site de référence de l'escarre*, www.escarre.fr, 2006
- [Frag, 2004] : A. Farag and A. Eid, *Local Quality Assessment of 3D Reconstructions from Sequence of Images: a Quantitative Approach*, Proceedings of Advanced Concepts for Intelligent Vision Systems, Bruxelles, Belgique, p. 161-168, 31 août - 3 septembre 2004
- [Faugeras, 1992] : O.D. Faugeras, Q.-T. Luong and S.J. Maybank, *Camera Self-Calibration: Theory and Experiments*, ECCV 1992, Lecture Notes in Computer Science, Santa Margherita Ligure, Italie, vol. 588, p. 321-334, 19-22 mai 1992
- [Faugeras, 1995] : O. Faugeras, *Stratification of 3-D vision: projective, affine, and metric representations*, Journal of the Optical Society of America, vol. 12, n° 3, p. 465-484, mars 1995
- [Fischler, 1981] : M.A. Fischler and R.C. Bolles, *Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography*, Communication of the Association for Computing Machinery, vol. 24, n° 6, p. 381-395, juin 1981
- [Gouet, 2000] : V. Gouet, *Mise en Correspondance d'Images en Couleur - Application à la Synthèse de Vues Intermédiaires*, Thèse de doctorat, Université Montpellier II Sciences et Techniques du Languedoc, 25 octobre 2000
- [Gruen, 2004] : A. Gruen, F. Remondino and L. Zhang, *Photogrammetric Reconstruction of the Great Buddha of Bamiyan, Afghanistan*, The Photogrammetric Record, vol. 19, n° 107, p. 177 - 199, septembre 2004
- [Hammarstedt, 2005] : P. Hammarstedt, P. Sturm and A. Heyden, *Closed-Form Solutions and Degenerate Cases for Camera Calibration with One-Dimensional Objects*, Proceedings of the 10th International Conference on Computer Vision, Beijing, Chine, p. 317-324, 17-20 octobre 2005
- [Harris, 1988] : C. Harris, and M. Stephens, *A Combined Corner and Edge Detector*, Proceedings of The Fourth Alvey Vision Conference, Manchester, Royaume-Uni, p. 147-151, 1988
- [Hartley, 1992] : R. Hartley, *Estimation of Relative Camera Positions for Uncalibrated Cameras*, ECCV'92 - Lecture Notes in Computer Science, vol. 588, p. 579-587, 1992
- [Hartley, 1994] : R. Hartley, *An Algorithm for Self Calibration from Several Views*, Conference on Computer Vision and Pattern Recognition, Seattle, USA, p. 908-912, 21-23 juin 1994
- [Hartley, 1997a] : R. Hartley and P. Sturm, *Triangulation*, Computer Vision and Image Understanding, vol. 68, n° 2, p. 146-157, novembre 1997

- [Hartley, 1997b] : R. Hartley, *In Defense of the 8-points Algorithm*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, n° 6, p. 580-593, juin 1997
- [Hartley, 2004] : R. Hartley, A. Zisserman, *Multiple View Geometry in Computer Vision*, seconde édition, Cambridge University Press, ISBN 0-521-54051-8, 2004
- [Heikkilä, 1997] : J. Heikkilä and O. Silvén, *A Four-step Camera Calibration Procedure with Implicit Image Correction*, Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97), San Juan, Puerto Rico, p. 1106-1112, 17-19 juin 1997
- [JEITA, 2002] : Technical Standardization Committee on AV & IT Storage Systems and Equipment, *Exchangeable image file format for digital still cameras: Exif Version 2.2*, Japan Electronics and Information Technology Industries Association, avril 2002
- [Jones, 1995] : B. F. Jones and P. Plassmann, *An Instrument to Measure the Dimensions of Skin Wounds*, IEEE Transactions on Biomedical Engineering, vol. 42, n° 5, p. 464 - 470, Mai 1995
- [Koeffer, 2004] : M.-J. Koeffer, *Actualisation de la prise en charge des escarres : application dans un hôpital gériatrique.*, Thèse de doctorat, Université Louis Pasteur de Strasbourg, 2004
- [Koenig, 2006] : E. Koenig, *Evaluation de systèmes de vision 3D*, Mémoire de Master Recherche, Polytech' Orléans - Laboratoire Vision et Robotique, septembre 2006
- [Krouskop, 2002] : T. A. Krouskop, R. Baker and M. S. Wilson, *A Non Contact Wound Measurement System*, Journal of Rehabilitation Research and Development, vol. 39, n° 3, p. 337-346, mai/juin 2002
- [Lavest, 1998] : J.M. Lavest, M. Viala and M. Dhome, *Do we Really Need an Accurate Calibration Pattern to Achieve Reliable Camera Calibration?*, Proceedings of the European Conference on Computer Vision, Freiburg, Allemagne, vol. 1, p. 158-174, juin 1998
- [Longuet-Higgins, 1981] : H. Longuet-Higgins, *A Computer Algorithm for Reconstructing a Scene from two Projections*, Nature, vol. 293, n° 5828, p. 133-135, septembre 1981
- [Loops, 1999] : C. Loops and Z. Zhang, *Computing Rectifying Homographies for Stereo Vision*, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Fort Collins, USA, vol. 1, p. 125-131, 23-25 juin 1999
- [Lourakis, 1999] : M. Lourakis and R. Deriche, *Camera Self-Calibration Using the Kruppa Equations and the SVD of the Fundamental Matrix: From Point Correspondences to 3D Measurements*, Rapport de Recherche INRIA, n° 3748, août 1999
- [Lourakis, 2000a] : M.I.A. Lourakis and R. Deriche, *Camera Self-Calibration Using the Kruppa Equations and the SVD of the Fundamental Matrix: The Case of Varying Intrinsic Parameters*, Rapport de Recherche INRIA, n° 3911, mars 2000
- [Lourakis, 2000b] : M.I.A. Lourakis, S.V. Tzurbakis, A.A. Argyros and S.C. Orphanoudakis, *Using Geometric Constraints for Matching Disparate Stereo Views of 3D Scenes Containing*

Planes, Proceedings of the International Conference on Pattern Recognition (ICPR'00), Barcelone, Espagne, vol. 1, p. 419-422, 3-8 septembre 2000

[Lourakis, 2004] : M.I.A. Lourakis and A.A. Argyros, *The Design and Implementation of a Generic Sparse Bundle Adjustment Software Package Based on the Levenberg-Marquardt Algorithm*, Rapport de Recherche, Heraklion, Crête, Grèce, n° 340, août 2004

[Lourakis, 2005] : M.I.A. Lourakis, *levmar: Levenberg-Marquardt nonlinear least squares algorithms in C/C++*, www.ics.forth.gr/~lourakis/levmar/, février 2005

[Lourakis, 2006] : M.I.A. Lourakis, *sba: A Generic Sparse Bundle Adjustment C/C++ Package Based on the Levenberg-Marquardt Algorithm*, www.ics.forth.gr/~lourakis/sba/, juin 2006

[Lowe, 2004] : D. G. Lowe, *Distinctive image features from scale-invariant keypoints*, International Journal of Computer Vision, vol. 60, n° 2, p. 91-110, Novembre 2004

[Lu, 2004] : X. Lu and R. Manduchi, *Wide Baseline Feature Matching Using the Cross-Epipolar Ordering Constraint*, 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'04), Washington DC, USA, vol. 1, p. 16-23, 27 juin - 2 juillet 2004

[Ma, 2000] : Y. Ma, R. Vidal, J. Kosecka and S. Sastry, *Kruppa Equation Revisited: its Renormalization and Degeneracy*, ECCV 2000, Dublin, Irlande, p. 561-577, 26 juin 2000

[Malian, 2002] : A. Malian, F.A. van den Heuvel and A. Azizi, *A Robust Photogrammetric System for Wound Measurement*, International Archives of Photogrammetry and Remote Sensing, Corfu, Grèce, vol. 34, Part 5, p. 264 - 269, 2-6 Septembre 2002

[Malis, 2001] : E. Malis and A. Bartoli, *Euclidean Bundle Adjustment Independent on Camera Intrinsic Parameters*, Rapport de Recherche INRIA, n° 4377, décembre 2001

[Matas, 2002] : J. Matas, O. Chum, M. Urban and T. Pajdla, *Robust Wide Baseline Stereo from Maximally Stable Extremal Regions*, Proceedings of The 13th British Machine Vision Conference (BMVC 2002), Cardiff, Royaume Unis, p. 384-393, 2-5 septembre 2002

[Matas, 2003] : J. Matas, P. Bilek, O. Chum, *Rotational Invariants for Wide-baseline Stereo*, International Research Report, n° MS03-146, mars 2003

[MAVIS-II] : P. Plassmann, *MAVIS-II - Measurement of Area and Volume Instrument System*,

<http://www.comp.glam.ac.uk/pages/staff/pplasma/MedImaging/Projects/Wounds/Mavis-II/INDEX.HTML>, 2006

[Microsoft, 2001] : Microsoft Research, *Stereo Matcher*, <http://research.microsoft.com/research/downloads/default.aspx#StereoMatcher>, 2001

[Mikolajczyk, 2005] : K. Mikolajczyk and C. Schmid, *A Performance Evaluation of Local Descriptors*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, n° 10, p. 1615-1630, octobre 2005

- [Montesinos, 1998] : P. Montesinos and V. Gouet and R. Deriche, *Differential Invariants for Color Images*, Proceedings of 14 th International Conference on Pattern Recognition, Brisbane, Australie, p. 838-840, Août 1998
- [Novotni, 2001] : M. Novotni and R. Klein, *A Geometric Approach to 3D Object Comparison*, Proceedings of the International Conference on Shape Modeling and Applications, Genova, Italie, p. 167-175, 7-11 mai 2001
- [Oram, 2001] : D. Oram, *Rectification for Any Epipolar Geometry*, Proceedings of the 12th British Machine Vision Conference (BMVC 2001), p. 653-662, 10-13 septembre 2001
- [Pollefeys, 1998] : M. Pollefeys, R. Koch and L van Gool, *Self-Calibration and Metric Reconstruction in spite of Varying and Unknown Intrinsic Camera Parameters*, International Conference on Computer Vision, Bombay, Inde, p. 90-95, 4-7 janvier 1998
- [Pollefeys, 1999] : M. Pollefeys and L. Van Gool, *Stratified Self-Calibration with the Modulus Constraint*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 21, n° 8, p. 707-724, août 1999
- [Pollefeys, 2002] : M. Pollefeys, L. Van Gool, M. Vergauwen, K. Cornelis, F. Verbiest and J. Tops, *Video-to-3D*, Photogrammetric Computer Vision - IAPRS, Graz, Autriche, vol. 34, n° 3A, p. 252 - 257, 9-13 septembre 2002
- [Pritchett, 1998] : P. Pritchett and A. Zisserman, *Wide Baseline Stereo Matching*, International Conference on Computer Vision (ICCV'98), Bombay, Inde, p. 754-760, 4-7 janvier 1998
- [Ricard, 2005] : J. Ricard, D. Coeurjolly and A. Baskurt, *Generalizations of Angular Radial Transform for 2D and 3D Shape Retrieval*, Pattern Recognition Letters, vol. 26, n° 14, p. 2174-2186, Octobre 2005
- [Sampson, 1982] : P.D. Sampson, *Fitting Conic Sections to 'Very Scattered' Data: An Iterative Refinement of the Bookstein Algorithm*, Computer Graphics and Image Processing, vol. 18, n° 1, p. 97-108, janvier 1982
- [Schaffalitzky, 2001] : F. Schaffalitzky and A. Zisserman, *Viewpoint Invariant Texture Matching and Wide Baseline Stereo*, Proceedings of the 8th International Conference on Computer Vision (ICCV 2001), Vancouver, Canada, p. 636-643, 2001
- [Scharstein, 2002] : D. Scharstein and R. Szeliski, *A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms*, International Journal of Computer Vision, vol. 47, n° 1-3, p. 7-42, avril 2002
- [Schmid, 1996] : C. Schmid, *Appariement d'Images par Invariants Locaux de Niveaux de Gris*, Thèse de doctorat, Institut National Polytechnique de Grenoble, 2 juillet 1996
- [Schmid, 2000] : C. Schmid, R. Mohr and C. Bauckhage, *Evaluation of Interest Point Detectors*, International Journal of Computer Vision, vol. 37, n° 2, p. 151-172, 2000

- [Slama, 1980] : C.C. Slama, *Manual of Photogrammetry*, Fourth Edition, ISBN 0937294012, 1980
- [Sturm, 1997] : P. Sturm, *Vision 3D non Calibrée - Contributions à la Reconstruction Projective et Etude des Mouvements Critiques pour l'Auto-calibrage*, Thèse de doctorat, Institut National Polytechnique de Grenoble, 17 décembre 1997
- [Sturm, 1999] : P. Sturm and S. Maybank, *On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications*, Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA, p. 432-437, juin 1999
- [Sturm, 2001] : P. Sturm, *On Focal Length Calibration from Two Views*, Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Kauai, Hawaï, vol. 2, p. 145-150, 8-14 décembre 2001
- [Sturm, 2002] : P. Sturm, *Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length*, Image and Vision Computing, vol. 20, n° 5-6, p. 415-426, mars 2002
- [Sturm, 2005] : P. Sturm, Z.L. Cheng, P.C.Y. Chen and A.N. Poo, *Focal Length Calibration from Two Views: Method and Analysis of Singular Cases*, Computer Vision and Image Understanding, vol. 99, n° 1, p. 58-95, juillet 2005
- [Torr, 1997] : P. Torr and D. Murray, *The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix*, International Journal of Computer Vision, vol. 24, n° 3, p. 271-300, septembre 1997
- [Torr, 2002a] : P. Torr, *A Structure and Motion Toolkit in Matlab: Interactive adventures in S and M*, n° MSR-TR-2002-56, juin 2002
- [Torr, 2002b] : P. Torr, *Bayesian Model Estimation and Selection for Epipolar Geometry and Generic Manifold Fitting*, International Journal of Computer Vision, vol. 50, n° 1, p. 35-61, octobre 2002
- [Torr, 2003] : P. Torr and A. Fitzgibbon, *Invariant Fitting of Two View Geometry or "In Defiance of the 8-Point Algorithm"*, Proceedings of British Machine Vision Conference, Norwich, Angleterre, p. 83-92, 9-11 septembre 2003
- [Triggs, 1997] : B. Triggs, *Autocalibration and the Absolute Quadric*, Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition, Porto Rico, USA, p. 609-614, 17-19 juin 1997
- [Tsai, 1986] : R.Y. Tsai, *An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision*, Proceedings of the Conference on Computer Vision and Pattern Recognition, Miami Beach, Floride, USA, p. 364-374, 22-26 juin 1986
- [Tuytelaars, 2000] : T. Tuytelaars and L. Van Gool, *Wide Baseline Stereo Matching based on Local, Affinely Invariant Regions*, The 11th British Machine Vision Conference (BMVC 2000), Bristol, Royaume-Uni, 11-14 septembre 2000

- [Urbanek, 2001] : M. Urbanek, R. Horaud and P. Sturm, *Calibration of Digital Amateur Cameras*, Rapport de recherche INRIA, n° 4214, juin 2001
- [Valdés, 2006] : A. Vladés, J. Ronda and G. Gallego, *The Absolute Line Quadric And Camera Autocalibration*, International Journal of Computer Vision, vol. 66, n° 3, p. 283-303, mars 2006
- [Vincent, 1991] : L. Vincent and P. Soille, *Watershed in Digital Spaces: an Efficient Algorithm Based on Immersion Simulations*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 13, n° 6, p. 583-598, juin 1991
- [Vranié, 2005] : D.V. Vranié and D. Saupe, *3D Shape Descriptor based on 3D Fourier Transform*, ACM Computing Surveys, vol. 37, n° 4, p. 345-387, décembre 2005
- [Xu, 1996] : G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion, and Object Recognition: A Unified Approach*, Kluwer Academic Publishers, ISBN 0-7923-4199-6, 1996
- [Zeller, 1996] : C. Zeller and O. Faugeras, *Camera Self-Calibration from Video Sequences: the Kruppa Equations Revisited*, Rapport de Recherche INRIA, n° 2793, février 1996
- [Zhang, 1994] : Z. Zhang, R. Deriche, O. Faugeras and Q.-T. Luong, *A Robust Technique for Matching Two Uncalibrated Images Through the Recovery of the Unknown Epipolar Geometry*, Rapport de Recherche INRIA, n° 2273, mai 1994
- [Zhang, 1998] : Z. Zhang, *Determining the Epipolar Geometry and its Uncertainty: A Review*, International Journal of Computer Vision, vol. 27, n° 2, p. 161-198, mars 1998
- [Zhang, 1999] : Z. Zhang, *Flexible Camera Calibration by Viewing a Plane from Unknown Orientations*, International Conference on Computer Vision (ICCV'99), Kerkyra, Corfu, Grèce, p. 666-673, 20-25 septembre 1999
- [Zhang, 2000] : Z. Zhang, *A Flexible New Technique for Camera Calibration*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, n° 11, p. 1330-1334, novembre 2000
- [Zhang, 2004] : Z. Zhang, *Camera Calibration with One-Dimensional Objects*, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 26, n° 7, p. 892-899, juillet 2004

Bibliographie de l'auteur

Publications nationales :

Y. Lucas, S. Treuillet, B. Albouy, *ESCALE – Une innovation dans le suivi thérapeutique des escarres par l'imagerie 3D couleur*, L'escarre – Revue officielle de l'association PERSE, Techni Media Services, n°25, p. 36-37, mars 2005

Y. Lucas, S. Treuillet, B. Albouy, J.C. Pichaud, *RAVIR : Des Systèmes de Vision Active pour le Relevé Anatomique 3D Appliqué à la Santé*, J3eA - Vol. 4, Hors-Série 3 http://www.bibsciences.org/bibsup/j3ea/full_HS/vol4_HS3/5/pdf/j3ea2005805.pdf, 2005.

J.C. Pichaud, Y. Lucas, S. Treuillet, B. Albouy, H. Wannous, H. Barre, *ESCALE : ESCarre Analyse Lisibilité Evaluation, Imagerie 3D couleur par simple appareil photo numérique – fruit de l'expérience d'une collaboration médico-scientifique*, L'escarre – Revue officielle de l'association PERSE, n°30, juin 2006.

Communications internationales :

B. Albouy, D. Birov, Y. Lucas, S. Treuillet, *Fundamental Matrix Estimation Revisited In Respect To 3D Reconstruction Accuracy*, CompSysTech'04 – Rousse, Bulgarie, 17-18 juin 2004.

B. Albouy, S. Treuillet, Y. Lucas and D. Birov. *Fundamental matrix estimation revisited through a global 3D reconstruction framework*, ACIVS – Bruxelles, p. 185-192, 31 août - 3 sept. 2004.

B. Albouy, S. Treuillet, Y. Lucas and J.C. Pichaud, *Volume Estimation from Uncalibrated Views Applied to Wound Measurement*, ICIAP – Cagliari, LNCS 3617, p. 945-952, septembre 2005.

Y. Lucas, S. Treuillet, B. Albouy, H. Wannous, J.C. Pichaud, *3D and Color Wound Assessment Using a Simple Digital Camera*, EPUAP 2006 – Berlin.

B. Albouy, E. Koenig, S. Treuillet, Y. Lucas, *Accurate 3D Structure Measurements from Two Uncalibrated Views*, ACIVS 2006 – Anvers, p. 1111-1121, 18-21 septembre 2006.

Communications nationales :

B. Albouy, *Relevé 3D couleur d'escarres par un appareil photo numérique*, JSTIM FRGBM – Nancy, p. 3-4, 21-23 mars 2005.

B. Albouy, S. Treuillet, Y. Lucas, *Reconstruction tridimensionnelle d'escarres par stéréovision couleur non calibrée*, ORASIS – Fournols, <http://orasis2005.univ-bpclermont.fr/papers/020.pdf>, 24-27 mai 2005.

B. Albouy, S. Treuillet, Y. Lucas, *Mesure volumétrique d'escarres à partir de vues stéréoscopiques non calibrées*, GRETSI – Louvain-la-Neuve, 6-9 sept. 2005.

B. Albouy, S. Treuillet, Y. Lucas, L. Thauvin, *Suivi thérapeutique d'escarres par imagerie 3D couleur non calibrée*, CAPTEURS 2005 – Bourges, 19-20 oct. 2005.

Travaux primés :

B. Albouy, *Relevé 3D couleur d'escarres par un appareil photo numérique*, JSTIM – Nancy, 21-23 mars 2005. 3^{ème} Prix de la conférence de la meilleure communication.

B. Albouy, *Principes de la reconstruction tridimensionnelle par stéréovision*, 4^{ème} prix Trophée e-student 2004-2005, Ecole des Mines de Nantes, http://www.e-doceo.net/home/trophees_e-students/presentation.php, 22 avril 2005.

TITRE : Relevé Volumétrique par Stéréovision à partir de Deux Vues non Calibrée – Application au Suivi Thérapeutique d'Escarres

RESUME : Cette thèse est consacrée au développement d'une chaîne complète de reconstruction 3D à partir de deux images prises à main levée par un appareil photo numérique du commerce. L'application visée est le relevé volumétrique d'escarres, pour lequel la stéréovision non calibrée est une approche inédite qui offre beaucoup de flexibilité dans la prise des vues. Les différents algorithmes ont été choisis et optimisés en vue de réduire les erreurs sur le modèle 3D inféré. En particulier, la méthode itérative de mise en correspondance semi-dense développée permet d'obtenir un nombre important d'appariements fiables sur des points de vues éloignés. Basée sur une contrainte épipolaire de plus en plus resserrée, cette méthode affine la géométrie épipolaire et l'auto-calibrage avec un indice de confiance de plus en plus élevé au fil des itérations. À la fin du processus, les appariements sont encore multipliés par corrélation affine à l'intérieur des triangles homologues. La position des points de vues déterminée par l'autocalibrage permet ensuite de reconstruire un modèle 3D par triangulation. Ainsi, deux à trois minutes suffisent pour reconstruire environ 3 000 points à partir d'une paire d'images sur des ordinateurs actuels. Diverses expérimentations évaluent la qualité de la reconstruction de façon objective et quantitative, en comparant les modèles 3D reconstruits à des vérités terrains. Elles permettent notamment de déterminer la configuration optimale des vues, obtenue pour une vergence de 25° et un rapport de distance entre vues de 1,2. Il en découle une erreur sur le calcul volumétrique d'environ 3 %, ce qui est suffisant pour l'application de suivi thérapeutique visée.

MOTS-CLEFS : Stéréovision, reconstruction 3D, mise en correspondance d'images, auto-calibrage, ajustement de faisceaux, évaluation quantitative.

TITLE: Accurate Volume Estimation from Two Uncalibrated Views Applied to Wound Healing.

RESUME: Firstly, a 3D reconstruction chain from two wide baseline photographs taken with an uncalibrated camera is drawn, in order to assess the wound volume. The different algorithms of this chain have been chosen and optimized so as to decrease the volume measurement error. At the beginning of the process, a half-dense iterative matching method, which is based on a more and more tightened epipolar constraint, is applied. Then, a self calibration process is run to work out relative point-of-view positions. The 3D model is triangulated from these locations.

Secondly, the quality of the whole 3D reconstruction chain is assessed thanks to some experimentations, which compare the reconstructed 3D models with ground truths. These experiments notably allow to work out the optimal point-of-view configuration, which is obtained for a 25° vergence and for a point-of-view distance ratio of 1,2. It results from it that the computed volume error is 3 %, which is sufficient for the aimed therapeutic application.

KEY-WORDS: Stereovision, 3D reconstruction, image matching, self calibration, bundle adjustment, quantitative assessment.

DISCIPLINE : Sciences et Technologies Industrielles

Laboratoire Vision et Robotique
Ecole Nationale Supérieure d'Ingénieurs de Bourges
10, boulevard Lahitolle
18020 BOURGES CEDEX