AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact : ddoc-theses-contact@univ-lorraine.fr

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

Département de formation doctorale en informatique **École doctorale IAEM Lorraine**
UFR STMIA

Système de Réalité Augmentée basé sur l'observation de structures planes : conception et évaluation

THÈSE

présentée et soutenue publiquement le ...

pour l'obtention du

Doctorat de l'université Henri Poincaré – Nancy 1

(spécialité informatique)

par

Javier Flavio VIGUERAS-GOMEZ

Équipe d'accueil : ISA, puis MAGRIT (LORIA, Nancy)

Composition du jury

Examinateurs :	Marie-Odile	BERGER	Directeur
	Jean-Marc	LAVEST	Rapporteurs
	Eric	MARCHAND	
	Dominique	MÉRY	Référent interne
	Michel	DEVY	
	Pierre	GURDJOS	Invité

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thloria.

Á mes parents et à ma sœur.

Table des matières

1 Réalité Augmentée : problématique et état de l'art **1**

1.1 Introduction 1

1.2 Etat de l'art sur les systèmes de RA 2

1.2.1 Calcul de la pose 2

1.2.2 Initialisation des systèmes de RA 3

1.2.3 Perception des systèmes de RA 4

1.3 Organisation du manuscrit 4

2 Description du système de RA utilisé **7**

2.1 Motivation des systèmes planaires 7

2.2 Le modèle sténopé 8

2.3 Relations de base : géométrie planaire 9

2.3.1 Homographies 9

2.3.2 Géométrie d'une vue d'une structure plane : les homographies scène-image . . 9

2.3.3 Géométrie de deux vues d'une structure plane : les homographies inter-images 10

2.3.4 Homographie induite par un plan : déplacement relatif entre deux vues 11

2.3.5 Stratification des matrices d'homographies 11

2.4 Méthode de suivi planaire 12

2.5 Approche utilisée : méthode de suivi multi-planaire 12

2.5.1 Schéma général 13

2.5.2 Méthode de calcul du point de vue par observation de plusieurs plans 13

2.5.3 Évaluation de l'approche multi-planaire de calcul du point de vue 14

3 Perception et stabilité dans un système de RA **17**

3.1 Critères d'évaluation d'un système de RA 17

3.1.1 L'erreur de mise en correspondance 18

3.1.2 L'incertitude sur les paramètres de calibrage 18

3.1.3 L'erreur de positionnement 3D 19

3.1.4 L'évaluation visuelle 20

3.2 Choix du modèle de caméra 24

3.2.1 Classement des modèles de caméra 24

3.2.2 Modèle de projection perspective 25

3.2.3 Modèles de projection affine et para-perspectif 27

3.2.4 Modèle de projection orthographique à l'échelle . . . 29
3.2.5 Modèle Euclidien . . . 31
3.2.6 Comparaison entre différents modèles de caméras . . . 31
3.3 Approches pour améliorer la stabilité . . . 32
3.3.1 Approche multi-image : l'ajustement de faisceaux . . . 32
3.3.2 Construction de cartes et localisation simultanées . . . 33
3.3.3 Filtrage itératif . . . 34
3.3.4 Sélection du modèle de mouvement . . . 34
3.4 Protocole de validation expérimentale . . . 35
3.5 Estimation statistique d'incertitudes . . . 36
3.5.1 Propagation directe . . . 36
3.5.2 Propagation inverse . . . 36

4 Calcul stable du point de vue de la caméra en utilisant la sélection de modèles 39
4.1 Introduction . . . 39
4.2 Schéma général . . . 40
4.3 Utilisation de méthodes de stabilisation . . . 40
4.3.1 Objectif . . . 40
4.3.2 État de l'art . . . 41
4.3.3 Notre approche de la sélection de modèles . . . 41
4.3.4 Résultats expérimentaux . . . 42
4.4 Amélioration de la sélection du modèle . . . 43
4.4.1 Méthode . . . 43
4.4.2 Ajustement . . . 44
4.4.3 Résultats . . . 44
4.5 Conclusion . . . 46

5 Étude pratique des erreurs de calibrage en RA 53
5.1 Introduction . . . 53
5.2 Protocole expérimental . . . 54
5.3 Choix du modèle de caméra . . . 54
5.3.1 Cas du modèle sténopé sans distorsion . . . 55
5.3.2 Prise en compte de la distorsion radiale . . . 57
5.3.3 Validation du modèle retenu . . . 57
5.4 Influence du calibrage . . . 58
5.4.1 Interprétation analytique . . . 58
5.4.2 Perturbation des intrinsèques . . . 58
5.4.3 Influence du point principal . . . 59
5.4.4 Influence de la distance focale . . . 63
5.4.5 Influence de la distorsion radiale . . . 63
5.5 Conclusion . . . 63

6 Méthodes de calcul des paramètres intrinsèques de la caméra **67**

6.1 Méthodes de calibrage interne 67

6.1.1 Techniques photogrammétriques 68

6.1.2 Auto-calibrage 68

6.1.3 Calibrage basé plans 69

6.1.4 Techniques basées sur des entités spécifiques dans la scène 70

6.2 Méthode des points de fuite 70

6.3 Calibrage interne basé plans 73

6.3.1 Relations de base du calibrage planaire 74

6.3.2 Méthode linéaire basée sur l'image de la conique absolue 77

6.3.3 Auto-calibrage itératif par filtrage de Kalman 78

6.3.4 Calibrage planaire basé sur la contrainte des lignes centrales 79

6.3.5 Auto-calibrage planaire basé sur la contrainte des lignes centrales 82

7 Extension et évaluation des méthodes de calibrage interne **87**

7.1 Pondération statistique des contraintes de calibrage 87

7.1.1 Pondération des contraintes associées à l'image de la conique absolue 90

7.1.2 Pondération des contraintes correspondant aux lignes centrales 90

7.1.3 Pondération pour la méthode d'auto-calibrage par lignes centrales 91

7.2 Intégration de structures multi-planaires aux méthodes de calibrage interne 92

7.2.1 Extension de la méthode de calibrage itératif 92

7.2.2 Extension de la méthode d'auto-calibrage par lignes centrales 93

7.3 Cadre d'évaluation des méthodes de calibrage interne 93

7.3.1 Évaluation sur des données de synthèse 94

7.3.2 Évaluation sur des images réelles 95

7.3.3 Extraction et mise en correspondance des points d'intérêt 97

7.3.4 Estimation des homographies et de leurs covariances 98

7.3.5 Vérité terrain 100

7.4 Comparaison des méthodes de calibrage 100

7.4.1 Méthode des points de fuite 100

7.4.2 Méthode de Sturm-Maybank 103

7.4.3 Méthode de Noirfalise 106

7.4.4 Calibrage basé sur la contrainte des lignes centrales 108

7.4.5 Calibrage monoplan vs. multiplan : un cas d'étude pratique 112

7.5 Discussion 113

8 Conclusion **117**

8.1 Apports de la thèse 117

8.2 Perspectives 118

Annexes **121**

A Filtre de Kalman étendu **123**

Bibliographie **125**

Table des figures

1.1 Continuum Réalité-Virtualité, classification de Milgram [MDG+95]. 1

2.1 Modèle sténopé de la caméra. Paramètres (a) extrinsèques et (b) intrinsèques. 8
2.2 Relations homographiques entre un plan et ses projections (scène-image) et entre les images. 10
2.3 Schéma général de la méthode de suivi planaire [SFZ00]. 12
2.4 Schéma général de la méthode de suivi multi-planaire [SB02a]. 13
2.5 Mire de calibrage composée par des surfaces planes. 14
2.6 La composante translationnelle t_x calculée au cours de la séquence en utilisant un, deux ou trois plans. 15
2.7 La composante translationnelle t_y calculée au cours de la séquence en utilisant un, deux ou trois plans. 16

3.1 Projection des cibles et des marqueurs du modèle de la scène sur une image réelle de la scène. À gauche, l'image augmentée complète ; à droite, les images amplifiées des trois marqueurs (gauche, central et droit) et de leurs correspondantes marques circulaires ajoutées. 21
3.2 Exemple de déplacement des points d'intérêt d'une scène suite à deux types de mouvement différents : (a) une rotation panoramique et (b) une translation parallèle à l'axe X du plan image. 22
3.3 Comparaison entre l'estimation faite au moyen d'une méthode de recalage séquentiel [SB02a] et la valeur réelle d'une des composantes du point de vue (t_x). L'écartement graduel entre les deux courbes met en évidence le problème de dérive des paramètres. 23
3.4 (a) Hiérarchie des modèles de caméra : (b) perspective, (c) affine et (d) orthographique à l'échelle. 26
3.5 Représentation géométrique des projections perspective, para-perspective et orthographique à l'échelle [HCD94]. 26

4.1 Schéma général de la méthode de recalage multi-planaire basée sur la sélection de modèle 40
4.2 La translation t_z calculée avec et sans le modèle prédictif. 43
4.3 Comparaison des critères de sélection sur une séquence synthétique bruitée $\sigma = 0.5$. . . . 46
4.4 Comparaison des critères de sélection sur une séquence synthétique bruitée $\sigma = 1$.. 47
4.5 Sélection du modèle en utilisant deux ou trois vues sur la séquence synthétique ($\sigma = 1$). . 47
4.6 Composante T_Z calculée en utilisant la sélection sur deux ou trois images pour la séquence synthétique. 48
4.7 Séquence de la mire : sélection de modèle avec deux ou trois images. 48
4.8 Séquence de la mire : la composante T_X calculée. 49
4.9 Séquence de la pièce : sélection du modèle. 49
4.10 Séquence de la pièce : la translation en x calculée avec la sélection de modèle. 50
4.11 Un cube incrusté sur l'image 200 : (a) sans sélection de modèle ; (b) : avec sélection . . . 50
4.12 Quelques exemples de la scène augmentée. La première colonne montre le suivi des points dans les trois plans utilisés, la second colonne montre la scène augmentée. 51

5.1 Le dispositif de contrôle à 4 degrés de liberté et la scène utilisée. 55

5.2 Valeurs des paramètres intrinsèques et intervalles de confiances, obtenus dans chaque image de la séquence, sans utiliser (première ligne) puis en utilisant (deuxième ligne) le modèle de distorsion. 56
5.3 Les valeurs du *gold standard* minimisent l'erreur d'alignement à l'intérieur et à l'extérieur du volume de mesure. 59
5.4 Valeurs de (λ_1, λ_2) obtenues dans chaque image de la séquence. 60
5.5 Points principaux obtenus en utilisant les paramètres de distorsion $L0$ à $L4$. 60
5.6 Influence du point principal sur l'alignement 2D. 61
5.7 (a) Erreur sur β obtenue avec le point principal PP2. (b) Erreur sur T_x obtenue avec la distance focale F2. 62
5.8 Évolution sur la séquence de (a) $|u - u_0|$ et (b) $1/Z$ pour le marqueur *gauche*. 62
5.9 Influence de la distance focale sur l'alignement 2D. 64
5.10 Influence des paramètres de distorsion sur l'alignement 2D. 65

6.1 Points de fuite obtenus à partir de l'image d'un rectangle ou d'un quadrillage. 71
6.2 Relation entre les points de fuite et le point principal. 72
6.3 La conique absolue, les points cycliques, et leurs images correspondantes. 75
6.4 Illustration du théorème de Poncelet [GP01]. 80
6.5 Contrainte de la ligne centrale. 82

7.1 Point principal estimé par la méthode des lignes centrales pondérées. 91
7.2 Génération d'images de synthèse : un motif carré. 94
7.3 Quelques images de la séquence de la mire. 96
7.4 Exemples des images réelles, utilisées dans le cadre d'évaluation expérimental 96
7.5 Séquence du *Paysage2* : Images clés pour les sous-séquences avec (a) trois plans et (b) un seul plan. Exemples de prises de vues pour l'ensemble avec (c) trois plans et (d) un plan. . 97
7.6 (a) Racine carré de la trace des matrices de covariance des homographies estimées directement par rapport à l'image clé (première image) et par composition de matrices, pour les séquences *Mire* et *Paysage*. (b) Nombre d'appariements correspondant. 99
7.7 Erreurs sur la distance focale en fonction de la position de la caméra (données de synthèse). 101
7.8 Erreurs sur la distance focale en fonction de la position de la caméra, lorsque la position du point principal a une erreur de $+20$ pixels sur u_0 et -20 pixels sur v_0. Écarts inférieurs à 1% ($\bullet$), écarts entre 1% et 10% positifs ($\cdot$) et négatifs ($\times$). 102
7.9 Estimation de la distance focale en fonction de l'erreur sur le point principal. Écarts inférieurs à 1% ($\bullet$), entre 1% et 10 % ($+$), et lorsque $f^2 < 0$ (distance focale imaginaire, $\circ$). 102
7.10 Méthode de Sturm-Maybank : paramètres intrinsèques calculés (f, u_0, v_0), lorsque τ est connu ; 104
7.11 Paramètres intrinsèques obtenus avec la méthode de Noirfalise pour la séquence *Paysage*, lorsque les homographies inter-images sont calculées (a) de forme itérative ou (b) si la mise en correspondance des points de calibrage a été directement effectuée par rapport à la première image. 108
7.12 Point principal estimé par la méthode des lignes centrales, en utilisant un seul plan ou plusieurs. 109
7.13 Lignes centrales pour les différents plans de la mire de calibrage, obtenues sur la séquence complète : (a) plan de la base, (b) plan à droite, (c) plan à gauche (l'image du fond correspond à la vue #38 de la séquence). 110
7.14 Point principal estimé par la méthode des lignes centrales, image par image et par filtrage de Kalman. 111
7.15 Comparaison des méthodes de calibrage interne, en utilisant les trois plans de la mire. . . 114

Chapitre 1

Réalité Augmentée : problématique et état de l'art

1.1 Introduction

La réalité augmentée (RA) a comme objectif l'insertion d'éléments virtuels générés par ordinateur sur une scène réelle vue par un utilisateur. En RA, les images composées peuvent ainsi afficher des informations que l'utilisateur ne peut percevoir par ses propres sens mais qui lui permettent d'améliorer sa performance lorsqu'il effectue certaines tâches.

Milgram et al [MDG+95] décrivent une taxonomie qui indique la relation entre la Réalité Augmentée et la Réalité Virtuelle (RV). Dans la réalité virtuelle, l'utilisateur est immergé dans un monde synthétique, et sa perception est isolée de l'environnement réel. Ils définissent le continuum Réalité-Virtualité (figure 1.1) de la manière suivante : le monde réel et un environnement complètement virtuel sont les deux extrémités de ce continuum, dont la région intermédiaire est nommée Réalité Mixte. Dans la réalité augmentée prédomine la perception du monde réel : les objets virtuels sont superposés à la scène réelle de manière à former une vue augmentée de la scène.

La réalité virtuelle est un domaine en constante évolution et des innovations aussi bien technologiques qu'académiques, dans les domaines de la vision par ordinateur et de la synthèse d'images, permettent d'avoir des environnements augmentés de plus en plus réalistes.

D'après [Azu97], un système de réalité augmentée doit posséder trois caractéristiques :

1. les objets virtuels sont intégrés dans un environnement réel, formant une vue augmentée du monde réel ;
2. la composition d'images est faite en temps réel et le système doit permettre une interactivité (l'utilisateur peut se déplacer dans l'environnement) ;
3. les objets virtuels sont modélisés en trois dimensions et positionnés dans un repère 3D associé à la scène.

Cette définition exclut les effets spéciaux réalisés pour certains films, car les vues augmentées ne sont pas obtenues en temps réel mais dans une étape de post-production.

Le premier système de réalité augmentée peut être attribué à Ivan Sutherland de l'université d'Harvard, dans les années 60 [Sut68]. Il a cependant fallu attendre les années 90 pour que la réalité augmentée



FIG. 1.1 – Continuum Réalité-Virtualité, classification de Milgram [MDG+95].

prenne son essor et soit considérée comme une discipline à part entière. L'intéret d'utiliser l'environnement de l'utilisateur plutot que de le resynthétiser est alors apparu comme un moyen efficace pour apprendre des procédures de maintenance ou pour guider les gestes d'un utilisateur.

Les applications potentielles du concept d'augmentation d'images sont très nombreuses [ABB+01]. Sans viser à l'exhaustivité, on peut citer les systèmes d'assistance à la maintenance [RSKM98, PHMG06], les systèmes d'assistance aux gestes médicaux-chirurgicaux [MVFCM03] et plus particulièrement les systèmes d'aide aux guidage d'outils (ponction, cathétarisation) [MFD+02, GKB+05], les systèmes d'apprentissage en conditions réalistes, sans oublier l'industrie du jeu [OSYT99], les applications militaires [JBL+00] et tout ce qui gravite autour du terme d'héritage culturel [VKA+01].

1.2 Etat de l'art sur les systèmes de RA

La réalité augmentée est un domaine pluri-disciplinaire qui nécessite la mise en commun de compétences dans des domaines très variés :

- **Vision par ordinateur** pour les aspects de suivi, de calcul de pose et également pour l'acquisition d'un modèle de la scène
- **Graphisme et visualisation** pour les problèmes de rendu des objets virtuels et pour la prise en compte des interactions lumineuses entre le monde réel et les objets virtuels. Ce domaine regroupe des recherches variées concernant le rendu lui même, l'inversion de phénomènes lumineux pour restituer correctement les ombrages sur les objets ajoutés [GCS+03], la mise en oeuvre de dispositifs de visualisation permettant d'améliorer la compréhension de la scène augmentée [BSN06]...
- **Interface** car il est nécessaire d'interagir avec le monde augmenté [BJK+06, BGL05]. Cette nécessaire interaction se situe aussi bien au niveau de l'initialisation du système (pose initiale, paramètres intrinsèques, re-initialisation après un arrêt de l'application) qu'au niveau de l'interaction avec les objets du monde augmenté.
- **La gestion de l'information** : concevoir une application de RA en vraie grandeur nécessite souvent de gérer et d'organiser de grandes quantités d'information. Dans le cas des systèmes de maintenance en particulier, il est nécessaire de savoir quelle information visualiser à un moment donné et comment aider l'utilisateur à y accéder facilement. Dans ce genre d'application, l'organisation et l'accès à l'information ont une importance primordiale, sans doute plus que la superposition cohérente de l'information dans le champ de vision de l'utilisateur.

D'autres domaines plus technologiques interviennent comme la mise en oeuvre de dispositifs d'acquisition sans fil, l'amélioration de l'autonomie des machines et la conception de dispositifs de visualisation (casque, lunette, miroirs,...) performants.

Le calcul de la pose de la caméra est évidemment la brique de base d'un système de réalité augmentée puisqu'il permet d'assurer la consistance géométrique de l'objet ajouté avec son environnement. Le problème du calcul de pose a ainsi focalisé l'essentiel des recherches en RA pendant longtemps conjointement avec le problème du calcul des inter-reflexions lumineuses entre objets réels et virtuels. Ce n'est que récemment que la problématique complète de construction de systèmes de réalité augmentée a été abordée en prenant en compte des applications effectives, et non plus jouet, et en y intégrant en particulier les aspects d'interaction avec le système et le choix de modes et de dispositifs de visualisation. Enfin des études sont actuellement menées afin de prouver, notamment dans les domaines médicaux, la valeur ajoutée des systèmes de réalité augmentée.

Dans cette thèse, nous ne nous intéressons qu'aux contributions de la vision par ordinateur aux systèmes de réalité augmentée. Nous proposons dans ce qui suit un état de l'art concernant les thématiques sur lesquelles la vision par ordinateur a une place particulièrement importante. Ces thèmes concernent les aspects d'initialisation des systèmes de RA, le problème du positionnement et enfin l'évaluation perceptive des systèmes de RA.

1.2.1 Calcul de la pose

En réalité augmentée, le calcul de la pose doit être causal, ce qui ne permet pas d'appliquer des méthodes de type *structure from motion* pour lesquelles plusieurs images de la scène doivent être dis-

ponibles en même temps. Deux catégories de techniques peuvent être employées pour le positionnement séquentiel : les techniques basées capteurs et les techniques basées modèle.

Les techniques basées capteurs reposent sur des mesures fournies par différents types de capteurs [Azu97] : magnétiques, acoustiques, inertiels, GPS, ... Ces méthodes présentent l'avantage d'être robustes aux mouvements abrupts de la caméra, mais les défauts d'être sensibles aux perturbations de l'environnement ou de portée limitée à un volume de taille réduite.

Les techniques basées modèle reposent sur l'appariement de primitives 3D de type points, courbes ou plans avec leurs homologues mesurées dans l'image. De nombreuses méthodes ont été proposées pour obtenir le point de vue de la caméra à partir de tels appariements [Low85, HJL+89, DD95, QL98, SB98, MC02].

Ces approches de suivi sont généralement plus précises que les procédés basés sur des capteurs, car elles reposent sur la minimisation directe de l'erreur d'alignement dans les images à augmenter. En revanche, elles sont moins robustes aux mouvements abruptes de la caméra car ces mouvements ne permettent pas de suivre les indices 2D entre images consécutives. D'autre part, il faut souligner que l'acquisition des connaissances 3D de l'environnement est une étape contraignante et parfois fastidieuse, qui n'est d'ailleurs pas toujours envisageable en pratique (par exemple dans des environnements vastes).

Pour tirer profit des avantages de chacune de ces techniques, la tendance actuelle est plutôt de coupler une approche basée vision avec une assistance basée capteur [SHC+96, YNA99, KD02, RLB+02, SAYT01, SUY04, ASB06].

Cette thèse s'appuie sur le système de positionnement multiplanaire développé dans l'équipe MAGRIT [SB02a] qui sera décrit dans le chapitre suivant, mais les résultats obtenus s'appliquent à la plupart des systèmes de positionnement développés par ailleurs.

1.2.2 Initialisation des systèmes de RA

Par initialisation d'un système de RA, nous entendons les étapes qui précèdent le calcul séquentiel de la pose et l'application de RA elle-même. Durant cette étape, les paramètres intrinsèques de la caméra sont en général fournis (la majorité des applications s'effectuent à focale constante) ainsi qu'une estimée initiale de la pose de la caméra. Le système de calcul séquentiel de pose peut alors démarrer à partir de ces données qui permettent de décrire complètement la caméra en position initiale. Ces étapes sont renouvelées à chaque fois que le calcul de pose est en échec au cours de l'application.

L'impact de cette étape d'initialisation est souvent peu abordé dans les papiers. Elle est pourtant importante puisqu'elle introduit des connaissances a priori sur la caméra qui seront conservées durant toute l'application. De plus, pour les systèmes fonctionnant par composition des déplacements relatifs de la caméra, un mauvais calcul de la pose initiale introduit un biais dans la localisation.

L'étape d'initialisation est le plus souvent réalisée à partir de l'observation d'une structure simple, par exemple un marqueur, qui est introduit dans la scène au début de l'application. Les coins du rectangle ou la contrainte d'orthogonalité des supports servent à estimer les intrinsèques de la caméra. Il faut cependant noter que le calcul est souvent fait à partir d'hypothèse restrictives sur la caméra, à savoir que le point principal est situé au milieu de l'image. Par ailleurs, la précision de la détection des indices images nécessaire au calcul des intrinsèques dépend de la taille du marqueur et de sa distance à l'objet. Enfin, la présence de distorsions au niveau de l'optique de la caméra peut également influer sur le calcul des intrinsèques. Il est donc a priori souhaitable d'étudier l'impact réel des procédures courantes d'initialisation sur le fonctionnement du système de RA, ce qui à notre connaissance, n'a jamais été fait.

Insérer des éléments dans l'environnement pour faciliter l'initialisation est une démarche utilisable en environnements industriels contrôlés. Elle devient contraignante voir inenvisageable lorsqu'on considère des environnements vastes ou extérieurs. Il devient alors nécessaire d'initialiser le système sur des structures présentes dans la scène. Les travaux sur la calibration automatique de caméra à partir de structures planes texturées [JSN04] vont dans ce sens et ont prouvé leur efficacité. Il est cependant nécessaire d'estimer leur pertinence dans un cadre pratique et d'estimer la fiabilité des paramètres estimés en fonction du contexte d'observation de la ou des structures planes afin de répondre aux questions suivantes : peut on déterminer des paramètres fiables à partir d'une seule structure ? à quelle distance doit se situer cette structure ? Est il préférable d'utiliser plusieurs structures disponibles dans la scène ? Ces questions, fondamentales dans la mise en oeuvre d'un système de RA, sont abordées dans la suite de cette thèse.

1.2.3 Perception des systèmes de RA

La réussite d'un système de RA est avant tout liée à la bonne compréhension qu'un individu a de la scène augmentée. Cette compréhension est évidemment vitale lorsqu'on aborde des applications médicales ou on aimerait avoir des garanties sur les images augmentées avant de les utiliser pour effectuer un geste médical précis [NPSA04].

Une bonne compréhension des scènes augmentées passe évidemment par un calcul de pose fiable et précis. Elle ne se limite cependant pas à cela puisque des aspects de visualisation rentrent en ligne de compte : le rendu de l'objet et la prise en compte plus ou moins réaliste des interactions physiques et lumineuses entre l'objet et la scène sont très importantes pour la compréhension de la scène [FGR93] : il est par exemple bien connu que la simulation des ombres portées de l'objet virtuel sur le monde réel améliore grandement la perception de la scène. Le problème de la perception de la profondeur dans les images augmentées a fait l'objet de diverses études [RMAR02]. Dans le cas des applications médicales où la perception de la profondeur est primordiale, l'étude de dispositifs et de modes de visualisation les plus aptes à permettre l'interprétation de la profondeur de la scène font actuellement l'objet d'études pratiques et théoriques approfondies. On y trouve l'étude de systèmes de visualisation classiques comme les Head mounted Display [RF00, SBHN06] ainsi que l'utilisation de dispositifs de visualisation nouveaux, tel que l'usage d'un miroir virtuel [BSN06].

Le cadre des applications de RA diffère cependant des applications médicales. L'observation de la scène augmentée dans le cadre médical est plus statique qu'il ne l'est dans les applications classique où l'utilisateur est libre de se mouvoir dans un environnement non délimité. Dans ce cas, la bonne perception dynamique de la scène devient un des critères importants de jugement d'un système de RA : la présence de sautillements ou de glissements dans l'image empêche la bonne perception de la scène augmentée. Ces problèmes d'instabilité de l'objet ajouté ont des origines diverses. La première est évidemment la difficulté à obtenir des correspondances modèles/image ou image/image. Cependant, même dans le cas où l'étape de correspondance se passe bien, on note des instabilités, particulièrement lorsque le mouvement de l'utilisateur est particulier (position fixe, mouvement en rotation). Les instabilités viennent alors du fait que le modèle de mouvement est trop général et les degrés de liberté supplémentaires par rapport au mouvement effectivement réalisés sont particulièrement bruités. Ce constat a été fait en particulier par Matsunaga et Kanitani [CM00] qui ont utilisé la sélection de modèle pour choisir un modèle de mouvement approprié. Dans le même ordre d'idée, afin de réduire le nombre de degrés de libertés des problèmes, divers travaux en réalité augmentée visent à utiliser des modèle de caméras moins généraux que le modèle projectif complet, par exemple affine [KV96]. L'étude de D. Cobzas [CJ03] vise par exemple à comparer l'impact visuel de l'utilisation de divers modèles de caméras (affine, projectif) sur les scènes augmentés.

Nous reviendrons sur ces problèmes de perception dans ce manuscrit à divers niveaux, d 'abord pour définir des critères permettant d'évaluer un système de de RA puis en travaillant sur la sélection de modèles pour améliorer la stabilité de l'incrustation.

1.3 Organisation du manuscrit

Le manuscrit est organisé de la façon suivante.

Le chapitre 2 décrit le système de positionnement développé et utilisé au sein du projet MAGRIT, ainsi que les fondements mathématiques de la géométrie planaire ;

Le chapitre 3 analyse les facteurs pouvant affecter la cohérence dans la perception des images augmentées (choix du modèle de caméra, sources d'imprécisions) et établit des critères permettant d'évaluer la précision des méthodes de calibrage ou de positionnement. Différentes techniques visant à améliorer la précision et la robustesse du calcul du point de vue sont aussi présentées.

Parmi ces techniques, différentes méthodes de sélection de modèles sont examinées au chapitre 4 et utilisées pour améliorer la stabilité de notre système de positionnement multiplan.

Au chapitre 5, la pertinence du modèle sténopé de caméra et l'influence du calibrage sur la cohérence des images augmentées (selon les critères définis au chapitre 3) sont évaluées de manière expérimentale.

Les deux derniers chapitres avant la conclusion sont dédiés au calcul des paramètres intrinsèques de

la caméra basé sur des structures planes : le chapitre 6 présente les fondements du calibrage monoplan et quelques approches existantes ; le chapitre 7 présente deux extensions que nous proposons à ces méthodes en utilisant des approches multiplans et une pondération statistique des données d'entrée. Les différentes méthodes présentées sont comparées de manière expérimentale.

Chapitre 2

Description du système de RA utilisé

2.1 Motivation des systèmes planaires

Les applications de la RA requièrent que la scène augmentée soit continûment mise à jour en fonction des déplacements de l'utilisateur sur la scène. Cela signifie que le point de vue de la caméra doit être calculé en permanence et en temps réel, et dans le cas des méthodes de positionnement basé vision, en fonction des éléments que la caméra observe sur la scène.

Le choix d'un système de RA doit prendre en compte les conditions pratiques de l'application tels que : l'environnement d'utilisation, la simplicité de mise en œuvre et les moyens matériels nécessaires pour le système (objets de calibrage, modèles préalables).

Dans le cas idéal, un système de RA devrait être capable de fonctionner dans tout type d'environnement, sans préparation préalable de la scène, et permettre à l'utilisateur de se déplacer librement dans l'environnement.

Dans le passé, de nombreuses approches ont cherché à faciliter le calcul de la position de la caméra en préparant la scène avec des marqueurs spéciaux et en suivant les positions de ces marqueurs [KB99, Nav03]. D'autres techniques utilisent la connaissance de certaines structures de la scène pour construire un modèle qui sera suivi de manière à obtenir le point de vue de la caméra [SB98, DC99, MC02]. Ainsi Drummond [DC99] utilise un modèle filaire partiel de la scène pour le calcul de la pose. Cependant, ces approches limitent la flexibilité des systèmes de RA ou nécessitent d'acquérir des connaissances métriques sur la scène. Il est donc important de développer des méthodes fonctionnant dans des environnements qui ne requiérent pas d'être préparés et qui ne demandent qu'une faible participation de l'utilisateur pour la construction du modèle de scène utilisé.

Les méthodes appelés *structure from motion* permettent de retrouver la structure de la scène en même temps que le mouvement de la caméra, sans aucune information préalable sur l'environnement observé sauf l'hypothèse de rigidité de la scène. Cependant, il n'est pas toujours possible de calculer précisement le point de vue de la caméra en utilisant ces méthodes, car les estimations peuvent s'avérer très sensibles aux erreurs de mesure des indices.

S'il est difficile d'acquérir certaines données métriques sur l'environnement, il est par contre assez simple d'acquérir un modèle planaire d'une scène. Une pièce d'intérieur peut être ainsi facilement modélisée par le plan du sol décrit par l'équation $Z = 0$, le support des murs étant modélisé s par des plans orthogonaux $X = 0$ et $Y = 0$.

Les plans sont des éléments géométriques fréquents dans les environnements fabriqués par l'homme. Les approches de positionnement basées plans consistent à suivre un ensemble d'indices coplanaires et à les caractériser la transformation géométrique reliant ces indices entre deux vues (cette transformation est appelée *homographie*).

Les approches basées plans sont plus flexibles que les systèmes basés modèles classiques, car la mise en correspondance entre les plans peut être effectué à partir de leurs projections 2D sur les images, au lieu d'exiger une connaissance des positions 3D d'un modèle de la scène.

Dans les méthodes planaires, la mise en correspondance des points d'intérêt est contrainte par l'hy-

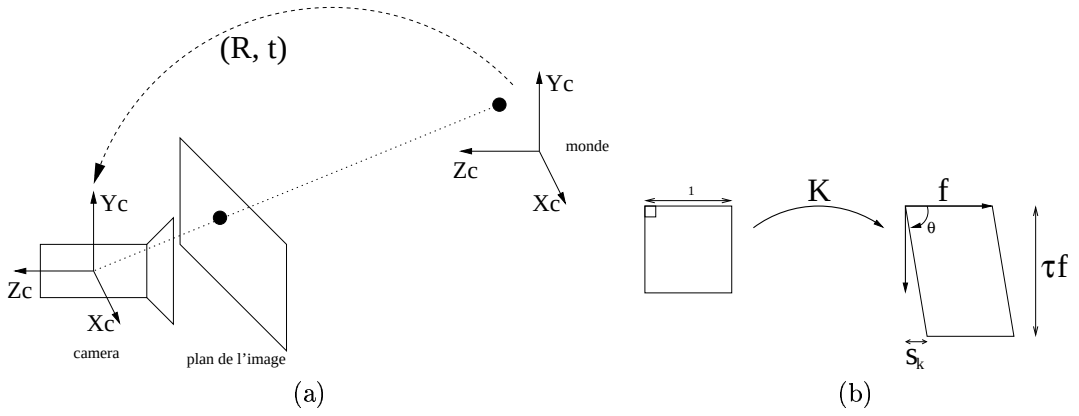


FIG. 2.1 – Modèle sténopé de la caméra. Paramètres (a) extrinsèques et (b) intrinsèques.

pothèse de coplanarité et elle est plus robuste que les relations utilisées pour des points quelconques. La robustesse de ces approches est due au fait que les projections des indices coplanaires sur deux vues d'une scène sont reliées par une correspondance biunivoque, à la différence de la géométrie épipolaire [LF97a] qui ne désigne pas une correspondance unique mais un ensemble de positions possible où un indice détecté dans une image peut être aperçu sur une autre vue.

Les techniques planaires figurent dans une position intermédiaire entre les méthodes basées modèles et les méthodes de *structure from motion*. Elles sont en général plus flexibles que les premières et plus robustes que les dernières.

Les méthodes de positionnement basées plans semblent une approche avantageuse pour les applications de la RA parce qu'elles ne requièrent ni l'utilisation de marqueurs spéciaux ni de préparer la scène. Ces approches conviennent à une grande diversité de scènes réelles où on trouve des surfaces planes texturées, autant en intérieur qu'en extérieur : bureaux, chambres, environnements de type urbain, façades de batîments, terrains plans, etc.

Pour ces raisons, une méthode de positionnement basée sur l'observation de plans a été développée au sein de l'équipe. Comme la précision du positionnement à partir d'un seul plan est souvent insuffisante, ce système intègre la connaissance de plusieurs plans présents dans la scène.

Par la suite, nous passons en revue les fondements théoriques utilisés dans cette thèse (sections 2.2 et 2.3), nous présentons la méthode basée plans employée (section 2.5), et enfin des expériences montrent la fiabilité de cette approche et son intérêt pour des applications en temps réel (section 2.5.3).

2.2 Le modèle sténopé

Le modèle sténopé d'une caméra (figure 2.1) projette un point 3D $\mathbf{x} = (x, y, z)$ en un pixel de coordonnées $\mathbf{u} = (u, v)$, suivant l'équation :

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim \mathbf{K}\,(\mathbf{R} \quad \mathbf{t}) \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix},$$

La matrice $\mathbf{P} = \mathbf{K}\,(\mathbf{R} \quad \mathbf{t})$ de dimensions 3×4 est connue comme la *matrice de projection* de la caméra.

Dans ce modèle, $\mathbf{R}$ et $\mathbf{t}$ sont respectivement la matrice de rotation et le vecteur de translation d'une transformation rigide (figure 2.1a) qui permet de passer du repère du monde au repère de la caméra : $\mathbf{x}_c = (\mathbf{R} \quad \mathbf{t})\,\tilde{\mathbf{x}}$. Les six paramètres qui définissent cette transformation (3 angles pour caractériser la matrice de rotation et 3 composantes du vecteur de translation) sont appelés *paramètres extrinsèques*.

La matrice des paramètres intrinsèques $\mathbf{K}$ permet de transformer les coordonnées $\mathbf{x}_c$ en coordonnées pixel sur un repère 2D qu'on appelle *repère de la rétine* ou *plan de l'image* (figure 2.1b) :

$$\tilde{\mathbf{u}} = \mathbf{K}\,\mathbf{x}_c$$

La matrice $\mathbf{K}$ est caractérisée par 5 paramètres :
- la distance focale f mesurée en taille horizontale d'un pixel,
- les deux coordonnées du point principal (u_0, v_0),
- le rapport d'aspect τ entre les dimensions d'un pixel : hauteur/largeur,
- le coefficient d'obliquité s_k entre les axes vertical et horizontal du pixel.

$$\mathbf{K} = \begin{pmatrix} f & s_k & u_0 \\ 0 & \tau f & v_0 \\ 0 & 0 & 1 \end{pmatrix}$$

Des travaux antérieurs ont montré que l'angle entre les deux axes d'un pixel est généralement très proche de 90° (c'est à dire, $s_k \approx 0$) et que le rapport d'aspect est en général invariable aux changements de zoom et à la mise au point de la caméra [EVF93, Wil94].

Lorsque ces deux paramètres sont connus, on peut appliquer une transformation affine sur le modèle sténopé pour obtenir un modèle avec des pixels carrés ($\tau = 1$, $s_k = 0$), appelé le modèle de *caméra naturelle* [LZ99], dont la matrice de paramètres intrinsèques comporte exclusivement la distance focale et les coordonnées du point principal.

2.3 Relations de base : géométrie planaire

2.3.1 Homographies

Une *homographie* (appelée aussi *transformation projective*) est une fonction de transfert h de l'espace projective P^2 vers le même espace, de telle manière que trois points colinéaires $\mathbf{x}_1$, $\mathbf{x}_2$ et $\mathbf{x}_3$ quelconques, sont toujours menés vers trois points $h(\mathbf{x}_1)$, $h(\mathbf{x}_2)$ et $h(\mathbf{x}_3)$, également colinéaires [HZ00].

Une fonction de transfert h est une homographie dans l'espace projective P^2 si et seulement s'il existe une matrice $\mathbf{H}$ de dimensions 3×3 de telle manière que tout point $\mathbf{x} \in P^2$ satisfasse l'égalité $h(\mathbf{x}) = \mathbf{H}\mathbf{x}$.

Les transformations homographiques préservent la colinéarité mais elles ne conservent en général ni les distances ni les angles. À présent, nous rappelons les propriétés fondamentales des transformations projectives des éléments liées aux structures planes.

2.3.2 Géométrie d'une vue d'une structure plane : les homographies scène-image

Les points sur un plan Π de la scène sont liés aux points sur le plan image par une *transformation projective plane* (dites *homographie plan à plan*). La représentation de cette transformation ainsi que ses propriétés sont bien connues dans la littérature de la géométrie projective [SK52] et la vision par ordinateur [MZ92][HZ00].

Sans perte de généralité, on choisit un système de référence $\mathcal{R}_w$ dans l'espace, de telle manière qu'un point $\mathbf{X}_\Pi$ quelconque situé sur le plan Π ait les coordonnées homogènes $\tilde{\mathbf{X}}_\Pi = (X, Y, 0, 1)^T$, c'est à dire, l'équation du plan est $\{\Pi : Z = 0\}$. Si le repère du monde $\mathcal{R}_w$ est lié au repère de la caméra $\mathcal{R}_c$ par une transformation rigide $(\mathbf{R} \quad \mathbf{t})$, l'équation du plan dans le repère de la caméra est

$$\mathbf{r}_3 \cdot (\mathbf{X}'_c - \mathbf{t}) = 0 \tag{2.1}$$

où $(\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{r}_3)$ sont les vecteurs colonnes de la matrice de rotation $\mathbf{R}$.

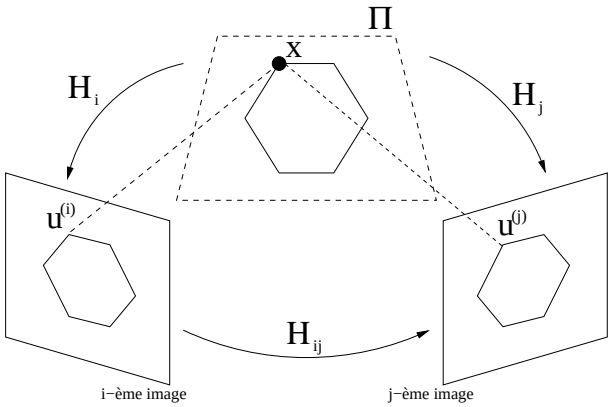


FIG. 2.2 – Relations homographiques entre un plan et ses projections (scène-image) et entre les images.

Étant donnée la matrice de paramètres intrinsèques de la caméra $\mathbf{K}$, la projection $\tilde{\mathbf{u}}$ du point $\tilde{\mathbf{X}}_\Pi = (X, Y, 0, 1)^T$ sur le plan image est déterminée par :

$$\tilde{\mathbf{u}} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \sim \mathbf{K} \begin{pmatrix} \mathbf{R} & \mathbf{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 0 \\ 1 \end{pmatrix} = \mathbf{K} \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{pmatrix} \begin{pmatrix} X \\ Y \\ 1 \end{pmatrix} \tag{2.2}$$

De cette manière, les coordonnées du point $\tilde{\mathbf{x}} = (X, Y, 1)^T$ (exprimées dans le repère 2D du plan Π) sont liées au point image $\tilde{\mathbf{u}}$ par la matrice 3×3 :

$$\mathbf{H} \sim \mathbf{K} \begin{pmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \mathbf{t} \end{pmatrix} \tag{2.3}$$

appelée l'*homographie scène-image*, tel que $\tilde{\mathbf{u}} \sim \mathbf{H}\tilde{\mathbf{x}}$ transforme tout point du plan Π en son point correspondant de l'image.

Les homographies possèdent 8 degrés de liberté et elles sont définies à un facteur d'échelle près. Elles peuvent être déterminées à partir d'au moins 4 points connus sur le plan Π et leurs projections sur le plan image.

Singularités des homographies scène-image

Une transformation homographique est bijective lorsque la matrice $\mathbf{H}$ correspondante n'est pas singulière, ce qui veut dire que $\det(\mathbf{H}) \neq 0$. Comme $\det(\mathbf{K}) = \tau f^2 \neq 0$ (si l'optique n'a pas de dégénérescences), une singularité sur l'homographie scène-image $\mathbf{H}$ se produit uniquement si $\det(\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}) = 0$. Comme

$$\det(\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}) = (\mathbf{r}_1 \times \mathbf{r}_2) \cdot \mathbf{t} = \mathbf{r}_3 \cdot \mathbf{t} = 0$$

et si l'on se reporte à l'équation (2.1), on déduit que $\mathbf{r}_3 \cdot \mathbf{X}'_c = 0$, ce qui veut dire que le centre de projection de la caméra appartient au plan Π. En conséquence, les singularités dans les homographies scène-image apparaissent uniquement si le plan Π est vu par sa tranche depuis le centre optique de la caméra.

2.3.3 Géométrie de deux vues d'une structure plane : les homographies inter-images

Un point $\mathbf{x}$ appartenant au plan Π est projeté sur l'image i dans la position $\tilde{\mathbf{u}}^{(i)} \sim \mathbf{H}_i\tilde{\mathbf{x}}$, et sur l'image j dans $\tilde{\mathbf{u}}^{(j)} \sim \mathbf{H}_j\tilde{\mathbf{x}}$. Lorsque la matrice $\mathbf{H}_j$ n'est pas singulière, le point $\mathbf{x}$ est rendu par la transformation $\tilde{\mathbf{x}} \sim \mathbf{H}_i^{-1}\tilde{\mathbf{u}}^{(i)}$ (d'après l'image i), et par conséquent, les correspondances sur ces deux images sont liées par une troisième transformation $\tilde{\mathbf{u}}^{(j)} \sim \mathbf{H}_j\mathbf{H}_i^{-1}\tilde{\mathbf{u}}^{(i)}$.

La matrice $\mathbf{H}_{ij} = \mathbf{H}_j\mathbf{H}_i^{-1}$ (nommée *homographie inter-image*) relie les projections sur deux vues des points d'un plan dans l'espace, c'est à dire, le plan induit une homographie entre les deux vues (figure 2.2).

La transformation entre les projections sur deux vues de points coplanaires est bijective si aucun des deux centres optiques des caméras n'est contenu dans le plan observé. Les homographies inter-images ont aussi 8 degrés de liberté, et peuvent être calculées à partir des correspondances sur les deux images d'au moins 4 points.

2.3.4 Homographie induite par un plan : déplacement relatif entre deux vues

Les homographies inter-images dépendent de la transformation géométrique reliant les deux repères caméras et des paramètres intrinsèques des caméras. Nous supposons que la position et l'orientation de la caméra sont connus pour la première image de la séquence ; sans perte de généralité, nous pouvons attribuer $(\mathbf{R}_1 \quad \mathbf{t}_1) = (\mathbf{I} \quad \mathbf{0})$ aux paramètres intrinsèques de la première image.

Étant données deux matrices de projection $\mathbf{P}_1 = \mathbf{K}_1[\mathbf{I}|\mathbf{0}]$ et $\mathbf{P}_2 = \mathbf{K}_2[\mathbf{A}|\mathbf{a}]$ et un plan défini par le vecteur v tel que $\mathbf{v}^T\mathbf{x} + 1 = 0$, alors l'homographie induite par ce plan s'exprime sous la forme [HZ00] :

$$\mathbf{H} = \mathbf{K}_2(\mathbf{A} - \mathbf{a}\mathbf{v}^T)\mathbf{K}_1^{-1}$$

où $\mathbf{A}$ et $\mathbf{a}$ sont les matrices de rotation et le vecteur de translation entre les deux vues, et $\mathbf{K}_i$ est la matrice des paramètres intrinsèques associée à l'image i.

2.3.5 Stratification des matrices d'homographies

Les matrices d'homographie entre un plan de la scène et l'image contient l'information nécessaire pour récuperer les propriétés métrique et affine du plan dans la vue correspondante.

D'après [HZ00], une matrice d'homographie $\mathbf{H}$ est décomposable par le produit de trois matrices :

$$\mathbf{H} = \mathbf{H}_p\mathbf{H}_a\mathbf{H}_s$$

où $\mathbf{H}_p$ est sa composante projective, $\mathbf{H}_a$ est sa composante affine, et $\mathbf{H}_s$ est sa composante euclidienne.

La composante projective $\mathbf{H}_p$ est en général déterminée par une transformation qui transfère la ligne de fuite du plan sur la vue donnée vers la droite à l'infini. Une transformation qui satisfait cette propriété est :

$$\mathbf{H}_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \lambda & 1 \end{pmatrix}$$

où $(\mu, \lambda, 1)^T$ correspond à la ligne de fuite associée au plan et est définie donc à deux dégrés de liberté.

Les transformations affines préservent la colinéarité des points et les proportions entre distances. Les propriétés affines de l'homographie peuvent être dissociées en deux composantes : une matrice triangulaire supérieure déterminée à deux degrés de liberté

$$\mathbf{H}_a = \begin{pmatrix} \alpha & \beta & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

et une similarité dans l'espace euclidien 2D

$$\mathbf{H}_s = \begin{pmatrix} s\mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix}$$

avec quatre dégrés de liberté : s un facteur d'échelle, un angle déterminant la rotation 2D $\mathbf{R}$, et deux composantes du vecteur de translation $\mathbf{t}$.

Étape de pré-traitement :

1. Indiquer manuellement sur la première image, la région plane à suivre.
2. Calculer la matrice de paramètres intrinsèques $\mathbf{K}$.
3. Calculer la matrice d'homographie entre la scène et l'image 0 : $\mathbf{H}^{(0)}$

Calcul de la matrice d'homographie entre l'image $i-1$ et i pour $i > 0$:

1. Détecter les points d'intérêt de l'image i appartenant au plan de référence.
2. Mettre en correspondance les points d'intérêt entre les images $i-1$ et i.
3. Calculer l'homographie induite par le plan entre les vues $i-1$ et i de manière robuste (en utilisant la méthode RANSAC [FB80]) : $\mathbf{H}_{i-1,i}$.
4. Calculer la matrice $\mathbf{H}^{(i)}$ de manière itérative : $\mathbf{H}^{(i)} = \mathbf{H}_{i-1,i}\mathbf{H}^{(i-1)}$.
5. Calculer le point de vue à partir de $\mathbf{H}^{(i)}$.

FIG. 2.3 – Schéma général de la méthode de suivi planaire [SFZ00].

2.4 Méthode de suivi planaire

Pluieurs systèmes de positionnement utilisant la présence de structures planes ont été proposés. Mendelsohn et Daniilidis [MDB98] utilisent la présence de marqueurs plans pour le recalage en RA. Simon et al [SFZ00] ont ensuite décrit un système de positionnement de la caméra basé sur l'observation d'une surface plane naturelle. Leur méthode requiert qu'un plan de la scène soit toujours visible pendant le suivi.

Dans cette méthode, dont le schéma général est rappelé en figure 2.3, le calcul du déplacement relatif entre deux vues est totalement automatique et robuste en raison de la contrainte homographique. Si le point de vue de la caméra est connu dans la première vue, la position de la caméra peut être calculée pour chaque vue de la séquence en utilisant seulement les homographies entre images images consécutives. De cette manière, on obtient une méthode séquentielle pour calculer automatiquement le point de vue de la caméra le long de la séquence.

Ils proposent une démarche simple pour la phase de pré-traitement : le calcul de la première homographie scène-image peut être obtenu à l'aide d'un contour rectangulaire et ce même contour permet de calculer la distance focale lorsque les coordonnées du point principal et le rapport d'aspect sont connus [HZ00, SFZ00].

Le processus de suivi est décrit par la boucle suivante : (i) la détection des points d'intérêt (ii) et leur mise en correspondance entre deux images consécutives, (iii) le calcul de la homographie induite par le plan sur les deux vues, (iv) le calcul de la homographie entre la scène et l'image courante par composition des matrices entre vues, (v) le calcul du point de vue de la caméra à partir de la homographie scène-image [HZ00, SFZ00].

2.5 Approche utilisée : méthode de suivi multi-planaire

Dans nos travaux, nous reprenons la méthode de suivi proposée par Simon et Berger [SB02a] basée sur l'observation de plusieurs plans dans la scène, qui étend le principe du suivi d'un seul plan [SFZ00] décrit dans la section 2.4.

Nous avons pris cette approche comme point de départ pour la recherche des stratégies améliorant la précision et la stabilité du calcul du point de vue, ce qui sera le sujet d'étude du chapitre 4.

Cette section décrit les grandes lignes de la méthode de suivi proposée par [SB02a] (fig. 2.4) et présente à la fin une comparaison entre les approches de calibrage mono-planaire et multi-planaire.

Étape de pré-traitement :

1. Fournir l'équation des plans observés et utilisés pour le recalage.
2. Calculer la matrice de projection pour la première image $\mathbf{P}^{(0)}$,

Calcul de la matrice de projection $\mathbf{P}^{(i)}$ pour $i > 0$:

1. Détecter les points d'intérêt de l'image i.
2. Déterminer l'ensemble des points d'intérêt en correspondance entre les images $i-1$ et i pour chaque plan observé.
3. Calculer les homographies induites par chaque plan entre $i-1$ et i.
4. Estimer le déplacement relatif $\mathbf{T}_{i-1,i}$ entre deux vues consécutives, à partir des n homographies obtenues.
5. Calculer la matrice $\mathbf{P}^{(i)}$ à partir de la matrice de projection précédente $\mathbf{P}^{(i-1)}$ et le déplacement estimé : $\mathbf{P}^{(i)} = \mathbf{T}_{i-1,i}\mathbf{P}^{(i-1)}$.

FIG. 2.4 – Schéma général de la méthode de suivi multi-planaire [SB02a].

2.5.1 Schéma général

La position de la caméra pour la première image de la séquence est déterminée, le plus souvent à l'aide d'un poster rectangulaire fixé dans la scène, en utilisant les contraintes d'orthogonalité. On suppose que la structure multi-planaire de la scène est décrite par un ensemble de polygones 3D $\mathcal{L}_p (1 \leq p \leq n)$. On suppose aussi que les équations des plans de la scène $\{v_k : v_k \cdot \mathbf{x} + 1 = 0\}$ sont déterminés préalablement (par exemple, à l'aide de la méthode décrite en [SB02b]).

Ces étapes de pré-traitement accomplies, le processus de suivi décrit une boucle comportant cinq étapes : (i) les points d'intérêt sont extraits et (ii) mis en correspondance entre deux images, (iii) les homographies correspondant à chacun des plans observés sont calculées, (iv) à partir des homographies des plans, on estime le déplacement relatif entre deux vues, (v) le point de vue de la caméra est estimé par la composition entre la position de la caméra à l'image précédente et le déplacement relatif calculé.

L'algorithme d'estimation du déplacement relatif entre deux vues est décrite ci-dessous.

2.5.2 Méthode de calcul du point de vue par observation de plusieurs plans

Étant donnés les paramètres intrinsèques et un ensemble de points en correspondance entre deux images, une méthode classique pour calculer le point de vue est de minimiser l'erreur moyenne entre les points en correspondance par rapport aux paramètres de la transformation [CM00] :

$$\mathbf{A}, \mathbf{a} = \arg\min J(\mathbf{A}, \mathbf{a}) = \frac{1}{N} \sum_{j=1}^{N} \|x'_j - Z(Hx_j)\|^2$$

où Z est la fonction de normalisation qui transforme la troisième composante d'un vecteur 3D en 1. Nos expérimentations ont montré que la précision obtenue grâce à l'observation d'un seul plan n'est pas suffisante pour obtenir une bonne impression visuelle de la scène augmentée. En effet, la précision dépend des positions relatives de la caméra par rapport au plan observé et également du nombre de points en correspondance déterminés sur ce plan. De plus, les erreurs s'accumulent au fil du temps puisque la matrice de projection à l'étape i est obtenue par composition de tous les déplacements relatifs la précédant. Pour de longues séquences, l'accumulation des erreurs peut ainsi conduire le processus de calcul à diverger.

Pour cette raison, [SB02a] ont proposé d'utiliser plusieurs plans pour le recalage. Ceci donne en effet plus d'informations sur l'espace tridimensionnel considéré et va conduire à réduire la variabilité des paramètres estimés. Ceci permet également de prendre en compte des environnements plus larges et des mouvements plus complexes. En effet, lors de certains mouvements, le plan observé peut devenir peu visible ou conduire à des points mal détectés. L'utilisation de plusieurs plans permet d'apporter une solution à ce problème, l'un au moins des plans considérés étant généralement bien visible dans l'image.

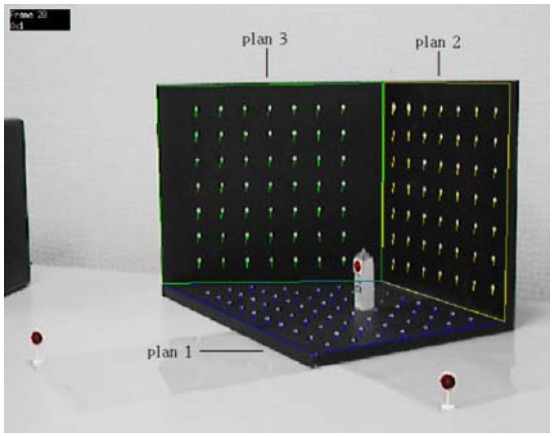


FIG. 2.5 – Mire de calibrage composée par des surfaces planes.

# Image	Mouvement
0 - 19	Rotation et Translation
20 - 29	Stationnaire
30 - 39	Translation selon l'axe des Y
40 - 49	Translation selon l'axe des X
50 - 64	Panoramique
65 - 69	Stationnaire

TAB. 2.1 – Description des mouvements de la caméra au cours de la séquence.

Dans le cadre multi-planaire, la fonction de coût utilisée est alors :

$$J(\mathbf{A}, \mathbf{a}) = \frac{1}{N_1 + \cdots + N_n} \sum_{k=1}^{n} \sum_{j=1}^{N_k} \|x'_{kj} - Z(H_k x_{kj})\|^2$$

où n est le nombre de plans, N_k est le nombre de points en correspondance pour le plan k, H_k est l'homographie planaire induite par le plan k ($H_k = K(A - av_k^T)K^{-1}$), et Z est la fonction de normalisation qui transforme la troisième composant d'un vecteur 3D en 1.

Une méthode classique pour optimiser cette fonction non linéaire est la méthode de Newton mais elle est sensible à l'estimée initiale. C'est pourquoi nous utilisons plutôt la méthode de Levenberg-Marquardt qui est plus stable et presque aussi rapide que la méthode de Newton.

2.5.3 Évaluation de l'approche multi-planaire de calcul du point de vue

Afin de prouver l'efficacité de la méthode de calcul du point de vue basée sur l'observation de plusieurs plans (figure 2.4, proposée en section 2.5.1), nous avons d'abord considéré une séquence d'images synthétiques utilisant les coordonnées des cibles d'une mire de calibrage réelle (figure 2.5) et les projetant sur les images sous différentes positions.

La scène est donc constituée de trois plans d'équations plan 1, $X - 0.577Y = 0$; plan 2, $Y = 0$; et plan 3, $Z = 0$. Le mouvement de la caméra dans cette séquence est décrit dans la la table 2.1. De nombreux types de mouvements sont considérés : translationnels, panoramiques et stationnaires. Cette séquence sera utilisée tout au long de ce chapitre pour valider nos différents apports.

La méthode a été testée en bruitant les points images avec différents écart-types. Les figures 2.6 et 2.7 montrent les composantes translationnelles t_x et t_y calculées avec un bruit additif d'écart type 0.5, ainsi que les valeurs réelles de t_x et t_y. Ces graphiques montrent que l'utilisation de plusieurs plans améliore notablement la précision du point de vue calculé. Plus précisément, ces graphiques montrent l'amélioration de la précision lorsque un, deux ou trois plans sont utilisés pour le calcul du point de vue.

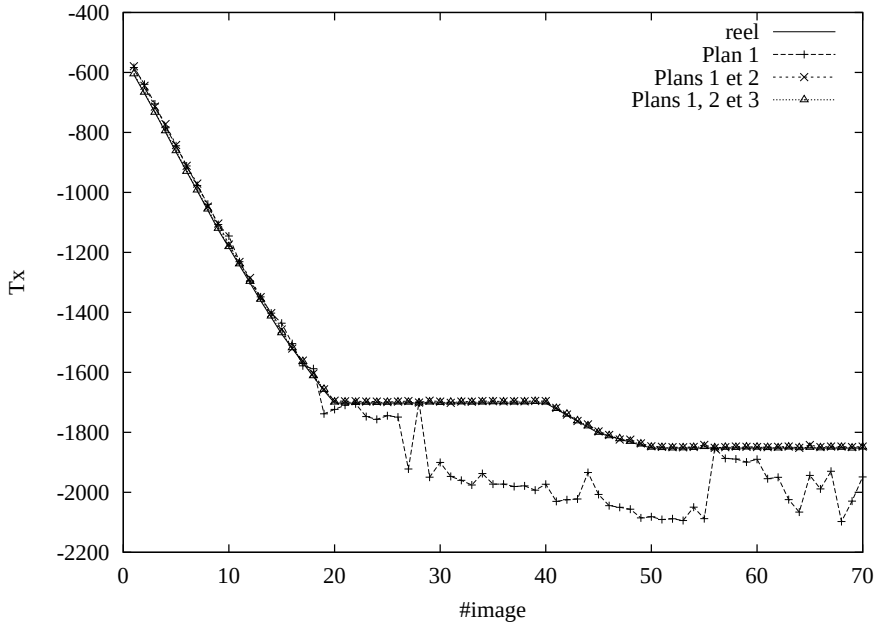


FIG. 2.6 – La composante translationnelle t_x calculée au cours de la séquence en utilisant un, deux ou trois plans.

Ces expérimentations prouvent que l'utilisation d'un seul plan conduit à des estimations assez instables du point de vue. L'utilisation de deux ou trois plans accroît considérablement la précision et la stabilité de l'estimation.

Afin de quantifier les performances de l'algorithme en terme de temps de calcul, la table 2.2 fournit le temps passé dans les différentes étapes de l'algorithme pour une image de la mire de calibrage de taille 765×576 : extraction et mise en correspondance des points d'intérêt (nous utilisons ici l'algorithme MIC [TH98] pour l'extraction des points), extraction des points appartenant aux structures planaires grâce a un algorithme de type RANSAC et enfin calcul du point de vue. Environ 500 points d'intérêt sont extraits dans chaque image. Parmi ceux ci, seuls une centaine appartiennent aux plans considérés et entrent dans le processus de calcul du point de vue. Le temps total est donc d'environ 66 ms, ce qui conduit à un calcul du point de vue pouvant traiter environ 16 images par seconde.

MIC	30,33 ms
Mise en correspondance	18,62 ms
RANSAC	12,13 ms
Calcul du point de vue	4,8 ms
Total	65,88 ms

TAB. 2.2 – Temps de calcul pour le processus complet pour un Pentium III 900 Mhz (moyenne réalisée sur 100 images).

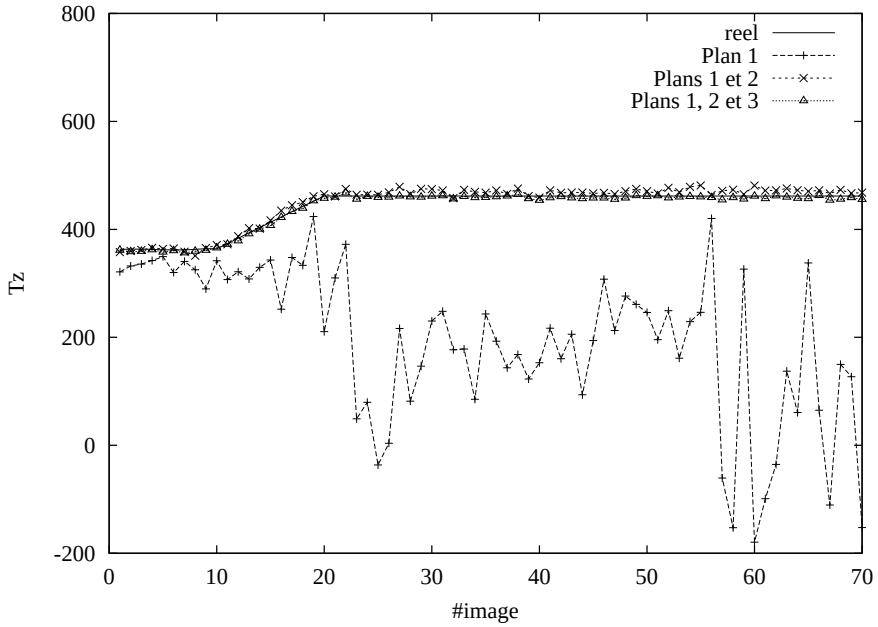


FIG. 2.7 – La composante translationnelle t_y calculée au cours de la séquence en utilisant un, deux ou trois plans.

Chapitre 3

Perception et stabilité dans un système de RA

Les méthodes de recalage par vision, de même que la fusion d'images réelles et virtuelles, requièrent une modélisation du processus d'acquisition d'image qui soit conforme à la réalité. Le choix du modèle de caméra et des méthodes utilisées pour en déterminer les paramètres est donc un choix très important dans la conception d'un système de RA, puisque des inconsistances dans le processus de recalage ou de fusion peuvent générer pour l'observateur des effets nuisant l'impression visuelle. Par exemple, si la caméra est stationnaire et que les objets virtuels sont sensés être immobiles dans les images de la séquence augmentée, une petite erreur sur les paramètres de la caméra peut donner l'impression à l'observateur que les objets ajoutés tremblent dans la scène.

Il est nécessaire d'analyser les facteurs qui peuvent affecter la cohérence dans la perception des images augmentées, ainsi que d'établir des critères mesurant la performance des systèmes de RA. Ceux-ci sont les objectifs de ce chapitre.

D'abord, nous présentons les critères qui nous permettront d'évaluer la performance d'un système de RA. Dans la section 3.2, nous évoquons quelques modèles de caméras souvent utilisés dans les applications de vision par ordinateur, et nous discutons de l'adéquation de ces modèles aux différentes conditions d'utilisation. La section 3.3 décrit certaines techniques, abordées dans la littérature du domaine, qui permettent d'incorporer des connaissances a priori (par exemple, sur les mouvements de la caméra ou la structure de la scène) afin d'améliorer la stabilité des résultats. Dans la section 3.4, nous proposons une démarche expérimentale pour comparer les résultats issus des systèmes de recalage étudiés dans nos travaux selon les critères abordés dans ce chapitre. À la fin du chapitre, nous faisons un rappel de la méthode d'estimation d'incertitudes de paramètres, l'un de critères d'évaluation cités dans ce chapitre.

3.1 Critères d'évaluation d'un système de RA

Pour examiner les facteurs qui affectent la performance d'un système de recalage, il est fondamental d'établir des critères objectifs pour l'évaluation et la comparaison des modèles et des paramètres de calibrage. Cependant, une mesure unique ne suffit pas pour comparer la performance des divers systèmes, lorsque leurs objectifs pratiques sont différents. Des critères mieux adaptés peuvent être obtenus en prenant en compte les exigences particulières à chaque application.

Des mesures qui nous aident à juger la qualité des résultats issus d'un processus de recalage ont déjà été proposés. Certaines mesures sont amplement utilisées dans le domaine de la vision par ordinateur ; par exemple : l'erreur de positionnement de la caméra, l'erreur de reprojection, l'incertitude des paramètres de calibrage, et l'erreur de reconstruction 3D. D'autres critères de qualité sont orientés vers les fonctionnalités spécifiques des systèmes de RA, tels que : l'erreur d'alignement (mise en correspondance entre les objets de synthèse et les indices de la scène) ou l'impression visuelle. Ci-dessous nous présentons quatre de ces critères qui nous permettront de comparer et d'évaluer les différentes techniques et hypothèses de calibrage, dans le cadre des applications de la RA, étudiés dans cette thèse :

- l'erreur de mise en correspondance,
- l'incertitude des paramètres de calibrage,
- l'erreur de positionnement de la caméra,
- et l'impression visuelle.

3.1.1 L'erreur de mise en correspondance

L'erreur de mise en correspondance est le critère le plus utilisé pour évaluer la précision d'un système de recalage. Le but de ce critère est de mesurer l'adéquation d'un ensemble de correspondances obtenues à partir des images, soit des appariements entre les images (2D-2D) soit des appariements modèle-image (si les positions 3D des primitives sont connues), par rapport à une contrainte géométrique donnée. Ces contraintes sont généralement imposées pour le modèle de la caméra, la structure de la scène (e.g. rigidité, coplanarité), et parfois pour le type de mouvement de la caméra (e.g. déplacement linéaires, rotations).

Une mesure fréquemment utilisée est l'*erreur de reprojection*. Lorsque les positions dans l'espace d'un ensemble de n points $\{\mathbf{X}_i\}$ $(1 \leq i \leq n)$ sont connues, on peut calculer les distances entre les points détectés $\{\mathbf{x}_i\}$ et les positions estimées des points projetés sur le plan image $\{\hat{\mathbf{x}}_i = \mathbf{P}\mathbf{X}_i\}$, au moyen de la matrice de projection $\mathbf{P}$ de la caméra. L'erreur de reprojection est déterminée par la racine de la moyenne des distances au carré :

$$r = \sqrt{\frac{1}{n}\sum_i ||\hat{\mathbf{x}}_i - \mathbf{x}_i||^2}$$

L'écart entre les points prédits et les points détectés peut être aussi mesuré par d'autres types de distances que la distance Euclidienne : la valeur absolue, la distance de Mahalanobis, des estimateurs robustes, etc. D'autres contraintes peuvent être aussi employées lorsque le modèle 3D n'est pas connu :

- la contrainte épipolaire, qui établit une relation entre les observations d'un point 3D sur deux vues, à l'aide de la matrice fondamentale [LH81, MF92, WHA89] ;
- des transformations homographiques, qui déterminent une relation bijective entre les points d'une vue et leurs correspondants sur une deuxième image quand :
  - les points observés sur la scène sont coplanaires [LV94] ;
  - la caméra n'effectue que des rotations [Har94a] ou des variations du zoom [AHR01].

3.1.2 L'incertitude sur les paramètres de calibrage

Le calcul des paramètres de la caméra est généralement effectué en minimisant une fonction d'erreur de mise en correspondance. Plusieurs facteurs peuvent amener des imprécisions dans le calcul de ces paramètres, entre autres :

- la sensibilité des paramètres par rapport aux erreurs de mesure des primitives sur les images et aux erreurs de modélisation ;
- la quantité de données d'entrée (mesures sur les primitives) employées dans l'estimation ;
- la distribution des primitives sur l'image.

Comme ces éléments changent normalement entre différentes vues, la précision du calcul tend à varier aussi entre différentes acquisitions.

Pour certaines applications, il est indispensable de connaître en permanence la fiabilité des résultats obtenus. Par exemple, il est très important pour un chirurgien d'être informé de la précision du repérage effectué en moyen d'un système de RA, pendant une intervention assistée ou une chirurgie robotisée [NPSA04], car une manœuvre chirurgicale incorrecte due à une erreur de calcul du positionnement, pourrait avoir des conséquences graves sur le patient.

Un critère fréquemment employé pour évaluer la fiabilité des valeurs estimées est le calcul de la matrice de covariances des paramètres. Pour effectuer le calcul des matrices de covariances de manière empirique, il est nécessaire d'obtenir plusieurs échantillons d'une même observation (ce qui n'est pas faisable pour certaines modalités de mesure). Une méthode courante consiste à remplacer la distribution empirique des mesures d'entrée par une fonction analytique de probabilité. Cette fonction est la première composante d'un modèle du processus d'estimation d'incertitudes. La deuxième partie consiste à évaluer l'erreur

induite sur une estimation numérique par des erreurs faites sur les mesures. Quand les paramètres sont obtenus au moyen de la minimisation d'une fonction d'erreur à partir d'un ensemble de données, il est donc possible d'estimer la covariance de l'ensemble de paramètres de la solution à partir de la covariance dans les données.

Plusieurs méthodes permettent d'évaluer la covariance des paramètres, parmi lesquelles on trouve :

- des techniques de simulation, comme la méthode Monte Carlo [HH64] ou la méthode *bootstrap* [Efr86], qui font des inférences empiriques basées sur la génération artificielle d'un processus aléatoire ; ces méthodes sont d'usage assez général mais requièrent normalement un échantillonnage exhaustif pour atteindre une estimation satisfaisante de la distribution des paramètres.
- des méthodes analytiques qui établissent une relation explicite entre la distribution d'erreur des observations et les dispersions sur les paramètres, comme la propagation statistique des incertitudes [Bev69, Kan96] qui est basée sur le développement en séries de Taylor de premier ordre de la fonction d'erreur pour obtenir les équations de propagation d'incertitudes.

Pour des applications interactives, l'évaluation de fiabilité doit être réalisée continuellement et en temps réel, puisque les conditions de mesure peuvent changer au cours de leur utilisation (e.g. variations dans la distribution des primitives sur l'image ou le nombre d'observations, perturbations sur l'éclairage,...). La méthode de propagation d'erreurs semble donc plus adaptée à ce type d'applications grâce à sa simplicité de calcul et au fait qu'elle apporte aussi des estimations fiables des covariances des paramètres de la caméra [CRZ97, Kan96] quand les conditions de simplification à l'ordre 1 sont satisfaites (voir section 3.5).

3.1.3 L'erreur de positionnement 3D

L'erreur de correspondance entre les primitives projetées et observées n'est pas toujours un critère d'évaluation suffisant. Quelques applications requièrent aussi une bonne connaissance du point de vue de la caméra, comme dans le cas où le contrôle d'un outil dépend de l'analyse d'images faite par un système de vision (comme dans certaines manœuvres robotisées) ou quand l'utilisateur doit interagir avec des éléments de l'entourage pour réaliser une tâche de manipulation. Par exemple, [Nic04] requiert de suivre soigneusement la position et l'orientation dans l'espace d'une aiguille pendant une intervention chirurgicale assistée.

Une erreur de reprojection faible ne garantit pas toujours que le point de vue soit estimé de manière précise, car la minimisation de cette erreur peut être issue d'une compensation entre certains paramètres estimés, par exemple :

- entre le point principal et les angles de rotation [AHR01],
- entre la distance focale et un déplacement le long de l'axe optique [CM00],
- entre les translations parallèles au plan image et des rotations dans les mêmes directions (voir section 3.1.4).

Une évaluation objective du point de vue nécessite la connaissance du positionnement de la caméra dans la scène. Quelquefois, on sait que la caméra est sujette à des mouvements restreints (e.g. si elle est posée sur un trépied, se déplace sur un rail ou une surface plane) ou grossièrement connus, et cela permet de vérifier que les paramètres extrinsèques obtenus satisfont ces contraintes.

Mais une évaluation plus complète demande une comparaison entre les paramètres extrinsèques estimés et des mesures fines du point de vue acquises par d'autres moyens (par exemple, en utilisant des capteurs de positionnement [RDB00]), ou imposées par un dispositif qui contrôle les déplacements de la caméra.

Certaines difficultés doivent être considérées lorsqu'on cherche à comparer le point de vue obtenu par des méthodes de vision et d'autres mesures du positionnement :

D'une part, les paramètres extrinsèques doivent être extraits du modèle de caméra utilisé. Cette extraction demande souvent de connaître certaines variables (e.g. les paramètres intrinsèques de la caméra pour le modèle sténopé, ou le rapport entre la profondeur moyenne et la distance focale [DD95]), mais des erreurs dans ces valeurs requises peuvent introduire des erreurs dans le processus d'extraction des paramètres extrinsèques. Donc, il est nécessaire d'analyser comment ces imprécisions peuvent affecter l'estimation du positionnement de la caméra pour connaître la fiabilité des paramètres extrinsèques obtenus par des méthodes de vision.

D'autre part, la comparaison entre les estimations du point de vue obtenues par des méthodes vision et les données réelles du positionnement ne peuvent pas toujours être effectués de manière directe. Il existe une transformation rigide entre les systèmes de coordonnées de la caméra et des capteurs du positionnement (ou de l'outil de contrôle), qui doit être calculée pour rapporter les valeurs à comparer vers un repère commun. La détermination de cette transformation est aussi soumise à des erreurs de calcul potentielles qui peuvent entraîner des résultats inexacts dans les comparaisons du positionnement estimé et réel.

3.1.4 L'évaluation visuelle

On considère souvent que les paramètres d'un modèle qui ajustent au mieux un ensemble d'observations sont ceux qui minimisent la fonction d'erreur de correspondance sur les données d'entrée. Néanmoins, les paramètres obtenus de cette manière représentent une estimation localement adaptée à l'ensemble d'entrées, et la fonction d'erreur évaluée sur le même ensemble n'apporte aucune information sur l'adéquation des paramètres pour des données hors de cet ensemble.

Puisque l'un des buts des applications de RA est de présenter à l'œil humain des compositions cohérentes entre les images réelles et les objets de synthèse, l'évaluation visuelle est une façon évidente d'expertiser la qualité d'un système de RA. En effet, le système de perception humain est très sensible pour détecter les incohérences qui pourraient se présenter dans les images composées.

On peut constater deux types principaux de problèmes visuels dans les images augmentées issus du processus d'estimation de paramètres :

- des *erreurs statiques*, perceptibles individuellement sur chaque image comme des problèmes d'alignement entre les objets virtuels et les éléments de la scène ;
- et des *erreurs dynamiques*, qui se révèlent seulement en observant le flux ininterrompu des images augmentées dans une séquence, manifestant des inconsistances tels que des tremblements ou des sautillements des objets ajoutés.

Erreurs statiques

Les raisons des incohérences de type *statique* les plus fréquentes sont : les erreurs de positionnement des objets de synthèse par rapport aux éléments de la scène réelle, les manques d'accord entre les propriétés projectives des images acquises et le modèle de caméra utilisé pour l'incrustation des objets virtuels.

Certains problèmes statiques sont la conséquence d'erreurs dans le processus de calcul de paramètres, car ces estimations sont parfois très sensibles aux mesures des primitives employées comme référence. La figure 3.1 montre une image augmentée dont la scène réelle contient une mire de calibrage. Les cibles de la mire (les disques blancs placés sur les plans de la mire) ont été utilisées pour obtenir des correspondances 3D-2D, de telle façon que les paramètres de la caméra soient récupérés en minimisant l'erreur de reprojection. L'écart quadratique moyen (erreur RMS) de la reprojection sur les cibles est d'environ 0,2 pixels. D'autres marqueurs sur la scène dont les positions dans l'espace sont connues ont été utilisés pour évaluer la cohérence visuelle : trois disques noirs ont été installés, un marqueur central posé sur une petite tour au milieu de la distribution des cibles de référence et deux panneaux latéraux placés hors de la distribution des cibles. Les circonférences et les centres correspondant à ces trois marqueurs ont été projetés en utilisant les paramètres estimés de la caméra. La circonférence associée au marqueur central se projette bien sur le cercle correspondant, mais les projections des deux panneaux latéraux se trouvent décalées par rapport aux positions des vrais marqueurs.

Même si l'erreur de correspondance évaluée sur les primitives de référence est réduite, il est possible que les paramètres estimés de la caméra soient localement adaptés, c'est à dire, qu'ils relient correctement des primitives de la scène avec ses projections seulement dans une région de l'image, mais que les projections ne soient pas correctes hors de celle-ci. Cela peut ne pas avoir de répercussions visibles pour les objets proches de ces primitives, mais les indices incrustés dans une zone éloignée des primitives suivies peuvent apparaître considérablement écartés des positions attendues. Ce problème apparaît fréquemment si l'information apportée par les primitives mesurées n'est pas suffisante pour inférer l'emplacement correct des objets incrustés dans toutes les autres positions de l'image, par exemple, quand les primitives de référence n'occupent qu'une portion réduite de l'image.

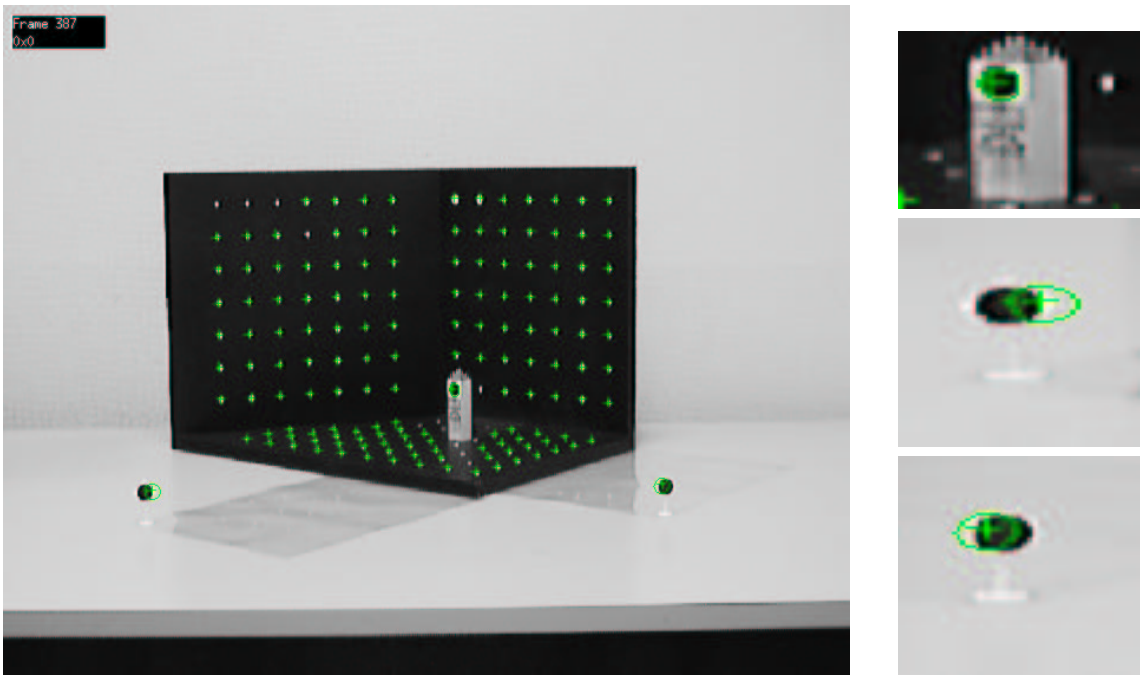


FIG. 3.1 – Projection des cibles et des marqueurs du modèle de la scène sur une image réelle de la scène. À gauche, l'image augmentée complète ; à droite, les images amplifiées des trois marqueurs (gauche, central et droit) et de leurs correspondantes marques circulaires ajoutées.

D'autres problèmes de type statique sont dus au choix incorrect du modèle de caméra utilisé. Par exemple, dans les images où l'effet de perspective est important, l'utilisation d'un modèle simplifié (e.g. affine, orthographique à l'échelle, etc.) révélera la manque d'accord entre la géométrie de la scène réelle et celle attribué aux objets de synthèse.

Plusieurs systèmes de RA existants montrent des résultats très satisfaisants pour la fusion d'images lorsque les images composées sont observées individuellement. Cependant, les incohérences visuelles apparaissent lorsque l'on regarde la séquence d'images en continu. Cela signifie qu'une évaluation fiable de la perception doit être effectué en prenant en compte les facteurs temporels du système.

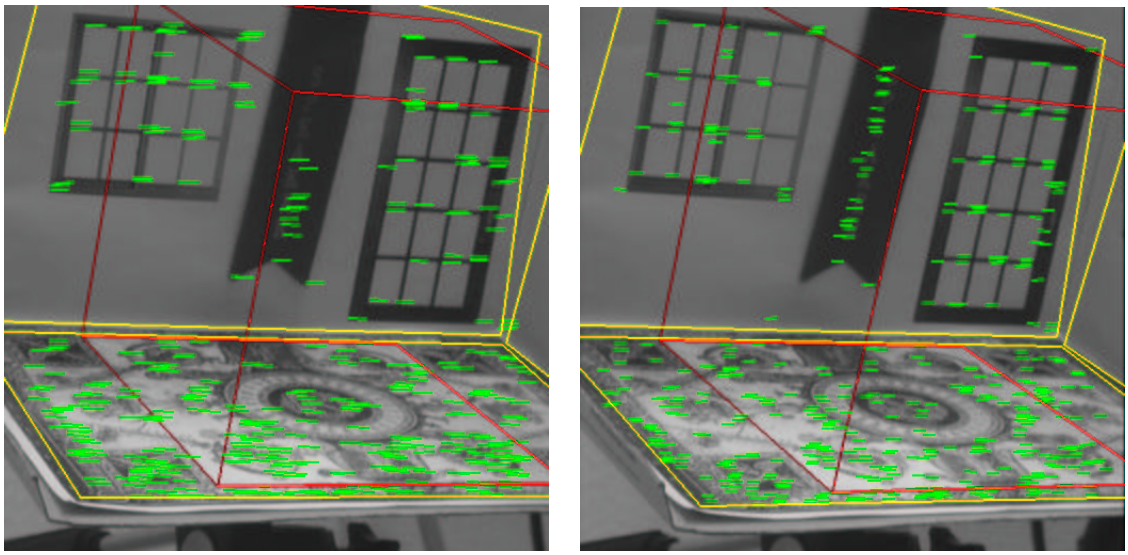
Erreurs dynamiques

Lorsqu'on ajoute des objets de synthèse sur les images de la scène, même si les superpositions pouvaient paraître cohérentes dans chaque vue séparément, l'observation continue d'une séquence d'images augmentée peut révéler des effets visuels gênants tels que des sautillements des objets virtuels. Ainsi, l'observation d'images en succession nous permet d'examiner davantage les aspects temporels des systèmes de RA.

Les effets *dynamiques* qui se manifestent plus fréquemment dans ces observations sont :

Tressaillements et glissements. Les problèmes de stabilité dans l'estimation des paramètres de la caméra entraînent souvent des inconsistances visuelles dans les images augmentées : les objets de synthèse semblent trembler, glisser, et parfois sursauter, par rapport à la scène réelle. Ces phénomènes sont particulièrement visibles quand les mouvements de la caméra entre les images consécutives sont réduits. L'origine de ces problèmes peut être de diverses natures :

- Lorsque les déplacements de la caméra sont parfois restreints à un nombre réduit de degrés de liberté, si le modèle de mouvement de la caméra est considéré général, les degrés de liberté additionnels contribuent à minimiser la fonction d'erreur censée évaluer l'adaptation des paramètres aux données, mais ils tendent aussi à s'ajuster au bruit entachant les mesures.



(a) (b)

FIG. 3.2 – Exemple de déplacement des points d'intérêt d'une scène suite à deux types de mouvement différents : (a) une rotation panoramique et (b) une translation parallèle à l'axe X du plan image.

Par exemple, lorsque la caméra est stationnaire, les paramètres calculés devraient en principe rester constants. Néanmoins, les mesures acquises sur les observations présentent de faibles variations d'une image à l'autre, ce qui se traduit par des fluctuations sur l'ensemble des paramètres de la caméra, et en conséquence les objets affichés au moyen de ces paramètres semblent osciller sur l'image.

- Autre cause possible de ces phénomènes sont les compensations dans l'estimation du mouvement de la caméra. Par exemple, l'ambiguïté entre de petites rotations et de petites translations parallèles au plan image : visuellement, les trajectoires suivies par les primitives observées semblent très similaires pour les deux types de mouvement (voir figure 3.2), surtout si la différence de profondeur des éléments observés est petite par rapport à la distance caméra-scène. Et les erreurs de détection des primitives compromettent d'ailleurs la précision de calcul, en entraînant souvent des compensations entre les angles de rotation et le vecteur de translation, pour mieux s'ajuster aux mesures obtenues.
- Un choix incorrect du modèle de caméra peut aussi produire ces phénomènes. Plusieurs modèles simplifiés utilisés comme approximations du modèle sténopé requièrent d'être employés uniquement si certaines conditions sur la scène observée ou sur les propriétés internes de la caméra sont satisfaites (voir section 3.2). Cependant, si la caméra se déplace vers une position où ces conditions ne sont pas satisfaites, les objets incrustés sembleront s'écarter progressivement des positions attendues.

Dérive. Le besoin fondamental des applications de la RA est d'être capable d'estimer les paramètres du modèle de caméra continuellement avec une bonne précision même pendant des périodes de temps arbitrairement longues et pour des mouvements quelconques de l'utilisateur. Dans la pratique, cette condition n'est pas toujours satisfaite et cela peut dépendre de la méthode utilisée pour effectuer l'étalonnage : la qualité des résultats obtenus avec les systèmes basés modèle dépend principalement de la position de la caméra par rapport à l'objet utilisé et les mouvements de la caméra sont limités dans le but d'observer toujours un ensemble de caractéristiques de l'objet de calibrage suffisant pour faire l'étalonnage.

D'autres systèmes, plus autonomes, utilisent la rigidité de la scène pour établir des contraintes qui nous permettent de calculer le déplacement relatif de la caméra effectué entre un couple d'images (e.g. la contrainte épipolaire [WHA89, MF92], des homographies pour les régions coplanaires [SFZ00] ou les caméras rotatives [Har94a]). Ces systèmes calculent le positionnement de la caméra par la composition des mouvements relatifs entre les vues d'une séquence d'images. Cependant, il existe des erreurs dans l'estimation des déplacements relatifs, en raison du bruit sur les mesures ou des erreurs d'approximation, même si ces erreurs individuellement semblent modestes par rapport au déplacement effectué. En raison

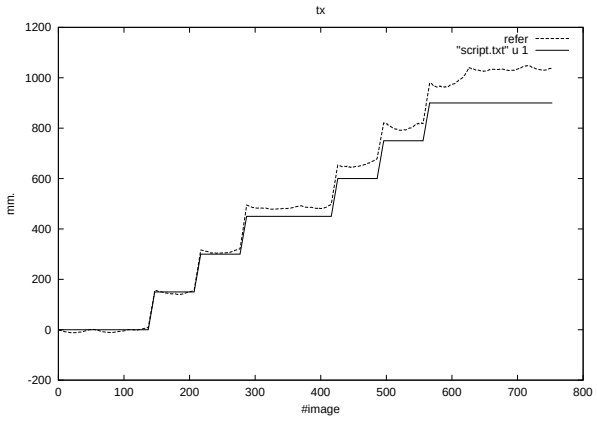


FIG. 3.3 – Comparaison entre l'estimation faite au moyen d'une méthode de recalage séquentiel [SB02a] et la valeur réelle d'une des composantes du point de vue (t_x). L'écartement graduel entre les deux courbes met en évidence le problème de dérive des paramètres.

du caractère incrémentiel de ces systèmes, les erreurs sur les déplacements relatifs s'accumulent au fur et à mesure occasionnant une dérive progressive des paramètres de la caméra (figure 3.3). En conséquence, on observe des erreurs d'accord visuel entre les objets de synthèse et de la scène, irrégularités qui augmentent lorsqu'on prolonge la durée d'utilisation du système.

Certaines techniques ont été proposées pour remédier à ce problème :

- Certaines approches cherchent à réduire les erreurs d'estimation des déplacements relatifs, soit en régularisant les paramètres au moyen d'un modèle de prédiction de mouvement (e.g. par filtrage de Kalman [Gag99, DM02]), soit en réduisant le nombre de degrés de liberté du modèle de mouvement relatif [Tor97, CM00] lorsqu'on juge que le vrai déplacement ne change que certains paramètres (e.g. rotations, translations).
- D'autres techniques visent à calculer périodiquement le point de vue à l'aide des marqueurs ou des éléments de référence sur la scène dont la structure est connue. Ils utilisent habituellement des approches séquentielles pour la mise à jour du point de vue, ce qui présente des avantages lorsque les éléments de référence ne sont pas totalement visibles dans les images, ou quand l'approche séquentielle est requise pour répondre à une exigence de vitesse de calcul ; de telle façon que les marqueurs sont seulement utilisés pour mettre à jour éventuellement les estimations.
  D'autres applications ont lieu dans des environnements où il n'est pas possible d'installer des marqueurs ni de faire une reconstruction complète de la scène par avance. Dans ces cas, une reconstruction séquentielle en ligne est souhaitable. Des nouvelles approches s'orientent vers une obtention causale de la structure et du mouvement [CFJS02, DM02]. Ces approches sont dédiées au calcul du positionnement de la caméra et non précisément à la reconstruction de la scène, et le résultat est plutôt un modèle géométrique sans structure (une carte d'éléments facilement identifiables de la scène) souvent formé par un ensemble de points.
- Quelques méthodes ont été récemment développées pour faire la mise en correspondance de points pour des déplacements longs, à l'aide de descripteurs affines de régions [FTG01] ou affines et invariants à l'échelle (comme la méthode SIFT [Low04]). Ces approches permettent le suivi de primitives entre deux vues éloignées d'une même scène, en évitant l'accumulation d'erreurs due à l'estimation incrémentielle des paramètres.

Latence. Dans les applications en temps réel, d'autres incorrections proviennent du délai entre le processus d'acquisition d'images avec la caméra et le rendu des objets de synthèse, ce qui peut occasionner un décalage entre les objets ajoutés par rapport aux éléments de la scène. Ce phénomène est plutôt lié aux problèmes de synchronisation entre les dispositifs physiques (d'acquisition d'images et d'affichage) et le temps requis par les processus de recalage, d'incrustation d'images et de visualisation.

Néanmoins, dans cette thèse nous considérons que la visualisation des images augmentées est effectuée

sur une sortie vidéo ; c'est à dire, l'affichage est fait sur le moniteur auquel sont superposées l'image de synthèse et l'image du monde réel par des techniques de mixage vidéo. Dans ce cas, il n'existe pas de problème de latence car l'image acquise par la caméra (et employée pour le processus de recalage) est la même que celle utilisée pour le mixage.

3.2 Choix du modèle de caméra

Le modèle de caméra établit la relation analytique entre les coordonnées 3D du monde et leurs coordonnées 2D correspondantes sur l'image. Cette relation doit être fidèle au vrai processus de formation d'images.

Diverses représentations de caméras peuvent être introduites dans un système basé vision : modèles perspectifs, affines, ou Euclidiens, entre autres. Comme les modèles de caméras sont basés sur des hypothèses géométriques différentes, chaque modèle possède des caractéristiques particulières. Pour examiner l'adéquation d'un modèle géométrique déterminé, il est nécessaire d'évaluer si le modèle s'ajuste aux besoins des applications du système.

Dans cette section, nous évoquons les modèles de caméra les plus fréquemment signalés dans la littérature, leurs caractéristiques géométriques, leurs propriétés qui nous permettent de les considérer davantage pour certains types d'applications, ainsi que quelques exemples de systèmes de RA qui emploient ces modèles. À la fin de cette section, nous discutons des critères à prendre en compte pour mieux choisir le modèle de la caméra en fonction des besoins de chaque application.

3.2.1 Classement des modèles de caméra

Modèle général de projection

Le modèle générique de caméra est défini comme la transformation de l'espace projective $\mathbb{P}^3$ vers l'espace projective $\mathbb{P}^2$. Les modèles de caméra décrits dans cette section peuvent être représentés par une matrice $\mathbf{P}_g$ de dimensions 3×4, telle que les points $\tilde{\mathbf{X}}_i = (x_i, y_i, z_i, 1)^T$ de l'espace 3D sont liés à leurs projections correspondantes $\tilde{\mathbf{U}}_i = (u_i, v_i, 1)^T$ sur le plan image par la relation :

$$\tilde{\mathbf{U}}_i \sim \mathbf{P}_g \tilde{\mathbf{X}}_i \tag{3.1}$$

Lorsque la matrice $\mathbf{P}_g$ est exprimée en vecteurs lignes, les coordonnées des projections sur le plan image peuvent être écrites comme :

$$\begin{pmatrix} u_i \\ v_i \\ 1 \end{pmatrix} \sim \begin{pmatrix} \mathbf{p}_1^T \\ \mathbf{p}_2^T \\ \mathbf{p}_3^T \end{pmatrix} \tilde{\mathbf{X}}_i \Longrightarrow \begin{pmatrix} u_i \\ v_i \end{pmatrix} = \frac{1}{\mathbf{p}_3 \cdot \tilde{\mathbf{X}}_i} \begin{pmatrix} \mathbf{p}_1 \cdot \tilde{\mathbf{X}}_i \\ \mathbf{p}_2 \cdot \tilde{\mathbf{X}}_i \end{pmatrix} \tag{3.2}$$

Le terme $(\mathbf{p}_3 \cdot \tilde{\mathbf{X}}_i)^{-1}$ (appelé de *division perspective*) sert à représenter les positions $\tilde{\mathbf{U}}_i$ de l'espace projectif $\mathbb{P}^2$ en leurs coordonnées homogènes. Ce terme joue un rôle important dans la caractérisation du modèle de la caméra, car il dénote le rapport entre la profondeur d'un point 3D et sa projection, ce qui nous permet d'établir les propriétés géométriques de chaque modèle de caméra, comme on expliquera dans les sections suivantes.

Approximations de la projection perspective

Le modèle de projection *perspective* est le modèle le plus général suivant la relation (3.1). Bien que ce modèle soit le plus souvent cité dans la littérature du domaine, des simplifications de ce modèle doivent parfois être imposées afin d'obtenir des modèles plus simples ou plus maniables dans des situations particulières.

Les autres modèles de caméra (*affine* et *orthographique à l'échelle*) abordés dans cette section peuvent être considérés comme des approximations :

- numériques (développement de la transformation perspective en séries de Taylor limité à un ordre précis) :

  - le modèle de caméra *affine* est une approximation au premier ordre,
  - tandis que le modèle *orthographique à l'échelle* est une approximation d'ordre zéro ;
- géométriques : en faisant intervenir des transformations plus simples, telles que des projections parallèles vers un plan auxiliaire (voir figures 3.4 et 3.5), souvent en utilisant un point de l'espace comme référence (e.g. un point de la scène représentant la profondeur moyenne).

Hiérarchie des modèles

Les modèles de caméra perspective, affine, et orthographique à l'échelle, peuvent être organisés hiérarchiquement (figure 3.4-a), en trouvant le modèle de projection perspective au sommet de la classification.

Le modèle affine (section 3.2.3) introduit les contraintes $p_3^T = (0,0,0,1)^T$ sur les coefficients de la matrice de projection $\mathbf{P}_g$. Ainsi, le modèle affine peut être considéré comme une sous-classe du modèle de projection perspective, et en même temps une approximation numérique et géométrique de ce modèle.

Au même titre, la projection orthographique à l'échelle (section 3.2.4) est un cas particulier d'une caméra affine, et une approximation des deux représentations antérieures.

Des détails sur ces modèles sont décrits par la suite (sections 3.2.2 à 3.2.4).

Unification des modèles projectif, affine, et orthographique à l'échelle

Certains auteurs ont proposé des modèles de caméras généraux incorporant les modèles cités ci-dessus : Sturm [Stu97b] propose une formulation unifiée des modèles perspectif et affine, dont le repère de la caméra a été déplacé du centre de projection vers le point principal (inspiré du travail de Szeliski et Kang [SK96]), ce qui permet d'étendre le modèle sténopé pour gérer les caméras ayant le centre de projection à l'infini (caméras *affines*, voir section 3.2.3). Cette formulation est équivalente à celle de [Lin99] qui propose une représentation générale de la projection incorporant les trois modèles : *perspective*, *affine*, et *orthographique à l'échelle*. Dans cette représentation, la matrice $\mathbf{P}_g$ est décomposée comme :

$$\mathbf{P}_g = \mathbf{K}_g \mathbf{E}_g = \begin{pmatrix} \alpha_u & s_k & \lambda\beta_u + \mu u_0 & (1-\mu)\gamma_u \\ 0 & \alpha_v & \lambda\beta_v + \mu v_0 & (1-\mu)\gamma_v \\ 0 & 0 & \mu & 1-\mu \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \tag{3.3}$$

La matrice de *paramètres extrinsèques* $\mathbf{E_g}$ (définie avec six degrés de liberté) transfère les points du repère du monde au repère de la caméra.

La matrice *généralisée de paramètres intrinsèques* $\mathbf{K}_g$ est une prolongation à une matrice 3×4 de la matrice de paramètres intrinsèques du modèle sténopé. On distingue deux types de paramètres dans cette matrice : les paramètres intrinsèques de la caméra $\{\alpha_u, \alpha_v, u_0, v_0, \beta_u, \beta_v, \gamma_u, \gamma_v, s_k\}$ et les paramètres modaux $\{\mu, \lambda\}$ qui servent à choisir l'un des modèles de caméra de la façon suivante :

Modèle	λ	μ
Projection perspectif pure	0	1
Projection affine	1	0
Projection orthographique à l'échelle	0	0

Cependant, ces modèles généralistes ont été utilisé uniquement comme une description unifiée des différents modèles de caméra, mais des méthodes d'étalonnage spécifiques à ces représentations n'ont pas encore été développés. Pourtant, les modèles de caméra ici abordés sont normalement utilisés de manière indépendante dans les différentes applications.

3.2.2 Modèle de projection perspective

Le modèle de caméra le plus fréquemment utilisé dans les systèmes de vision par ordinateur est le modèle de projection perspective (ce modèle a été présenté en détail dans le chapitre 2). L'hypothèse géométrique fondamentale de ce modèle consiste à supposer que tous les rayons qui lient un point dans l'espace avec sa projection correspondante sur le plan image concourent vers un point, nommé le *centre de projection perspective* (figure 3.4-b).

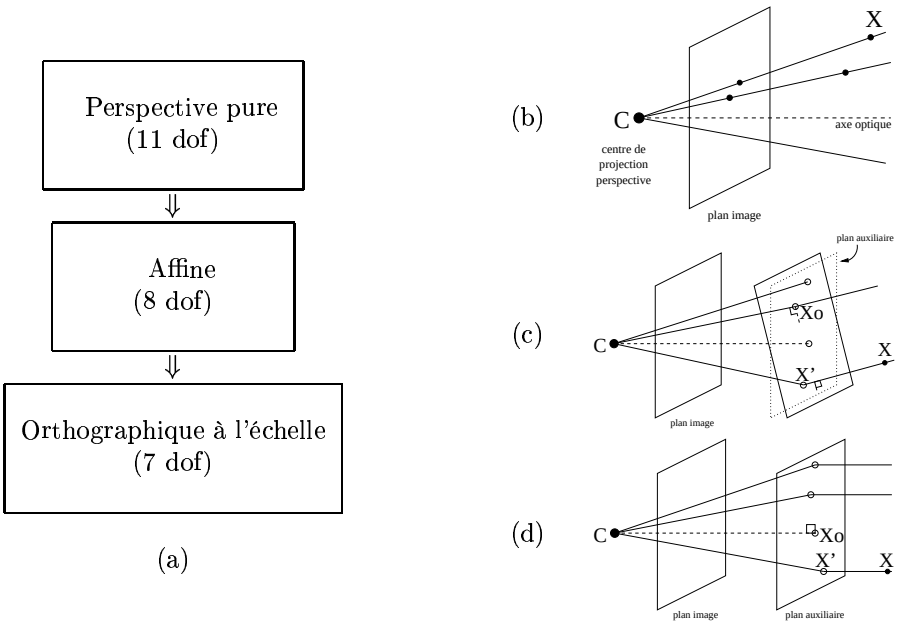


FIG. 3.4 – (a) Hiérarchie des modèles de caméra : (b) perspective, (c) affine et (d) orthographique à l'échelle.

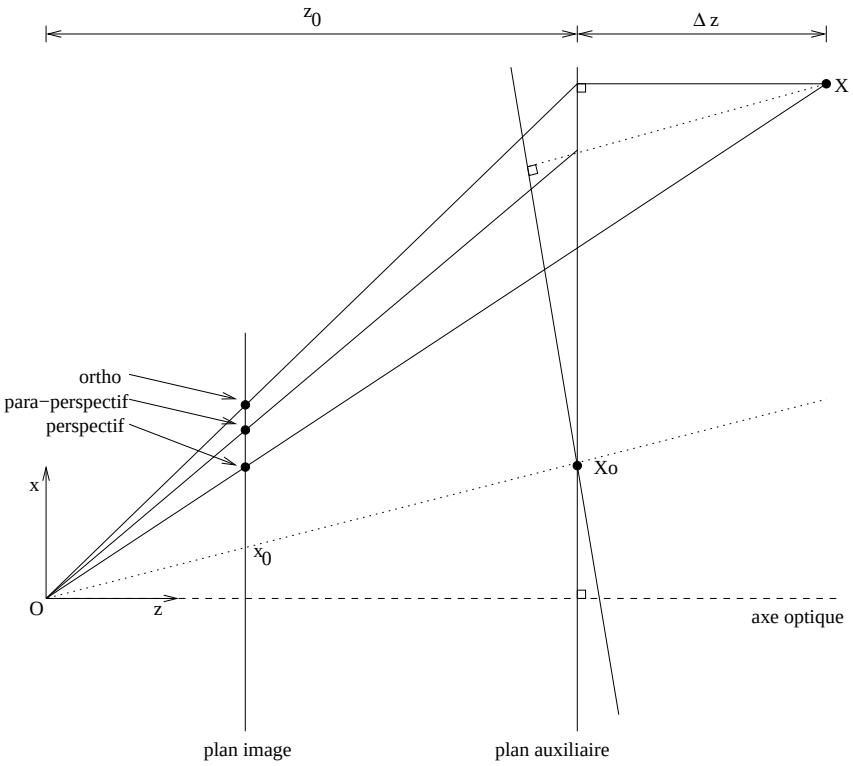


FIG. 3.5 – Représentation géométrique des projections perspective, para-perspective et orthographique à l'échelle [HCD94].

De manière générale, les coordonnées d'un point dans l'espace $\mathbf{X}_i$ et sa projection sur le plan image $\mathbf{U}_i$ sont liés par la matrice 3×4 de projection $\mathbf{P}$ selon la relation : $\tilde{\mathbf{U}}_i \sim \mathbf{P}\tilde{\mathbf{X}}_i$ (éq. 3.1). Comme les coefficients de la matrice $\mathbf{P}$ sont déterminés à un facteur d'échelle près, le modèle de projection perspective a donc 11 degrés de liberté.

L'emploi du modèle de projection perspective présente plusieurs avantages :

- le modèle décrit une approche satisfaisante du processus de formation d'images pour un grand nombre de dispositifs optiques ;
- comme la matrice de projection perspective collecte à la fois les paramètres intrinsèques de la caméra et le point de vue, toute variation soit dans le mouvement de la caméra soit dans ses paramètres internes peut être considérée par ce modèle ;
- il existe dans la littérature une grande diversité d'algorithmes pour l'étalonnage de ce modèle, ainsi que des analyses comparatives de ces méthodes.

On trouve des approches proposant des solutions directes en résolvant un système linéaire d'équations comme les méthodes de Faugeras-Toscani [FT87] ou de Hall [HTMS82] ; des méthodes linéaires itératives pour améliorer le conditionnement des systèmes d'équations et la réduction du biais dans les estimations, tel que l'approche de Kanatani [Kan93b] ; il existe aussi des solutions basées sur des méthodes d'optimisation non linéaires (e.g. [JM01] en utilisant la méthode de Levenberg-Marquardt) généralement plus précises (mais souvent sensibles à l'estimation initiale requise) et qui nous permettent d'incorporer divers types de mesures d'adéquation : les erreurs de reprojection au carré, des estimateurs robustes (médianes, moindres carrés réduits, M-estimateurs [Hub81]).

Plusieurs systèmes de RA utilisent la projection perspective pour effectuer le suivi de la caméra et la fusion d'images. Par exemple, la méthode de Seo et Hong [SH00], où tous les coefficients de la matrice de projection de la caméra virtuelle sont estimés à chaque prise de vue, à partir des correspondances 3D-2D d'un ensemble de points.

Malgré la flexibilité de ce modèle à s'adapter à différents types de dispositifs d'acquisition d'images et de leurs applications, le nombre de paramètres à estimer pour la matrice de projection dépasse souvent le nombre de degrés de liberté nécessaires pour représenter les manœuvres de la caméra dans des situations courantes (e.g. lorsque les paramètres internes de la caméra ne varient pas). Donc, l'estimation simultanée des onze paramètres du modèle entraîne des problèmes de stabilité dans les calculs qui se traduisent en problèmes de positionnement de la caméra virtuelle et en conséquence de la perception des images augmentées.

3.2.3 Modèles de projection affine et para-perspectif

Le modèle de projection para-perspective a été d'abord énoncé et utilisé par Ohta [iOMS81] comme une approximation de la projection perspective dans le problème de reconstruction 3D. Il fut ensuite étudié par d'autres auteurs tels que Aloimonos [Alo90], Horaud *et al.* [HDL97], Lingrand *et al.* [LV96].

Approximation numérique de la projection perspective

Le modèle para-perspectif est une *approximation de premier ordre* par série de Taylor de la fonction de projection perspective.

En général, lorsque une fonction $\mathbf{g}(\mathbf{X})$ est évaluée autour d'une valeur de référence $\mathbf{X}_0$, son approximation de premier ordre est calculée par : $\mathbf{g}(\mathbf{X}) \approx \mathbf{g}(\mathbf{X}_0) + \nabla\mathbf{g}(\mathbf{X}_0) \cdot (\mathbf{X} - \mathbf{X}_0)$.

Étant donné un point de la scène $\mathbf{X} = (x, y, z)^T$ et le point de référence $\mathbf{X}_0 = (x_0, y_0, z_0)^T$ dans le repère de la caméra, ainsi que la fonction $\mathbf{U} = \mathbf{g}(\mathbf{X})$ associée à la projection perspective de $\mathbf{X}$ sur le plan de l'image (d'après l'équation 3.2), le résultat de l'approximation limitée au premier ordre de la série de Taylor pour cette fonction est :

$$\mathbf{U} \approx \begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{\alpha_u^{'}}{z_0}\left(x - \frac{(z-z_0)x_0}{z_0}\right) + \frac{s_k^{'}}{z_0}\left(y - \frac{(z-z_0)y_0}{z_0}\right) + u_0 \\ \frac{\alpha_v^{'}}{z_0}\left(y - \frac{(z-z_0)y_0}{z_0}\right) + v_0 \end{pmatrix} \tag{3.4}$$

où $\alpha_u^{'}$, $\alpha_v^{'}$ et $s_k^{'}$ sont les paramètres intrinsèques du modèle perspectif pure.

Alors, on obtient que les coordonnées (u, v) de la projection para-perspective d'un point sont exprimées comme une fonction linéaire des coordonnées 3D correspondantes (x, y, z).

Approximation géométrique

De manière géométrique, le résultat (3.4) peut être interprété comme une transformation à deux stades (voir figures 3.4-c et 3.5) :

1. Le point $\mathbf{X}$ est d'abord projeté sur un plan *auxiliaire* (parallèle au plan image et passant par le point de référence $\mathbf{X}_0$) au moyen d'une projection parallèle dont la direction est déterminée par $\overline{\mathbf{C}\mathbf{X}_0}$ (le rayon partant du centre optique de la caméra vers le point de référence $\mathbf{X}_0$). La projection de $\mathbf{X}$ sur le plan auxiliaire est donc le point intermédiaire $\mathbf{X}'$:

$$\mathbf{X}^{'} = \mathbf{X} + \left(1 - \frac{z}{z_0}\right)\mathbf{X}_0 = \left(x - \frac{x_0}{z_0}(z - z_0)\,, \quad y - \frac{y_0}{z_0}(z - z_0)\,, \quad z_0\right)^T$$

2. Le point $\mathbf{X}^{'}$ est ensuite projeté sur le plan image en utilisant une projection perspective. Comme le plan auxiliaire est parallèle au plan image, cette deuxième étape est équivalent à une transformation affine 2D des deux premières coordonnées du point $\mathbf{X}^{'}$.

Équivalence avec le modèle de projection affine

D'après l'équation (3.4), et en utilisant la représentation (3.3), le modèle para-perspectif peut être défini par la matrice :

$$\mathbf{P}_q = \mathbf{K}_q\mathbf{E}_q = \begin{pmatrix} \alpha_u & s_k & \beta_u & \gamma_u \\ 0 & \alpha_v & \beta_v & \gamma_v \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} = \begin{pmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{3.5}$$

où les correspondances 3D-2D satisfont la relation $\tilde{\mathbf{U}}_i \sim \mathbf{P}_q\tilde{\mathbf{X}}_i$, tel que :

$$\begin{pmatrix} u_i \\ v_i \end{pmatrix} = \begin{pmatrix} \mathbf{q}_1 \cdot \tilde{\mathbf{X}}_i \\ \mathbf{q}_2 \cdot \tilde{\mathbf{X}}_i \end{pmatrix} \tag{3.6}$$

Tout modèle ayant la forme (3.6) est dénommé *affine* et possède la propriété de préserver la co-linéarité des points projetés et les proportions des distances de leurs projections. Donc, le modèle de caméra para-perspective appartient à l'ensemble de transformations affines. D'ailleurs, Basri [Bas94] a démontré l'équivalence entre les transformations para-perspectives et les transformations affines. Une transformation affine quelconque peut donc être considérée comme une version *non calibrée* d'une caméra para-perspective [WB96b] (dont la matrice de paramètres intrinsèques $\mathbf{K}_q$ n'est pas explicitement attribuée).

Caractéristiques du modèle de projection affine

Les principales caractéristiques de ce modèle sont :
- une caméra affine a 8 degrés de liberté ;
- la projection affine est une transformation strictement linéaire dans les coordonnées inhomogènes des correspondances 3D-2D ;
- comme la sous-matrice 3×3 gauche de la matrice de projection $\mathbf{P}_q$ est singulière, le centre de projection du modèle est placé à l'infini et les rayons de projection de tous les points dans l'espace sont parallèles [HZ00].

Exemples d'applications

Le modèle para-perspectif a été utilisée par plusieurs auteurs comme une approximation de la projection perspective [Alo90, HCD94] pour les problèmes de reconstruction de formes à partir de textures ou de contours, ainsi que pour l'estimation du positionnement de la caméra.

Par rapport au problème de positionnement, Poelman et Kanade [PK94] ont suggéré une méthode pour le calcul linéaire des coefficients du modèle para-perspectif, et l'obtention du point de vue d'après ces coefficients à l'aide d'une méthode de factorisation de la matrice $\mathbf{P}_q$ (éq. 3.5) du modèle. Horaud *et al.* [HDL97] proposent autrement un algorithme itératif pour estimer le point de vue (lorsque les paramètres intrinsèques du modèle sont fixés) en utilisant des approximations para-perspectives de la projection.

Kutulakos et Vallino [KV96] proposent un système de RA qui utilise une représentation affine générale de la caméra. Ils exploitent la linéarité du modèle affine de même que la réduction du nombre de degrés de liberté par rapport au modèle sténopé, pour obtenir un système facile à mettre en œuvre, dans le sens que très peu de connaissances *a priori* sont exigées : on ne requiert pas d'information de la caméra ni de la structure de la scène. Le système requiert que quatre points non coplanaires soient indiqués par l'utilisateur sur une paire d'images au début de l'application, puis de les suivre tout au long des déplacements effectués par la caméra. Ces points sont utilisés pour former une base du système de coordonnées, composée par trois vecteurs non coplanaires. Ensuite, pour le rendu des objets virtuels, il suffit d'exprimer les coordonnées 3D des objets comme une combinaison linéaire des vecteurs dans la base, car les projections résultantes sont des combinaisons linéaires des projections des vecteurs de la base.

Restrictions du modèle affine comme approximation du modèle perspective

La déviation de la projection affine par rapport à la projection perspective est donnée par :

$$\mathbf{x}_{aff} - \mathbf{x}_{persp} = \frac{\Delta z}{z_0} (\mathbf{x}_{persp} - \mathbf{x}_0)$$

où $\mathbf{x}_{aff}$ est la projection du point $\mathbf{X}$ selon le modèle affine, $\mathbf{x}_{persp}$ est la projection du même point selon le modèle perspectif, $\mathbf{x}_0$ est la projection sur le plan image du point de référence, et Δz est la différence de profondeur entre le point $\mathbf{X}$ de la scène et le plan d'approximation [HZ00].

D'après cette relation, on déduit que l'écart entre les projections affine et perspective d'un point est réduit si :

- les différences Δz de profondeur (ou *relief*) des points observés sont faibles par rapport à la profondeur z_0 du plan d'approximation ;
- les distances entre les points 3D observés et la droite auxiliaire $\overline{\mathbf{CX}_0}$ sont réduites.

L'utilisation du modèle affine est donc similaire au modèle perspectif seulement si les conditions décrites ci-dessus sont satisfaites.

La première condition dépend de la distribution dans l'espace des points de la scène par rapport à la position de la caméra, et du choix du point de référence (ordinairement choisi comme la position moyenne des points de la scène). Le modèle affine n'est pas une approximation convenable du modèle perspectif, si la scène contient des éléments à des niveaux de profondeur très différents.

La deuxième condition dépend de la position des points autour de la droite auxiliaire. Pour cette raison, les images acquises par une caméra avec une distance focale grande respectent mieux ces conditions (car leur champ de vue est réduit) que dans le cas où la distance focale serait plus petite.

3.2.4 Modèle de projection orthographique à l'échelle

Cette approximation de la projection perspective, a été introduit dans les travaux de Dementhon et Davis [DD92, DD95] et Soatto [SP95]. On rencontre également ce modèle sous les noms de projection *perspective faible* ou *orthographique graduée*. Il s'agit d'un cas particulier de l'approximation para-perspective où le point de référence se trouve sur l'axe optique de la caméra.

Approximation numérique de la projection perspective

Du point de vue numérique, le modèle orthographique à l'échelle estime le terme de division perspective de l'équation (3.2) par une *approximation d'ordre zéro* de sa série de Taylor, $(\mathbf{p}_3 \cdot \tilde{\mathbf{X}}_i)^{-1} \approx (\mathbf{p}_3 \cdot \tilde{\mathbf{X}}_0)^{-1}$,

de telle manière que :

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} \frac{\alpha_u' x + s_k' y}{z_0} + u_0 \\ \frac{\alpha_v' y}{z_0} + v_0 \end{pmatrix}$$

Cela signifie que la profondeur $z_i = z_0 + \Delta z_i$ (par rapport au plan image) de n'importe quel point sera estimée constante et égal à z_0. Ceci est équivalent à une projection orthogonale des points de la scène sur le plan auxiliaire $Z = z_0$, suivi d'un changement d'échelle sur l'image résultante (ce qui donne son nom au modèle).

Caractéristiques du modèle

Concrètement, il s'agit d'un cas particulier de la projection para-perspective dont la droite auxiliaire est l'axe optique de la caméra (voir fig. 3.4-d).

Ce modèle a 7 degrés de liberté, et sa matrice correspondante (selon la représentation 3.3) a la forme :

$$\mathbf{P}_o = \begin{pmatrix} \alpha_u & s_k & 0 & \gamma_u \\ 0 & \alpha_v & 0 & \gamma_v \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix}$$

L'approximation du modèle de projection perspective par une projection orthographique à l'échelle est valable dans conditions analogues à l'approximation affine (voir section 3.2.3). Ceci est traduit dans le contexte particulier d'une orthographie à l'échelle par :

- les variations de relief des points dans la scène doivent être réduites au regard de la profondeur z_0 du plan auxiliaire ;
- les écartements entre les points observés de la scène et l'axe optique de la caméra doivent être petits.

Un algorithme de calcul du positionnement basé sur le modèle orthographique à l'échelle

Comme les conditions d'approximation décrites ci-dessus limitent le nombre de situations pour lesquelles les modèles approximatifs sont comparables au modèle de projection perspective, une utilisation moins contrainte de ces représentations consiste à utiliser l'un des modèles simplifiés pour obtenir une représentation initiale de la caméra et raffiner cette première transformation jusqu'à atteindre une paramétrisation perspective.

Dementhon et Davis [DD95] emploient ce principe pour calculer le point de vue de la caméra, en utilisant la connaissance d'au moins quatre points non coplanaires de la scène et de leurs projections sur l'image. Leur approche introduit la réécriture des équations de la projection perspective (3.2) en fonction de l'ensemble de résidus entre les projections prédites par le modèle perspectif et les projections prédites par son approximation orthographique à l'échelle (un facteur résiduel pour chaque point). Étant connus ces résidus, on peut calculer le point de vue par des méthodes linéaires directs. La première estimation du point de vue est faite en supposant que le vrai modèle de projection est orthographique à l'échelle (c'est à dire, en fixant les résidus à zéro). Le nouveau point de vue est employé pour recalculer les résidus, qui sont à leur tour utilisés pour raffiner l'estimation du point de vue, et ainsi successivement jusqu'à la convergence.

Cet algorithme a l'avantage de se passer d'une estimée initiale du point de vue. D'après quelques tests [DD95, SP95], l'algorithme converge normalement vers des valeurs proches du véritable point de vue, même si les résidus sont élevés, à condition que les points utilisés de la scène se trouvent approximativement distribués autour de l'axe focal.

Néanmoins, pour cette méthode, il est nécessaire de connaître au préalable le rapport entre la distance focale et la profondeur du plan auxiliaire, pour obtenir l'échelle du modèle. D'ailleurs, des erreurs dans la détermination de cette échelle peuvent occasionner des déviations sur le point de vue calculé [MC00a].

3.2.5 Modèle Euclidien

Plusieurs applications de vision par ordinateur ont pour finalité l'obtention d'information métrique telles que : la trajectoire de la caméra (e.g. dans la navigation robotisée), la reconstruction tridimensionnelle d'objets ou la mesure de la taille de certains éléments de la scène observée.

Dans les différents cas, la solution est liée à l'obtention de la transformation entre le système de référence désigné dans la scène (le repère du monde) et le repère de la caméra, dont la transformation observée est Euclidienne. Le modèle de cette transformation a donc 6 degrés de liberté et est de la forme :

$$\begin{pmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^T & 1 \end{pmatrix} \tilde{\mathbf{X}}$$

Ce modèle correspond à la matrice des paramètres extrinsèques du modèle général de projection (éq. 3.3).

Les algorithmes de calcul du point de vue par des techniques de vision se classifient essentiellement en deux groupes :

- Des méthodes où la matrice de projection générale est d'abord déterminée et ensuite les paramètres extrinsèques sont extraits (normalement par factorisation de cette matrice) : par exemple, la méthode de Faugeras-Toscani [FT87] pour le modèle de projection perspective, ou la méthode de Poelman-Kanade [PK94] pour le modèle para-perspectif.
- D'autres méthodes utilisent l'hypothèse courante que les paramètres intrinsèques sont connus au préalable au moyen d'un processus de pré-calibrage, en estimant uniquement le positionnement de la caméra. Quan et Lan [QL98] précisent que le point de vue peut être calculé de manière unique et estimé linéairement à partir de quatre points de référence en position générale et connue. D'autres méthodes proposent un calcul direct du point de vue à partir de correspondances entre points (Horaud [HCLL89]) ou entre lignes (Lowe [Low92]).
  Une compilation et analyse des différentes techniques a été proposée par Haralick [HLON91].

3.2.6 Comparaison entre différents modèles de caméras

Le choix du modèle de caméra requiert aussi de mesurer la justesse avec laquelle le modèle prédit la position sur l'image où un point donné sera placé. Une mesure typiquement utilisée en RA est l'erreur de reprojection. Cobzas et Jägesrsand [CJ03] ont proposé une étude basée sur ce critère d'évaluation pour analyser comment le choix du modèle de la caméra peut affecter la précision du recalage. Ils comparent expérimentalement les modèles affine, Euclidien, orthographique à l'échelle, et perspectif, en utilisant deux ensembles de points sur la scène, l'un pour l'obtention des paramètres de chaque modèle, et l'autre pour valider l'erreur de reprojection entre les positions détectées et attendues.

D'après cette évaluation, les erreurs de reprojection pour les modèles affine et projectif sont inférieures à celles obtenues avec les modèles Euclidien et orthographique à l'échelle. Les auteurs expliquent que l'erreur élevée obtenue pour le modèle Euclidien peut être due à une erreur sur le processus d'auto-calibrage ; c'est à dire, à d'autres facteurs que le modèle de caméra utilisé.

Toutefois, les distributions des primitives utilisées par [CJ03] dans leurs expériences ne sont pas suffisamment variées pour mettre en évidence les propriétés des différents modèles : l'adéquation des modèles approximatifs de la projection perspective varie en fonction des différents niveaux de zoom de la caméra, de la distance entre la caméra et l'objet de calibrage, ainsi que des différences de profondeur dans les objets de calibrage observés.

Lorsque l'erreur de reprojection est utilisée comme critère d'évaluation, le modèle de projection perspective se présente comme une approximation satisfaisante d'une grande variété de dispositifs d'acquisition d'images.

Les autres modèles de cette section, qui peuvent être considérés à la fois comme des approximations numériques ou géométriques du modèle de projection perspective, semblent plus stables grâce à la diminution du nombre de paramètres à estimer, à leur linéarité, et permettent en général un calcul plus robuste de leurs paramètres même avec un faible nombre de données [SP95, DD95, KV96].

Néanmoins, la validité des approximations du modèle de projection perspective est limitée et dépend principalement des facteurs suivants :

Le type de la scène. Les primitives de référence ne doivent pas disposer des différences de relief importantes par rapport à la distance entre l'objet et la caméra.

La position de la caméra. Les points utilisés pour l'étalonnage du modèle ne peuvent pas être trop éloignés de l'axe focal.

Pour obtenir des résultats plus précis dans le calcul du point de vue, Soatto et Perona [SP95] suggèrent que, lorsqu'on utilise des modèles simplifiés de caméras, l'objet de calibrage doit être visé approximativement au centre de l'image, ainsi les points de calibrage se trouveront près de l'axe optique de la caméra. Cependant, cette stratégie peut être inadéquate dans les applications où les mouvements de la caméra par rapport à la scène ne sont pas contraints, comme c'est le cas pour la RA.

Les paramètres intrinsèques de la caméra. Les images acquises avec une caméra dont la distance focale est grande préservent la condition d'un champ de vue réduit. Les distances des points de référence à l'axe focal sont alors petites par rapport aux distances entre les points et le plan image.

Lorsqu'on peut garantir ces conditions, les modèles simplifiés constituent une approximation satisfaisante de la projection perspective. Cependant, les applications qui requièrent une grande flexibilité, soit dans le type de scène, soit les mouvements de la caméra ou les variations des paramètres intrinsèques, nécessitent de considérer un modèle plus général tel que le modèle perspectif.

Dans le choix du modèle de caméra, outre la mesure d'adéquation du modèle aux images acquises, on doit également prendre en compte les caractéristiques des méthodes employées pour l'estimation des paramètres : si les approches des solutions sont linéaires ou des techniques d'optimisation non linéaires sont nécessaires ; si on requiert d'effectuer des calculs itératifs ou s'il existe des solutions directes. Il est aussi recommandable d'évaluer la robustesse des estimations par rapport aux données en entrée et pour les méthodes itératives, il faut examiner la sensibilité de la méthode aux solutions initiales ainsi que les conditions de convergence de la méthode.

3.3 Approches pour améliorer la stabilité

La cohérence temporelle est fréquemment affectée par l'utilisation de méthodes de calcul qui estiment invariablement la totalité des paramètres, même lorsqu'un nombre réduit de variables est concerné.

Il est donc important d'utiliser toute information additionnelle qui contribue à améliorer la stabilité : connaissance sur les mouvements de la caméra, contraintes sur les paramètres internes ou sur la structure de la scène, etc. Par exemple, lorsque la caméra se déplace sur un dispositif robotique, il est possible d'inclure des prédictions sur la vitesse et la trajectoire de la position de la caméra. Ou bien, on peut profiter de la présence de structures planes dans plusieurs types de scènes (tant à l'extérieur qu'à l'intérieur) pour utiliser des transformations plan à plan dans le mouvement rigide.

Dans cette section, nous citons différentes stratégies qui ont été adaptées aux systèmes de RA pour intégrer des connaissances complémentaires.

3.3.1 Approche multi-image : l'ajustement de faisceaux

Bien qu'il existe des systèmes de suivi d'une caméra basés sur l'analyse d'une seule image ou d'un couple d'images, certaines approches suggèrent d'utiliser un plus grand nombre d'images afin d'exploiter les multiples observations des primitives pour améliorer la stabilité des résultats.

La méthode d'ajustement de faisceaux [TMHF00] utilise les observations de primitives sur plusieurs images pour calculer simultanément la structure de la scène ainsi que les paramètres de la caméra.

Dans cette méthode, on minimise une fonction d'erreur globale qui a comme variables, les paramètres de la caméra à chaque prise de vue ainsi que les positions 3D des points suivis. Une fonction fréquemment employée est la vraisemblance de la reconstruction des positions 3D à partir des correspondances de primitives dans les images [BS01], qui est normalement la somme des erreurs de reprojection au carré. D'ailleurs, on peut incorporer certaines informations à cette fonction d'erreur, comme des contraintes sur les primitives (e.g. coplanarité [Bar00]) et sur les paramètres (si certains paramètres restent constants au long de la séquence ou si certains valeurs sont connues a priori [TMHF00]).

Les résultats obtenus au moyen de cette méthode semblent assez robustes grâce à l'emploi des correspondances de primitives sur plusieurs images. Les multiples observations effectuées nous permettent d'explorer différentes distributions des indices ainsi que d'améliorer la mise en correspondance des primitives.

Le nombre de variables à calculer simultanément augmente en fonction du nombre d'images et de primitives suivies, et par conséquent, la complexité du système d'équations à résoudre devient aussi plus grand. Bien qu'il existe des mises en œuvre utilisant des méthodes efficaces d'optimisation et de manipulation de matrices *creuses* (employées fréquemment pour cette technique), la quantité de données à traiter au moyen de cette méthode est contraint par la vitesse de calcul demandée au système. Il existe donc un compromis entre le temps de calcul et le nombre d'images à traiter, mais en même temps un nombre d'images plus élevé peut contribuer à améliorer la précision.

Lorsqu'on souhaite utiliser les paramètres résultants d'un ajustement de faisceaux pour effectuer la superposition d'objets virtuels sur les images de la scène réelle (la même séquence utilisée pour effectuer l'ajustement), la phase d'ajustement doit toujours précéder la phase d'augmentation d'images, car la reconstruction des points de la scène et le calcul des paramètres de la caméra sont obtenus simultanément. Dès lors, cette technique est employée régulièrement dans les méthodes hors ligne ; par exemple, pour la réalisation d'effets spéciaux pour le cinéma en phase de post-production. Mais son utilisation dans les systèmes en temps réel (tels que les applications de la RA) est souvent limitée à effectuer la reconstruction de la scène comme une étape préliminaire du processus, en employant par la suite une méthode de suivi de la caméra basée modèle.

3.3.2 Construction de cartes et localisation simultanées

Certaines systèmes de RA doivent être utilisés dans des environnements où la structure n'est pas connue auparavant. Une possibilité pour affronter cette limitation consiste à reconstruire la scène au fur et à mesure du déroulement de l'application et utiliser le modèle obtenu de la scène pour se repérer dans l'environnement. Cependant, cette stratégie doit être réduite en complexité, lorsque les processus de reconstruction et de localisation doivent se développer en temps réel.

Si le système est contraint en temps de calcul, seul un nombre réduit de points (dits *points clés*) peut être calculé. En réalité, l'ensemble des points clés récupérés est limité et dépourvu de structure. Cependant, le principal but de cette stratégie n'est pas l'obtention d'une reconstruction précise de la scène mais l'emploi de cet ensemble de points clés (appelé *carte causale*) comme une démarche intermédiaire pour l'estimation du positionnement de la caméra.

Quelques approches [GRS+02, DM02] ont été proposées pour estimer la position de certains points clés de la scène et, en même temps, employer ces points pour estimer la localisation de la caméra.

Au départ de l'application, la position de la caméra peut être estimée, dans les toutes premières images de la séquence, soit par une localisation basée modèle [DMM03] soit par la construction d'une carte hors ligne utilisée ensuite pour la localisation en ligne.

Après cela, la mise à jour du positionnement de la caméra s'effectue de manière itérative en deux étapes :

- Dans la première étape, le positionnement de la caméra est estimé en utilisant la carte des points clés.
- Dans l'étape d'apprentissage, les positions des points clés de la carte causale sont raffinées, en utilisant leurs positions précédentes et leurs incertitudes estimées, ainsi que la localisation calculée de la caméra. Dans cette étape, des nouveaux points de l'image peuvent être détectés et leurs positions 3D calculées pour les additionner à la carte.

À la différence des méthodes qui réalisent la reconstruction de la scène exclusivement dans la phase d'initialisation, l'étape d'apprentissage de ces techniques n'est pas contrainte à la phase d'initialisation, mais c'est un processus qui se répète tout au long de l'utilisation du système pour raffiner les estimations des points clés, et surtout pour introduire des nouveaux points de référence à la carte lorsque des nouveaux indices apparaissent dans les images. Ainsi, cette propriété facilite la conception des systèmes plus autonomes.

Cependant, le faible nombre de points clés sur la carte, ainsi que les incertitudes dans l'estimation de leurs positions dans l'espace peuvent entraîner des erreurs dans le calcul du positionnement de la caméra,

que pour les applications de la RA, se traduisent par des inconsistances dans les images augmentées.

3.3.3 Filtrage itératif

Un autre moyen fréquemment utilisé pour aborder le problème de stabilisation du point de vue est le filtrage itératif par inférence statistique, le filtre de Kalman étant la méthode la plus connue dans ce genre. Cette technique nous permet d'incorporer des hypothèses de mouvement pour régulariser la trajectoire de la caméra.

En général, le filtrage itératif combine les estimations faites par un modèle d'évolution des paramètres (*phase de prédiction*) avec les nouvelles mesures qui sont incorporées de manière incrémentielle pour rectifier les prédictions (*phase de correction*).

Pour les systèmes de RA, cette caractéristique itérative du filtre a permis la mise en œuvre d'implémentations pour aider à la prédiction des positions des primitives dans les images [Gag99] et le calcul stable des paramètres de positionnement de la caméra en temps réel [Gen92, DM02].

Un détail important à prendre en considération dans la conception d'un filtre itératif est la spécification du modèle de mouvement dans la phase de prédiction. La performance de l'algorithme dépend considérablement de la capacité de ce modèle à s'ajuster approximativement aux vrais déplacements de la caméra.

L'utilisation d'un modèle de prédiction du mouvement peut s'avérer appropriée pour certaines situations, si l'évolution du mouvement est contrôlée ou approximativement connue [Oli97]. Comme dans la navigation robotique [DMM03], où les déplacements satisfont certaines hypothèses fréquemment imposées : par exemple, que la vitesse ou l'accélération soient localement constantes. Cependant, plusieurs auteurs ont signalé que le choix d'un modèle de mouvement n'est pas toujours facile pour certaines applications de la RA [Low92], car l'utilisateur peut déplacer la caméra de manière arbitraire et parfois imprévisible [RDLW96], par exemple, lorsque la caméra est portée à la main ou au-dessus de la tête de l'utilisateur, attachée à un casque.

3.3.4 Sélection du modèle de mouvement

L'estimation du point de vue de la caméra est la plupart du temps effectuée sur l'ensemble des paramètres du modèle géométrique utilisé ; mais dans plusieurs situations, le vrai mouvement de la caméra peut être contraint à la variation d'un sous-ensemble de ces paramètres (par exemple, dans le cas panoramique, ou la caméra fixe). Si le type de mouvement effectué par la caméra à chaque déplacement était connu, il serait possible d'améliorer la stabilité de l'estimation du point de vue en calculant seulement les paramètres concernés par le mouvement effectué. Néanmoins, en pratique, nous ne connaissons pas la façon dont la caméra se déplace dans la scène, et la plupart du temps, la seule information disponible pour l'apprendre sont les images acquises.

Une stratégie pour choisir le type de mouvement le plus adéquat pour un déplacement particulier consiste à tester plusieurs modèles de mouvement avec différentes restrictions (par exemple, en estimant seulement les paramètres de rotation, ou de translation) et comparer l'adéquation de chacun de ces modèles au déplacement des primitives entre les images acquises en mesurant, par exemple, l'erreur de correspondance. Cependant, avec cette méthode, le modèle le plus complexe sera toujours mieux adapté aux données, puisque les variables supplémentaires permettent au modèle de s'ajuster non seulement au vrai déplacement, mais aussi au bruit sur les données.

La sélection de modèles [Aka74, Boz87] est une méthode d'inférence statistique qui permet d'identifier parmi une série de modèles, celui qui s'adapte le mieux à un ensemble de données bruitées. Les critères de sélection de modèles prennent normalement en considération deux termes : une mesure de précision (par exemple, l'erreur de correspondance) et une mesure de complexité du modèle. Puisqu'un modèle complexe tend à s'adapter mieux à l'ensemble de données que les modèles plus simples, le terme de complexité établit une estimation statistique pour évaluer si les différences entre les mesures de précision sont dues uniquement à l'augmentation du nombre de degrés de liberté, ou bien si les variables supplémentaires servent à caractériser davantage le mouvement.

Le problème de stabilisation dans le calcul du point de vue en utilisant la sélection de modèles a été abordé par Matsunaga et Kanatani [CM00]. Leur idée est de classifier les mouvements de la caméra par

modèle (stationnaire, panoramique, zoom, ...) de façon à ce qu'un certain nombre de paramètres soit fixé si cela semble adéquat.

Dans cette thèse, nous reprenons l'approche de sélection de modèles pour stabiliser la trajectoire de la caméra dans la méthode itérative multi-planaire (présentée dans le chapitre 2). Dans le chapitre 4 nous examinons les performances de différents critères de sélection, et nous montrons les résultats concernant l'utilisation de la sélection de modèles pour notre méthode itérative.

3.4 Protocole de validation expérimentale

Avec l'objectif de distinguer les facteurs qui affectent la performance d'un système de RA dans des applications réelles, nous proposons un cadre d'évaluation expérimentale qui nous permet de mesurer les critères de qualité sous différentes situations pratiques. Le cadre expérimental a pour buts :

- d'identifier un modèle de caméra qui soit adapté à différentes conditions d'application ;
- d'établir la validité de certaines hypothèses de simplification des modèles ;
- de déterminer les facteurs qui affectent la cohérence visuelle des images augmentées, ainsi que le calcul du positionnement de la caméra.

Dans les deux sections précédentes, nous avons mentionné l'importance d'utiliser un modèle adéquat de caméra ainsi que d'effectuer une estimation stable des paramètres de la caméra. Normalement, la réduction du nombre de degrés de liberté augmente la stabilité des paramètres calculés. Toutefois, l'utilisation de modèles simplifiés requiert d'examiner si cette amélioration en termes de stabilité n'est pas obtenue au détriment de la précision des paramètres de la caméra. Certains travaux analysent les répercussions des erreurs de calibrage pour le problème du calcul simultané de structure et mouvement [CP00, AF98, WHA89, ZK01]. Ces études s'intéressent spécifiquement à la manière dont les erreurs sur les paramètres intrinsèques affectent le processus de reconstruction 3D à partir de deux vues.

Cependant, il n'y a pas de travaux qui analysent la façon dont l'impression visuelle et le calcul du point de vue sont affectés par les erreurs de pre-calibrage, ou par les hypothèses de simplification souvent utilisées, comme celle de la fixation du point principal au milieu de l'image. Ainsi, une partie de nos travaux est dédiée à vérifier la validité de ces hypothèses fréquemment formulées pour simplifier le calcul des paramètres de la caméra, ainsi qu'à examiner l'incidence des erreurs d'estimation de la distance focale et le point principal sur la perception et le calcul du point de vue.

Nous avons décidé de faire des expérimentations sur des séquences d'images réelles, notamment pour prendre en compte des perturbations existantes dans les situations pratiques, mais qui ne peuvent pas être facilement représentées par le modèle de la caméra.

Pour estimer les paramètres de la caméra de façon précise, les images ont été acquises en visant une mire de calibrage (comme celle de la figure 3.1) ; puisque nous connaissons les positions 3D des cibles de la mire, et du fait qu'il existe des méthodes de détection fine de projection de cercles, il est possible d'établir des correspondances 3D-2D précises de ces points. Les paramètres de calibrage sont alors obtenus image par image, en minimisant l'erreur de reprojection pour les cibles de la mire.

La façon dont les critères proposés sont évalués dans nos expériences est décrite ci-dessous.

L'erreur de reprojection. En dehors des points de référence utilisés pour l'estimation des paramètres de la caméra, des petits panneaux de validation ont été placés dans la scène, l'un parmi la distribution de cibles de référence, les autres hors de cette distribution. L'erreur de reprojection est mesurée pour chaque panneau en comparant les centres détectés et leur projection calculée au moyen des paramètres de calibrage obtenus.

L'erreur de positionnement. Dans nos expérimentations, nous contrôlons les mouvements de la caméra à l'aide d'une table robotisée. Dans ces conditions, les déplacements de la caméra sont connus à chaque acquisition et peuvent être comparés au point de vue estimé.

L'incertitude des paramètres. La méthode de propagation d'erreurs permet d'obtenir les matrices de covariances des paramètres dans chaque image de la séquence.

Ces matrices nous apportent en général une mesure de la fiabilité des paramètres calculés : elles nous permettront d'établir des intervalles de confiance dans l'estimation des paramètres, de déterminer les configurations les moins sensibles aux erreurs de détection de primitives, de mesurer les corrélations entre certains paramètres du modèle de la caméra, d'identifier les paramètres qui affectent plus sévèrement les critères d'évaluation proposés pour évaluer la performance du système, ainsi que de détecter les paramètres critiques du modèle de la caméra.

L'évaluation visuelle. Pour effectuer une évaluation visuelle des résultats obtenus, nous ajoutons des cones virtuels sur les disques réels des panneaux de validation. Cela nous permet d'examiner la concordance entre les objets réels et virtuels, et ainsi de juger de la qualité perceptive des images augmentées, notamment pour constater sur l'ensemble de la séquence les problèmes dynamiques dus aux erreurs de modélisation ou aux erreurs introduites sur les données d'entrée.

3.5 Estimation statistique d'incertitudes

Dans la présente section, nous présentons une brève revue de la théorie de propagation d'erreurs, en suivant la méthodologie proposée dans [HZ00] et [Kan96].

3.5.1 Propagation directe

Soient $X \in \mathbb{R}^M$ un vecteur aléatoire de moyenne $\overline{X}$ et matrice de covariance Σ; $f : \mathbb{R}^M \to \mathbb{R}^N$ une fonction différentiable dans le voisinage de $\overline{X}$; et $Y = f(X)$. Donc, l'approximation de premier ordre de $f(X)$ par séries de Taylor est exprimée comme :

$$f(\overline{X} + \Delta X) = f(\overline{X}) + J\Delta X + O(\Delta X^2) \tag{3.7}$$

où $\Delta X = X - \overline{X}$, et J est la matrice jacobienne (les dérivées partielles de premier ordre de f par rapport à X) évaluée en $\overline{X}$.

La moyenne de Y est donc estimée comme :

$$\begin{aligned} E[Y] \approx E[f(\overline{X}) + J(X - \overline{X})] &= E[f(\overline{X})] + JE[X - \overline{X}] \\ E[Y] &\approx f(\overline{X}) + \mathbf{0} \\ E[Y] &\approx f(\overline{X}) \end{aligned}$$

La matrice de covariances de Y est définie par : $V[Y] = E[(Y - \overline{Y})(Y - \overline{Y})^T]$. D'après l'équation 3.7, $Y - \overline{Y} \approx (f(\overline{X}) + J(X - \overline{X})) - f(\overline{X}) = J(X - \overline{X})$. Alors,

$$\begin{aligned} V[Y] \approx E[J(X - \overline{X})(X - \overline{X})^T J^T] &= JE[(X - \overline{X})(X - \overline{X})^T]J^T \\ V[Y] &\approx J\Sigma J^T \end{aligned}$$

En résumé, d'après les approximations de premier ordre de la fonction f, Y est une variable aléatoire de moyenne $f(\overline{X})$ et matrice de covariances $J\Sigma J^T$.

3.5.2 Propagation inverse

Le calcul de paramètres d'un modèle est souvent effectué en utilisant l'estimateur du maximum de vraisemblance. La méthode peut être décrite comme : soit X une variable soumise à une loi f de paramètres Θ, tels que $\overline{X} = f(\overline{\Theta})$ ($\overline{\Theta}$ sont les vrais paramètres qui décrivent l'ensemble de données non bruitées $\overline{X}$) et $X = f(\overline{\Theta}) + W$, où W est un vecteur aléatoire de moyenne $\mathbf{0}$ et de covariance Σ. La méthode du maximum de vraisemblance postule que l'estimation $\hat{\Theta}$ devrait être celle qui maximise la probabilité d'obtenir les valeurs observées sur X.

En général, l'optimisation est faite sur le logarithme de la vraisemblance, ce qui équivaut à minimiser le terme $-2 \log L$, qui pour des distributions de probabilité Gaussiennes peut être exprimé comme une

fonction d'erreur de la forme $\|X - f(\Theta)\|_\Sigma$, c'est à dire, la distance de Mahalanobis entre les données X et les estimations $f(\Theta)$ par rapport à la matrice de covariances Σ.

En utilisant une série de Taylor de premier ordre, f peut être exprimée comme :

$$f(\Theta) = f(\overline{\Theta}) + J_\Theta(\Delta\Theta) + O(\Delta\Theta^2)$$

où J_Θ est le jacobien de la fonction f par rapport à l'ensemble de paramètres Θ.

La minimisation du critère $C = \|X - f(\Theta)\|_\Sigma$ doit satisfaire les conditions d'un minimum local, c'est à dire :

- $\frac{\partial C}{\partial \Theta} = \mathbf{0}$;
- $\frac{\partial^2 C}{\partial \Theta^2}$ est symétrique définie positive.

D'après la première condition, l'estimation $\hat{\Theta}$ doit satisfaire :

$$\frac{\partial}{\partial \Theta}\|X - f(\Theta)\|_\Sigma = \mathbf{0}$$

$$\Longrightarrow -2J_\Theta^T\Sigma^{-1}\left((X - \overline{X}) - J_\Theta(\hat{\Theta} - \overline{\Theta})\right) = \mathbf{0}$$

$$J_\Theta^T\Sigma^{-1}J_\Theta(\hat{\Theta} - \overline{\Theta}) = J_\Theta^T\Sigma^{-1}(X - \overline{X})$$

$$\hat{\Theta} = \overline{\Theta} + \left(J_\Theta^T\Sigma^{-1}J_\Theta\right)^{-1}J_\Theta^T\Sigma^{-1}(X - \overline{X})$$

Cela signifie que l'estimateur $\hat{\Theta}$ peut être exprimé aussi par une transformation affine des données X :

$$\hat{\Theta} = \overline{\Theta} + A(X - \overline{X})$$

où $A = \left(J_\Theta^T\Sigma^{-1}J_\Theta\right)^{-1}J_\Theta^T\Sigma^{-1}$. Alors, les estimations utilisées pour la propagation directe des erreurs (voir section 3.5.1) peuvent être appliquées, de telle façon qu'on obtient :

$$\begin{aligned} E[\hat{\Theta}] &= \overline{\Theta} \\ V[\hat{\Theta}] = A\Sigma A^T &= (J_\Theta^T\Sigma^{-1}J_\Theta)^{-1} \end{aligned}$$

La deuxième condition des minima locaux équivaut à ce que

$$\frac{\partial^2}{\partial \Theta^2}\|X - f(\Theta)\|_\Sigma = 2J_\Theta^T\Sigma^{-1}J_\Theta$$

soit symétrique définie positive. Cette condition est satisfaite lorsque Σ est symétrique définie positive, et J_Θ a le rang égal à la dimension du vecteur de paramètres Θ.

Si ces conditions ne sont pas satisfaites, on dit que le système qui défini l'estimateur $\hat{\Theta}$ est de rang déficient, donc on utilise l'estimation de covariances pour systèmes sur-paramétrés [HZ00]. Dans ce cas, la variance est calculée en utilisant l'opération de pseudo-inversion de Moore-Penrose $(\cdot^+)$ telle que :

$$V[\hat{\Theta}] = A\Sigma A^T = (J_\Theta^T\Sigma^+J_\Theta)^+$$

Chapitre 4

Calcul stable du point de vue de la caméra en utilisant la sélection de modèles

En raison du caractère interactif de nombreuses applications de RA, l'estimation des paramètres extrinsèques de la caméra doit être effectuée de manière séquentielle en temps réel, ou proche de l'être. Cependant, les points de vue calculés sont souvent affectés par des fluctuations statistiques qui se traduisent par des phénomènes d'oscillations ou de glissement des objets ajoutés, ce qui nuit à l'impression visuelle de la scène augmentée. Dans le cadre d'environnements de type multiplanaire, nous considérons dans ce chapitre le problème du calcul du point de vue pour des applications de nature séquentielle et nous proposons d'utiliser des méthodes de sélection de modèles pour améliorer la précision et la stabilité des trajectoires obtenues.

4.1 Introduction

Bien que le problème du calcul du point de vue ait reçu beaucoup attention dans la communauté vision, on est loin d'une solution précise et robuste dans le cadre séquentiel. Dans le contexte de la RA, la qualité du calcul du positionnement de la caméra a des implications sur l'impression visuelle de l'utilisateur sur la scène augmentée. Malgré les progrès réalisés dans le domaine du calcul du point de vue, des fluctuations statistiques sur le point de vue conduisent à des effet de sautillements ou à des effets de glissement de l'objet incrusté dans son environnement. Ce problème est particulièrement perceptible quand le mouvement de la caméra est lent, le bruit sur les primitives extraites conduisant à des fluctuations assez grandes du point de vue. Ce problème de stabilisation de la caméra a été considéré dans [CM00]. L' idée de Kanatani et Matsuaga était de classifier les mouvements de la caméra par un certain nombres de modèles (stationnaire, panoramique, général,...) de façon à fixer un certain nombre de paramètres. Ainsi, la stabilité et la précision du point de vue s'améliore a priori puisque le nombre de degrés de liberté de la fonction à optimiser est moindre.

Dans nos travaux, nous reprenons cette idée et nous apportons les contributions suivantes :

(i) poursuivant l'approche de Matsunaga et Kanatani, nous étudions les performances d'un certain nombre de critères de sélection de modèles différents de ceux envisagés dans leur approche (ii) nous examinons et comparons une variation de la méthode de sélection de modèles qui utilise la persistance temporelle du choix d'un modèle sur plus de deux vues.

Le schéma général de notre approche du calcul du point de vue est présenté en section 4.2. La méthode utilisant la structure multi-planaire de la scène pour calculer le point de vue a été décrite en section 2.5.2. Les tests des critères de sélection ainsi que la méthode de sélection du modèle sont décrites dans les sections 4.3 et 4.4. Enfin, des résultats de scènes augmentées viennent conclure ce chapitre.

4.2 Schéma général

La méthode de recalage que nous proposons dans ce chapitre est basé sur l'algorithme général de recalage multiplanaire (figure 2.4) proposé dans le chapitre 2, étendu par une stratégie de détection du type de mouvement de la caméra appuyée sur la sélection statistique de modèles. Nous supposons que les paramètres intrinsèques de la caméra sont constants et sont déterminés préalablement.

Le processus de recalage décrit une boucle comportant quatre étapes : (i) les points d'intérêt sont extraits et mis en correspondance entre deux images, (ii) les homographies correspondant à chacun des plans observés sont calculées en utilisant une estimation robuste de type RANSAC, (iii) le mouvement de la caméra est choisi en utilisant les critères de sélection de modèles, (iv) le point de vue de la caméra est réestimé en considérant le modèle de mouvement sélectionné. Dans la suite de ce chapitre, chacune de ces étapes est décrite en détail.

Étape de prétraitement :

1. Calcul des paramètres intrinsèques de la caméra.
2. Fournir l'équation des plans observés et utilisés pour le recalage.
3. Calculer la matrice de projection pour la première image $\mathtt{P}^0$,

Calcul de la matrice de projection $\mathtt{P}^i$ pour $i > 0$:

1. Déterminer l'ensemble des points d'intérêt en correspondance entre les images $i-1$ et i pour chaque plan observé.
2. Calculer les homographies induites par chaque plan entre $i-1$ et i.
3. Déterminer le meilleur modèle de mouvement selon le critère de sélection qui réalise un compromis entre précision et simplicité du modèle.
4. Calculer la matrice $P^{(i)}$ à partir de $P^{(i-1)}$ et des homographies en utilisant le modèle sélectionné.

FIG. 4.1 – Schéma général de la méthode de recalage multi-planaire basée sur la sélection de modèle

4.3 Utilisation de méthodes de stabilisation

4.3.1 Objectif

Dans le chapitre 2 nous avons présenté une méthode pour le calcul du positionnement de la caméra. Nous avons montré que la précision du calcul du point de vue est améliorée en considérant une démarche multiplanaire au lieu d'une approche basée sur l'observation d'un seul plan (section 2.5.3).

Néanmoins, on observe très souvent des fluctuations importantes sur les paramètres calculés ce qui se traduit pour l'observateur par l'impression de voir l'objet ajouté osciller ou bien glisser dans la scène. Ces phénomènes sont particulièrement visibles quand le mouvement de la caméra est lent. Par le passé, plusieurs papiers ont préconisé l'utilisation du filtre de Kalman pour prédire et stabiliser les trajectoire. Un tel filtre n'est cependant pas toujours avantageux pour la RA, en particulier parce que les modèles dynamiques considérés en général (modèle à vitesse ou accélération constante) ne sont pas satisfaisants pour décrire les mouvements humains.

Poursuivant les travaux de Matsunaga and Kanatani[CM00] et [TFZ98], nous cherchons à utiliser les méthodes de sélection de modèles pour réduire les fluctuations sur les paramètres calculés et améliorer ainsi l'impression visuelle de la scène augmentée. L'idée sous jacente à la sélection de modèles part de la constatation suivante : un modèle d'ordre élevé approche toujours mieux un ensemble de données qu'un modèle d'ordre inférieur. Cependant, les modèles d'ordre élevé approximent en fait une partie du bruit qu'ils sont censés pourtant éliminer. Un modèle d'ordre élevé, bien que théoriquement plus précis, est donc en fait moins stable aux perturbations aléatoires des données. Une bonne méthode de sélection de modèles doit donc réaliser un compromis entre précision et stabilité. Le principe des méthodes de

sélection de modèles est donc d'exiger que le modèle choisi explique bien les données et ait en même temps la structure la plus simple possible.

Dans le cas du calcul du point de vue, les homographies (une pour chaque plan observé) sont paramétrées par 9 paramètres (R, t, u_0, v_0, f). Si les mouvements où les intrinsèques sont contraints (par exemple si on considère certains mouvements comme les panoramiques, les mouvements stationnaires ou les zooms), les homographies auront un nombre de degrés de liberté moindre et donc un nombre de paramètres à estimer plus petit. La stabilité du point de vue sera donc améliorée si le bon modèle de mouvement est utilisé. Il est donc nécessaire de déterminer le mouvement réel de la caméra en utilisant seulement les observations, c'est-à-dire dans notre cas, deux ou trois images consécutives.

4.3.2 État de l'art

De nombreux critères de sélection de modèles pour réaliser un compromis entre le résidu de l'approximation et le degré du modèle ont été proposés dans la littérature [BS98]. Tous sont la somme d'un terme mesurant la précision du modèle et d'un terme mesurant sa complexité. La plupart d'entre eux sont basés sur des critères statistiques ou sur des critères issus de la théorie de l'information. Parmi eux, les plus utilisés sont sans doute le critère géométrique d'Akaike (AIC) ainsi que le critère de description de longueur minimale (MDL). Le critère AIC peut être vu comme une approximation d'un critère entropique (la distance de Kullbak-Leibler), alors que le critère MDL favorise le modèle dont la description en terme de bits est minimale :

$$G_{AIC} = \hat{J} + 2k\epsilon^2$$
$$G_{MDL} = \hat{J} - k\epsilon^2 log\epsilon^2$$

où k est le nombre de degrés de liberté du mouvement. Le niveau de bruit ϵ^2 est habituellement estimé à partir du résidu $\hat{J}$ (celui correspondant au modèle d'ordre le plus élevé) [CM00][Boz87]). Quel que soit le critère considéré, la valeur associée à chaque modèle est calculée comme la somme du résidu et du facteur de complexité de chaque modèle $E_{critere} = \hat{J} + \epsilon^2 c(M_k)$. L'usage d'un modèle trop complexe est donc pénalisé.

Cette approche a déjà été utilisée par Kanatani[CM00] pour le problème de la calibration à partir d'un seul motif plan observé. Il y classifie les mouvements en 6 catégories, quatre d'entres elles correspondant au cas d'une focale fixe :

Mouvement	**Paramètres connus**	**Variables**
stationnaire	$\mathbf{A}_i = \mathbf{A}_{i-1}, \mathbf{a}_i = \mathbf{a}_{i-1}$	–
panoramique	$\mathbf{a}_i = \mathbf{a}_{i-1}$	$\mathbf{A}_i$
$t-$ prédit	$t_i = 2t_{i-1} - t_{i-2}, \mathbf{a}_i = \mathbf{a}_{i-1}$	$\mathbf{A}_i$
modèle général	–	$\mathbf{A}_i, \mathbf{a}_i$

Dans le cas du modèle prédit, la position de la caméra est extrapolée linéairement par $t_i = 2t_{i-1} - t_{i-2}$, et l'optimisation n'est effectuée que sur la rotation.

Dans l'approche de Kanatani, seuls deux critères de sélection sont étudiés, G_{AIC} et G_{MDL}, en considérant un seul motif plan pour le recalage. Cependant, il existe beaucoup d'autres critères de sélection, en particulier ceux utilisant la matrice de covariance ou la matrice d'information sur les paramètres calculés.

4.3.3 Notre approche de la sélection de modèles

Nous suggérons d'utiliser conjointement les méthodes de sélection de modèles et la stratégie multiplanaire pour accroître la précision et la stabilité des paramètres calculés. La sélection de modèles a été utilisée dans de nombreux domaines. Cependant, il n'existe aucun critère qui soit reconnu comme meilleur dans tous les cas, comme le montrent plusieurs papiers évaluant ces critères pour des problèmes différents : approximation par des polynômes [Boz87], fusion de données surfaciques [BS98], sélection de mouvement [Tor97], détection de primitives géométriques [Kan02].

Pour cette raison, nous avons commencé par comparer différents critères de sélection. Nous avons plus particulièrement considéré les critères qui prennent en considération la matrice de covariance sur les

paramètres estimés ($V(\theta_k)$) et la matrice d'information de Fisher ($I(\theta) = E(\frac{\partial}{\partial\theta}J(X|\theta))^t\frac{\partial}{\partial\theta}J(X|\theta))$). En effet, des critères tels qu'AIC sont seulement des approximations asymptotiques de critères considérant la matrice de covariance ou la matrice d'information. Nous espérons donc que de tels critères amélioreront la sélection de modèles.

Critère	Terme de complexité
Akaike AIC [Aka74]	$2k$
Bozdogan CAIC [Boz87]	$k(\log n + 1)$
Bozdogan CAICF [Boz87]	$k(\log n + 2) + \log \lvert\mathbf{I}(\theta_k)\rvert$
Schwarz BIC [Sch78]	$2k \log n$
Kanatani gMDL [CM00]	$-k \log \epsilon^2$

4.3.4 Résultats expérimentaux

Les expériences que nous décrivons ont été conduites sur la séquence synthétique décrite en section 2.5.3 (utilisée pour évaluer la performance de l'approche multiplanaire) avec différents niveaux de bruit et sur des séquences réelles. Par souci de simplicité dans la présentation des résultats, nous avons affecté un entier à chaque type de mouvement : 0 pour le mouvement stationnaire, 1 pour le panoramique, 2 pour le mouvement général (rotation + translation) et 3 pour le mouvement prédit.

Utilisation du modèle prédictif

Dans de nombreuses expérimentations, les variations des paramètres sont relativement petites et certaines parties peuvent donc apparaître linéaires par morceau à première vue. Le modèle prédictif ayant beaucoup moins de paramètres que le modèle général, le modèle prédictif est alors sélectionné dès que le niveau de bruit s'élève. La sélection de ce modèle trop restrictif peut alors conduire à la divergence du processus. A titre d'exemple, la figure 4.2 montre le comportement du processus de calcul du point de vue pour la même séquence que précédemment quand on prend en compte ou non le modèle prédictif dans la liste de modèle. Il apparaît clairement sur cette figure que la présence du modèle prédictif tend à faire diverger le processus du point de vue attendu. Ces résultats ont été obtenus en considérant le critère CAICF, les autres critères donnant des résultats similaires. Pour ces raisons, nous n'utilisons pas dans la suite de modèle prédictif. Il est préférable en effet d'avoir quelques fluctuations dans les paramètres plutôt que de faire diverger le processus.

Comparaison des critères de sélection

Afin de comparer les critères, nous avons utilisé la séquence synthétique bruitée. Pour chaque image i, nous avons appliqué l'étape de sélection de modèle. Comme cela requiert la minimisation de la fonction de coût, nous avons utilisé la véritable valeur du point de vue à l'étape $i-1$ comme initialisation afin d'éviter que les problèmes éventuels de dérive ne perturbent l'évaluation des modèles. Le véritable modèle étant connu, nous montrons dans les tables 4.1 et 4.2 le pourcentage de modèles correctement choisis pour chacun des critères évalués. Le terme *+ complexe* signifie qu'un modèle d'ordre plus élevé que le vrai modèle a été choisi. Alors que le terme *- complexe* signifie qu'un modèle d'ordre moins élevé, donc trop restrictif, a été choisi.

Ces deux tables montrent que pour une valeur de bruit modérée, la plupart des critères se comportent bien, c'est-à-dire qu'un modèle d'ordre supérieur ou égal au véritable modèle est presque toujours choisi. On note cependant que les critères AIC et $gMDL$ ont tendance à produire des modèles trop généraux, ce qui n'est pas très intéressant dans une optique de stabilisation.

Lorsque le bruit augmente (table 4.2), les performances de certains critères se dégradent nettement. Cependant on peut noter que les deux critères CAIF et CAICF se comportent le mieux : ils se comportent très bien dans le cas stationnaire, n'ont que peu tendance à sélectionner un modèle trop général dans le cas panoramique et se comportent très honorablement dans le cas général. Le critère CAICF apparaît comme supérieur au critère CAIC, puisqu'il induit moins de sélections de modèles d'ordre inférieur que CAIC. C'est la raison pour laquelle les expérimentations qui suivent seront faites en utilisant le critère CAICF.

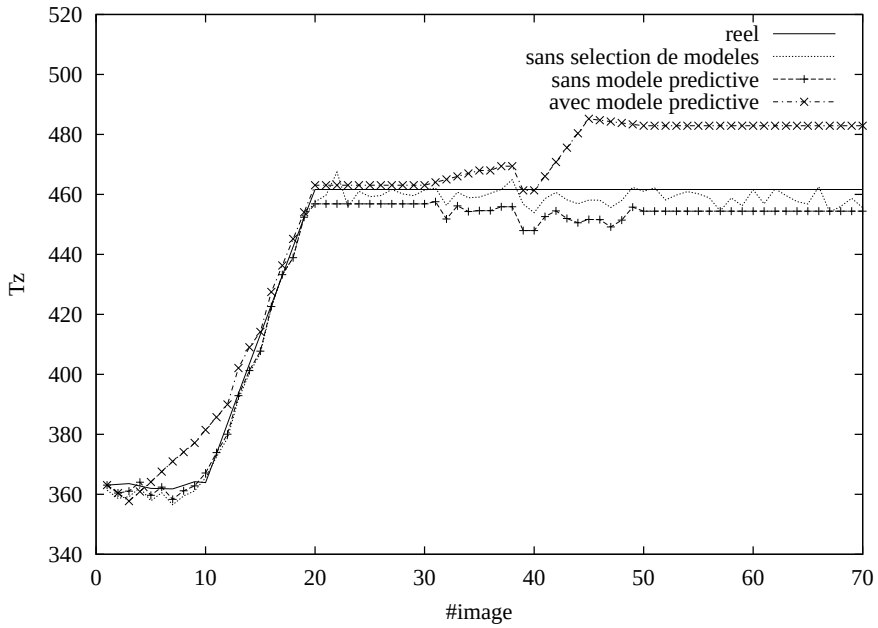


FIG. 4.2 – La translation t_z calculée avec et sans le modèle prédictif.

Ces résultats tendent donc à montrer que l'introduction de la matrice d'information sur les paramètres calculés améliore la sélection du modèle.

Les figures 4.3 et 4.4 comparent les performances des critères de sélection de modèles avec le véritable modèle de mouvement de la séquence. L'axe des x correspond au numéro des images dans la séquence et l'axe des y le modèle de mouvement choisi. On peut constater dans ces graphiques que certaines petites translations sont étiquetées par le processus en modèle panoramique. Ceci est dû au fait que ces deux types de mouvement sont assez difficiles à discerner quand les mouvements sont faibles. Mais comme la complexité du modèle panoramique est moindre que celle de la translation (qui rentre dans le modèle général), le mouvement panoramique est détecté. Une solution pour remédier à ce problème serait d'introduire un modèle translationnel. Cependant, ce mouvement apparaît rarement dans les séquences tournées librement. Nous ne l'avons donc pas introduit dans la liste des modèles.

4.4 Amélioration de la sélection du modèle

4.4.1 Méthode

Cette méthode améliore très sensiblement la stabilité des points de vue et la qualité visuelle des incrustations. Cependant, il peut subsister des erreurs dans l'étiquetage du mouvement, par exemple la confusion entre panoramique et translation dans le cas de petits mouvements. La répétition de choix inappropriés de mouvements pouvant conduire à la divergence du processus, nous souhaitons améliorer le processus de sélection. Pour cela, nous proposons d'utiliser la cohérence temporelle des modèles détectés et d'utiliser donc plus de deux vues pour valider la sélection du modèle. Ceci a pour effet d'augmenter la complexité du processus (les points en correspondance doivent être suivis dans plus de deux images) mais cela contribue à éviter qu'un modèle inapproprié soit choisi.

Dans notre approche, un modèle ne sera donc validé que si ce choix est le même pour au moins 3 images consécutives, ce qui présuppose qu'un modèle de mouvement persiste dans au moins trois images consécutives. Dans le cas où des modèles différents seraient trouvés, le modèle le plus général d'entre eux sera choisi, sachant qu'il est préférable de choisir un modèle plus général plutôt que préférer un modèle trop restrictif qui peut conduire à la divergence. Plus précisément, notre algorithme est le suivant :

mouvement	critère	$\sigma = 0.3$		
		- complexe	**correct**	+ complexe
static	AIC	-	83.1%	16.9%
	CAIC	-	98.7%	1.3%
	CAICF	-	100.0%	0.0%
	BIC	-	100.0%	0.0%
	gMDL	-	77.5%	22.5%
pan	AIC	0.0%	85.3%	14.7%
	CAIC	0.0%	99.3%	0.7%
	CAICF	0.0%	98.7%	1.3%
	BIC	0.0%	100.0%	0.0%
	gMDL	0.0%	84.7%	15.3%
general	AIC	0.0%	100.0%	-
	CAIC	1.5%	98.5%	-
	CAICF	1.3%	98.7%	-
	BIC	5.4%	94.6%	-
	gMDL	0.0%	100.0%	-

TAB. 4.1 – Pourcentage de bonnes sélections de modèles pour le niveau de bruit $\sigma = 0.3$.

1. Les paramètres de la caméra sont connus pour les images $i-1$ et $i-2$.
2. Sélectionner le modèle de mouvement $M_{i,i-1}$ entre l'image courante i et la précédente $i-1$.
3. Sélectionner le modèle de mouvement $M_{i,i-2}$ entre l'image courante i et l'image $i-2$.
4. Le modèle sélectionné est le modèle M' le plus simple tel que l'espace des paramètres de $M_{i,i-1}$ et $M_{i,i-2}$ soient des sous espaces de M'. Si les espaces sont emboîtés, ceci signifie que M' sera le plus général des deux modèles $M_{i,i-1}$ et $M_{i,i-2}$.

Cette méthode comporte toutefois un léger inconvénient quand on passe d'un modèle complexe à un plus simple, la première image de la transition est alors toujours affectée du modèle le plus complexe. La transition au modèle le plus simple se fait donc avec un temps de retard.

4.4.2 Ajustement

Une fois le modèle estimé sur la base de trois images, la position de la caméra courante est recalculée de la façon suivante en tenant compte des points en correspondance (x_{i-1}, x_i) et (x_{i-2}, x_i) entre les trois images :

$$\begin{aligned} J(\mathbf{A}, \mathbf{a}) &= \sum_{k=1}^{n} \sum_{j=1}^{N_k} \|x_{kj}^i - Z(H_k^{i,i-1} x_{kj}^{i-1})\|^2 \\ &+ \sum_{k=1}^{n} \sum_{j=1}^{N_k} \|x_{kj}^i - Z(H_k^{i,i-2} x_{kj}^{i-2})\|^2 \end{aligned}$$

où k désigne le plan observé et H_k^{ij} l'homographie induite par le k^{ieme} plan entre les images i et j..

4.4.3 Résultats

Nous présentons d'abord des résultats de sélection de modèles utilisant des triplets d'images sur la séquence synthétique additionnée d'un bruit de variance ($\sigma = 1.0$). La figure 4.5 montre le modèle sélectionné sur la séquence en utilisant deux ou trois images pour la sélection. La figure 4.6 montre la composante translationnelle T_Z calculée. Nous pouvons constater qu'entre les images 30 et 50, si nous utilisons seulement deux images pour la sélection, il y a un certain nombre de confusions entre modèle panoramique et modèle général. Ces problèmes s'atténuent visiblement quand les triplets d'images sont

motion	criterion	$\sigma = 1.0$		
		- général	**correct**	+ général
static	AIC	-	83.7%	16.3%
	CAIC	-	100.0%	0.0%
	CAICF	-	100.0%	0.0%
	BIC	-	100.0%	0.0%
	gMDL	-	0.0%	100.0%
pan	AIC	0.0%	86.7%	13.3%
	CAIC	0.0%	100.0%	0.0%
	CAICF	0.0%	97.3%	2.7%
	BIC	0.0%	100.0%	0.0%
	gMDL	0.0%	0.0%	100.0%
general	AIC	11.5%	88.5%	-
	CAIC	24.1%	75.9%	-
	CAICF	20.3%	79.7%	-
	BIC	33.6%	66.4%	-
	gMDL	0.0%	100.0%	-

TAB. 4.2 – Pourcentage de bonnes sélections de modèles pour le niveau de bruit $\sigma = 1.$.

utilisés pour la sélection. De manière générale, la probabilité de sélectionner un mauvais modèle décroît quand les triplets sont utilisés. Enfin le diagramme montrant T_Z prouve que l'utilisation de triplets améliore la précision du point de vue calculé par rapport à l'utilisation de deux vues.

Séquence réelle : la mire

Nous montrons ici des résultats concernant une séquence réelle de 260 images montrant une mire de calibration. La figure 4.7 montre le résultat de la sélection de modèle en utilisant deux ou trois images. L'utilisation de triplets permet d'éliminer un certains nombre de sélections erronées. La figure 4.8 montre la composante T_x calculée entre les images 130 et 170. Le point de vue calculé avec les triplets est plus précis qu'avec le processus à deux images.

Séquence de la pièce

Enfin, nous considérons une séquence de 225 images d'une pièce en modèle réduit. Cette séquence était constituée successivement d'un mouvement stationnaire, général, stationnaire, panoramique, stationnaire, panoramique et enfin stationnaire.

La figure 4.9 montre que l'utilisation de triplets améliore légèrement la sélection du modèle. Les images de 30-110 montrent que certains mouvement, identifiés à tort comme panoramique ou stationnaire quand on utilise deux vues, sont bien classifiés en modèle général quand on utilise trois vues.

La figure 4.10 montre que la sélection du point de vue améliore la stabilité de la trajectoire. Ce graphique montre les points de vue calculés lorsque le modèle général est toujours utilisé et lorsque la sélection de modèles est utilisée. Ce graphique montre clairement l'effet stabilisateur de la sélection de modèle puisque la section 0-30 est bien stationnaire. Les images de la figure 4.11 montrent également l'impact de la sélection sur la robustesse du processus : l'image (a) montre en effet la scène augmentée au bout de 200 images lorsqu'aucune sélection de modèle n'est utilisée (c'est à dire quand le modèle général est toujours utilisé). L'image (b) montre l'image augmentée quand la sélection est utilisée. Il est bien clair sur ces deux images que la sélection apporte beaucoup de robustesse et le lecteur pourra s'en convaincre en regardant sur le site
http ://www.loria.fr/~vigueras/orasis2003.html les vidéos complètes. De façon générale, la sélection de modèle réduit les variations aléatoires de certains paramètres du point de vue ce qui permet de diminuer les accumulations d'erreur en composant les mouvements relatifs ; ceci conduit donc au final à une évaluation plus robuste du point de vue.

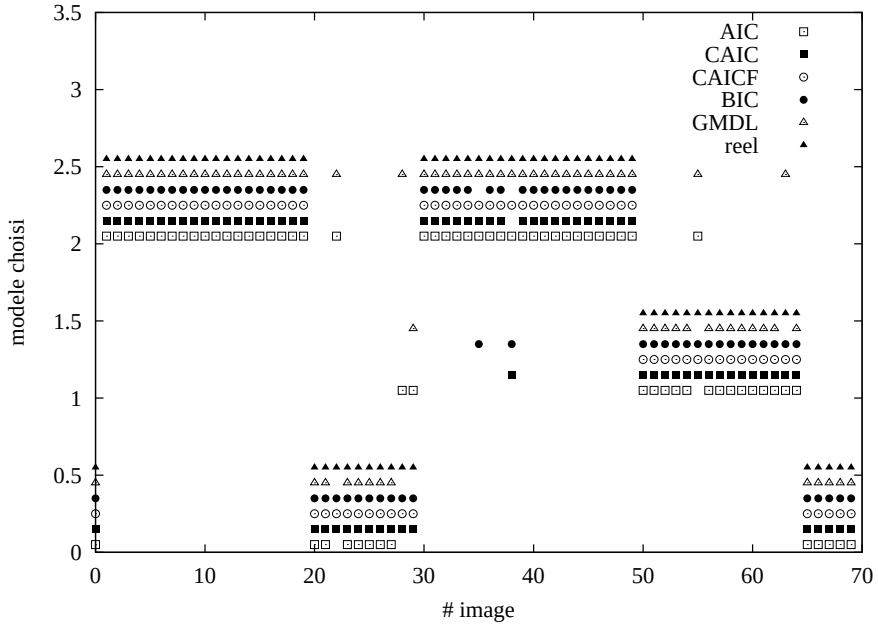


FIG. 4.3 – Comparaison des critères de sélection sur une séquence synthétique bruitée $\sigma = 0.5$.

Finalement, la figure 4.12 montre quelques exemples d'augmentation d'une scène avec un cube. La séquence complète est disponible sur notre page web. Un symbole dans le coin supérieur gauche de l'image indique le modèle sélectionné : la croix rouge désigne le modèle stationnaire, le cercle vert le panoramique et le carré bleu le modèle général.

4.5 Conclusion

Nous avons proposé dans ce chapitre plusieurs améliorations aux méthodes existantes de calcul séquentiel du point de vue pour des scènes multi-planaires : contexte multiplanaire pour le calcul du point de vue, test de divers critères de sélection de modèle pour améliorer la stabilité de la trajectoire, proposition d'un critère sur plus de deux images pour améliorer la sélection du modèle. Les résultats de cette étude montrent que cette méthode améliore de façon importante la précision et la stabilité de la trajectoire calculée. Le test de différents critères de sélection de modèles a mis en évidence que l'usage de critères impliquant l'information sur la covariance des paramètres calculés améliorait la précision et la robustesse des trajectoires calculées.

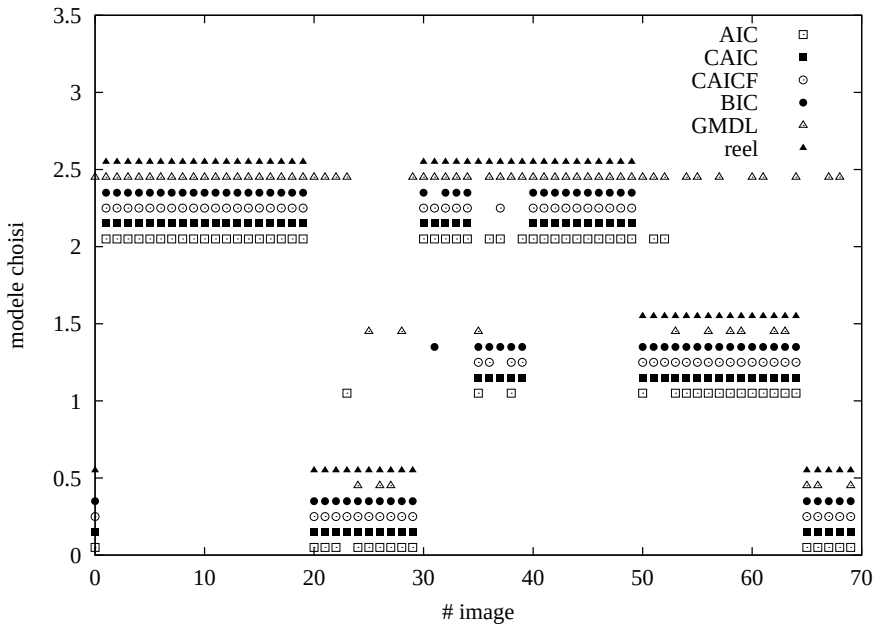


FIG. 4.4 – Comparaison des critères de sélection sur une séquence synthétique bruitée $\sigma = 1$..

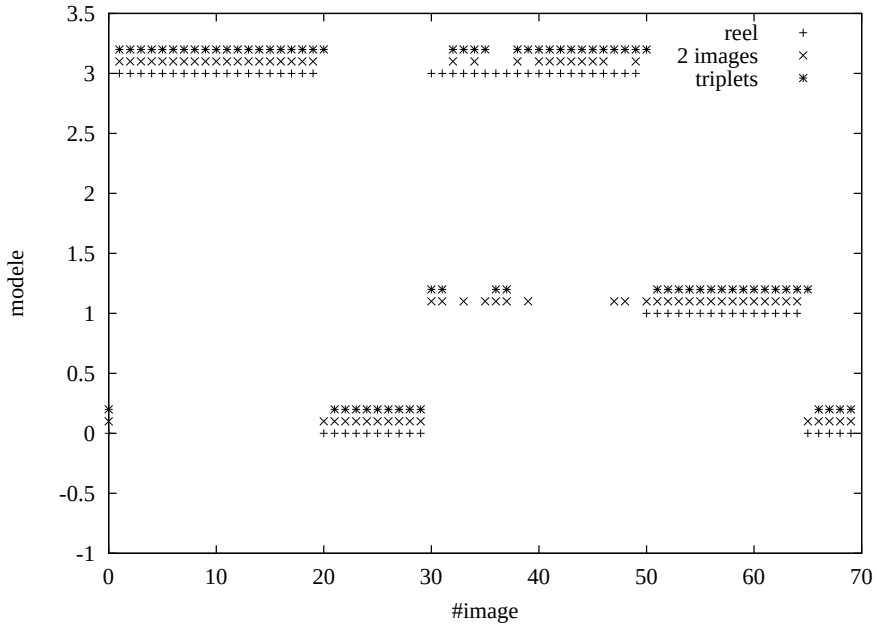


FIG. 4.5 – Sélection du modèle en utilisant deux ou trois vues sur la séquence synthétique ($\sigma = 1$).

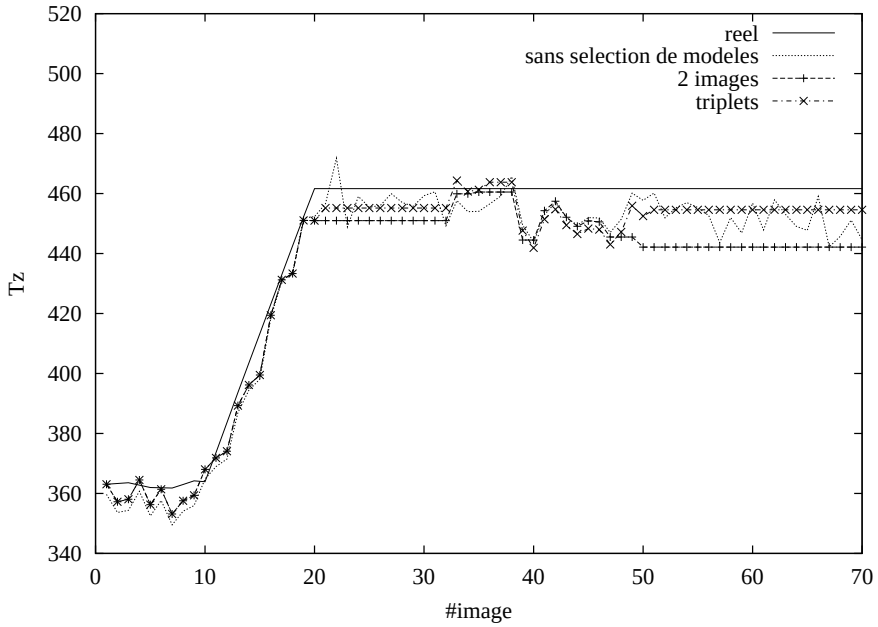


FIG. 4.6 – Composante T_Z calculée en utilisant la sélection sur deux ou trois images pour la séquence synthétique.

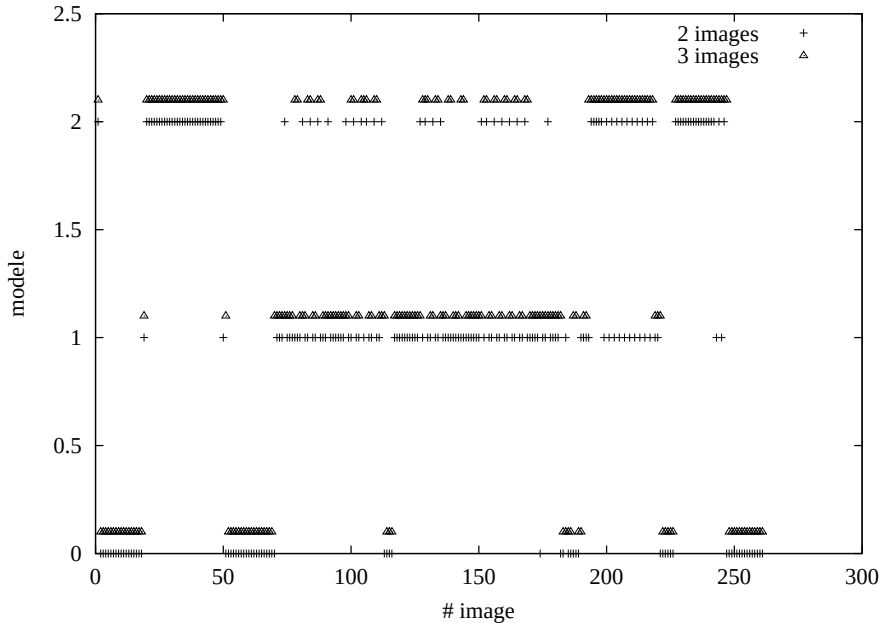


FIG. 4.7 – Séquence de la mire : sélection de modèle avec deux ou trois images.

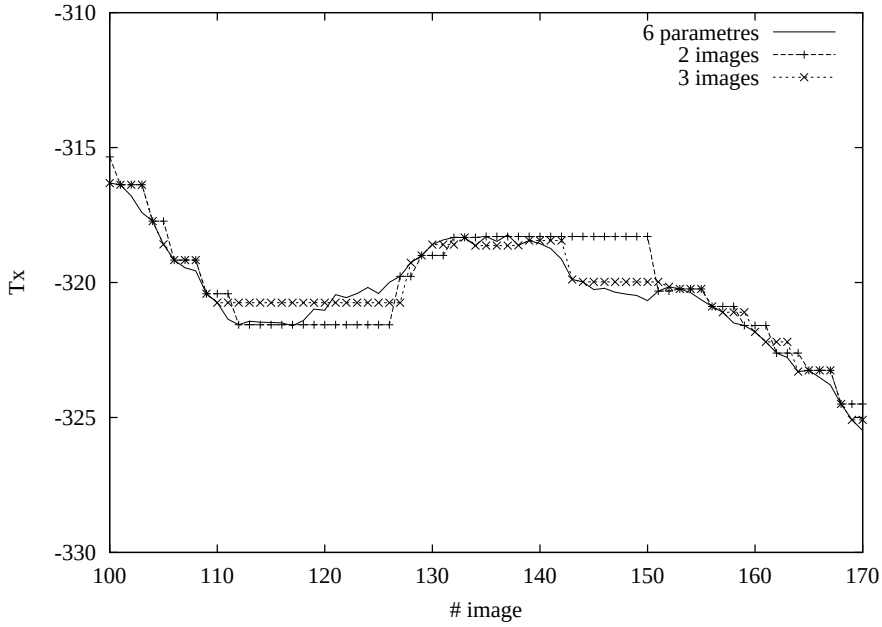


FIG. 4.8 – Séquence de la mire : la composante T_X calculée.

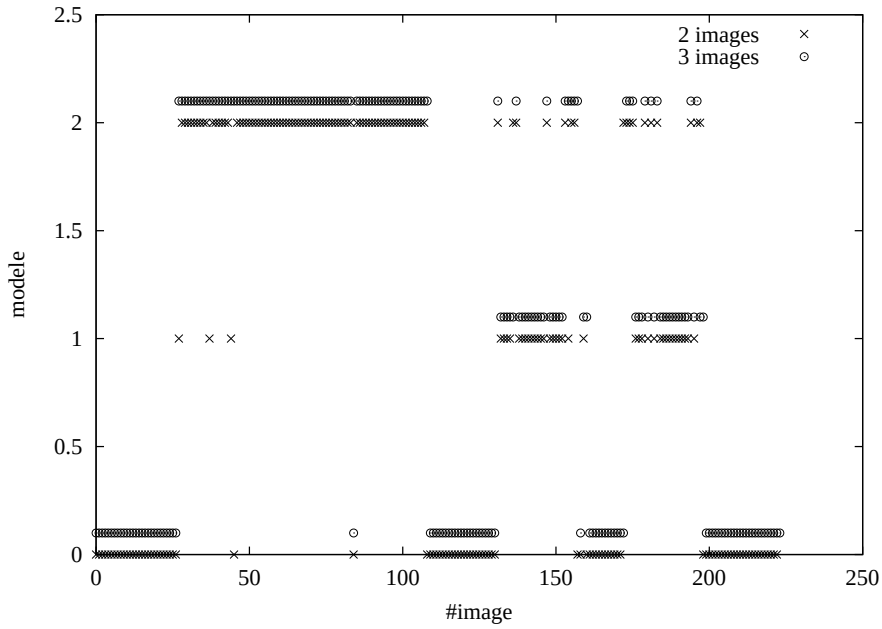


FIG. 4.9 – Séquence de la pièce : sélection du modèle.

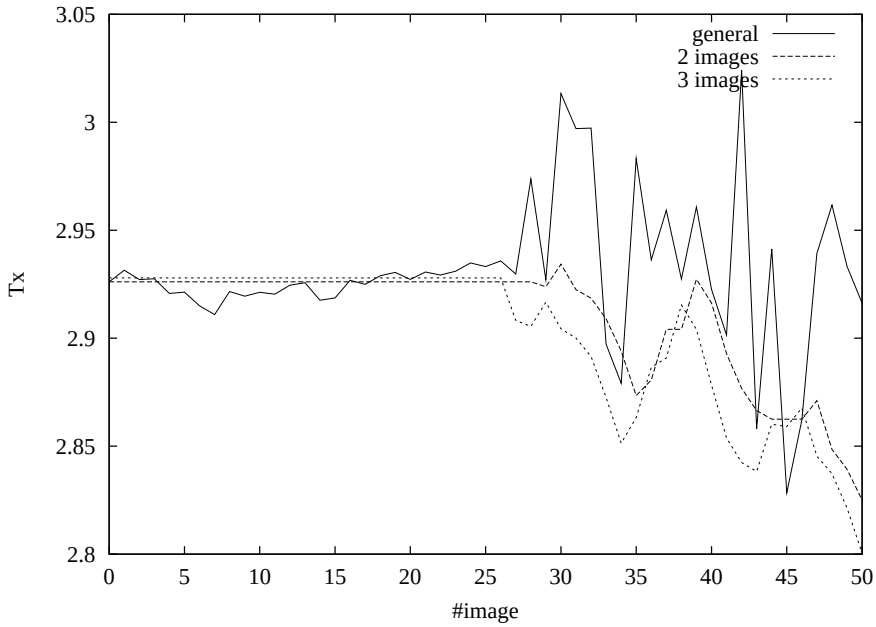


FIG. 4.10 – Séquence de la pièce : la translation en x calculée avec la sélection de modèle.

a. b.

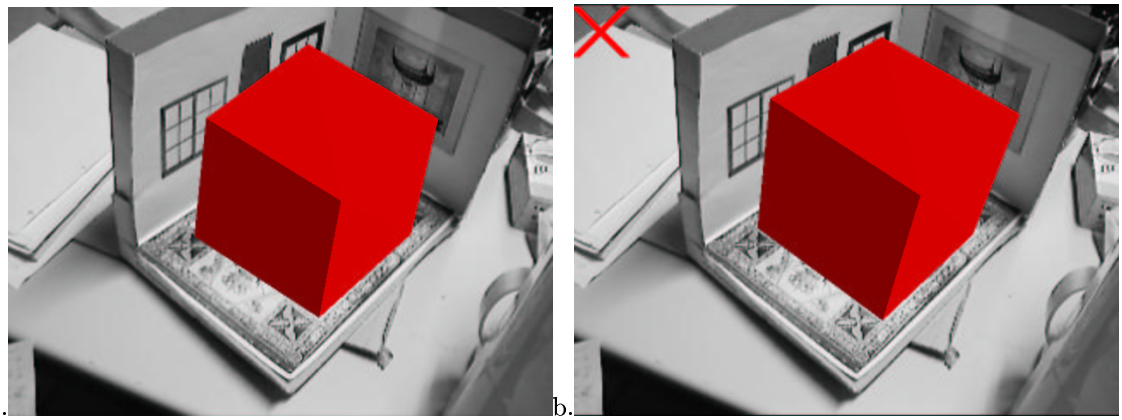


FIG. 4.11 – Un cube incrusté sur l'image 200 : (a) sans sélection de modèle ; (b) : avec sélection

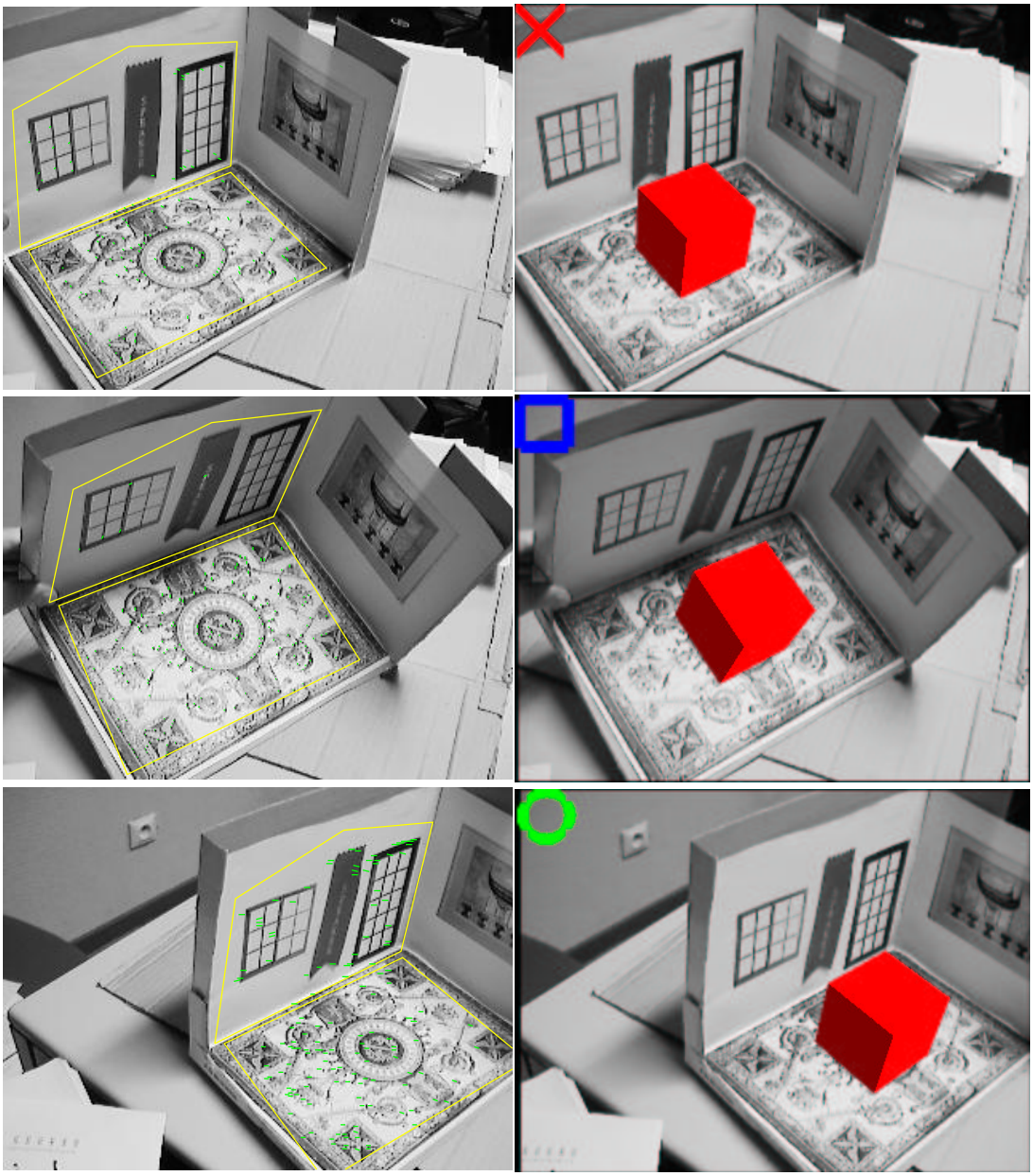


FIG. 4.12 – Quelques exemples de la scène augmentée. La première colonne montre le suivi des points dans les trois plans utilisés, la second colonne montre la scène augmentée.

Chapitre 5

Étude pratique des erreurs de calibrage en RA

Dans ce chapitre, nous évaluons la pertinence effective du modèle sténopé dans le cadre de la réalité augmentée. Nous mesurons la précision obtenue avec ce modèle en situation réelle, avec des paramètres intrinsèques éventuellement entâchés d'erreur. Nous avons utilisé pour cette étude des images réelles obtenues à partir de caméras de différentes qualités dont les déplacements sont contrôlés par ordinateur. Nous avons proposé un protocole expérimental qui nous a permis d'utiliser des critères d'évaluation spécifiques à la RA tels que l'erreur d'alignement et l'impression visuelle, en plus de l'erreur de positionnement plus couramment utilisée.

5.1 Introduction

Le modèle de caméra le plus souvent utilisé dans le problème de positionnement est le modèle sténopé, et les paramètres intrinsèques sont fréquemment calculés dans une étape préliminaire dite de *pre-calibrage*, puis considérés comme constants durant le processus de suivi. Des hypothèses simplificatrices comme celle de fixer le point principal au centre de l'image sont aussi souvent utilisées pour réduire le nombre de données requises par le système de calibrage et / ou la sensibilité du système face au bruit présent sur ces données.

L'objectif de ce chapitre est de regarder quels peuvent être les effets de telles approximations dans un contexte de réalité augmentée, et de proposer des recommandations pratiques pour obtenir une précision suffisante suivant le type d'application visé.

Plusieurs travaux ont été dédiés à l'analyse des effets d'un calibrage erroné sur le calcul simultané du mouvement de la caméra et de la structure de la scène [WCH92, Oli97, AF98, LVD98, HK02]. La plupart de ces travaux s'intéressent à l'étude théorique des effets d'un calibrage erroné sur l'estimation du mouvement, en supposant que le modèle sténopé est parfaitement respecté par la caméra. Ces travaux ont pu expliquer certaines difficultés liées au positionnement de caméra : par exemple, Agapito et al. ont montré dans [AHR01] qu'une erreur obtenue sur le point principal pouvait être compensée par une erreur sur les angles de rotation, sans que cela ne soit perceptible au niveau de la projection image. Cependant, ces travaux ne remettent généralement pas en cause le respect du modèle sténopé par les caméras réelles. D'autre part, ils n'intègrent pas certains aspects pratiques et numériques du calibrage : les paramètres sont calculés à partir d'un ensemble d'images et de primitives particulières, et leur validité est souvent restreinte au volume où les mesures sont prises [PJN99, CJ03]. Or, dans nombre d'applications, il peut être nécessaire d'obtenir un alignement précis hors du volume de mesure. Un exemple est donné en gynécologie [STN04], dans une application d'apprentissage de l'extraction d'un enfant au forceps : dans cet exemple, les parties invisibles d'un forceps, situées à l'intérieur d'un mannequin, sont affichées par-dessus le mannequin en utilisant un système de positionnement basé sur le suivi automatique de marqueurs rétro-réfléchissants, placés aux extrêmités visibles du forceps.

Tout comme Bougnoux dans [Bou98], bien que dans un contexte différent, nous analysons l'influence

des erreurs de calibrage en situation réelle. Nos résultats ont été obtenus avec deux caméras de qualités différentes, induisant plus ou moins de distorsion dans les images acquises : une caméra possédant une bonne optique et une simple webcam. Ces résultats montrent en particulier que fixer les paramètres intrinsèques n'a de sens que lorsque la distorsion radiale est corrigée, y compris pour des caméras présentant très peu de distorsion. Nos résultats ont été obtenus à l'aide d'une caméra fixée sur un dispositif de déplacement contrôlé par ordinateur, possédant 4 degrés de liberté. Une mire de calibrage est utilisée pour calculer les paramètres de la caméra sur toute une séquence, en supposant les paramètres intrinsèques fixes ou variables, puis corrects ou erronés. Enfin, trois marqueurs, dont les positions 3D sont connues, sont placés à des distances plus ou moins éloignées de la mire. Ce protocole nous permet d'analyser l'influence du calibrage en RA, en nous basant sur trois critères : les points de vue de la caméra (les points de vue calculés sont confrontés aux commandes envoyées au dispositif de déplacement), l'erreur d'alignement dans les images (les positions projetées des marqueurs sont comparées aux positions détectées dans la séquence) et l'impression visuelle (des vidéos d'incrustation accompagnent le chapitre).

Le plan du chapitre est le suivant : la section 5.2 présente plus en détail notre protocole expérimental ; en section 5.3, nous évaluons la pertinance de fixer les paramètres intrinsèques de la caméra avec le modèle sténopé classique, puis en tenant compte de la distorsion radiale ; la section 5.4 montre l'influence d'un calibrage erroné en RA et nos conclusions sont résumées en section 5.5.

5.2 Protocole expérimental

Le dispositif permettant de contrôler les mouvements de la caméra est constitué d'une table en translation à 2 degrés de liberté (appelée table XY), de résolution 0.48 mm, sur laquelle se déplace une unité à deux axes de rotation (*pan-tilt unit* ou PTU), de résolution 0.013 deg (Fig. 5.1). Une caméra est fixée à la PTU et une séquence de 753 images d'une scène calibrée est acquise à une résolution de 640×480 pixels. Dans cette séquence, la caméra suit un chemin en créneaux (fig. 5.1), et les commandes envoyées à la PTU sont choisies de manière à ce que la caméra vise à peu près le centre de la scène dans toutes les images. À trois reprises dans la séquence (au début, au milieu et à la fin), une rotation pure est appliquée à la caméra ayant pour effet un déplacement de la projection de la scène sur les bords de l'image.

La scène acquise est composée d'une mire de calibrage de taille $39 \times 26 \times 26$ cm et de trois marqueurs placés à différentes distances de la mire dont les coordonnées 3D des centres ont été mesurées dans le repère de la mire (Fig. 5.1). La mire de calibrage a une fonction double : elle permet d'une part d'obtenir les valeurs correctes des paramètres intrinsèques de la caméra (notre *gold standard*), et d'autre part de calculer les points de vue de la caméra à intrinsèques constants, éventuellement entachés d'erreur.

Les points de vue calculés peuvent être comparés à la vérité terrain (commandes envoyées à la table XY et à la PTU) en utilisant une transformation hand-eye calculée au préalable. Un marqueur, appelé marqueur *mire* dans la suite, est placé à l'intérieur de la distribution des points de la mire. Deux autres marqueurs (marqueur *gauche* situé à gauche en haut de la figure 5.1 et marqueur *droite* situé à droite) sont disposés à une distance de 27 et (respectivement) 19 cm de la mire. Ces marqueurs, qui sont des disques noirs apparaissant sur fond blanc, sont suivis automatiquement d'une image à l'autre de la séquence en utilisant une simple technique de corrélation.

Enfin, pour ne pas limiter la portée de nos résultats à un seul type de caméra, des tests ont été réalisés en utilisant trois optiques de qualités différentes : une optique de bonne qualité (une caméra numérique Sony DFW-VL500 avec une grande focale), une optique de qualité moyenne (la même caméra avec une plus petite focale) et une optique de qualité médiocre (la webcam Logitech QuickCam). Les focales et paramètres de distorsion mesurées sur ces optiques (voir détails plus loin) sont données en table 5.1. La dernière colonne du tableau indique l'erreur relative due à la distorsion radiale obtenue dans le coin en-haut à droite de l'image.

5.3 Choix du modèle de caméra

Une hypothèse souvent utilisée en RA est que les paramètres intrinsèques de la caméra restent constants au cours du temps. Fixer les intrinsèques permet d'éviter les phénomènes de compensations qui

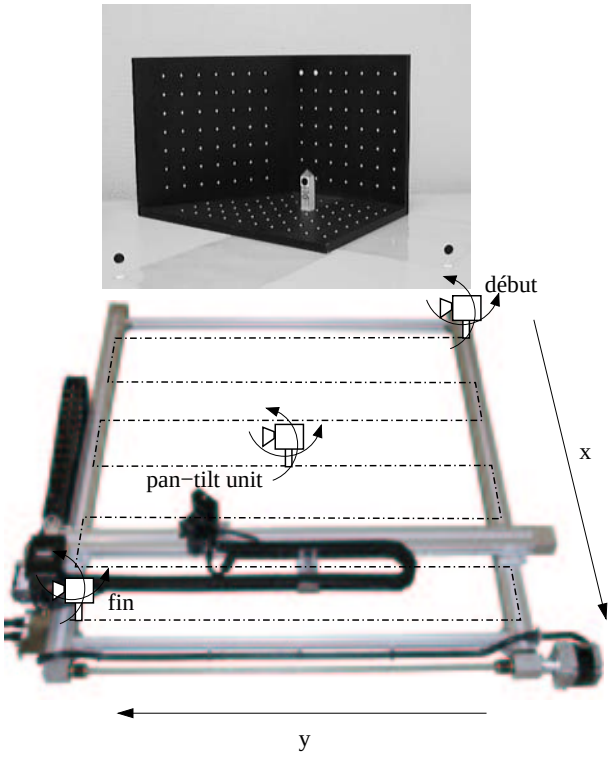


FIG. 5.1 – Le dispositif de contrôle à 4 degrés de liberté et la scène utilisée.

Caméra	f	$\lambda_1.10^7$	$\lambda_2.10^{13}$	c_x	c_y	Erreur
Logitech	824.4	1.492	-1.187	236.4	298.7	2.46%
Sony 1	1113.1	-1.417	2.480	303.0	290.2	1.56%
Sony 2	2325.2	-0.675	-0.153	311.2	312.5	0.95%

TAB. 5.1 – Distances focales et paramètres de distorsion des optiques utilisées dans nos expérimentations.

peuvent se produire entre les paramètres intrinsèques et extrinsèques [AHR01]. De telles compensations induisent généralement des fluctuations statistiques sur les paramètres calculés, ce qui peut générer un effet de tremblement au niveau des éléments virtuels ajoutés. Toutefois, il est important de vérifier que cette hypothèse est pertinente en pratique et qu'elle ne risque pas au contraire d'être source d'imprécision pour l'étape de positionnement. Dans cette section, nous montrons que fixer les paramètres intrinsèques est généralement source d'imprécision, sauf si la distorsion optique de la caméra est prise en compte.

5.3.1 Cas du modèle sténopé sans distorsion

Des travaux antérieurs ont montré que le calcul du rapport d'aspect utilisé dans le modèle sténopé est robuste aux changements de position et d'orientation de la caméra par rapport à la mire utilisée pour le calibrage [SM99], ainsi qu'aux zooms et changements de focus de la caméra [EVF93]. Dans la suite de ce chapitre, nous considérons donc que ce paramètre est connu et constant durant le suivi de la caméra, et nous utilisons pour nos tests sa valeur donnée par le *gold standard* (voir la section 5.3.3).

Afin d'évaluer en pratique la pertinence de l'hypothèse des paramètres intrinsèques constants, nous avons réalisé une minimisation à 9 paramètres de l'erreur de reprojection classique :

$$err_f(\Phi, \mathbf{p}) = \sum_j (u_j - \hat{u}_j)^2 + (v_j - \hat{v}_j)^2,$$

où $\Phi = [f, u_0, v_0]^t$ est le vecteur des paramètres intrinsèques, $\mathbf{p} = [\alpha, \beta, \gamma, t_x, t_y, t_z]^t$ le vecteur des

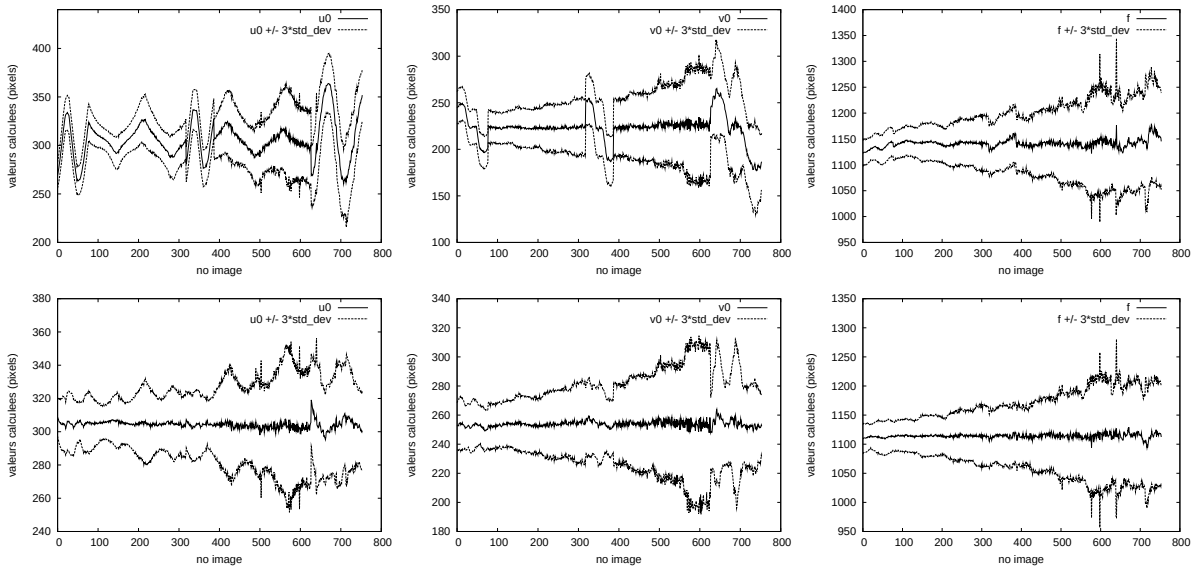


FIG. 5.2 – Valeurs des paramètres intrinsèques et intervalles de confiances, obtenus dans chaque image de la séquence, sans utiliser (première ligne) puis en utilisant (deuxième ligne) le modèle de distorsion.

		Gold standard	Distorsion non corrigée	Distorsion corrigée
Logi-tech	u_0	341.4	334.7 ± 11.4	341.3 ± 2.24
	v_0	261.3	273.0 ± 13.3	263.7 ± 1.48
	f	824.4	816.8 ± 16.8	825.8 ± 2.67
Sony 1	u_0	306.6	305.8 ± 19.0	304.3 ± 2.24
	v_0	253.0	223.2 ± 15.0	253.9 ± 2.10
	f	1113.1	1141.3 ± 7.73	1114.0 ± 3.64
Sony 2	u_0	303.4	304.1 ± 38.1	302.5 ± 5.17
	v_0	306.2	207.0 ± 35.6	306.9 ± 5.04
	f	2325.2	2429.8 ± 44.6	2326.8 ± 10.9

TAB. 5.2 – Moyennes et écarts-types des paramètres intrinsèques obtenus avec les trois optiques, sans et avec correction des distorsions.

paramètres extrinsèques (angles d'Euler et translation), (u_j, v_j) sont les projections des centres des cibles de la mire et $(\hat{u}_j, \hat{v}_j)$ les centres détectés. La méthode de Levenberg-Marquardt est utilisée pour minimiser itérativement cette fonction, avec une estimée initiale donnée par la méthode de Faugeras-Toscani [FT87].

La première ligne de la figure 5.2 montre les valeurs de u_0, v_0 et f obtenues dans chaque image de la séquence, en utilisant la lentille de qualité moyenne. Les intervalles de confiance correspondant aux valeurs des paramètres plus ou moins trois fois leur écart type. Ces écarts-types sont calculés par propagation inverse d'un bruit gaussien d'écart-type 0.5 pixels supposé sur la position détectée des cibles de la mire [HZ00]. On constate que les valeurs obtenues peuvent être très différentes d'une image à une autre : les valeurs de u_0 varient entre 260 et 364 pixels, celles de v_0 entre 177 et 265 pixels et celles de f entre 1122 et 1178 pixels. Pour u_0 et v_0, il n'existe d'ailleurs aucune droite horizontale passant à l'intérieur de tous les intervalles de confiance.

Des résultats similaires ont été obtenus avec les optiques de qualité basse et élevée : la table 5.2 montre la valeur moyenne et l'écart-type des paramètres intrinsèques calculés sur la séquence pour les trois optiques. Ces valeurs peuvent être comparées aux valeurs du *gold standard*, obtenues selon la méthode précisée en section 5.3.3 : on constate que certaines valeurs moyennes sont très éloignées des valeurs attendues, y compris pour l'optique de bonne qualité.

5.3.2 Prise en compte de la distorsion radiale

La même expérience a été réalisée en ajoutant au modèle de caméra une approximation à l'ordre 4 de la distorsion radiale. Les coordonnées $(\hat{u}', \hat{v}')$ de l'image corrigée sont obtenues à partir des coordonnées $(\hat{u}, \hat{v})$ observées, en appliquant la formule suivante :

$$\begin{pmatrix} \hat{u}' \\ \hat{v}' \end{pmatrix} = \begin{pmatrix} \hat{u} \\ \hat{v} \end{pmatrix} + (\lambda_1 r^2 + \lambda_2 r^4) \begin{pmatrix} \hat{u} - c_u \\ \hat{v} - c_v \end{pmatrix},$$

où $r = \sqrt{(\hat{u} - c_u)^2 + (\hat{v} - c_v)^2}$, λ_1 et λ_2 sont les premier et le deuxième paramètres de correction de distorsion et (c_u, c_v) est le centre de distorsion. En prenant les paramètres de correction de distorsion donnés par le *gold standard* et en corrigeant les coordonnées des centres des cibles à partir de ces valeurs, le minimum de la fonction de coût err_f est atteint pour des valeurs de paramètres intrinsèques plus stables que précédemment (cf. deuxième ligne de la figure 5.2) : u_0 varie entre 297 et 320 pixels, v_0 entre 246 et 261 pixels et f entre 1094 et 1130 pixels. De plus, pour chaque paramètre nous obtenons une large bande de droites horizontales traversant tous les intervalles de confiance. Enfin, les valeurs moyennes obtenues avec les trois optiques sont à présent bien plus proche des valeurs attendues, et les valeurs des écarts-types sont considérablement réduites (Tab. 5.2).

Ces résultats sont en accord avec les résultats synthétiques présentés dans [FM96]. Dans cette étude, 60 points 3D sont tirés aléatoirement dans une sphère de rayon 14 unités. La caméra est placée à différentes distances du centre de la sphère (en moyenne à 70 unités) avec des orientations aléatoires. Les points 3D sont projetés dans l'image et du bruit gaussien ainsi qu'une distorsion radiale sont appliqués aux projections. Lorsque les paramètres intrinsèques sont calculés sans tenir compte de la distorsion, des valeurs incorrectes sont obtenues pour ces paramètres, avec des intervalles de confiance qui n'incluent pas les bonnes valeurs.

5.3.3 Validation du modèle retenu

Utiliser le modèle sténopé de caméra sans tenir compte de la distorsion radiale conduit donc généralement à des valeurs incorrectes de paramètres intrinsèques (compensations locales dues à la trop grande simplicité du modèle). Corriger la distorsion radiale permet au contraire d'obtenir des valeurs d'intrinsèques moins sensibles à la position de la caméra par rapport à la mire de calibrage. Cependant, cela ne prouve pas que le modèle sténopé avec prise en compte de la distorsion radiale est suffisamment riche pour les applications de RA. Pour évaluer la pertinence de ce modèle en RA, nous mesurons à la fois l'erreur de positionnement 3D et l'erreur d'alignement 2D obtenues en utilisant ce modèle pour chaque image de la séquence.

Les paramètres intrinsèques et de distorsion utilisés dans cette validation sont obtenus à partir de 50 images de la mire de calibrage. Les positions de caméra ont été choisies de manière à ce que la mire apparaisse en grand dans chaque image, avec ses trois faces bien visibles. Les paramètres retenus sont ceux qui minimisent l'erreur de reprojection des centres des cibles dans toutes les images :

$$\min_{a,\Phi,\Lambda} \sum_{i=1}^{50} err_f(\Phi, \mathbf{p}_i(\Phi)),$$

où a est le rapport d'aspect et $\Lambda = [\lambda_1, \lambda_2, c_u, c_v]^t$ est le vecteur des paramètres de distorsion. $\mathbf{p}_i(\Phi)$ est le point de vue de la caméra qui minimise err_f dans l'image i pour un vecteur donné Φ de paramètres intrinsèques. En utilisant ces paramètres pour calculer les points de vue de caméra sur les 753 images de la séquence, nous obtenons à la fois :

- des points de vue très proches de la vérité terrain sur toute la séquence : l'erreur moyenne est indiquée en table 5.3 (les composantes T_x et T_y de la translation sont exprimées dans le repère de la table XY). Ces erreurs sont de l'ordre de la résolution de la table XY et de la PTU. Soulignons que le nombre n d'images utilisées pour le calcul des moyennes varie d'une optique à l'autre : suivant l'optique utilisée et la position de la caméra sur la table, la mire de calibrage peut en effet apparaître trop petite, trop grande ou hors du champ de vision dans certaines images, ne permettant pas de calculer le point de vue ;

	n	ΔT_x(mm)	ΔT_y(mm)	Δpan(deg)	$\Delta tilt$(deg)
Logitech	222	0.001	0.000	0.020	0.060
Sony 1	753	-0.165	-0.004	-0.020	0.014
Sony 2	480	0.000	0.000	-0.011	-0.027

TAB. 5.3 – Erreurs moyennes obtenues sur les paramètres du point de vue en utilisant des intrinsèques fixes et en corrigeant la distorsion radiale.

- des erreurs d'alignement très faibles pour les trois marqueurs (Fig. 5.3) : ces erreurs sont de l'ordre de 1 pixel et sont principalement dues aux erreurs d'incertitude dans la détection des marqueurs.

5.4 Influence du calibrage

Nous mesurons à présent la précision obtenue en pratique avec une distorsion corrigée mais un calibrage imprécis, et un calcul de point de vue reposant sur des appariements 3D-2D de primitives quelconques. Ces résultats s'appliquent à n'importe quel système de positionnement basé modèle, qu'il repose sur la détection automatique de marqueurs artificiels placés dans la scène [KB99, VKSN02, STN04] comme sur l'apprentissage et la reconnaissance d'indices naturels déjà présents dans la scène [GRS+02, LVTF03, SL04].

La précision est mesurée en terme d'erreur de positionnement et d'alignement 2D. Dans cette section, la lentille de qualité moyenne est utilisée, mais comme seules les images corrigées sont considérées, l'étude est largement indépendante du niveau de distorsion présent dans les images originales.

5.4.1 Interprétation analytique

Une interprétation analytique de certains de nos résultats peut être obtenue à l'aide de l'équation du flot optique, étendue à la prise en compte d'intrinsèques variables. Lorsqu'une caméra subit un déplacement $(\dot{\alpha}, \dot{\beta}, \dot{\gamma}, \dot{t_x}, \dot{t_y}, \dot{t_z})$ et une variation d'intrinsèques $(\dot{f}, \dot{u_0}, \dot{v_0})$ entre deux images, le mouvement horizontal apparent $\dot{x}$ d'un pixel de coordonnées $(x, y) = (u-u_0, v-v_0)$ centrées autour du point principal est donné par :

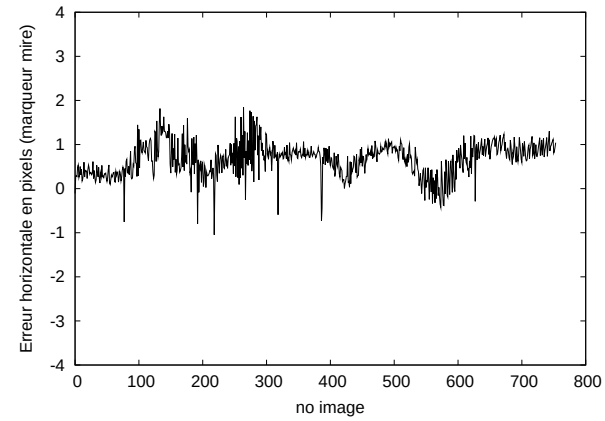
$$\dot{x} = \frac{x}{f}\dot{f} + \dot{u_0} + y\dot{\alpha} - f(1 + \frac{x^2}{f^2})\dot{\beta} - \frac{xy}{f}\dot{\gamma} - \frac{f}{Z}\dot{t_x} + \frac{x}{Z}\dot{t_z},$$

où Z est la profondeur du point dans le repère caméra (une expression similaire est obtenue pour $\dot{y}$). Cette expression peut être interprétée comme l'erreur d'alignement horizontale théoriquement obtenue lorsqu'un ou plusieurs paramètres p_i de la caméra sont erronés ($\dot{p_i} \neq 0$).

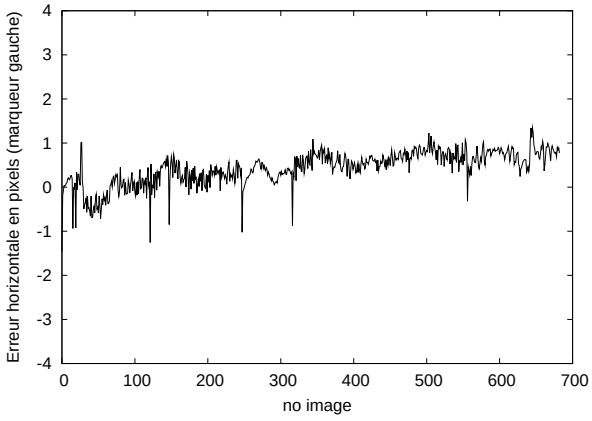
5.4.2 Perturbation des intrinsèques

Dans cette étude, nous voulons perturber les paramètres de calibrage de la caméra, mais sans toutefois sortir des marges d'erreurs habituellement observées en situation pratique. Nous nous basons pour cela sur le type d'erreurs que nous obtenons lorsque la caméra est calibrée à partir d'une seule vue d'une mire de calibrage, cette procédure de calibrage étant couramment utilisée.

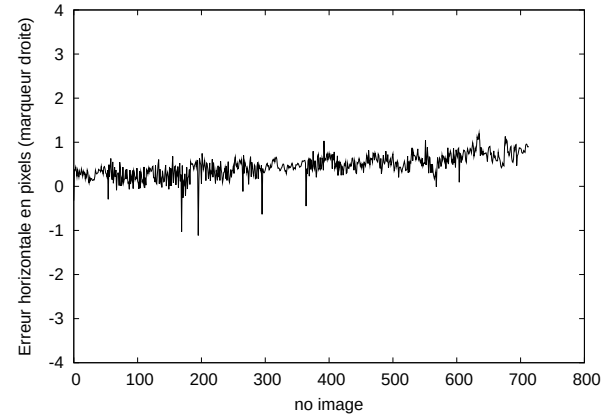
La figure 5.4 montre la distribution des valeurs de (λ_1, λ_2) obtenues en minimisant l'erreur de reprojection err_f suivant Λ, Φ et $\mathbf{p}$ dans chaque image de la séquence. Les valeurs de (λ_1, λ_2) correspondant au *gold standard* sont étiquetées L0 ; les valeurs que nous utilisons dans nos expérimentations sont étiquetées de L1 à L4 : elles correspondent aux valeurs obtenues au 5ème (L1), 95ème (L2), 10ème (L3) et 90ème (L4) centile des valeurs ordonnées de λ_1. Ces valeurs ne sont pas seulement utilisées pour mesurer l'influence des paramètres de distorsion : elles nous permettent aussi d'établir des marges d'erreurs réalistes pour la distance focale et le point principal.



marqueur *mire*



marqueur *gauche*



marqueur *droit*

FIG. 5.3 – Les valeurs du *gold standard* minimisent l'erreur d'alignement à l'intérieur et à l'extérieur du volume de mesure.

5.4.3 Influence du point principal

L'influence du point principal est évaluée en minimisant err_f suivant les six paramètres du point de vue, avec des valeurs de f et Λ issues du *gold standard*, et des valeurs de (u_0, v_0) modifiées. L'amplitude et la direction des erreurs introduites sur le point principal sont choisies de la manière suivante. La figure 5.5 montre l'ensemble des points principaux obtenus en minimisant err_f suivant Φ et $\mathbf{p}$ dans chaque image de la séquence, en utilisant les paramètres de distorsion L0 à L4 (pour plus de visibilité, une image sur 10 seulement est représentée). Cela donne une idée des erreurs pouvant être obtenues avec un calibrage mono-image et des paramètres de distorsion obtenus en pratique. Afin de tester des amplitudes

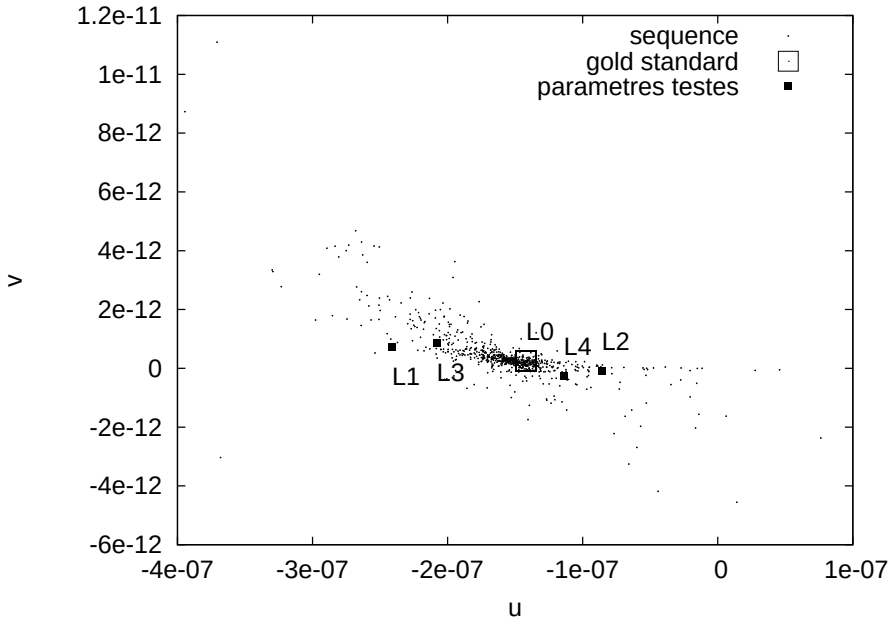


FIG. 5.4 – Valeurs de (λ_1, λ_2) obtenues dans chaque image de la séquence.

et directions d'erreur variées sans toutefois obtenir un trop grand nombre de cas à présenter, nous avons choisi de faire évoluer le point principal sur une spirale représentative du nuage de points et centrée autour du point principal PP0 issu du *gold standard*. Les points principaux ainsi choisis sont étiquetés de PP1 à PP5 en figure 5.5. Remarquons que PP2 est exactement au centre de l'image, ce qui nous permettra d'évaluer la pertinence de l'hypothèse couramment acceptée du point principal à cette position.

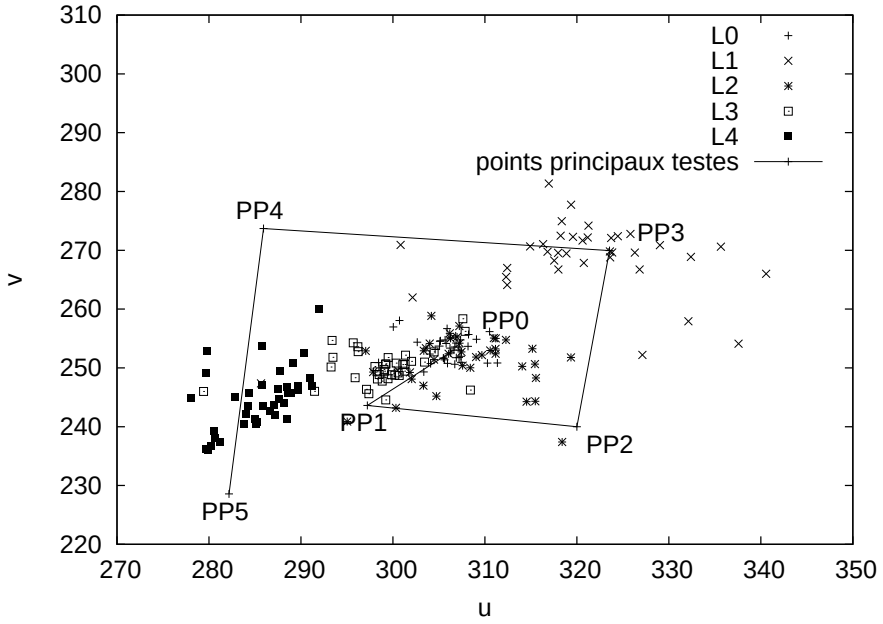


FIG. 5.5 – Points principaux obtenus en utilisant les paramètres de distorsion $L0$ à $L4$.

La figure 5.6 montre les erreurs d'alignement horizontal que nous obtenons sur les trois marqueurs en

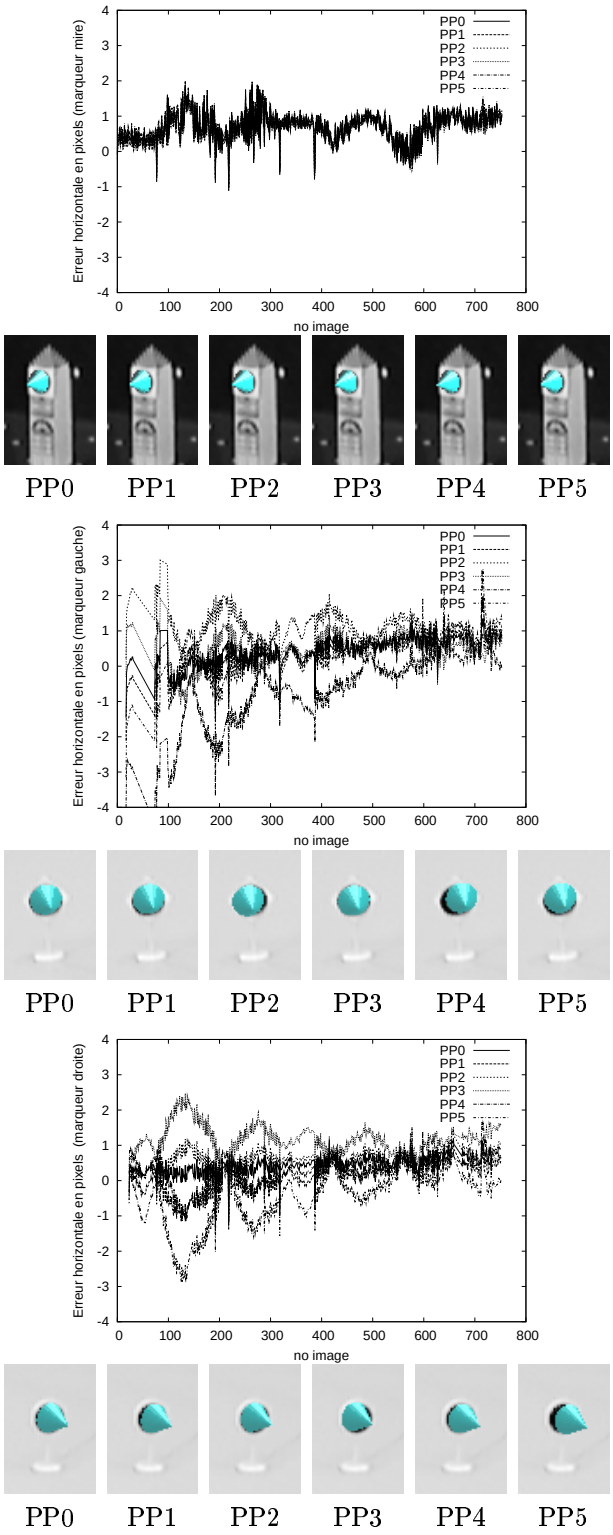
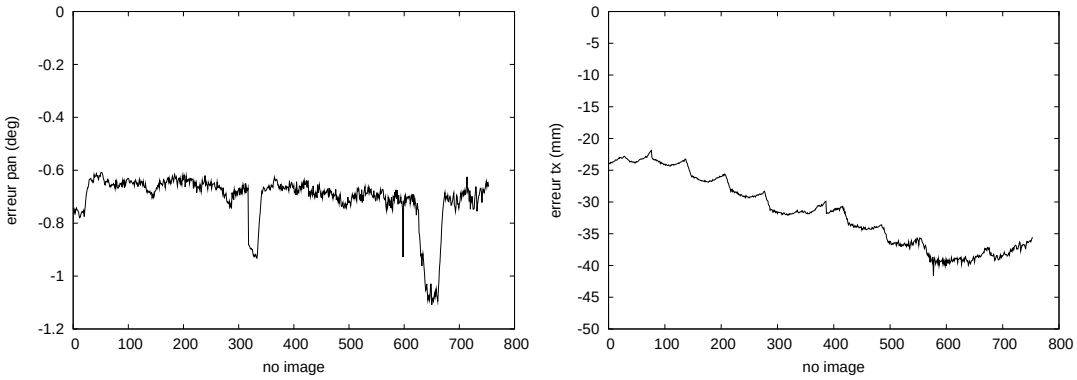


FIG. 5.6 – Influence du point principal sur l'alignement 2D.

utilisant les points principaux PP1 à PP5. Ces erreurs sont négligeables pour le marqueur *mire* placé au centre du volume de mesure. Cela est dû au fait que les erreurs introduites sur le point principal sont

compensées par des erreurs sur les angles β et α (pan et tilt) [AHR01], de manière à minimiser l'erreur de reprojection des primitives utilisées pour le positionnement. Ces compensations apparaissent en table 5.4.

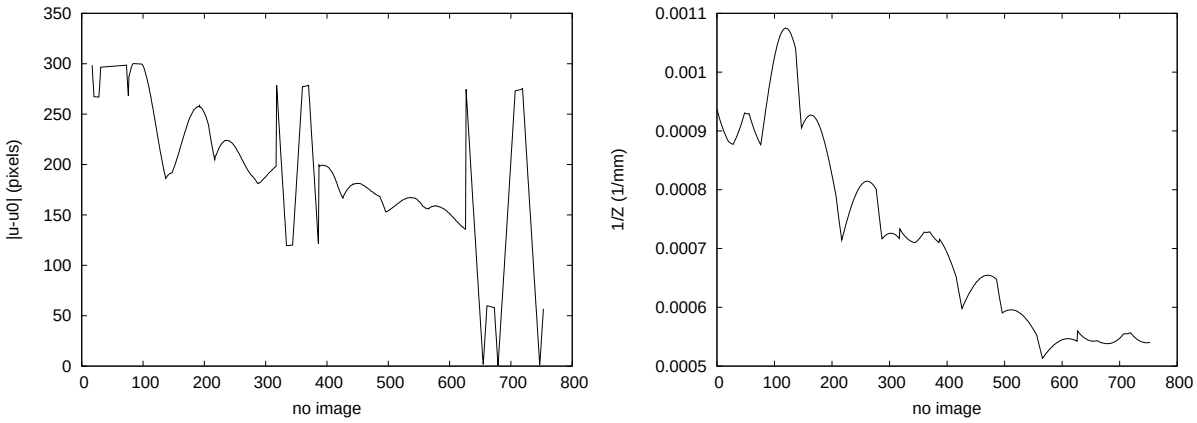


(a) (b)

FIG. 5.7 – (a) Erreur sur β obtenue avec le point principal PP2. (b) Erreur sur T_x obtenue avec la distance focale F2.

Cependant, les erreurs d'alignement sont plus importantes en dehors du volume de mesure : jusqu'à 4 pixels pour le marqueur *gauche* avec PP4 et jusqu'à 3 pixels pour le marqueur *droite* avec PP5 (Fig. 5.6). La forme sinueuse des courbes d'erreurs s'explique à l'aide de l'équation du flot optique : les erreurs de compensation entre $\dot{u_0}$ and $\dot{\beta}$ introduisent des erreurs d'alignement horizontale $\dot{x} \approx \dot{u_0} - f(1 + \frac{x^2}{f^2})\dot{\beta}$. Comme l'erreur sur le point principal $\dot{u_0}$ est constant le long de la séquence et la compensation observée sur $\dot{\beta}$ varie peu (voir la figure 5.7.(a)), la forme des courbes est principalement due aux rapprochements et éloignements successifs des marqueurs à la droite verticale passant par le point principal ($x^2 = (u - u_0)^2$ croît ou décroît, cf. Fig. 5.8.(a)).

Afin d'évaluer visuellement ces erreurs d'alignement, des cones virtuels dont les bases doivent être superposées aux marqueurs réels ont été projetés dans les images. Les résultats d'incrustation obtenus dans l'image 110 où les trois marqueurs sont bien visibles sont présentés en figure 5.6. Une vidéo montrant les incrustations obtenues sur toute la séquence avec le point principal PP5 (PP5.avi) peut être téléchargée à l'adresse `http ://www.loria.fr/~gsimon/rfia2006.html`. Cette vidéo peut être comparée à la vidéo basée sur les paramètres de calibrage issus du *gold standard* (GOLD.avi).



(a) (b)

FIG. 5.8 – Évolution sur la séquence de (a) $|u - u_0|$ et (b) $1/Z$ pour le marqueur *gauche*.

	ΔT_x(mm)	ΔT_y(mm)	Δpan(deg)	$\Delta tilt$(deg)
PP1	-0.22	-0.29	0.45	-0.49
PP2	-0.35	-0.14	-0.70	-0.66
PP3	0.88	0.22	-0.85	0.90
PP4	1.64	0.81	1.01	1.09
PP5	0.46	-1.00	1.20	-1.30
F1	23.46	1.05	-0.03	0.03
F2	-31.66	-1.41	-0.01	-0.01
F3	39.22	1.75	-0.03	0.04
F4	-47.41	-2.11	-0.00	-0.03
F5	54.98	2.45	-0.04	0.05
L1	-4.91	-0.18	-0.04	-0.02
L2	0.85	0.29	-0.01	0.01
L3	-1.66	0.17	-0.03	0.00
L4	0.23	0.04	-0.02	0.01
L5	7.73	1.00	-0.04	0.09

TAB. 5.4 – Erreurs moyennes obtenues sur les paramètres du point de vue lorsque des points principaux (PP1-PP5), distances focales (F1-F5) et paramètres de distorsion (L1-L5) erronés sont utilisés.

5.4.4 Influence de la distance focale

Nous évaluons à présent la sensibilité du positonnement face à des erreurs sur la focale. Les distances focales F1=1128, F2=1093, F3=1138, F4=1083 et F5=1148 sont choisies suivant le même principe que précédemment, sauf que la sélection se fait ici dans une seule dimension. Les résultats sont présentés en figure 5.9 (F0=1113.1 est la valeur du *gold standard*). À nouveau, l'influence des erreurs sur la focale est imperceptible à l'intérieur du volume de mesure, en raison du phénomène de compensation bien connu qui a lieu entre la focale et la translation le long de l'axe optique (voir table 5.4 et figure 5.7.(b)). Cependant, les erreurs d'alignement sont plus importantes en dehors du volume de mesure. La forme des courbes d'erreur s'explique par le fait qu'une erreur sur la focale induit une erreur horizontale $\dot{x} \approx x(\frac{\dot{f}}{f} + \frac{\dot{t_z}}{Z})$, dont la valeur absolue décroit (avec |x|) lorsque le marqueur se rapproche de la droite verticale passant par le point principal, mais aussi lorque la caméra s'éloigne des marqueurs (accroissement de Z) : la figure 5.8.(b) montre l'évolution de $1/Z$ pour le marqueur *gauche*. Une vidéo montrant les résultats d'augmentation obtenus avec la focale F5 est disponible sur notre site web (F5.avi).

5.4.5 Influence de la distorsion radiale

L'influence de la distorsion radiale sur les erreurs d'alignement est présentée en figure 5.10. L0 à L4 sont les ensembles de paramètres de correction de distorsion dont le choix est présenté en section 5.4.2 et L5 correspond à des valeurs nulles pour λ_1 et λ_2 (distorsions non corrigées). L'influence des distorsions est négligeable à l'intérieur du volume de mesure, et aussi très faible à l'extérieur, sauf pour L1 qui correspond au couple (λ_1, λ_2) le plus éloigné du *gold standard* (cf. Fig. 5.4), et pour le cas des distorsions non corrigées. La vidéo correspondant à ce dernier cas est disponible sur le site (L5.avi). Corriger les distorsions améliore donc généralement la précision de l'alignement de manière significative, même lorsque le calcul des paramètres de correction de distorsion n'est pas réalisé dans des conditions optimales (une seule vue de la mire placée dans une position et orientation arbitraires). Nous n'étudions pas ici de manière théorique les compensations possibles entre les paramètres de distorsion et les paramètres externes de la caméra. Dans nos expérimentations, elles concernent toutefois principalement la translation de la caméra le long de l'axe optique, comme en atteste le tableau 5.4.

5.5 Conclusion

Dans ce chapitre, nous avons présenté un ensemble de résultats expérimentaux permettant d'évaluer la pertinence du modèle sténopé en pratique et l'influence des erreurs de calibrage dans un contexte de

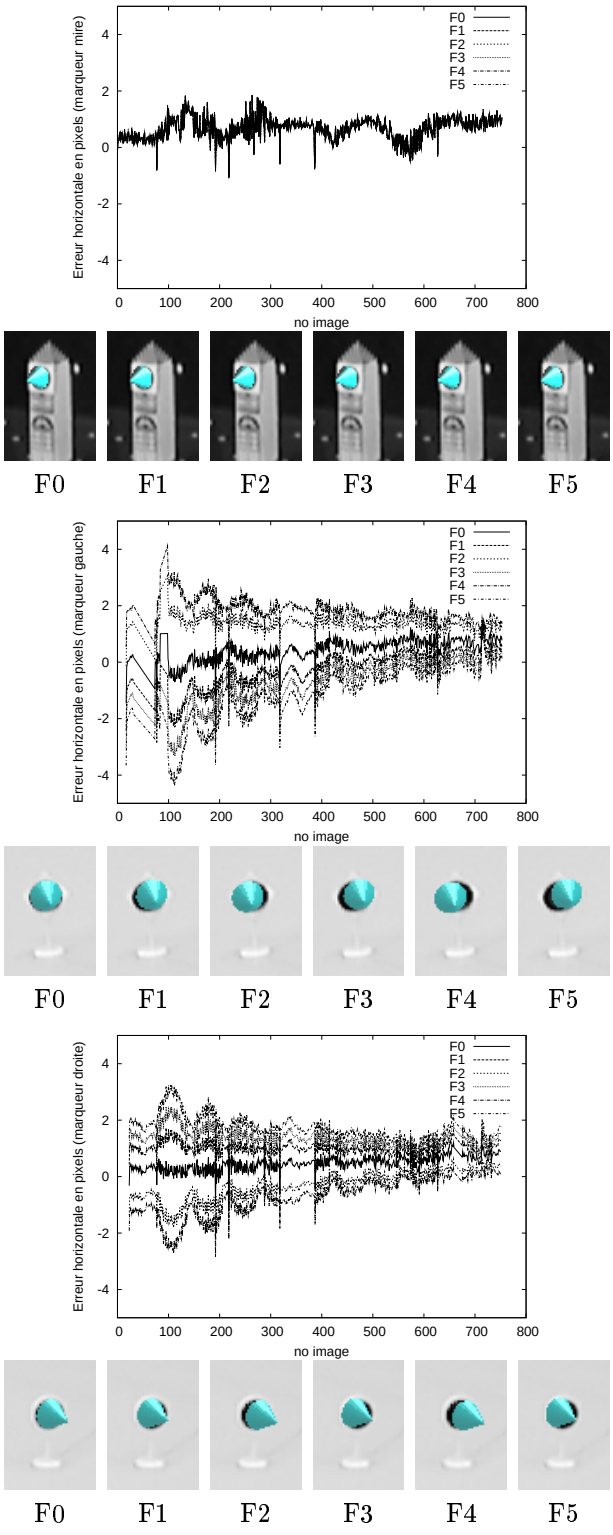


FIG. 5.9 – Influence de la distance focale sur l'alignement 2D.

RA. Nos conclusions sont les suivantes :

1. L'utilisation du modèle sténopé de caméra à intrinsèques fixes et distorsion radiale corrigée conduit

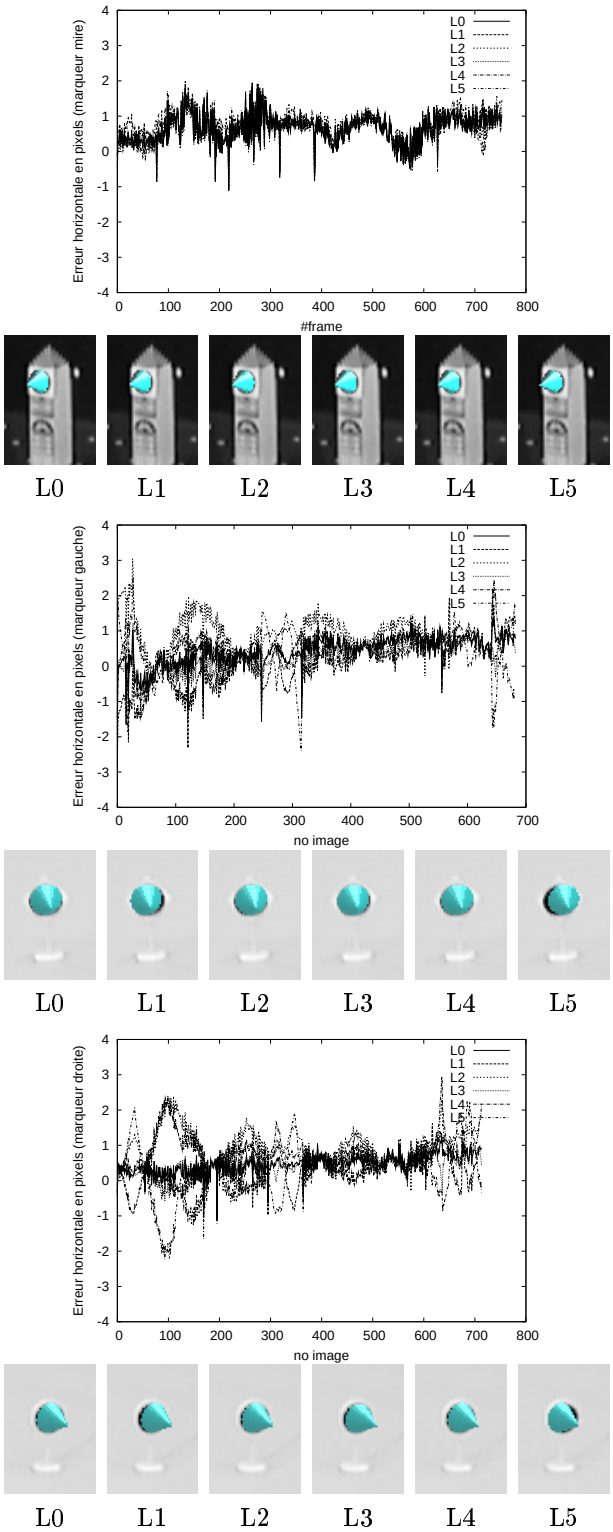


FIG. 5.10 – Influence des paramètres de distorsion sur l'alignement 2D.

à des résultats précis de positionnement 3D et d'alignement 2D : avec ce modèle, nous avons obtenu une excellente précision quelque soit l'optique utilisée, la position relative entre la caméra et la scène

filmée, et la distance entre l'objet ajouté et le volume de mesure.

2. À l'inverse, l'utilisation du modèle sténopé sans correction de distorsion aboutit à des résultats imprécis, y compris pour des optiques de bonne qualité (moins de 1% d'erreur dans un coin de l'image). Corriger les distorsions s'avère particulièrement nécessaire lorsqu'on cherche à déterminer les paramètres intrinsèques de la caméra, pour éviter des ajustements locaux dus aux insuffisances du modèle. Durant le processus de positionnement de la caméra, cela affecte principalement la précision de l'alignement en dehors du volume de mesure. Des solutions maétrielles et logicielles existent pour corriger la distorsion radiale à cadence vidéo, qui devraient être employées lorsqu'une précision élevée est requise en dehors du volume de mesure.
3. Fixer le point principal au centre de l'image ou utiliser des paramètres intrinsèques "imprécis" (obtenus à partir d'une seule vue de mire par exemple) ne nuit généralement pas de façon perceptible à la précision de l'alignement à l'intérieur du volume de mesure. Par contre, ces approximations se traduisent à l'extérieur du volume de mesure par un effet de glissement à basse fréquence des objets ajoutés, dû aux rapprochements et éloignements successifs de l'image de ces objets par rapport au point principal (en ce qui concerne les erreurs sur le point principal et la focale) et aussi de la position 3D de ces objets par rapport à la caméra (en ce qui concerne les erreurs sur la focale). Pour des applications qui requièrent une grande précision en dehors du volume de mesure, le calibrage doit donc être réalisée avec soin, en utilisant de préférence plusieurs images et une technique de type ajustement de faisceau.

Chapitre 6

Méthodes de calcul des paramètres intrinsèques de la caméra

En vue d'obtenir un système de RA dont la mise en œuvre du système de calibrage soit accessible pour les utilisateurs quelconques, nous avons examiné les méthodes de calibrage interne basées sur l'observation de modèles simples ou tirant avantage des indices déjà présents sur la scène. Nous nous intéressons notamment aux méthodes de calibrage basées sur l'observation de plans, dites *calibrage planaire* :

- d'une part, parce que les motifs plans sont faciles à construire et peu coûteux, par exemple : en imprimant une grille sur une feuille de papier ;
- d'autre part, puisqu'on trouve des surfaces planes texturées dans un grand nombre d'environnements où les applications de RA sont fréquentes : pièces d'habitation, bureaux, scènes architecturales et urbaines.

Dans ce chapitre, nous analysons les contraintes de calibrage qui peuvent être obtenues à partir d'un plan de la scène observée dans une image et nous passons en revue les principes et les méthodes connues de calibrage interne basé plans.

6.1 Méthodes de calibrage interne

Les méthodes de calibrage interne peuvent être classifiées :

- en fonction du nombre d'images requises [BAA98], on distingue ainsi les méthodes :
  - *passives* : les paramètres internes peuvent être estimés en utilisant une seule prise de vue, soit d'une mire de calibrage dont les positions des primitives observées sont données [Sla80, Tsa87] soit d'un objet où certaines caractéristiques géométriques sont connues (e.g. ensembles de droites parallèles [CT90, RCR99]).
  - *actives* : plusieurs images de la scène sont acquises depuis différents points de vue, en effectuant des mouvements de la caméra [MF92, DB93, Har94a, Bas95, Zha00].
- en fonction des connaissances sur la scène observée [Zha98], telles que :
  - *photogrammétriques* : en utilisant les points de la scène dont les positions 3D sont connues [Tsa87, FT87] ;
  - d'*auto-calibrage* : en déplaçant la caméra dans une scène statique, où la rigidité de la scène fournit les contraintes de calibrage, lorsque aucune autre information de la structure de la scène n'est disponible [Luo92, MF92, Har94c] ;
  - en visant des objets géométriques spéciaux (tels que cercles concentriques [KKK02], rectangles [BAA98], coniques [YSH00, HJ01]) dont on peut extraire des propriétés invariantes à leur position dans l'espace, en contraignant les paramètres intrinsèques de la caméra.

Dans cette section, nous présentons une revue des techniques de calibrage groupées selon ce deuxième critère, car nous nous intéressons entre autre aux aspects pratiques du calibrage interne (tels que l'estimation de paramètres à partir de modèles simples ou de surfaces texturées) à différents niveaux de connaissances de la scène.

6.1.1 Techniques photogrammétriques

Les premiers travaux amenant des solutions au problème de l'estimation des paramètres intrinsèques d'une caméra s'appuient sur l'observation de mires 3D [Bro71, Sla80, FT87, WCH92] : l'estimation des paramètres de la caméra est effectuée en observant un objet de calibrage dont la géométrie est connue. Les mires sont souvent formées de motifs facilement identifiables (polyèdres [CJ03], surfaces avec carrés [LL95] ou cercles [HS96] ou figure 7.4(a)) choisis pour définir des points d'intérêt qui peuvent être mesurés avec une bonne précision, et qui nous permettent d'obtenir les correspondances 3D-2D utilisées comme données d'entrée de ces méthodes. Ce type de méthodes appartiennent à une catégorie de calibrage appelée *photogrammétrique*, et elles effectuent en général le calcul simultané de tous les paramètres du modèle de la caméra.

Cependant, quelques auteurs [LVD98, AHR01] ont averti que l'estimation simultanée des paramètres intrinsèques et extrinsèques peut entraîner des résultats incorrects : lorsqu'il existe des erreurs sur un paramètre (par exemple, à cause des erreurs de mesure ou de modélisation), il est possible que des compensations sur les autres paramètres se produisent afin d'optimiser, dans l'ensemble, le critère employé pour l'étalonnage du modèle de caméra. Ces compensations ne se produisent pas toujours entre les mêmes paramètres ni de la même manière, car les corrélations entre les paramètres varient en fonction de plusieurs facteurs [LL95, Enc95] tels que : la géométrie de l'objet de calibrage, la distribution des primitives observées, la position et l'orientation de la caméra par rapport à l'objet de calibrage.

L'autre inconvénient de ce type de méthodes est le besoin d'utiliser des objets spéciaux pour le calibrage, car il est nécessaire de disposer d'une mire adéquate et spécifique aux conditions d'utilisation (e.g. type d'application, distance focale de la caméra, modèle de la caméra et de la distorsion). Les mires de calibrage de haute qualité sont en général rigoureusement usinées et les positions des motifs de référence placés sur la mire sont minutieusement mesurés. Elles doivent être fabriquées en matériaux mécaniquement stables dans le temps pour éviter que les éventuelles déformations dégradent les mesures données par le fabriquant [LVD98]. Dès lors, le coût de fabrication et de mesure des mires de calibrage est élevé.

Lavest *et al.* [LVD98] montrent que la précision du calibrage dépend principalement de la précision de détection des primitives sur les images, et moins de l'exactitude de la mesure des positions 3D correspondantes.

6.1.2 Auto-calibrage

Les techniques photogrammétriques requièrent des objets spéciaux pour réaliser le calibrage. Mais pour certaines applications, ces objets ne sont pas toujours disponibles [Stu97b] (e.g. applications dédiées au grand public) ou son emploi s'avère trop gênant (e.g. transporter la mire de calibrage, pour les applications mobiles). Une idée pour améliorer la souplesse d'utilisation consiste à remplacer la mire par une scène statique quelconque, et profiter de l'observation sur plusieurs prises de vue des indices naturels du décor.

Basées sur ce principe, les méthodes, appelées *d'auto-calibrage*, requièrent uniquement que des primitives, telles que points ou droites, soient détectées et mises en correspondance sur un ensemble d'au minimum trois images de la scène acquises depuis différents points de vue.

Une approche consiste à reconstruire la scène en même temps qu'on calcule les paramètres de la caméra [Har94b, HA96], comme le font les méthodes d'obtention de la structure à partir du mouvement (*structure from motion*).

Cependant, le stade de reconstruction de la scène n'est pas toujours indispensable. Au moyen de déplacer la caméra dans une scène statique, la rigidité de la scène permet d'établir des contraintes sur les paramètres intrinsèques de la caméra [FLM92, Har94c] en utilisant seulement les informations extraites de la séquence d'images : les correspondances des primitives détectées. Ces méthodes sont basées sur l'invariabilité aux déplacements de la projection de la *conique absolue* (voir section 6.3.2), lorsque les paramètres intrinsèques de la caméra sont constants [HZ00]. Ces approches sont classifiées en trois groupes par la façon dont la conique absolue est liée à la contrainte de rigidité établie par les correspondances des primitives observées :

- basées sur les équations de Kruppa [MF92, ZF96] ;

- par la stratification du problème dans leurs parties affine et euclidienne [PvG97] ;
- par l'identification de la quadrique absolue [Tri97, PMP+05].

Un système d'équations est dérivé de chacune de ces approches dont les inconnues sont les coefficients de l'expression analytique de la projection de la conique absolue, des coefficients qui sont une fonction des paramètres intrinsèques de la caméra (voir éqs. 6.8 et 6.9). Dès lors, la résolution du système d'équations établi équivaut à l'obtention du calibrage interne de la caméra.

Le principal avantage des méthodes d'auto-calibrage est la possibilité de calculer les paramètres intrinsèques de la caméra sans avoir besoin d'un objet de calibrage spécial, ni de connaissances *a priori* sur les positions ni la structure des éléments de la scène. Cette caractéristique rend ces approches assez convenables pour un grand nombre d'applications. Cependant, lors des comparaisons expérimentales, la précision des paramètres intrinsèques issus des méthodes d'auto-calibrage est en général inférieure à la précision des résultats obtenus par les techniques photogrammétriques [LF97b].

Les méthodes d'auto-calibrage ont d'autres contraintes importantes : les singularités et le mauvais conditionnement numérique entraînés pour certains types de déplacements de la caméra, empêchant d'obtenir des résultats fiables. Ces problèmes se produisent, par exemple, lorsque le centre optique est constant sur l'ensemble de vues (rotations pures) ou lorsque la matrice fondamentale entre chaque paire de vues est antisymétrique (e.g. translations pures), des mouvements qui sont fréquents dans certaines applications telles que la robotique mobile. Une taxonomie détaillée des mouvements critiques pour l'auto-calibrage se trouve dans les travaux de Sturm [Stu97a, Stu97b].

Des solutions à ces problèmes ont été envisagées, en développant des méthodes spécialisées qui estiment les paramètres intrinsèques à partir de mouvements particuliers. Par exemple, Du et Brady [DB93], Hartley [Har94a], et Basu [Bas95], ont proposé des méthodes basées sur l'acquisition de vues au moyen d'une caméra en rotation. En ce qui concerne les translations pures, Wiles et Brady [WB96a] ont présenté une méthode pour les cas où le mouvement du centre optique de la caméra est coplanaire, et Dron [Dro93] a conçu une méthode pour les cas où le déplacement de la caméra est uni-directionelle.

6.1.3 Calibrage basé plans

Calibrage planaire

Certains travaux dérivés des techniques photogrammétriques 3D montrent que le calibrage de la caméra peut être effectué en observant des primitives coplanaires dont les positions 2D sont connues [Tsa87, LT88, Zha00]. Ce type de méthodes, appelées de *calibrage planaire*, aident à améliorer la flexibilité du calcul des paramètres intrinsèques, grâce à la simplicité d'utilisation et de fabrication d'objets de calibrage plans (e.g. des grilles ou des damiers imprimées sur une feuille de papier) qui peuvent être acquis à un coût réduit.

Il est aussi possible de tirer profit de l'observation des structures planes déjà présentes sur la scène ; cependant, les positions exactes des indices sur ces structures planes ne sont pas toujours connues. Dans le cas le plus général, la seule information disponible c'est la région où les primitives coplanaires se trouvent. On ne connaît donc ni la position relative de la région plane par rapport à la caméra (appelée la *structure euclidienne du plan de référence*) ni sa position absolue dans la scène.

Certaines méthodes ont été développées pour effectuer le calibrage interne à partir de plusieurs vues d'un plan texturé quelconque. Ces méthodes se basent sur le fait que les transformations homographiques reliant les projections d'un ensemble de primitives coplanaires fournissent des contraintes sur les paramètres intrinsèques de la caméra [HZ00]. Pour certaines de ces méthodes, il est nécessaire de donner des informations sur la position ou la distribution des primitives coplanaires [Stu00, GP01], mais il est possible que cette information ne soit pas toujours disponible [Tri98b].

Auto-calibrage planaire

Des méthodes de calibrage interne basées sur l'observation d'une surface plane texturée dont la position réelle est inconnue, dites *d'auto-calibrage planaire*, ont été progressivement développées dans les dernières années [Tri98b, MC00b, NJS03, GS03]. Elles estiment simultanément les paramètres intrinsèques et la position relative de la surface plane par rapport à la caméra (dites *la rectification euclidienne*) en

s'appuyant sur les contraintes dérivées des homographies inter-images induites pour le plan observé. L'un des intérêts de l'auto-calibrage planaire est que les matrices des homographies induites pour le plan observé représentent une contrainte plus restrictive que la matrice fondamentale, et éliminent certaines singularités qui se présentent dans les méthodes classiques d'auto-calibrage.

Les méthodes d'auto-calibrage basées plans se trouvent dans un stade intermédiaire entre les méthodes photogrammétriques et les méthodes générales d'auto-calibrage : il n'est pas nécessaire de connaître les positions exactes d'un modèle 3D de calibrage, il suffit de déterminer les régions où on trouve des primitives coplanaires, ainsi que les homographies entre les images correspondantes à ces régions et qui peuvent être obtenues uniquement à partir des images acquises.

Les principales motivations pour considérer l'emploi des techniques basées plans sont :

- la construction de motifs plans est plus accessible que l'acquisition d'un volume de calibrage usiné avec des dimensions très rigoureuses ;
- on peut trouver des surfaces planes dans de nombreux environnements ;
- le calibrage planaire est plus flexible que les techniques photogrammétriques par rapport aux besoins matériels (les objets de calibrage), et en général plus robuste que les méthodes d'auto-calibrage [Zha00].

Des études sur les singularités et des comparaisons des méthodes de calibrage planaire ont été réalisées par Sturm [SM99] et Zollner [ZS03] pour les cas où la structure du plan de référence est donnée.

6.1.4 Techniques basées sur des entités spécifiques dans la scène

D'autres techniques de calibrage sont basées sur l'observation d'entités géométriques spécifiques dont on peut extraire des attributs liés aux paramètres intrinsèques de la caméra, qui sont invariants à leurs positions relatives par rapport à la caméra. Ces méthodes emploient des entités telles que : points de fuite correspondant aux orientations orthogonales [RCR99, CT90], rectangles [BAA98], cercles concentriques [KKK02], ellipses [HJ01] et coniques en général [YSH00].

Certaines de ces techniques requièrent que les éléments géométriques correspondant soient mis sur un motif plan (e.g. imprimés sur une feuille de papier) qui doit être observé par la caméra depuis plusieurs positions et orientations différentes [YSH00, HJ01, KKK02].

D'autres techniques profitent de la géométrie des éléments trouvés dans les scènes, tels que les objets en forme de parallélépipèdes (e.g. bâtiments, mobilier, boîtes) fréquemment présents dans les scènes architecturales ou dans certains espaces intérieurs.

Cependant, la flexibilité d'utilisation de ces méthodes dans des applications générales est conditionnée par la présence dans l'environnement des primitives de calibrage recherchées.

6.2 Méthode des points de fuite

Caprile et Torre [CT90] ont développé une méthode de calibrage en utilisant les propriétés perspectives des points de fuite. Dans leur méthode, la distance focale et le point principal sont récupérés à partir d'une seule image d'un prisme rectangulaire (le rapport d'aspect et l'angle entre les axes de la caméra sont supposés connus). Basés sur cette méthode, Cipolla *et al.* [RCR99] ont proposé une méthode qui tire profit des contraintes de parallélisme et d'orthogonalité présentes dans certaines scènes, notamment architecturales (e.g. bâtiments).

Dans la pratique, les points de fuite sont déterminés sur le plan image par l'observation d'un ensemble de lignes parallèles dans l'espace. Le point de fuite est donc défini par l'intersection des projections de ces droites (figure 6.1). Lorsque cet ensemble de droites dans l'espace n'est pas parallèle au plan image, les coordonnées du point de fuite sont finies ; sinon, le point de fuite se trouve à l'infini et ses coordonnées sont déterminées par l'orientation des lignes projetées.

Liebowitz *et al.* [LZ99] ont fait une analyse des dégénérescences de la méthode lors du calibrage à partir d'une ou deux vues et ont également établi les connaissances *a priori* nécessaires pour utiliser les contraintes imposées par cette méthode lorsqu'on parvient à obtenir un, deux, ou trois points de fuite (avec des coordonnées finies). Issues de cette analyse, d'autres travaux (parmi lesquels on trouve

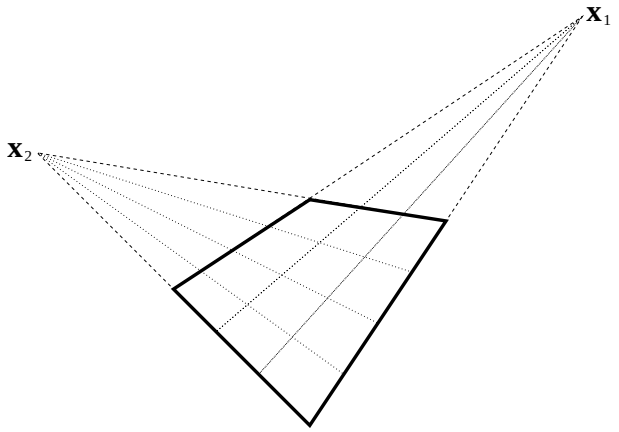


FIG. 6.1 – Points de fuite obtenus à partir de l'image d'un rectangle ou d'un quadrillage.

[BAA98, SFZ00]) proposent d'employer une technique simple pour l'estimation de la distance focale de la caméra à l'aide d'un contour rectangulaire.

Par la suite, nous citons les contraintes de calibrage issues des points de fuite, et en section 7.4.1 nous présentons les résultats de nos expériences montrant les limites de précision qui peuvent être atteintes avec cette méthode.

Fondements

Étant donnée une matrice de projection $\mathbf{P} = \mathbf{K}\begin{pmatrix}\mathbf{R} & \mathbf{t}\end{pmatrix}$ (où $\mathbf{K}$ est la matrice de paramètres intrinsèques et $\begin{pmatrix}\mathbf{R} & \mathbf{t}\end{pmatrix}$ est la transformation rigide qui ramène les coordonnées du monde vers les repère de la caméra), les points de fuite correspondant aux directions Ox, Oy et Oz ($\tilde{\mathbf{v}}_1$, $\tilde{\mathbf{v}}_2$ et $\tilde{\mathbf{v}}_3$, en coordonnées homogènes) sont déterminés par :

$$\lambda_i \tilde{\mathbf{v}}_i = \mathbf{P}\mathbf{e}_i \quad \forall i = 1,2,3$$

où $\mathbf{e}_i$ sont les points à l'infini associés à ces trois orientations : $\mathbf{e}_1 = (1,0,0,0)^T$, $\mathbf{e}_2 = (0,1,0,0)^T$ et $\mathbf{e}_1 = (0,0,1,0)^T$, et λ_i sont des facteurs d'échelle homogénéisant les coordonnées.

Ces relations peuvent être exprimées en terme de la matrice de paramètres intrinsèques de la caméra et de la matrice de rotation associées à $\mathbf{P}$, comme :

$$\begin{aligned}\begin{pmatrix}\lambda_1\tilde{\mathbf{v}}_1 & \lambda_2\tilde{\mathbf{v}}_2 & \lambda_3\tilde{\mathbf{v}}_3\end{pmatrix} &= \mathbf{P}\begin{pmatrix}\mathbf{e}_1 & \mathbf{e}_2 & \mathbf{e}_3\end{pmatrix} \\ &= \begin{pmatrix}\mathbf{p}_1 & \mathbf{p}_2 & \mathbf{p}_3\end{pmatrix} \\ &= \mathbf{K}\mathbf{R}\end{aligned}$$

où $\mathbf{p}_i$ est la i-ème colonne de la matrice $\mathbf{P}$.

La matrice de rotation a donc la forme :

$$\mathbf{R} = \begin{pmatrix}\mathbf{r}_1 & \mathbf{r}_2 & \mathbf{r}_3\end{pmatrix} = \begin{pmatrix}\lambda_1\mathbf{K}^{-1}\tilde{\mathbf{v}}_1 & \lambda_2\mathbf{K}^{-1}\tilde{\mathbf{v}}_2 & \lambda_3\mathbf{K}^{-1}\tilde{\mathbf{v}}_3\end{pmatrix}$$

où $\mathbf{r}_i$ $(i = 1,2,3)$ sont les vecteurs colonnes de $\mathbf{R}$.

Alors, en utilisant la propriété d'orthogonalité entre les trois vecteurs de la matrice de rotation, nous pouvons établir les contraintes suivantes :

$$\mathbf{r}_1^T\mathbf{r}_2 = \mathbf{r}_1^T\mathbf{r}_3 = \mathbf{r}_2^T\mathbf{r}_3 = 0$$

D'où

$$\left\{\begin{array}{rcl}\tilde{\mathbf{v}}_1^T\left(\mathbf{K}^{-T}\mathbf{K}^{-1}\right)\tilde{\mathbf{v}}_2 &=& 0 \\ \tilde{\mathbf{v}}_1^T\left(\mathbf{K}^{-T}\mathbf{K}^{-1}\right)\tilde{\mathbf{v}}_3 &=& 0 \\ \tilde{\mathbf{v}}_2^T\left(\mathbf{K}^{-T}\mathbf{K}^{-1}\right)\tilde{\mathbf{v}}_3 &=& 0\end{array}\right. \tag{6.1}$$

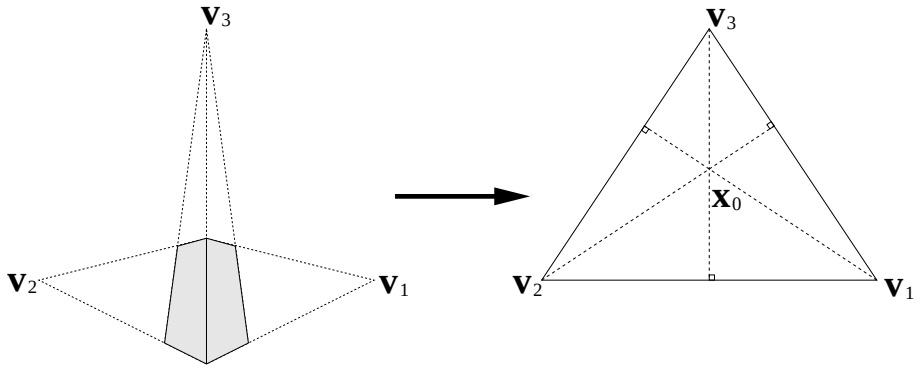


FIG. 6.2 – Relation entre les points de fuite et le point principal.

D'après ces équations, et comme l'angle entre les axes du plan image est supposé droit, chaque couple $(\mathbf{v}_i, \mathbf{v}_j)$ de points de fuite correspondant à des directions orthogonales, fournit une contrainte de la forme :

$$1 + \frac{1}{f^2}\left[(x_i - u_0)(x_j - u_0) + \frac{1}{\tau^2}(y_i - v_0)(y_j - v_0)\right] = 0 \tag{6.2}$$

où $\mathbf{x}_0 = (u_0, v_0)^T$ est le point principal, et $(x_i, y_i, 1)^T$ sont les coordonnées homogènes du point de fuite $\tilde{\mathbf{v}}_i$ (à condition que le point de fuite ait des coordonnées finies sur le plan image, c'est à dire, que les lignes droites utilisées pour déterminer le point de fuite ne soient pas parallèles au plan de l'image).

Trois points de fuite : calcul de la distance focale et du point principal

Caprile et Torre [CT90] ont démontré qu'il est possible d'estimer la distance focale ainsi que le point principal à partir d'une seule vue, lorsque trois points de fuite correspondants à des directions orthogonales sont connus. Leur méthode de calibrage est basée sur la contrainte (6.2). En supposant que la caméra est *naturelle* (définie selon [LZ99] comme un modèle de caméra dont le rapport d'aspect est unitaire et l'angle entre les axes des pixels est droit), les trois contraintes (6.1) sont récrites comme :

$$\begin{cases} (\mathbf{v}_1 - \mathbf{x}_0) \cdot (\mathbf{v}_2 - \mathbf{x}_0) = -f^2 \\ (\mathbf{v}_1 - \mathbf{x}_0) \cdot (\mathbf{v}_3 - \mathbf{x}_0) = -f^2 \\ (\mathbf{v}_2 - \mathbf{x}_0) \cdot (\mathbf{v}_3 - \mathbf{x}_0) = -f^2 \end{cases}$$

La soustraction des deux premières équations est :

$$(\mathbf{v}_1 - \mathbf{x}_0) \cdot (\mathbf{v}_2 - \mathbf{x}_0) - (\mathbf{v}_1 - \mathbf{x}_0) \cdot (\mathbf{v}_3 - \mathbf{x}_0) = 0$$

d'où en simplifiant, on obtient :

$$(\mathbf{v}_1 - \mathbf{x}_0) \cdot (\mathbf{v}_2 - \mathbf{v}_3) = 0$$

Cela signifie que le segment $\overrightarrow{\mathbf{v}_1\mathbf{x}_0}$ est perpendiculaire au segment $\overrightarrow{\mathbf{v}_2\mathbf{v}_3}$ (voir figure 6.2). De la même manière, on peut obtenir les contraintes d'orthogonalité pour le couple de segments $\overrightarrow{\mathbf{v}_2\mathbf{x}_0}$ et $\overrightarrow{\mathbf{v}_1\mathbf{v}_3}$, et entre $\overrightarrow{\mathbf{v}_3\mathbf{x}_0}$ et $\overrightarrow{\mathbf{v}_1\mathbf{v}_2}$. Ceci démontre que le point principal de la caméra se trouve sur l'orthocentre du triangle ayant comme sommets les trois points de fuite.

La procédure de calibrage est donc :

1. Les points de fuite correspondant aux trois directions orthogonales sont obtenus. Chaque point de fuite est estimé comme le point de convergence des projections de deux ou plusieurs lignes parallèles, placées sur les faces d'un parallélépipède (e.g. les arêtes d'un cube).
2. Si le rapport d'aspect τ n'est pas unitaire, les coordonnées des points de fuite sont transformées (en divisant leurs composantes y_i par τ) afin de satisfaire la contrainte $\tau = 1$.

3. Le point principal de la caméra naturelle $\mathbf{x}'_0 = (x'_0, y'_0)$ est calculé comme l'orthocentre du triangle formé par les trois points de fuite, et le point principal réel est obtenu en renormalisant leurs coordonnées ($x_0 = x'_0$, $y_0 = \tau y'_0$).
4. Le point principal étant connu, la distance focale est estimée d'après la formule :

$$f = \sqrt{-\left[(x_i - u_0)(x_j - u_0) + \frac{1}{\tau^2}(y_i - v_0)(y_j - v_0)\right]} \tag{6.3}$$

Deux points de fuite : calcul de la distance focale

L'expression (6.3) est utilisée dans [SFZ00] pour calculer la distance focale à partir de seulement deux points de fuite orthogonaux, et de la connaissance du point principal et du rapport d'aspect de la caméra. Alors, la distance focale peut être obtenue à partir d'un rectangle (figure 6.1) en utilisant la technique suivante :

1. Les points de fuite correspondant à deux directions orthogonales sont obtenus.
2. Le rapport d'aspect et le point principal sont supposés connus, et le coefficient d'obliquité est considéré nul (l'angle entre les axes est droit).
3. La distance focale est calculée par la formule (6.3).

C'est le seul cas où les contraintes des points de fuite peuvent être utilisées pour effectuer le calibrage interne d'une caméra en utilisant un motif planaire, au lieu d'un volume (comme c'est le cas décrit antérieurement), et il sera donc notre sujet d'analyse dans la suite de cette section.

Configurations critiques

L'équation (6.2) a été dérivée à partir des contraintes (6.1) sous l'hypothèse que les points de fuite aient des coordonnées finies. Les algorithmes décrits antérieurement peuvent être employés uniquement si aucun des points de fuite utilisés comme données d'entrée ne se trouve à l'infini.

Si cette condition n'est pas satisfaite, les contraintes (6.1) deviennent :

- lorsque seulement l'un des points de fuite se trouve à l'infini (soit $\mathbf{v}_i = (x_i, y_i, 0)^T$ et $\mathbf{v}_j = (x_j, y_j, 1)^T$) :

$$x_i(x_j - u_0) + \frac{y_i}{\tau^2}(y_j - v_0) = 0 \tag{6.4}$$

  Dans cette contrainte, le terme correspondant à la distance focale a été annulé, et seuls le point principal et le rapport d'aspect sont contraints par cette équation. Cette caractéristique a été utilisée par Gurdjos *et al.* [GP01] dans le but de découpler le calcul du point principal et de la distance focale, et de déterminer le lieu géométrique de la position du point principal (une ligne droite, car l'ordre de l'équation est linéaire). Dans la section 6.3.4, nous présentons de manière plus détaillée cette méthode.
- quand les deux points de fuite sont à l'infini (soit $\mathbf{v}_i = (x_i, y_i, 0)^T$ et $\mathbf{v}_j = (x_j, y_j, 0)^T$) :

$$x_i x_j + \frac{y_i y_j}{\tau^2} = 0$$

  Dans ce cas, le plan correspondant est parallèle au plan image, et seul le rapport d'aspect peut être estimé à partir de la connaissance de ces deux points.

6.3 Calibrage interne basé plans

Dans cette section, nous présentons les principes fondamentaux du calibrage planaire ainsi qu'une analyse de quelques méthodes de calibrage des paramètres intrinsèques basées plans.

Comme données d'entrée, on considère avoir acquis n vues d'un ensemble de primitives localisées sur un plan Π, en utilisant une caméra dont les paramètres intrinsèques sont supposés constants tout au long

de la séquence d'images. Le but de ces méthodes de calibrage est donc d'utiliser les n vues d'une structure plane pour calculer les cinq paramètres intrinsèques (f, τ, s_k, u_0 et v_0) associés à la caméra ayant pris cette séquence d'images.

Ici, nous examinons les relations entre les paramètres intrinsèques d'une caméra, et les homographies liant un ensemble de primitives coplanaires à leurs projections sur les images. Nous décrivons les contraintes dérivées de ces relations ainsi que la manière dont elles sont employées par les différentes méthodes de calibrage interne en observant une structure plane.

Les méthodes ici présentées ont été choisies parce qu'elles présentent soit :

- des solutions classiques de référence dans le domaine [SM99] (section 6.3.2),
- des solutions adaptées spécialement pour effectuer le calibrage interne de manière séquentielle [NJS03] (section 6.3.3),
- des approches qui séparent l'estimation des paramètres en différentes étapes, e.g. en découplant les paramètres intrinsèques pour éviter des compensations dans les calculs [GP01, GCP02, GS03] (sections 6.3.4 et 6.3.5).

Les deux premières méthodes emploient une contrainte classique, représentant les conditions d'orthogonalité et de normalisation de deux entités géométriques nommées *points cycliques*, qui seront décrits dans la section 6.3.1. L'autre approche est caractérisée par la détermination du lieu géométrique du centre de projection perspective, dont l'intérêt principal et de découpler l'estimation de la distance focale et le calcul du point principal (section 6.3.4).

Ces contraintes ont été utilisées pour les deux modalités de calibrage : l'obtention seulement des paramètres intrinsèques, et le calcul simultanée du calibrage interne et de la structure du plan de référence.

Le problème du calibrage interne classique à base de plans consiste à calculer les paramètres intrinsèques de la caméra à partir d'un ensemble de n vues d'une structure plane connue, en utilisant comme information d'entrée l'ensemble des homographies scène-images $\{\mathbf{H}_1, \cdots, \mathbf{H}_n\}$.

Lorsque la structure de la surface plane observée est inconnue, la détermination des matrices d'homographies scène-image n'est pas faisable. Néanmoins, les correspondances de points entre les images permettent de calculer des transformations homographiques entre les vues.

Pour le problème de l'auto-calibrage planaire, on dispose des homographies inter-images, soit entre toutes les images consécutives $\{\mathbf{H}_{i-1,i}, \forall i = 2, \cdots, n\}$ où référencées à une image clé $\{\mathbf{H}_{1,i}, \forall i = 2, \cdots, n\}$, mais la structure du plan de référence est ordinairement inconnue.

6.3.1 Relations de base du calibrage planaire

Contraintes induites par les homographies scène-image sur les paramètres intrinsèques

À partir des homographies scène-image, des contraintes sur les paramètres intrinsèques peuvent être établies au moyen de la propriété d'orthonormalité de la matrice de rotation des paramètres extrinsèques. Étant donnée la matrice de projection $\mathbf{P} = \mathbf{K}\,(\mathbf{R}\ \mathbf{t})$ et d'après l'équation (2.3), on obtient les relations suivantes :

$$\mathbf{K}^{-1}\mathbf{H} \sim (\mathbf{r}_1\ \mathbf{r}_2\ \mathbf{t}) \Rightarrow \begin{cases} \mathbf{r}_1 &= \lambda \mathbf{K}^{-1}\mathbf{h}_1 \\ \mathbf{r}_2 &= \lambda \mathbf{K}^{-1}\mathbf{h}_2 \\ \mathbf{t} &= \lambda \mathbf{K}^{-1}\mathbf{h}_3 \end{cases}$$

Ici, $(\mathbf{h}_1\ \mathbf{h}_2\ \mathbf{h}_3)$ sont les vecteurs colonnes de $\mathbf{H}$, et λ est un facteur scalaire.

Comme les vecteurs $\mathbf{r}_1$ et $\mathbf{r}_2$ correspondent aux deux premières colonnes d'une matrice de rotation, ils sont astreints aux conditions d'orthogonalité ($\mathbf{r}_1 \cdot \mathbf{r}_2 = 0$) et de normalisation ($\|\mathbf{r}_1\| = \|\mathbf{r}_2\| = 1$). Ainsi, ces deux conditions sont écrites en termes de $\mathbf{H}$ et de $\mathbf{K}$ comme :

$$\left\{ \begin{array}{ll} \text{orthogonalité} & \mathbf{r}_1 \cdot \mathbf{r}_2 = 0 \\ \text{normalisation} & \|\mathbf{r}_1\|^2 - \|\mathbf{r}_2\|^2 = 0 \end{array} \right\} \Rightarrow \left\{ \begin{array}{c} \mathbf{h}_1^T \mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 = 0 \\ \mathbf{h}_1^T \mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_1 - \mathbf{h}_2^T \mathbf{K}^{-T}\mathbf{K}^{-1}\mathbf{h}_2 = 0 \end{array} \right\} \tag{6.5}$$

L'emploi de certaines entités projectives abstraites, telles que la *conique absolue* et les *points cycliques*, a permis de donner une interprétation géométrique aux contraintes de calibrage planaire (6.5), ainsi que de mettre en évidence d'autres propriétés importantes liées à la rectification métrique de la scène [HZ00] (par

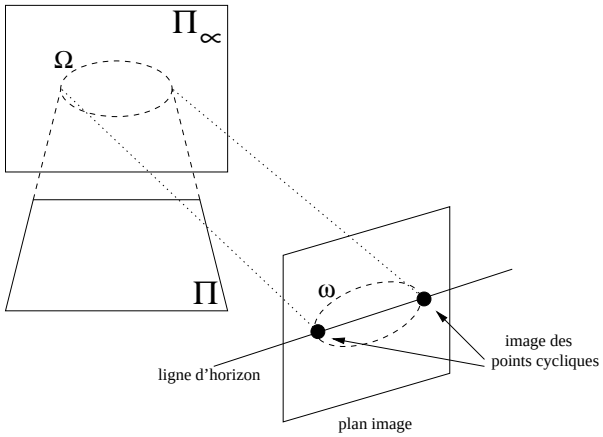


FIG. 6.3 – La conique absolue, les points cycliques, et leurs images correspondantes.

exemple, des entités invariantes aux transformations euclidiennes). Cette représentation a été exploitée dans l'implémentation de méthodes de calibrage planaire tels que [Tri98b, SM99, Zha00] entre autres.

Dans cette section, la *conique absolue* et les *points cycliques* sont présentés, ainsi que leur rapport avec les contraintes de calibrage planaire.

La conique absolue et les points cycliques

L'intersection de toutes les sphères de l'espace projectif $\mathbb{P}^3$ avec le plan à l'infini Π_∞ déterminent une conique Ω_∞, nommée la *conique absolue*. La conique absolue est une conique complexe dont les points $\mathbf{x} = (x_1, x_2, x_3, x_4)^T$ sont déterminés par les équations $x_1^2 + x_2^2 + x_3^2 = 0$ et $x_4 = 0$.

Comme les points de la conique se trouvent sur le plan $\Pi_\infty = (0, 0, 0, 1)^T$, leurs coordonnées $\mathbf{x}$ peuvent être exprimées par les vecteurs d'orientations $\hat{\mathbf{x}} = (x_1, x_2, x_3)^T$, de telle manière que la conique absolue est associée à la matrice $\Omega_\infty \sim \mathbf{I}_3$ (matrice identité de dimension 3×3) ; et les points appartenant à cette conique sont contraints par l'équation $\hat{\mathbf{x}}^T \Omega_\infty \hat{\mathbf{x}} = 0$.

Une propriété importante de la conique absolue est que Ω_∞ est une conique fixe sous une transformation projective $\mathbf{T}$ si, et seulement si, $\mathbf{T}$ est une transformation rigide [HZ00]. Donc, si Ω_∞ est invariante sous les transformations euclidiennes, sa position relative par rapport à une caméra en mouvement est fixe et la projection sur le plan image de cette caméra est aussi constante.

Points cycliques

Un cercle situé sur le plan Π peut être défini comme l'intersection entre Π et une sphère quelconque contenant le cercle. Alors, l'intersection du cercle avec le plan à l'infini est identique à l'intersection du plan Π avec la conique absolue Ω_∞ [Zha00] (voir fig. 6.3), puisque Ω_∞ a été définie comme l'intersection de toutes les sphères avec Π_∞.

Alors, Π intersecte le plan à l'infini dans une droite, et cette droite croise Ω_∞ en deux points, appelés *points cycliques* (figure 6.3). En fait, tous les cercles coplanaires intersectent la conique absolue Ω_∞ aux mêmes deux points cycliques, qui sont associés au plan Π.

Sans perte de généralité, on choisit que l'équation du plan soit $\{\Pi : Z = 0\}$ et que le système de coordonnées du monde soit lié au système de coordonnées de la caméra par la transformation rigide $(\mathbf{R} \quad \mathbf{t}) = (\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3 \ \mathbf{t})$. Tous les points sur l'intersection entre le plan Π et le plan à l'infini sont donc de la forme :

$$\hat{\mathbf{x}}_\infty = (\mathbf{r}_1 \ \mathbf{r}_2 \ \mathbf{r}_3 \ \mathbf{t}) \begin{pmatrix} X \\ Y \\ 0 \\ 0 \end{pmatrix} = X\mathbf{r}_1 + Y\mathbf{r}_2$$

Les points cycliques appartiennent aussi à la conique absolue et doivent satisfaire la contrainte $\hat{\mathbf{x}}_\infty^T \Omega \hat{\mathbf{x}}_\infty = 0$:

$$\begin{aligned} (X\mathbf{r}_1 + Y\mathbf{r}_2)^T \Omega (X\mathbf{r}_1 + Y\mathbf{r}_2) &= 0 \\ X^2 (\mathbf{r}_1^T \Omega \mathbf{r}_1) + Y^2 (\mathbf{r}_2^T \Omega \mathbf{r}_2) &= 0 \\ X^2 + Y^2 &= 0 \end{aligned}$$

ayant comme solutions $Y = \pm iX$; de telle manière que les points cycliques sont au nombre de deux et déterminés par les deux vecteurs complexes conjugués :

$$\hat{\mathbf{x}}_\infty \sim \mathbf{r}_1 \pm i\mathbf{r}_2 \tag{6.6}$$

Image de la conique absolue

L'image $\mathbf{u}$ d'un point $\mathbf{x}_\infty = (x_1, x_2, x_3, 0)^T$ du plan à l'infini est obtenue par $\mathbf{u} \sim \mathbf{K}\mathbf{R}\hat{\mathbf{x}}_\infty$. Ainsi, $\mathbf{H}_\infty = \mathbf{K}\mathbf{R}$ est l'homographie qui relie les points du plan Π_∞ à leurs projections sur l'image. L'orientation $\hat{\mathbf{x}}_\infty$ peut être obtenue d'après son image comme $\hat{\mathbf{x}}_\infty \sim \mathbf{R}^{-1}\mathbf{K}^{-1}\mathbf{u}$.

L'image ω de la conique absolue est donc obtenue à partir de la relation suivante :

$$\begin{aligned} \hat{\mathbf{x}}_\infty^T \Omega \hat{\mathbf{x}}_\infty &\sim \left(\mathbf{R}^{-1}\mathbf{K}^{-1}\mathbf{u}\right)^T \Omega \left(\mathbf{R}^{-1}\mathbf{K}^{-1}\mathbf{u}\right) = 0 \\ &\sim \mathbf{u}^T \mathbf{K}^{-T} \left(\mathbf{R}^{-T} \Omega \mathbf{R}^{-1}\right) \mathbf{K}^{-1}\mathbf{u} = 0 \\ &\sim \mathbf{u}^T \mathbf{K}^{-T} \mathbf{K}^{-1} \mathbf{u} = 0 \end{aligned}$$

Dès lors, l'image de la conique absolue Ω d'après la transformation projective $\mathbf{P}$ est déterminée par $\omega \sim \mathbf{K}^{-T}\mathbf{K}^{-1}$. Cette matrice est symétrique et elle est caractérisée par 6 coefficients (à un facteur près).

L'image de la conique absolue dépend uniquement des paramètres intrinsèques de la caméra, propriété qui la rend importante dans la formulation du problème de calibrage interne. L'image de la conique absolue est donc invariable aux transformations euclidiennes, ce qui signifie que pour une caméra en mouvement, ω est fixe tant que les paramètres intrinsèques sont aussi constants. Le problème du calcul des paramètres intrinsèques est donc équivalent à la détermination de ω [Stu97b][HZ00].

Rapport entre la conique absolue et les contraintes de calibrage planaire

D'après l'estimation de la position des points cycliques, obtenue dans (6.6), la projection des points cycliques sur le plan image est donnée par $\mathbf{u}_\infty \sim \mathbf{K}(\mathbf{r}_1 \pm i\mathbf{r}_2) = \mathbf{h}_1 \pm i\mathbf{h}_2$. Puisque les points cycliques d'un plan Π appartiennent à la conique absolue Ω, leurs projections doivent donc faire partie de l'image de la conique absolue ω, et satisfaire l'équation $\mathbf{u}_\infty^T \omega \mathbf{u}_\infty = 0$. En conséquence :

$$(\mathbf{h}_1 \pm i\mathbf{h}_2)^T \omega (\mathbf{h}_1 \pm i\mathbf{h}_2) = 0$$

d'où en séparant les parties réelle et imaginaire :

$$\left(\mathbf{h}_1^T \omega \mathbf{h}_1 - \mathbf{h}_2^T \omega \mathbf{h}_2\right) \pm 2i \left(\mathbf{h}_1^T \omega \mathbf{h}_2\right) = 0$$

Donc, les parties réelle et imaginaire du terme gauche de l'équation doivent être respectivement égales à zéro :

$$\begin{cases} \mathbf{h}_1^T \omega \mathbf{h}_1 - \mathbf{h}_2^T \omega \mathbf{h}_2 = 0 \\ \mathbf{h}_1^T \omega \mathbf{h}_2 = 0 \end{cases} \tag{6.7}$$

Ces deux équations sont identiques aux contraintes (6.5) présentés en section 2.3.1, et sont équivalentes aux restrictions d'orthogonalité et de normalisation des vecteurs $\mathbf{r}_1$ et $\mathbf{r}_2$ définissant une base pour le plan Π dans le repère de la caméra.

Les 8 coefficients d'une homographie scène-image contraignent la matrice $\mathbf{K}$ à l'égard des paramètres extrinsèques $(\mathbf{R} \quad \mathbf{t})$ (6 degrés de liberté), en fournissant ainsi 2 contraintes sur les paramètres intrinsèques pour chaque homographie. Un minimum de deux homographies scène-images différentes sont requises pour estimer (f, τ, u_0 et v_0), et trois homographies pour estimer les cinq paramètres intrinsèques.

Relation entre l'image de la conique absolue et les paramètres intrinsèques

L'angle entre les axes horizontaux et verticaux de l'image et le rapport d'aspect sont généralement stables (même pour de longues périodes d'utilisation de la caméra) [Wil94, Enc95]. Les variations de ces paramètres dues aux variables physiques externes, à la mise au point et même aux variations du zoom, sont usuellement négligeables. Comme l'angle entre les axes d'un pixel est en général de 90°, le coefficient d'obliquité est habituellement fixé à 0, supposition que nous utiliserons dans la suite de ce texte, sauf indication contraire.

Alors, si le terme d'obliquité est fixé à 0, $\omega_{12} = 0$ et

$$\omega = \begin{pmatrix} \omega_{11} & \omega_{12} & \omega_{13} \\ \omega_{12} & \omega_{22} & \omega_{23} \\ \omega_{13} & \omega_{23} & \omega_{33} \end{pmatrix} \sim \mathbf{K}^{-T}\mathbf{K}^{-1} = \frac{1}{(\tau f)^2} \begin{pmatrix} \tau^2 & 0 & -\tau^2 u_0 \\ 0 & 1 & -v_0 \\ -\tau^2 u_0 & -v_0 & \tau^2 f^2 + \tau^2 u_0^2 + v_0^2 \end{pmatrix} \tag{6.8}$$

Lorsque les coefficients de ω sont donnés (nous verrons plus loin comment les calculer), les paramètres intrinsèques de la caméra seront déterminés directement par les formules suivantes :

$$\tau = \sqrt{\frac{\omega_{11}}{\omega_{22}}} \qquad u_0 = \frac{\omega_{13}}{\omega_{11}} \qquad v_0 = \frac{\omega_{23}}{\omega_{22}} \qquad f = \sqrt{\frac{1}{\omega_{11}}\left(\omega_{33} - \frac{\omega_{13}^2}{\omega_{11}} - \frac{\omega_{23}^2}{\omega_{22}}\right)} \tag{6.9}$$

6.3.2 Méthode linéaire basée sur l'image de la conique absolue

Sturm et Maybank [SM99] ont développé une méthode linéaire de calibrage interne en exploitant l'invariabilité de la conique absolue aux transformations euclidiennes et la relation de cette conique avec les paramètres intrinsèques.

D'après la contrainte de la conique absolue sur les points cycliques (6.7), chaque homographie scène-image fournit exactement deux équations linéaires en terme des coefficients de ω :

$$\begin{cases} \begin{aligned} (h_{11}h_{12})\omega_{11} + (h_{11}h_{32} + h_{12}h_{31})\omega_{13} + (h_{21}h_{22})\omega_{22} \\ +(h_{21}h_{32} + h_{22}h_{31})\omega_{23} + (h_{31}h_{32})\omega_{33} \end{aligned} &= 0 \\ \begin{aligned} ({h_{11}}^2 - {h_{12}}^2)\omega_{11} + 2\left(h_{11}h_{31} - h_{12}h_{32}\right)\omega_{13} + ({h_{21}}^2 - {h_{22}}^2)\omega_{22} \\ +2\left(h_{21}h_{31} - h_{22}h_{32}\right)\omega_{23} + ({h_{31}}^2 - {h_{32}}^2)\omega_{33} \end{aligned} &= 0 \end{cases} \tag{6.10}$$

Ces équations sont utilisées pour construire un système linéaire homogène de la forme :

$$\mathbf{A}\mathbf{w} = \begin{pmatrix} \vdots \\ \mathbf{a}_i^T \\ \mathbf{b}_i^T \\ \vdots \end{pmatrix} \mathbf{w} = \mathbf{0} \tag{6.11}$$

$$\begin{aligned} \mathbf{a}_i^T &= (\quad h_{11}h_{12}, \quad h_{11}h_{32} + h_{12}h_{31}, \quad h_{21}h_{22}, \quad h_{21}h_{32} + h_{22}h_{31}, \quad h_{31}h_{32} \quad) \\ \mathbf{b}_i^T &= (\quad {h_{11}}^2 - {h_{12}}^2, \quad 2\left(h_{11}h_{31} - h_{12}h_{32}\right), \quad {h_{21}}^2 - {h_{22}}^2, \quad 2\left(h_{21}h_{31} - h_{22}h_{32}\right), \quad {h_{31}}^2 - {h_{32}}^2 \quad) \end{aligned}$$

où $\mathbf{a}_i$ et $\mathbf{b}_i$ sont les vecteurs définissant les deux contraintes de calibrage pour l'homographie scène-image $\mathbf{H}_i$, et $\mathbf{w} = \{\omega_{11}, \omega_{13}, \omega_{22}, \omega_{23}, \omega_{33}\}$

Algorithme

L'algorithme proposé par Sturm et Maybank [SM99] est basé sur la détermination de l'image de la conique absolue à partir de la résolution du système homogène d'équations linéaires (6.11).

Leur méthode fournit une solution au problème du calcul des paramètres intrinsèques à partir de la connaissance de n matrices d'homographies, considérant même la possibilité que certains paramètres intrinsèques varient au long de la séquence (cela est possible en ajoutant des nouveaux coefficients $\omega_{13}^{(i)}$

et $\omega_{23}^{(i)}$ si le point principal varie, ou $\omega_{33}^{(i)}$ si la distance focale varie, et les colonnes correspondantes dans la matrice du système).

L'algorithme est composé par les étapes suivantes :

1. Calcul des n matrices d'homographie.
2. Construction du système linéaire homogène (6.11) : $\mathbf{Aw} = \mathbf{0}$.
3. Renormalisation par colonnes ($\mathbf{A}' = \mathbf{AD}$) de la matrice $\mathbf{A}$, pour améliorer le conditionnement numérique du système. La renormalisation est effectué en suivant la méthode de stabilisation proposée par Golub [Gol89].
4. Résolution du système $\mathbf{A}'\mathbf{w}' = \mathbf{0}$, en utilisant la méthode de résolution par SVD [PFTV88, Gol89].
5. Obtention de $\mathbf{w}$ par substitution de $\mathbf{w} = \mathbf{Dw}'$.
6. Récupération des paramètres intrinsèques en utilisant les formules (6.9).

La connaissance *a priori* de certains paramètres intrinsèques permet de restreindre le nombre de termes dans les contraintes, de façon que les équations restent linéaires en fonction des coefficients de ω [SM99] :

- si le rapport d'aspect τ est connu, le nombre de coefficients est réduit au moyen de la substitution $\omega_{11} = \tau^2\omega_{22}$;
- la connaissance des coordonnées du point principal nous permet d'ajouter les contraintes $\omega_{13} = -u_0\omega_{11}$ et $\omega_{23} = -v_0\omega_{22}$.

Néanmoins, quand la distance focale est donnée, les substitutions requises pour limiter le nombre de variables ne sont plus linéaires en fonction de $\mathbf{w}$, dû à la complexité du terme $\omega_{33} = \tau^2 f^2 + \tau^2 u_0^2 + v_0^2$. Cependant, le cas où la distance focale est connue lorsque les autres paramètres sont inconnus, n'est pas une situation habituelle.

6.3.3 Auto-calibrage itératif par filtrage de Kalman

Noirfalise *et al.* [NJS03] ont proposé un système d'auto-calibrage de caméras en visant une structure plane. Leur approche est basée sur un filtrage de Kalman qui permet un traitement séquentiel des images acquises (le modèle du filtre de Kalman est présenté dans l'annexe A).

Comme la position du plan observé par rapport à la caméra n'est pas connue, il est nécessaire d'effectuer une rectification projective du plan. Leur approche suggère donc d'obtenir simultanément une base projective pour la première image de la séquence ainsi que les paramètres intrinsèques de la caméra.

À chaque image, une itération du processus de filtrage est réalisée ; les entrées à chaque itération sont les transformations homographiques des plans observés entre les images $i-1$ et i, les matrices de covariances de ces homographies, ainsi qu'une estimation initiale des paramètres intrinsèques et de leurs covariances correspondantes.

La *fonction de mesure* du filtre estime l'adaptation des paramètres proposés (paramètres intrinsèques et de rectification de la première image) aux équations de l'image de la conique absolue (6.7). Pour cela, on doit calculer les homographies entre les points sur le plan rectifié ($Z = 0$) et leurs observations sur les images, c'est à dire, les homographies scène-image $\mathbf{H}_{0i}$. Comme ces homographies sont inconnues, l'approche propose l'estimation des coefficients de l'homographie $\mathbf{H}_{01}$, entre le plan de la scène et son observation sur la première image, c'est à dire, d'effectuer la rectification du plan. Ainsi, les homographies $\mathbf{H}_{0i}$ utilisées dans les contraintes de la conique absolue sont obtenues comme $\mathbf{H}_{0i} = \mathbf{H}_{1i}\mathbf{H}_{01}$, tel que $\mathbf{H}_{1i}$ est l'homographie entre l'image i et la première image (ou l'image clé).

Homographie entre la scène et l'image clé

La matrice $\mathbf{H}_{01}$ est l'homographie permettant d'effectuer la reconstruction euclidienne du plan Π. Cette matrice peut être décomposée en ses parties projective $\mathbf{H}_p$, affine $\mathbf{H}_a$ et euclidienne $\mathbf{H}_e$ [HZ00], de telle manière que $\mathbf{H}_{01} = \mathbf{H}_p\mathbf{H}_a\mathbf{H}_e$. Lorsque la partie euclidienne correspond à une transformation de similarité sur le plan Π, les contraintes de calibrage restent constants lors de ces transformations.

Il suffit d'estimer les matrices correspondantes aux transformations affine et projective pour établir une rectification métrique du plan projeté. Ces matrices sont définies par :

$$\mathbf{H}_p = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \mu & \lambda & 1 \end{pmatrix} \qquad \mathbf{H}_a = \begin{pmatrix} \beta & \alpha & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$\mathbf{H}_{0,1} = \begin{pmatrix} \beta & \alpha & 0 \\ 0 & 1 & 0 \\ \mu\beta & \mu\alpha + \lambda & 1 \end{pmatrix} = \begin{pmatrix} a & b & 0 \\ 0 & 1 & 0 \\ c & d & 1 \end{pmatrix} \tag{6.12}$$

La rectification métrique du plan Π ne requiert que 4 paramètres. Comme une transformation homographique a 8 degrés de liberté, les autre 4 coefficients variables correspondent à la transformation $\mathbf{H}_e$ (deux paramètres pour une translation sur le plan, un angle de rotation, et un facteur d'échelle).

Algorithme

L'algorithme proposé par les auteurs est :

1. Initialisation des paramètres intrinsèques et de la matrice $\mathbf{H}_{01}$. Comme les auteurs proposent de commencer le calibrage dans une position fronto-parallèle, les paramètres $\{a = 1, b = c = d = 0\}$ définissent un ensemble proche de la solution attendue. Également, il est nécessaire de proposer des valeurs initiales de paramètres intrinsèques et des matrices de covariances de tout l'ensemble de paramètres.
2. Pour l'image courante i, la matrice $\mathbf{H}_{0i}$ est estimée comme $\mathbf{H}_{0i} = \mathbf{H}_{1i}\mathbf{H}_{01}$.
3. Étape de prédiction du filtre de Kalman : étant supposé que les paramètres intrinsèques sont constants et que les paramètres de rectification sont obtenus pour une image fixe, les valeurs prédites sont égales à celles estimées pour l'image précédente. Néanmoins, les matrices de covariances de cette étape doivent être mise à jour.
4. Étape de correction du filtre de Kalman : étant données les vecteurs $\mathbf{h}_1$ et $\mathbf{h}_2$, correspondants aux deux premières colonnes de la matrice $\mathbf{H}_{0i}$, la fonction d'erreur du filtre est estimée à l'aide des contraintes (6.7) :

$$\mathbf{z}_i = \begin{pmatrix} \mathbf{h}_1^T \omega \mathbf{h}_1 - \mathbf{h}_2^T \omega \mathbf{h}_2 \\ \mathbf{h}_1^T \omega \mathbf{h}_2 \end{pmatrix} \tag{6.13}$$

   La totalité de paramètres est mise à jour en utilisant les équations du filtre (voir annexe A).
5. Les pas 2-4 sont répétés lors de l'acquisition de chaque nouvelle image.

6.3.4 Calibrage planaire basé sur la contrainte des lignes centrales

Gurdjos et Payrissat [GP01] ont proposé une formulation alternative à l'image de la conique absolue pour les contraintes de calibrage planaire. Leur formulation est basée sur un théorème de Poncelet qui décrit une contrainte sur la position du centre de projection à partir de deux figures planes en correspondance perspective. D'après ce théorème, on déduit que le lieu géométrique du point principal est une ligne droite (nommée *ligne centrale*) et que sa position sur le plan image est indépendante de la distance focale. Cette propriété leur a permis d'implanter une méthode de calibrage planaire qui découple le calcul de la distance focale de l'estimation des autres paramètres intrinsèques ([GP01]).

Le principal intérêt de ce découplage est d'éviter les possibles compensations entre les paramètres dans le processus de calcul, ce qui peut devenir une source d'erreur dans le processus du calibrage et, postérieurement, dans le positionnement de la caméra (comme nous l'avons montré au chapitre 3).

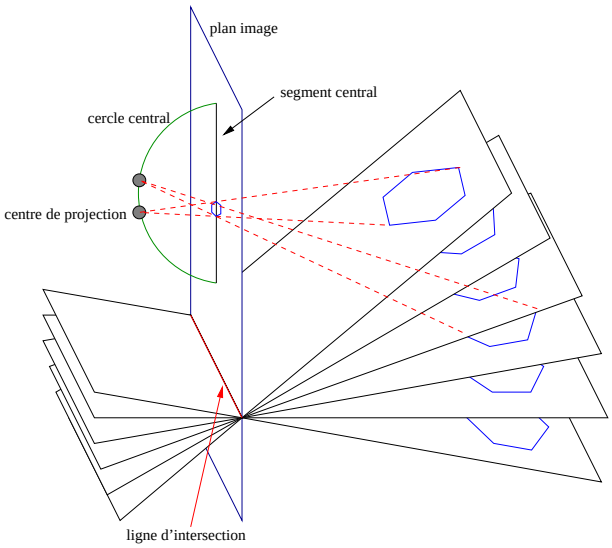


FIG. 6.4 – Illustration du théorème de Poncelet [GP01].

Théorème de Poncelet

Le théorème de Poncelet [Pon62, GP01] affirme que : *"Si une figure plane est la projection perspective d'une autre figure plane, ces figures restent en correspondance perspective lorsque le plan de la première tourne autour de son intersection avec le plan de la deuxième ; le centre de projection se trouve alors sur un cercle, dans une perpendiculaire du plan à cette intersection. En conséquence, la projection de ce cercle sur le plan de l'image est un segment de droite qui détermine le lieu géométrique du point principal"*.

D'après ce théorème, lorsqu'on connaît la position d'une figure sur une surface plane et sa projection sur l'image, le lieu géométrique des centres de projection est un cercle (appelé *cercle central*), et la projection du cercle sur le plan image est le segment d'une droite (nommée la *ligne centrale*). Donc, chaque transformation entre la scène et l'image d'une figure plane nous permet de contraindre la position du point principal de la caméra sur un segment de la ligne centrale correspondante. Ce principe est illustré dans la figure 6.4. Le point principal peut être donc calculé comme l'intersection de plusieurs lignes centrales.

La technique proposée par [GP01] permet d'effectuer le calcul des paramètres intrinsèques en deux étapes :

- l'estimation des coordonnées du point principal et le rapport d'aspect, en utilisant la contrainte des lignes centrales ;
- la détermination de la distance focale, en considérant que les autres paramètres intrinsèques sont donnés (e.g. en utilisant la méthode linéaire de Sturm-Maybank, section 6.3.2).

Expressions analytiques du cercle central et de la ligne centrale

Dans cette section, nous récapitulons les expressions algébriques qui définissent la contrainte de la ligne centrale, ainsi que leur rapport avec les contraintes de calibrage planaire (équations 6.11) dérivées dans la section 6.3.2. Ces éléments sont présentés en détail par les auteurs de cette méthode dans [GP01, GCP02].

D'après les équations (6.7), il est possible de déterminer une contrainte sur la position du point principal, qui soit indépendante de la distance focale. Comme le terme ω_{33} est le seul coefficient de ω qui dépend de la distance focale (voir équation 6.8), on utilise les deux équations (6.10) pour éliminer ce terme au moyen des opérations suivantes :

- chacune des deux équations (6.10) est multipliée par le dernier coefficient de l'autre équation (celui associé au terme ω_{33}) ;

- la différence entre les deux équations annule donc la composante correspondant à ω_{33} :

$$(h_{31}h_{32})\,\mathbf{b}^T\mathbf{w} - \left({h_{31}}^2 - {h_{32}}^2\right)\mathbf{a}^T\mathbf{w} = 0$$

d'où on obtient :

$$\begin{array}{r} (h_{11}h_{32} - h_{12}h_{31})\,(h_{11}h_{31} + h_{12}h_{32})\,\omega_{11} + \left(h_{31}^2 + h_{32}^2\right)(h_{11}h_{32} - h_{12}h_{31})\,\omega_{13} \\ + (h_{21}h_{32} - h_{22}h_{31})\,(h_{21}h_{31} + h_{22}h_{32})\,\omega_{22} + \left(h_{31}^2 + h_{32}^2\right)(h_{21}h_{32} - h_{22}h_{31})\,\omega_{23} \end{array} = 0 \quad (6.14)$$

En remplaçant les termes ω_{ij} par leurs expressions (éq. 6.8) en fonction des paramètres intrinsèques, on obtient l'équation de la ligne centrale associée à la matrice d'homographie $\mathbf{H}$:

$$\varphi_1\left(\tau^2 u_0\right) + \varphi_2\left(v_0\right) + \varphi_3\left(\tau^2\right) + \varphi_4 = 0 \quad (6.15)$$

où

$$\left\{\begin{array}{rcl} \varphi_1 & = & -\left(h_{31}^2 + h_{32}^2\right)(h_{11}h_{32} - h_{12}h_{31}) \\ \varphi_2 & = & -\left(h_{31}^2 + h_{32}^2\right)(h_{21}h_{32} - h_{22}h_{31}) \\ \varphi_3 & = & (h_{11}h_{32} - h_{12}h_{31})\,(h_{11}h_{31} + h_{12}h_{32}) \\ \varphi_4 & = & (h_{21}h_{32} - h_{22}h_{31})\,(h_{21}h_{31} + h_{22}h_{32}) \end{array}\right. \quad (6.16)$$

Étant donnée une matrice d'homographie non singulière, le seul cas pour lequel la ligne centrale ne peut pas être calculée, apparaît lorsque $h_{31} = h_{32} = 0$. Ce cas correspond à une homographie scène-image $\mathbf{H}$ où le plan correspondant est fronto-parallèle au plan image.

Calcul de la position du point principal

Comme les coefficients $\Phi = \{\varphi_1, \varphi_2, \varphi_3, \varphi_4\}$ sont déterminés uniquement à partir des homographies scène-images, il est possible à partir de plusieurs homographies de construire un système linéaire d'équations dont les inconnues sont les termes $\{\tau^2 u_0, v_0, \tau^2\}$. La solution de ce système d'équations nous permet d'obtenir directement le rapport d'aspect et la position du point principal.

Du point de vue géométrique, le point principal correspond à l'intersection des lignes centrales obtenues à partir des homographies différentes. Cependant, il est possible que par des erreurs de calcul des homographies scène-image, les lignes centrales obtenues $L(\Phi)$ ne s'intersectent pas exactement au même point.

La solution proposée par [GP01] consiste à calculer les coordonnées du point qui minimisent la somme des carrés des distances de ce point à chacune des lignes centrales. La distance du point $\hat{\mathbf{u}} = (\hat{u}, \hat{v})$ à la droite centrale $L(\Phi)$ déterminée par l'équation (6.15) est :

$$d\left(\hat{\mathbf{u}}, L\left(\Phi\right)\right) = \frac{\left|\varphi_1\hat{u} + \varphi_2\hat{v} + \left(\varphi_3\tau^2 + \varphi_4\right)\right|}{\sqrt{\varphi_1^2 + \varphi_2^2}} \quad (6.17)$$

En normalisant chacune des équations par le terme correspondant $\left(\varphi_1^2 + \varphi_2^2\right)^{-1/2}$, on obtient un système linéaire qui peut être résolu par la méthode de décomposition SVD dont la solution équivaut à minimiser la fonction d'erreur résiduelle suivante :

$$\sum_{i=1}^{n} d^2\left(\hat{\mathbf{u}}, L\left(\Phi_i\right)\right) \quad (6.18)$$

et le point principal est estimée par : $(u_0, v_0) = (\frac{1}{\tau^2}\hat{u}, \hat{v})$.

Algorithme

La méthode de calibrage en utilisant la contrainte des lignes centrales est constituée par les pas suivants :

1. Estimation des homographies scène-images pour une séquence d'images.

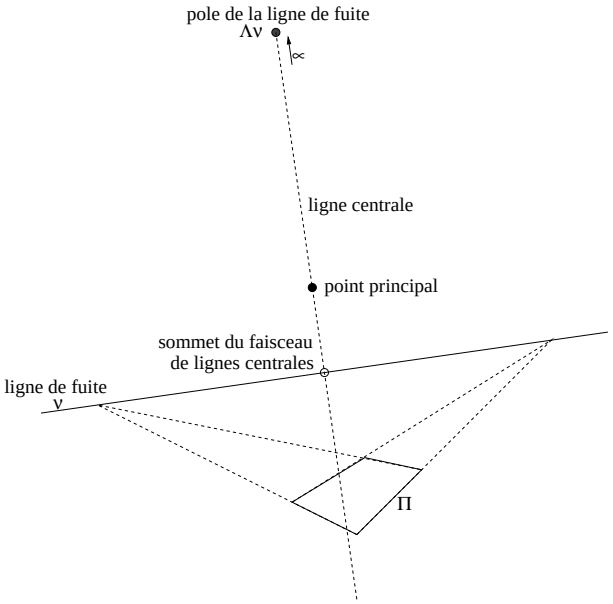


FIG. 6.5 – Contrainte de la ligne centrale.

2. Obtention des équations des lignes centrales associées à chaque homographie scène-image. Les droites sont déterminées par les coefficients Φ de l'équation (6.15), et normalisées par le terme $\left(\varphi_1^2 + \varphi_2^2\right)^{-1/2}$.
3. Calcul des coordonnées du point principal comme l'intersection des lignes centrales. La solution est obtenue à partir du système d'équations formé par les équations issues du pas 2 de l'algorithme. Le système linéaire est résolu par la méthode de décomposition SVD.

6.3.5 Auto-calibrage planaire basé sur la contrainte des lignes centrales

L'équation de la ligne centrale (6.15) ne peut être déterminée que si les homographies scène-images sont connues. Si la structure métrique du plan observé est inconnue, il faut récupérer la rectification métrique du plan en même temps que les paramètres intrinsèques.

Gurdjos et Sturm [GS03] ont proposé une solution à ce problème. Le but est d'obtenir le point principal comme l'intersection de plusieurs lignes centrales mais, les matrices d'homographies scène-image étant inconnues, les contraintes associées ne correspondent pas à une droite unique mais à une famille de droites paramétrées par les coefficients de la matrice de homographie scène-image de la première image $\mathbf{H}_{01}$ (6.12). Leur approche consiste donc à chercher les paramètres de $\mathbf{H}_{01}$ qui permettent la concurrence des lignes centrales et d'estimer simultanément la position du point principal par l'intersection de ces lignes.

Par la suite, nous passons en revue les propriétés et les éléments utilisés pour la construction des équations d'auto-calibrage basées sur la contrainte de la ligne centrale, suivi de la description de la méthode.

Relation entre la ligne centrale et la contrainte des points de fuite

Il existe une relation entre la contrainte des points de fuite (section 6.2) et la contrainte de la ligne centrale, basée sur le fait que la contrainte des points de fuite est issue de la condition d'orthogonalité de la matrice de rotation des paramètres extrinsèques (6.1) et que cette condition correspond aussi à la première des équations (6.7) dont la contrainte de la ligne centrale a été dérivée.

L'une des configurations critiques de la contrainte des points de fuite et celle dont l'un des points de fuite se trouve à l'infini. L'équation correspondant (6.4) est exprimée uniquement en fonction des

coordonnées du point principal et du rapport d'aspect. Étant donnée une matrice d'homographie $\mathbf{H}'$, cette même équation peut être obtenue à partir de la première des contraintes (6.10), si le coefficient qui multiplie ω_{33} est nul ($h'_{31}h'_{32} = 0$).

L'équation (6.14) est équivalente à la première contrainte dans (6.10), si les deux premières colonnes de la matrice $\mathbf{H}'$ sont :

$$\begin{pmatrix} h_{11}h_{31} + h_{12}h_{32} & h_{11}h_{32} - h_{12}h_{31} \\ h_{21}h_{31} + h_{22}h_{32} & h_{21}h_{32} - h_{22}h_{31} \\ h_{31}^2 + h_{32}^2 & 0 \end{pmatrix}$$

Cela est valable si $\mathbf{H}'$ est construite (d'après [GCP02]) comme :

$$\mathbf{H}' = \begin{pmatrix} h'_{11} & h'_{12} & h'_{13} \\ h'_{21} & h'_{22} & h'_{23} \\ h'_{31} & 0 & h'_{33} \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} h_{31} & h_{32} & 0 \\ h_{32} & -h_{31} & 0 \\ 0 & 0 & 1 \end{pmatrix} = \mathbf{HS}$$

pour n'importe quelle matrice d'homographie $\mathbf{H}$, où $\mathbf{S}$ est conçue en fonction des coefficients h_{31} et h_{32} de $\mathbf{H}$ ayant la propriété d'être une transformation de similarité (rotation et changement d'échelle) du système de coordonnées de référence sur le plan Π.

Du point de vue géométrique, la matrice $\mathbf{S}$ a la fonction de transformer le système de coordonnées original de telle manière que l'un des points de fuite, correspondant aux axes principaux du nouveau repère, se positionne à l'infini. De cette manière, la nouvelle équation permet le découplage entre $\{u_0, v_0, \tau\}$ et la distance focale :

$$\tau^2 h'_{12}\left(h'_{11} - h'_{31}u_0\right) + h'_{22}\left(h'_{21} - h'_{31}v_0\right) = 0 \tag{6.19}$$

et cela correspond à l'ensemble de paramètres $\Phi = \{-h'_{12}h'_{31}, -h'_{22}h'_{31}, h'_{12}h'_{11}, h'_{22}h'_{21}\}$ de l'équation de la ligne centrale (6.15).

Si le plan observé est fronto-parallèle au plan image, les deux points de fuite orthogonaux se trouvent à l'infini (voir section 6.2, *Configurations critiques*) et il n'existe pas donc un changement des coordonnées sur le plan Π qui nous permet d'établir une contrainte en termes du point principal.

En outre qu'une interprétation géométrique de la contrainte, cette identité ($\mathbf{H}' = \mathbf{HS}$) a été introduite en vue de simplifier les expressions algébriques des éléments présentés ci-dessous.

Sommet du faisceau de lignes centrales

La contrainte de la ligne centrale (6.15) peut être réécrite comme :

$$\left[\begin{pmatrix} 0 \\ \varphi_2 \\ \varphi_4 \end{pmatrix} + \tau^2 \begin{pmatrix} \varphi_1 \\ 0 \\ \varphi_3 \end{pmatrix}\right] \cdot \begin{pmatrix} u_0 \\ v_0 \\ 1 \end{pmatrix} = \left(\mathbf{d}_1 + \tau^2 \mathbf{d}_2\right) \cdot \tilde{\mathbf{p}}_0 = 0 \tag{6.20}$$

où $\tilde{\mathbf{p}}_0 = (u_0, v_0, 1)^T$ est la position du point principal en coordonnées homogènes, et d'après (6.19) :

$$\mathbf{d}_1 = \begin{pmatrix} 0 \\ \varphi_2 \\ \varphi_4 \end{pmatrix} = -h'_{22} \begin{pmatrix} 0 \\ h'_{31} \\ -h'_{21} \end{pmatrix} = -h'_{22}\left(\mathbf{h}'_1 \times \mathbf{e}_1\right) \qquad \mathbf{d}_2 = \begin{pmatrix} \varphi_1 \\ 0 \\ \varphi_3 \end{pmatrix} = h'_{12} \begin{pmatrix} -h'_{31} \\ 0 \\ h'_{11} \end{pmatrix} = h'_{12}\left(\mathbf{h}'_1 \times \mathbf{e}_2\right)$$

$\mathbf{h}'_i$ étant la i-ème colonne de la matrice $\mathbf{H}'$.

Le vecteur $\mathbf{d}_1 + \tau^2\mathbf{d}_2$ définit un faisceau de droites associé à la matrice $\mathbf{H}'$ ($= \mathbf{HS}$), paramétré par le rapport d'aspect τ. Alors, le point $\mathbf{p}_s$ qui satisfait les conditions $\mathbf{d}_1 \cdot \mathbf{p}_s = 0$ et $\mathbf{d}_2 \cdot \mathbf{p}_s = 0$, appartient à toute droite du faisceau $(\mathbf{d}_1 + \tau^2\mathbf{d}_2) \cdot \mathbf{x} = 0$. Le point $\mathbf{p}_s$, nommé *le sommet du faisceau de lignes centrales*, est donc orthogonal aux vecteurs $\mathbf{d}_1$ et $\mathbf{d}_2$ et déterminé par :

$$\mathbf{p}_s \sim \mathbf{d}_1 \times \mathbf{d}_2 \sim \mathbf{h}'_1$$

En fonction de la matrice d'homographie $\mathbf{H}$, la première colonne de la matrice $\mathbf{H}'$ correspond à :

$$\mathbf{h}'_1 = \mathbf{H}'\mathbf{e}_1 = \mathbf{HSe}_1 = \mathbf{H} \begin{pmatrix} h_{31} \\ h_{32} \\ 0 \end{pmatrix} = \mathbf{H} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} h_{31} \\ h_{32} \\ h_{33} \end{pmatrix} = \mathbf{H} \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \mathbf{H}^T\mathbf{e}_3$$

Dans leur explication, [GS03] utilisent une entité géométrique complexe nommée *conique duale des points cycliques* C^*_∞. Cette conique a la propriété de définir la famille de droites tangentielles à la conique absolue passant par les points cycliques et possède la forme $C^*_\infty \sim diag(1,1,0)$. La projection $\mathbf{\Sigma}$ de cette conique, lors d'une transformation homographique $\mathbf{H}$ entre le plan de la scène et l'image est $\mathbf{\Sigma} = \mathbf{H}C^*_\infty\mathbf{H}^T$. De cette manière, le sommet du faisceau de lignes centrales peut être exprimé par :

$$\mathbf{p}_s \sim \mathbf{\Sigma}\mathbf{e}_3$$

Étant donné que la ligne de fuite associé à l'homographie scène-image $\mathbf{H}$ a pour paramètres $\mathbf{l}_\infty = \mathbf{H}^{-T}\mathbf{e}_3$, on vérifie que :

$$\mathbf{l}^T_\infty\mathbf{p}_s = \left(\mathbf{e}^T_3\mathbf{H}^{-1}\right)\left(\mathbf{H}C^*_\infty\mathbf{H}^T\mathbf{e}_3\right) = \left(\mathbf{e}^T_3C^*_\infty\right)\mathbf{H}^T\mathbf{e}_3 = \mathbf{0}^T\mathbf{H}^T\mathbf{e}_3 = 0$$

ce qui signifie que $\mathbf{p}_s$ appartient à la ligne de fuite $\mathbf{l}_\infty$ du plan Π (figure 6.5), c'est à dire, $\mathbf{p}_s$ est l'intersection entre la ligne de fuite et la ligne centrale.

Pôle de la ligne de fuite

Les paramètres de la ligne de fuite, en terme des coefficients de $\mathbf{H}$, sont déterminés par :

$$\mathbf{l}_\infty \sim \mathbf{H}^{-T}\mathbf{e}_3 \sim \begin{pmatrix} h_{21}h_{32} - h_{22}h_{31} \\ -h_{11}h_{32} + h_{12}h_{31} \\ h_{11}h_{22} - h_{12}h_{21} \end{pmatrix}$$

Les deux premières composantes de ce vecteur sont semblables aux paramètres φ_1 et φ_2 de la ligne centrale (éqs. 6.15 et 6.16) à un facteur près. Ces deux paramètres déterminent la pente de la ligne centrale.

En fait, une façon de préciser une droite consiste à fournir deux points appartenant à la droite. Si l'un de ces points est l'intersection entre la droite concerné et la droite à l'infini (c'est à dire, l'intersection de toutes les droites parallèles à celle-ci), ce point-ci caractérise la pente de la droite.

D'après les équations (6.20) et (6.16), l'intersection entre la ligne centrale $\mathbf{l}_c = \mathbf{d}_1 + \tau^2\mathbf{d}_2$ et la droite à l'infini est le point $\mathbf{p}_l$:

$$\mathbf{p}_l \sim \mathbf{l}_c \times \mathbf{e}_3 \sim \begin{pmatrix} \tau^2\varphi_1 \\ \varphi_2 \\ \tau^2\varphi_3 + \varphi_4 \end{pmatrix} \times \mathbf{e}_3 = \begin{pmatrix} \varphi_2 \\ -\tau^2\varphi_1 \\ 0 \end{pmatrix} \sim \begin{pmatrix} h_{21}h_{32} - h_{22}h_{31} \\ \tau^2\left(-h_{11}h_{32} + h_{12}h_{31}\right) \\ 0 \end{pmatrix}$$

Comme la géométrie intrinsèque de la caméra transforme la conique duale des points cycliques C^*_∞ en $\Lambda = \mathbf{K}C^*_\infty\mathbf{K}^T \sim diag(1,\tau^2,0)$, alors, la position de $\mathbf{p}_l$ peut s'exprimer comme :

$$\mathbf{p}_l \sim \Lambda\mathbf{l}_\infty$$

Toute droite $\mathbf{l}$ contenant le point $\mathbf{p}_l$ satisfait donc $\mathbf{l}\cdot\mathbf{p}_l = \mathbf{l}^T\Lambda\mathbf{l}_\infty = 0$, et cela représente la famille de droites orthogonales à la ligne de fuite par rapport à la géométrie intrinsèque $\mathbf{K}$ de la caméra. Alors, le point $\mathbf{p}_l$ est l'intersection de toutes les droites orthogonales à la ligne de fuite par rapport à $\mathbf{K}$, dit *le pôle de la ligne de fuite.*

Contrainte de calibrage

En fonction des éléments décrits ci-dessus, la contrainte de calibrage peut être précisée de la manière suivante : la ligne centrale $\mathbf{l}_c$ contient le point principal $\tilde{\mathbf{p}}_0$ et passe par le sommet $\mathbf{p}_s$ du faisceau de droites $(\mathbf{d}_1 + \tau^2\mathbf{d}_2)\cdot\mathbf{x} = 0$, c'est à dire, $\mathbf{l}_c \sim \mathbf{p}_s \times \tilde{\mathbf{p}}_0$, et elle est orthogonale à la ligne de fuite $\mathbf{l}_\infty$ (figure 6.5) :

$$\mathbf{l}^T_c\Lambda\mathbf{l}_\infty = (\mathbf{p}_s \times \tilde{\mathbf{p}}_0)\cdot\mathbf{p}_l = 0$$

On peut dire également que le sommet du faisceau de droites, le point principal et le pôle de la ligne de fuite sont colinéaires.

Soit $\mathbf{M}$ la matrice dont les colonnes sont les points $\mathbf{p}_s$, $\mathbf{p}_l$ et $\tilde{\mathbf{p}}_0$. La nullité du déterminant de $\mathbf{M}$ indique que les trois points sont colinéaires et la contrainte de calibrage peut être donc écrite comme :

$$\det(\mathbf{M}) = \det(\mathbf{p}_s \mid \mathbf{p}_l \mid \tilde{\mathbf{p}}_0) = 0 \tag{6.21}$$

Contrainte de la ligne centrale à partir des homographies inter-vues

Les données d'entrée pour la méthode d'auto-calibrage planaire basée sur la contrainte de la ligne centrale sont les matrices d'homographies entre les images acquises. Ces homographies sont calculées par rapport à une image clé (dans notre représentation, nous l'appelons l'image 1) : $\mathbf{H}_{1i}$ est l'homographie qui relie les projections correspondantes entre l'image 1 et l'image i des points de la scène appartenant à un plan Π.

La matrice $\mathbf{H}_{01}$ est l'homographie entre le plan de la scène et sa projection sur la image clé, de telle manière que l'homographie $\mathbf{H}_{0i}$ entre le plan de la scène et la image i est obtenue comme $\mathbf{H}_{0i} = \mathbf{H}_{1i}\mathbf{H}_{01}$.

Le sommet du faisceau des lignes centrales associé à l'image i est donc représenté par :

$$\mathbf{p}_{s,i} \sim \mathbf{H}_{0i} C_\infty^* \mathbf{H}_{0i}^T \mathbf{e}_3 = \mathbf{H}_{1i} \left(\mathbf{H}_{01} C_\infty^* \mathbf{H}_{01}^T \right) \mathbf{H}_{1i}^T \mathbf{e}_3 = \mathbf{H}_{1i} \mathbf{\Sigma}_1 \mathbf{H}_{1i}^T \mathbf{e}_3$$

où $\mathbf{\Sigma}_1$ est l'image de la conique duale des points cycliques pour la première image.

Le pôle de la ligne de fuite correspondant à l'image i est :

$$\mathbf{p}_{l,i} \sim \Lambda \mathbf{l}_{\infty,i} = \Lambda \mathbf{H}_{0i}^{-T} \mathbf{e}_3 = \Lambda \mathbf{H}_{1i}^{-T} \left(\mathbf{H}_{01}^{-T} \mathbf{e}_3 \right) = \Lambda \mathbf{H}_{1i}^{-T} \mathbf{l}_{\infty,1}$$

où $\mathbf{l}_{\infty,1}$ est la ligne de fuite du plan Π à l'image clé.

En utilisant la représentation (6.12) de la matrice $\mathbf{H}_{01}$, on obtient :

$$\begin{aligned} \mathbf{\Sigma}_1 &\sim \begin{pmatrix} \alpha^2+\beta^2 & \alpha & \mu\left(\alpha^2+\beta^2\right)+\alpha\lambda \\ \alpha & 1 & \mu\alpha+\lambda \\ \mu\left(\alpha^2+\beta^2\right)+\alpha\lambda & \mu\alpha+\lambda & \mu^2\beta^2+(\mu\alpha+\lambda)^2 \end{pmatrix} \\ \mathbf{l}_{\infty,1} &\sim \begin{pmatrix} -\mu & -\lambda & 1 \end{pmatrix}^T \end{aligned}$$

Alors, la contrainte de calibrage (6.21) prend la forme :

$$\begin{gathered} \det\left(\mathbf{M}_i\right) = 0 \\ \mathbf{M}_i = \left(\mathbf{H}_{1i} \mathbf{\Sigma}_1 \mathbf{H}_{1i}^T \mathbf{e}_3 \mid \Lambda \mathbf{H}_{1i}^{-T} \mathbf{l}_{\infty,1} \mid \tilde{\mathbf{p}}_0 \right) \end{gathered} \tag{6.22}$$

La méthode de Gurdjos-Sturm

Le problème d'auto-calibrage planaire en utilisant la contrainte de la ligne centrale consiste donc à obtenir les coordonnées du point principal et les coefficients de la matrice $\mathbf{H}_{01}$ qui satisfont la restriction (6.22) pour toutes les images contenant des projections du plan de calibrage. Une solution à ce problème consiste à minimiser la somme des carrés des déterminants (6.22) sur l'ensemble de vues, c'est à dire, chercher le minima de la fonction : $\sum_{i=1}^n \left(\det(\mathbf{M}_i)\right)^2$.

Au lieu d'utiliser le déterminant de $\mathbf{M}$ comme critère de pénalisation de la contrainte de calibrage, [GS03] proposent d'employer la distance euclidienne du point principal à la ligne centrale comme mesure d'erreur, d'une forme analogue à celle proposée pour la méthode de lignes centrales (équation 6.17) lorsque les positions des points d'intérêt sur le plan de calibrage sont connues (section 6.3.4) :

La distance d'un point $\tilde{\mathbf{p}} = (x, y, 1)^T$ à la droite $\mathbf{l}_i = \mathbf{p}_{s,i} \times \mathbf{p}_{l,i}$ est déterminée par :

$$d(\tilde{\mathbf{p}}, \mathbf{l}_i) = \frac{|\mathbf{l}_i \cdot \tilde{\mathbf{p}}|}{\sqrt{\mathbf{l}_i^T \Lambda \mathbf{l}_i}}$$

et la fonction d'erreur suggérée est donc :

$$e = \sum_{i=1}^n \left(\frac{\mathbf{l}_i \cdot \tilde{\mathbf{p}}}{\sqrt{\mathbf{l}_i^T \Lambda \mathbf{l}_i}} \right)^2 = \sum_{i=1}^n \frac{\left(\det\left(\mathbf{M}_i\right)\right)^2}{\mathbf{l}_i^T \Lambda \mathbf{l}_i} \tag{6.23}$$

de telle manière que :

$$\left\{ \tilde{\mathbf{p}}_0, \hat{\alpha}, \hat{\beta}, \hat{\mu}, \hat{\lambda} \right\} = \arg \min_{\{\tilde{\mathbf{p}}, \alpha, \beta, \mu, \lambda\}} \sum_{i=1}^n \frac{\left(\left(\mathbf{p}_{s,i} \times \mathbf{p}_{l,i} \right) \cdot \tilde{\mathbf{p}} \right)^2}{\left(\mathbf{p}_{s,i} \times \mathbf{p}_{l,i} \right)^T \Lambda \left(\mathbf{p}_{s,i} \times \mathbf{p}_{l,i} \right)}$$

Comme les homographies sont estimées à un facteur près, dans l'emploi de det ($\mathbf{M}$) comme critère de pénalisation, la mesure d'erreur peut varier en fonction de l'échelle choisie pour représenter les homographies, étant donné que la pénalisation de la contrainte devrait être invariante car les homographies sont en principe les mêmes. Dans ce cas-ci, la procédure habituelle consiste à normaliser les homographies inter-vues de telle manière que $\|\mathbf{H}_{1i}\|_F = 1$.

En contrepartie, la distance euclidienne entre le point principal et la ligne centrale établit une interprétation géométrique du critère de pénalisation et permet par ailleurs de balancer les contraintes de telle façon que la fonction d'erreur soit invariante vis à vis des facteurs d'échelle sur les homographies.

Algorithme

Description de la méthode de Gurdjos et Sturm [GS03] :

- Obtention des n matrices d'homographie entre les images et l'image clé : $\{\mathbf{H}_{1i}\}$.
- Spécification de la fonction d'erreur (6.23), basée sur la colinéarité du sommet du faisceau des lignes centrales, le pôle de la ligne de fuite, et le point principal de la caméra.
- Obtention des paramètres intrinsèques de la caméra $\{u_0, v_0, \tau\}$ et les paramètres de l'homographie scène-image associée à l'image clé $\{\alpha, \beta, \mu, \lambda\}$ par la minimisation de la fonction d'erreur citée ci-dessus.

L'algorithme d'estimation non-linéaire employée par [GS03] et dans nos expériences est la méthode de Levenberg-Marquardt [PFTV88]. Les valeurs initiales requises pour cette approche sont dans le cas des paramètres intrinsèques : $\tau = 1$, et (u_0, v_0) au centre de l'image. Dans le cas de la matrice $\mathbf{H}_{01}$, une connaissance approximative de ses paramètres est nécessaire afin d'éviter la convergence vers des minima locaux. À ce sujet, une disposition initiale du plan de calibrage fronto-parallèle au plan de l'image permet une estimation approchée directe : $\beta = 1$, $\alpha = \mu = \lambda = 0$.

Chapitre 7

Extension et évaluation des méthodes de calibrage interne

Dans ce chapitre, nous présentons nos contributions aux techniques de calibrage :

- Nous proposons l'emploi de méthodes statistiques pour pondérer les contraintes de calibrage, en fonction de la précision de détection de primitives (section 7.1).
- Nous analysons les contraintes de calibrage qui peuvent être obtenues à partir de plusieurs plans de la scène observés dans une seule image ou dans une séquence d'images, envisageant l'extension des méthodes de calibrage planaire interne au cas où plusieurs plans sont observés dans les images (section 7.2).

En section 7.3, nous proposons une démarche expérimentale pour comparer les méthodes de calibrage interne planaire abordées dans le chapitre 6 et les développements que nous suggérons dans ce chapitre, et montrons les résultats issus de nos expériences évaluant la fiabilité et la précision des différentes méthodes de calibrage présentées (section 7.4).

Les scénarios proposés pour notre étude correspondent à l'emploi des moyens matériels suivants : un motif simple de calibrage, une ou plusieurs surfaces planes texturées.

7.1 Pondération statistique des contraintes de calibrage

La plupart des méthodes de calibrage s'appuient sur la minimisation de fonctions d'erreurs résiduelles. Fréquemment, les fonctions à minimiser sont formées par des erreurs algébriques pénalisant le non-respect de la contrainte de calibrage. Parmi ces fonctions, la somme des carrés des erreurs résiduelles est la plus utilisée. Le but de ces méthodes est donc d'estimer les paramètres Θ qui minimisent une fonction d'erreur résiduelle de la forme :

$$e\left(\Theta, \{\mathbf{x}_i\}\right) = \sum_{i=1}^{n} d^2\left(\Theta, \mathbf{x}_i\right) \tag{7.1}$$

où $\mathbf{x}_i$ sont les données d'entrée, et $d(\Theta, \mathbf{x}_i)$ mesure l'erreur sur la contrainte de calibrage (voir tableau 7.1).

Conditionnement d'un système linéaire

La détermination des paramètres intrinsèques par la méthode de Sturm-Maybank et la méthode des lignes centrales se ramène donc à la résolution d'un système d'équations linéaires homogènes qu'on représente par $\mathbf{A}\Theta = 0$.

Il arrive parfois qu'une petite erreur sur les coefficients de la matrice $\mathbf{A}$ entraîne une grande variation sur les paramètres calculés Θ. À ce sujet, Golub [Gol89] propose une stratégie pour réduire le conditionnement numérique d'un système linéaire. Ce pre-conditionnement consiste à normaliser les vecteurs

Méthode	Paramètres Θ	Contraintes	Erreur résiduelle $d(\Theta, \mathbf{x}_i)$
Sturm-Maybank (section 6.3.2)	$\mathbf{w} = \begin{pmatrix} \omega_{11} \\ \omega_{13} \\ \omega_{22} \\ \omega_{23} \\ \omega_{33} \end{pmatrix}$	$\begin{pmatrix} \mathbf{a}^T \\ \mathbf{b}^T \end{pmatrix} \mathbf{w} = \mathbf{0}$	$\left\| \begin{pmatrix} \mathbf{a} \cdot \mathbf{w} \\ \mathbf{b} \cdot \mathbf{w} \end{pmatrix} \right\|$
Des lignes centrales (section 6.3.4)	$\{u_0, v_0, \tau\}$	$\begin{pmatrix} \tau^2\varphi_1 \\ \varphi_2 \\ \tau^2\varphi_3 + \varphi_4 \end{pmatrix} \cdot \begin{pmatrix} u_0 \\ v_0 \\ 1 \end{pmatrix} = 0$	$\frac{\lvert\varphi_1(\tau^2 u_0)+\varphi_2 v_0+(\varphi_3\tau^2+\varphi_4)\rvert}{\sqrt{\varphi_1^2+\varphi_2^2}}$
Des lignes centrales et rectification du plan (section 6.3.5)	$\{u_0, v_0, \tau\}$ $\{\alpha, \beta, \mu, \lambda\}$	$(\mathbf{p}_s \times \mathbf{p}_l) \cdot \begin{pmatrix} u_0 \\ v_0 \\ 1 \end{pmatrix} = 0$	$\frac{(\mathbf{p}_s \times \mathbf{p}_l) \cdot \tilde{\mathbf{p}}_0}{\sqrt{(\mathbf{p}_s \times \mathbf{p}_l)^T \Lambda (\mathbf{p}_s \times \mathbf{p}_l)}}$

TAB. 7.1 – Contraintes et erreurs résiduelles pour les méthodes de calibrage.

colonnes de la matrice $\mathbf{A}$ afin d'obtenir un nouveau système linéaire $\mathbf{A}'\Theta' = 0$ avec un coefficient de conditionnement numérique plus petit que pour le système original, ce qui rend la solution plus stable aux variations des coefficients de $\mathbf{A}$.

Le pre-conditionnement adéquat du système d'équations est une procédure qui permet la réduction du problème de stabilité par rapport aux erreurs sur les coefficients de la matrice du système, mais ce n'est pas une approche toujours efficace, puisque :

- Cette stratégie considère que les perturbations sur les coefficients de $\mathbf{A}$ sont homogènes et indépendants, ce qui n'est généralement pas le cas pour des systèmes issus des contraintes de calibrage abordées.
- Le pre-conditionnement algébrique ne prend pas en compte les distributions des erreurs de détection des points d'intérêt, ni la manière dont ces erreurs se propagent vers la détermination d'autres entités (e.g. les homographies, les équations des lignes centrales). Par exemple, la variabilité dans l'estimation des coefficients d'une matrice d'homographie dépend de plusieurs facteurs, tels que le nombre de correspondances observées entre les images, de leur position et leur distribution sur l'image, etc.

L'estimation des paramètres par la minimisation de la fonction (7.1) n'est donc pas statistiquement optimale, et si on l'utilise arbitrairement, les erreurs de détection des primitives observées peuvent entraîner des erreurs significatives dans les résultats ou des problèmes de délimitation du domaine de convergence de la solution.

Propagation d'erreurs et pondération statistique

Comme la méthode de propagation d'erreurs (section 2.5) nous permet de prendre en compte les incertitudes des données d'entrée pour évaluer la variabilité des résultats obtenus, nous proposons d'intégrer cette méthode aux contraintes de calibrage de chacune des méthodes présentées dans ce chapitre.

Les contraintes de la conique absolue (6.7) et des lignes centrales (6.15) sont linéaires dans les paramètres à estimer (ω et le point principal, respectivement). En revanche, la contrainte d'auto-calibrage des lignes centrales (6.21) est linéaire par rapport aux coordonnées du point principal $\tilde{\mathbf{p}}_0$, mais elle est non-linéaire par rapport aux quatre paramètres déterminant l'homographie de la première image ($\{\alpha, \beta, \mu, \lambda\}$) qui sont des variables à estimer aussi pour ce problème et contenues dans la spécification de la droite centrale ($\mathbf{p}_s \times \mathbf{p}_l$).

En résumé, les trois contraintes de calibrage se basent sur l'un de ces principes :

- l'incidence de la solution sur un ensemble d'équations (6.7),
- la concurrence d'un ensemble de droites (6.15, 6.21).

Ces contraintes peuvent être exprimées par la forme commune $\overline{\mathbf{q}} \cdot \overline{\Theta} = 0$, où $\overline{\mathbf{q}}$ sont les données d'entrée en absence du bruit et $\overline{\Theta}$ sont les vrais paramètres du modèle (voir tableau 7.1). Lorsque des données

bruitées ($\mathbf{q}_i = \overline{\mathbf{q}}_i + \Delta\mathbf{q}_i$) et un ensemble de paramètres ($\Theta = \overline{\Theta} + \Delta\Theta$) sont connus, l'erreur sur la contrainte est :

$$\begin{aligned} \epsilon_i = \Theta \cdot \mathbf{q}_i &= \Theta \cdot (\overline{\mathbf{q}}_i + \Delta\mathbf{q}_i) \\ &= \Theta \cdot \overline{\mathbf{q}}_i + \Theta \cdot \Delta\mathbf{q}_i \\ &= \Theta \cdot \Delta\mathbf{q}_i \end{aligned}$$

La variance des erreurs résiduelles ϵ_i est donc donnée par :

$$\begin{aligned} \sigma_i^2 &= E[\epsilon_i^2] \\ &= E[\Theta^T \Delta\mathbf{q}_i \Delta\mathbf{q}_i^T \Theta] \\ &= \Theta^T V[\mathbf{q}_i]\Theta \end{aligned} \tag{7.2}$$

Les variables aléatoires $\rho_i = \epsilon_i/\sigma_i$ suivent approximativement une distribution de moyenne 0 et d'écart-type unitaire. En supposant que les variables sont indépendantes et suivent une loi gaussienne, l'estimation du maximum de vraisemblance de la probabilité conjointe des variables ρ_i est associée à la minimisation de la fonction suivante :

$$\sum_{i=1}^{n} \left(\frac{\epsilon_i}{\sigma_i}\right)^2 = \sum_{i=1}^{n} \frac{(\Theta \cdot \mathbf{q}_i)^2}{\Theta^T V[\mathbf{q}_i]\Theta} \tag{7.3}$$

La variance de l'erreur résiduelle apporte une mesure de la fiabilité d'employer une donnée particulière (e.g. les primitives sur un plan donné, ou dans une image) dans les processus de calibrage, et nous permettent de différencier, par exemple, les meilleures dispositions entre la caméra et la scène pour effectuer le calibrage, ou comparer la qualité des estimations faites en utilisant des surfaces différentes.

Moindres carrés itérés repondérés

Comme les paramètres à estimer Θ se trouvent à l'intérieur du dénominateur ($\Theta^T V[\mathbf{q}_i]\Theta$), la méthode d'optimisation à utiliser doit être non-linéaire. Si le dénominateur était considéré constant (z_i), la fonction à minimiser serait de la forme :

$$\sum_{i=1}^{n} \frac{1}{z_i} (\Theta \cdot \mathbf{q}_i)^2 = \Theta^T \left(\sum_{i=1}^{n} \frac{1}{z_i} \mathbf{q}_i \mathbf{q}_i^T \right) \Theta$$

Dans ce cas, les paramètres Θ qui minimisent la fonction correspondent au vecteur propre associé à la valeur propre la plus petite de la matrice :

$$\mathbf{Q} = \sum_{i=1}^{n} \left(\frac{1}{z_i} \mathbf{q}_i \mathbf{q}_i^T \right)$$

Afin d'éviter l'utilisation de méthodes d'optimisation non-linéaires, il existe des approches [Kan93a] de solution basées sur le principe de considérer les quotients comme constants, de telle manière à obtenir une pondération itérative, où les quotients z_i sont traités comme constants à chaque itération et mis à jour avant la prochaine itération, jusqu'à la convergence de la solution. Les solutions intermédiaires des paramètres sont trouvées par la décomposition SVD de la matrice $\mathbf{Q}$ à chaque itération.

Triggs [Tri98a] montre que les estimations effectuées avec ce principe correspondent à une estimation de première ordre des valeurs des paramètres. La plupart du temps, la différence n'est pas significative sauf si la fonction présente des variations fortes dans l'espace des paramètres.

Données requises pour l'estimation des coefficients de pondération

Les données d'entrée $\mathbf{q}_i$ sont déterminées : par les vecteurs $\mathbf{a}$ et $\mathbf{b}$, pour la méthode de Sturm-Maybank ; par le vecteur $\Phi = (\phi_1, \phi_2, \phi_3, \phi_4)$, pour la méthode des lignes centrales ; et par les vecteurs $\mathbf{p}_s$ et $\mathbf{p}_l$, pour la méthode des lignes centrales avec rectification du plan.

Pour les deux premières méthodes, les entrées dépendent uniquement des homographies scène-image et sont calculées dès le début de l'optimisation de la fonction d'erreur résiduelle. De la même manière, nous allons calculer les covariances de $\mathbf{q}_i$ dès le début car elles sont constantes, et peuvent être estimées à partir des homographies scène-images et leurs covariances correspondantes.

Ainsi, pour les deux premières méthodes, nous avons besoin d'estimer d'abord les homographies scène-image $\mathbf{H}_{0}i$ et les covariances correspondantes $Cov[\mathbf{H}_{0i}]$.

La stratégie à adopter pour la méthode des lignes centrales avec rectification du plan doit être différente, car les vecteurs $\mathbf{p}_s$ $(= (\mathbf{H}_{1i}\mathbf{H}_{01})\, diag(1,1,0)\, (\mathbf{H}_{01}^T\mathbf{H}_{1i}^T)\, \mathbf{e}_3)$ et $\mathbf{p}_l$ $(= diag(1,\tau^2,0)\mathbf{H}_{1i}^{-T}\mathbf{v})$ ne dépendent pas seulement des homographies inter-images. Ils dépendent aussi des paramètres de rectification du plan pour la première image, c'est à dire, de la matrice $\mathbf{H}_{01}$ qui est aussi l'un des ensembles de variables à estimer dans le processus de calibrage.

Les détails de l'implémentation de la pondération statistique pour chaque méthode sont présentés par la suite. Avant cela, nous faisons une brève revue de l'estimation de la matrice de covariances d'une transformation homographique.

7.1.1 Pondération des contraintes associées à l'image de la conique absolue

Pour les contraintes de la conique absolue, la matrice $\mathbf{Q}$ associée à la fonction d'erreur résiduelle est :

$$\mathbf{Q} = \sum_{i=1}^{n} \left(\frac{\mathbf{a}_i^T\mathbf{a}_i}{z(\mathbf{a}_i)} + \frac{\mathbf{b}_i^T\mathbf{b}_i}{z(\mathbf{b}_i)} \right)$$

où $z(\mathbf{a}_i)$ et $z(\mathbf{b}_i)$, sont les variances estimées de $(\mathbf{a}_i \cdot \mathbf{w})$ et $(\mathbf{b}_i \cdot \mathbf{w})$.

Ceci est équivalent à construire une matrice diagonale $\mathbf{D} = diag\left((z(\mathbf{a}_1))^{-1/2}, (z(\mathbf{b}_1))^{-1/2}, \cdots \right)$ qui multiplie à gauche la matrice $\mathbf{A}$ du système homogène original (éq. 6.11) afin d'obtenir une matrice de système $\mathbf{A}' = \mathbf{DA}$. De cette manière, la matrice associée à la fonction d'erreur est $\mathbf{Q} = \mathbf{A^T D^2 A} = \mathbf{A'^T A'}$.

En réalité, le vecteur propre recherché sur $\mathbf{Q}$ pour obtenir les solutions peut être obtenu sans construire cette matrice, car ce vecteur peut être acquis directement par décomposition SVD de la matrice $\mathbf{A}'$.

Il reste à déterminer les coefficients de la matrice diagonale $\mathbf{D}$. Par l'équation (7.2), $z(\mathbf{a}_i) = \mathbf{w}^T Cov[\mathbf{a}_i]\mathbf{w}$ et $z(\mathbf{b}_i) = \mathbf{w}^T Cov[\mathbf{b}_i]\mathbf{w}$. Comme les vecteurs $\mathbf{a}_i$ et $\mathbf{b}_i$ sont exclusivement fonction de l'homographie scène-image $\mathbf{H}_{0i}$, leurs matrices de covariances peuvent être estimées par la méthode de propagation directe d'incertitudes de la façon suivante :

$$Cov[\mathbf{a}_i] \approx \mathbf{J_a} Cov[\mathbf{H}_{0i}] \mathbf{J_a}^T$$
$$Cov[\mathbf{b}_i] \approx \mathbf{J_b} Cov[\mathbf{H}_{0i}] \mathbf{J_b}^T$$

où $\mathbf{J_a}$ et $\mathbf{J_b}$ sont des matrices 5×9 qui correspondent au jacobien des vecteurs $\mathbf{a}$ (ou $\mathbf{b}$) par rapport aux coefficients de l'homographie ; c'est à dire :

$$\mathbf{J_a} = \begin{pmatrix} \frac{\partial a_1}{\partial h_{11}} & \cdots & \frac{\partial a_1}{\partial h_{33}} \\ \vdots & \ddots & \vdots \\ \frac{\partial a_5}{\partial h_{11}} & \cdots & \frac{\partial a_5}{\partial h_{33}} \end{pmatrix}$$

7.1.2 Pondération des contraintes correspondant aux lignes centrales

Pour la contrainte de la ligne centrale, chaque droite est associée à une homographie scène-image $\mathbf{H}_{0i}$ par le vecteur 3D $\mathbf{l}_i = (\tau^2\varphi_1^{(i)}, \varphi_2^{(i)}, \tau^2\varphi_3^{(i)} + \varphi_4^{(i)})$, de telle manière qu'on cherche le point principal $\mathbf{p_0}$ satisfaisant les contraintes $\mathbf{l}_i \cdot \mathbf{p_0} = 0$, pour tout $\mathbf{l}_i$.

La matrice $\mathbf{Q}$ associée à la fonction pondérée d'erreur résiduelle pour estimer l'intersection d'un ensemble de lignes droites est :

$$\mathbf{Q} = \sum_{i=1}^{n} \left(\frac{\mathbf{l}_i^T\mathbf{l}_i}{\tilde{\mathbf{p}}_0^T Cov[\mathbf{l}_i]\tilde{\mathbf{p}}_0} \right)$$

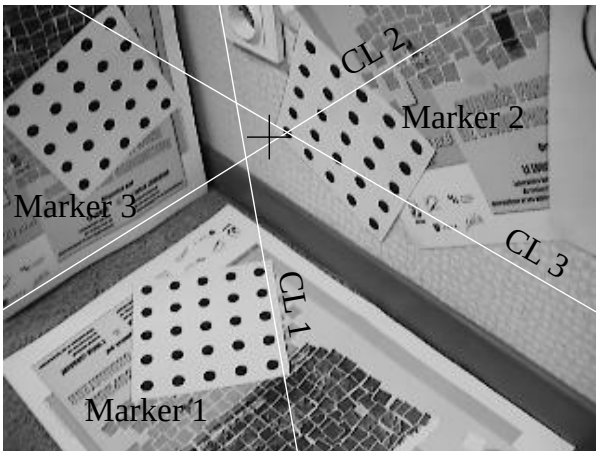


FIG. 7.1 – Point principal estimé par la méthode des lignes centrales pondérées.

La covariance des droites $\mathbf{l}_i$ est approchée par $Cov[\mathbf{l}_i] = \mathbf{J_l} Cov[\mathbf{H}_{0i}]\mathbf{J_l}^T$, où la matrice du jacobien adjoint à la droite est

$$\mathbf{J_l} = \begin{pmatrix} \tau^2 \frac{\partial \phi_1}{\partial h_{11}} & \cdots & \tau^2 \frac{\partial \phi_1}{\partial h_{33}} \\ \frac{\partial \phi_2}{\partial h_{11}} & \cdots & \frac{\partial \phi_2}{\partial h_{33}} \\ \tau^2 \frac{\partial \phi_3}{\partial h_{11}} + \frac{\partial \phi_4}{\partial h_{11}} & \cdots & \tau^2 \frac{\partial \phi_3}{\partial h_{33}} + \frac{\partial \phi_4}{\partial h_{33}} \end{pmatrix}$$

Lorsque le rapport d'aspect τ est connu, la matrice de covariance de chaque droite est calculée à la première itération et gardée constante par la suite. Cependant, cette contrainte nous permet de calculer simultanément les trois paramètres intrinsèques (u_0, v_0, τ). Dans cette situation, il serait nécessaire d'estimer à nouveau le jacobien de chaque droite $\mathbf{l}_i$ et sa matrice de covariances. Comme la valeur du rapport d'aspect est toujours un paramètre assez stable, et normalement près de 1, nous avons comparé les résultats issus de deux expériences :

- utilisant la valeur $\tau = 1$ pour le calcul du jacobien et des matrices de covariances, puis en gardant constants ces estimations ;
- en mettant à jour à chaque itération le jacobien et les covariances.

Nos résultats indiquent que les différences entre ces deux stratégies sont négligeables. La distance entre les points principaux estimés dans les deux expériences est toujours inférieure à une centième de pixel.

Comme le fondement géométrique de la méthode des lignes centrales est le calcul de l'intersection d'un ensemble de droites, la différence entre la méthode originale et la méthode par pondération statistique peut être illustré graphiquement (figure 7.1) de la manière suivante :

Pour le cas où une seule vue de trois plans est utilisée pour effectuer le calibrage, le point principal estimé par la méthode originale est le centre du cercle inscrit au triangle formé par les trois lignes centrales. Par contre, lorsque la méthode de pondération est employée, le point principal estimé se déplace vers les lignes centrales qui ont été calculées avec une meilleure précision.

7.1.3 Pondération pour la méthode d'auto-calibrage par lignes centrales

D'après l'expression (7.3), la version pondérée de la méthode d'auto-calibrage par lignes centrales confère une fonction d'erreur résiduelle de la forme :

$$\sum_{i=1}^{n} \frac{(\det(\mathbf{M}_i))^2}{J_{M_i} Cov[\mathbf{H}_{1i}] J_{M_i}^T} \tag{7.4}$$

où le jacobien du déterminant de chaque matrice $\mathbf{M}_i$ (6.22) en fonction des coefficients de l'homographie $\mathbf{H}_{1i}$ est $J_{M_i} = \left\{ \frac{d \det(\mathbf{M}_i)}{d \mathbf{H}_{1i}} \right\}$.

L'approche de solution que nous proposons est basée sur la technique de moindres carrés itérés (voir section 7.1). L'estimation du facteur de pondération $z_i = J_{M_i} Cov[\mathbf{H}_{1i}] J_{M_i}^T$ est effectué une fois par itération et gardé constant lors de l'optimisation non-linéaire de la nouvelle fonction : $\sum_{i=1}^{n} \frac{(\det(\mathbf{M}_i))^2}{z_i}$.

En raison de la forme polynomiale de cette fonction par rapport aux paramètres à estimer, la convergence nécessite entre 5 et 10 itérations de la méthode de Levenberg-Marquardt. Une fois qu'un minimum de cette fonction est obtenue, les dénominateurs z_i sont mis à jour, et la procédure d'optimisation est répétée sur la fonction actualisée.

De manière analogue à la méthode originale proposée par Gurdjos et Sturm [GS03], les résultats sont aussi assez sensibles aux paramètres initiaux des matrices de rectification.

7.2 Intégration de structures multi-planaires aux méthodes de calibrage interne

Puisque la plupart de méthodes existantes se focalisent sur le problème du calcul des paramètres intrinsèques en observant un seul plan sur plusieurs vues, nous avons décidé d'étendre certaines de ces méthodes à l'utilisation des plusieurs plans visibles sur une séquence d'images, et d'analyser la précision et la stabilité des résultats suite à ce développement.

Les méthodes de calibrage interne présentés dans ce chapitre se divisent en deux catégories :

- les méthodes de calibrage planaire (sections 6.3.2 et 6.3.4), où les données d'entrée sont des matrices scène-image, et dont le but est d'estimer seulement les paramètres intrinsèques de la caméra ;
- les méthodes de auto-calibrage planaire (sections 6.3.3 et 6.3.5), où les données d'entrée sont des matrices inter-images, et dont il est nécessaire d'estimer l'équation du plan (ou la matrice d'homographie) pour une image clé plus les paramètres intrinsèques de la caméra.

Comme les méthodes de calibrage planaire sont basées sur les homographies scène-images, leur extension à l'observation de plusieurs régions planes à chaque vue est immédiate : puisque chaque homographie qui ramène un ensemble de points coplanaires dans l'espace vers le plan image permet d'obtenir les contraintes (6.7) sur l'image de la conique absolue, indépendamment de la disposition du plan duquel l'homographie procède.

Pour ces méthodes, l'intégration de nouveaux plans de calibrage s'effectue de la façon suivante :

- Pour la méthode de Sturm-Maybank, il suffit d'ajouter les deux contraintes de calibrage de la conique absolue, représentées par $\mathbf{a}$ et $\mathbf{b}$, de la même manière qu'on le fait lorsqu'on intègre une nouvelle vue d'un plan unique.
- Pour la méthode des lignes centrales, la ligne centrale correspondante est intégrée à l'ensemble des droites intersectées pour calculer le point principal.

Cependant, lorsqu'on ne connaît que les correspondances inter-images, il est nécessaire d'estimer l'homographie scène-image pour une image clé. Dans le cas multiplanaire, l'homographie scène-image de chacun des plans de référence doit être estimée pour l'image clé.

Les méthodes de calibrage itératif proposée par Noirfalise [NJS03] et d'auto-calibrage planaire proposée par Gurdjos et Sturm [GS03] estiment d'ailleurs la rectification projective d'après la première image, du plan observé durant la séquence. Il y a donc à estimer les paramètres intrinsèques de la caméra et quatre paramètres pour la rectification de chaque plan de calibrage.

Nous développons dans les sections suivantes l'extension de ces deux méthodes, en utilisant plusieurs plans par image.

7.2.1 Extension de la méthode de calibrage itératif

Les matrices $\{\mathbf{H}_{01}^{(1)}, \cdots, \mathbf{H}_{01}^{(N_p)}\}$ de la forme (6.12) sont associées à chacun des N_p plans de calibrage, correspondant aux matrices d'homographie (inconnues) plan-image pour une vue clé.

Les $4N_p$ paramètres $\{a_i, b_i, c_i, d_i\}$ $(i = 1, \cdots, N_p)$ définissant les matrices $\mathbf{H}_{01}^{(i)}$ sont estimés simultanément avec les paramètres intrinsèques à l'aide du filtre de Kalman. Pour effectuer cela, la fonction d'erreur pour le cas multiplanaire est de dimension $2N_p$, représentant les $2N_p$ contraintes de calibrage fournies par les N_p plans.

L'intégration de plusieurs plans de calibrage impose aussi que les vecteurs solutions soient de dimension $4N_p+3$ (lors du calcul du point principal et de la distance focale uniquement) et les matrices de covariances du système soient de dimensions $(4N_p+3)\times(4N_p+3)$.

Pour le calibrage basé sur un plan unique, Noirfalise *et al.* ont proposé de débuter l'enregistrement de la séquence de calibrage dans une position où le plan image de la caméra soit presque fronto-parallèle au plan observé, de manière à proposer que l'homographie scène-image initiale soit la matrice identité ($\mathbf{H}_{01}^{(1)} \sim \mathbf{I}_{3\times 3}$).

Cependant, dans le cas multiplanaire on espère tirer parti des diverses dispositions des plans par rapport à la caméra et il est alors difficile de proposer des paramètres initiaux pour les homographies $\mathbf{H}_{01}^{(i)}$ qui ne soient pas très éloignées de leurs véritables valeurs au risque de converger vers une solution qui est un minimum local.

7.2.2 Extension de la méthode d'auto-calibrage par lignes centrales

Le problème à resoudre est celui du calcul simultané des 2 coordonnées du point principal (et le rapport d'aspect τ s'il est inconnu) et des paramètres liés aux homographies scène-image des N_p plans de calibrage pour une image clé. L'homographie scène-image correspondant au plan j est déterminée par 4 paramètres : $\{\alpha^{(j)}, \beta^{(j)}, \mu^{(j)}, \lambda^{(j)}\}$.

Notre approche consiste à estimer les $4N_p+2$ paramètres en minimisant la somme des fonctions d'erreurs résiduelles attribuées à chacun des N_p plans, c'est à dire :

$$\sum_{j=1}^{N_p}\left(\sum_{i=1}^{n}\frac{\left(\det\left(\mathbf{M}_i^{(j)}\right)\right)^2}{J_{M_i^{(j)}}Cov[\mathbf{H}_{1i}^{(j)}]J_{M_i^{(j)}}^T}\right)$$

où $\mathbf{M}_i^{(j)} = \left(\mathbf{H}_{1i}^{(j)}\mathbf{\Sigma}_1{\mathbf{H}_{1i}^{(j)}}^T\mathbf{e}_3 \mid \Lambda{\mathbf{H}_{1i}^{(j)}}^{-T}\mathbf{l}_{\infty,1} \mid \tilde{\mathbf{p}}_0\right)$ et $J_{M_i^{(j)}} = \left\{\frac{d\det(\mathbf{M}_i^{(j)})}{d\mathbf{H}_{1i}^{(j)}}\right\}$.

Dû au caractère non-linéaire de cette fonction, nous avons choisi d'effectuer l'estimation des paramètres par la méthode de moindres carrés itérés, de telle manière que les facteurs de pondération $z_i^{(j)} = J_{M_i^{(j)}}Cov[\mathbf{H}_{1i}]J_{M_i^{(j)}}^T$ sont tenus constants à chaque itération du processus, en minimisant à chaque fois par la méthode de Levenberg-Marquardt une fonction de la forme :

$$\sum_{j=1}^{N_p}\sum_{i=1}^{n}\frac{\left(\det\left(\mathbf{M}_i^{(j)}\right)\right)^2}{z_i^{(j)}}$$

7.3 Cadre d'évaluation des méthodes de calibrage interne

Afin de réaliser une étude comparative entre les méthodes de calibrage interne présentées dans ce chapitre, nous avons conduit plusieurs expérimentations dans le but de confronter les performances de ces méthodes, lorsqu'on emploie des ressources matérielles telles que des objets de calibrage plans et simples ou des surfaces planes texturées quelconques appartenant à la scène.

Les méthodes de calibrage ici étudiées requièrent en entrée les données suivantes :

- des points de fuite correspondant à des directions orthogonales, pour les méthodes proposées par [CT90, RCR99], section 6.2.
- un ensemble de matrices d'homographies, pour les méthodes de calibrage planaire (section 6.3) :
  - soit des homographies entre l'image et la scène ;
  - soit des homographies inter-images, entre chaque image et une image *clé*, employées par les méthodes d'auto-calibrage planaire.

Nous avons généré différents ensembles de données d'expérimentation, afin de fournir les informations d'entrée :

- des données de synthèse, en simulant un motif simple : un contour rectangulaire ;
- des images réelles :

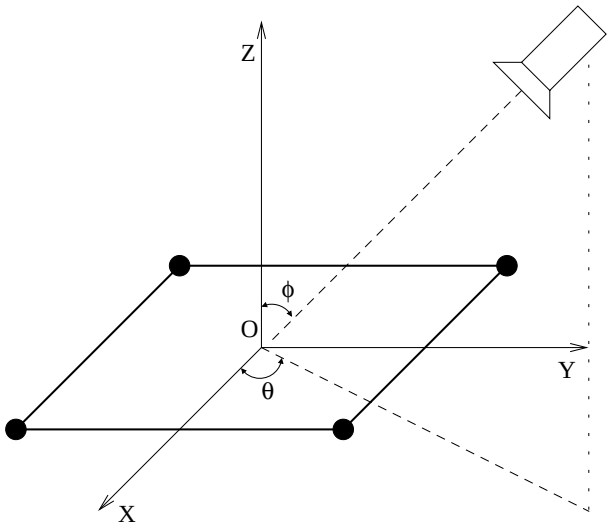


FIG. 7.2 – Génération d'images de synthèse : un motif carré.

- en utilisant les marqueurs d'une mire de calibrage ;
- en utilisant des indices naturels extraits de surfaces planes texturées.

Les résultats issus des méthodes de calibrage interne sont : les paramètres intrinsèques du modèle sténopé : les cinq coefficients de la matrice $\mathbf{K}$, sous réserve de la connaissance *a priori* de certains paramètres ; et pour les méthodes d'auto-calibrage planaire, des descripteurs euclidiens du plan de calibrage. Pour quelques méthodes, nous proposons aussi d'estimer les écarts-types théoriques des paramètres intrinsèques au moyen de la méthode de propagation d'incertitudes (section 2.5).

Le protocole expérimental suivi pour les différents types de données est décrit par la suite.

7.3.1 Évaluation sur des données de synthèse

Nous générons artificiellement des points de calibrage en attribuant des valeurs prédéfinies aux paramètres intrinsèques de la caméra et aux points de vue.

Les paramètres intrinsèques calculés au moyen des méthodes de calibrage ici abordées, sont comparés avec les valeurs employées dans la génération des données de synthèse.

Les résultats des expériences conduites à l'aide de ces simulations cherchent à évaluer notamment :

- les limites de précision dans l'estimation des paramètres intrinsèques qui peuvent être atteintes par les différentes méthodes de calibrage ;
- la sensibilité des paramètres intrinsèques estimés, par rapport au niveau de bruit de détection des primitives ;
- l'influence sur le calibrage interne des positions relatives de la caméra par rapport au motif observé ;
- l'influence des erreurs dans la connaissance *a priori* de certains paramètres (e.g. point principal, rapport d'aspect) sur les paramètres à calculer ;

Simulation d'un motif carré

Afin d'analyser le comportement des méthodes de calibrage interne basé sur l'observation de motifs simples, tels que un carré et une grille, nous avons simulé des images (de taille 640×480) lors de déplacements d'une caméra visant ce type de motifs.

La configuration utilisée dans ces expériences est un carré de taille 10×10 placé sur le plan xy, dont le centre est situé à l'origine du système de coordonnées de la scène : le motif est un carré, dont nous utilisons les quatre coins comme points de calibrage (figure 7.2).

Nous considérons que la caméra se déplace entre deux sphères de rayons donnés (de 40 et 80, resp.) dont les centres coïncident avec le centre du motif, pour simuler que le motif est toujours visé au milieu de l'image. Parce que les rotations de la caméra autour de l'axe focal équivalent aux rotations de l'image autour du point principal, nous spécifions donc les positions de la caméra uniquement avec deux angles (figure 7.2) dans nos simulations : l'angle d'inclinaison par rapport à l'axe focal (ϕ), et le déplacement angulaire par rapport au plan xz (θ).

Nous avons simulé plusieurs ensembles de données correspondant à ce motif afin d'évaluer la performance des méthodes de calibrage par rapport à différents facteurs : le niveau du bruit ajouté, la distance entre la caméra et le motif et la distance focale.

Dans notre étude, nous prenons en compte deux hypothèses différentes sur la position du point principal de la caméra :

- leurs coordonnées étant connues de manière exacte ;
- ayant une erreur par rapport à la vraie position.

Ceci nous permet d'évaluer l'influence des erreurs du point principal sur l'estimation de la distance focale, ainsi que de déterminer les configurations image-scène qui sont plus sensibles à cette erreur.

Nous avons utilisé ces valeurs pour générer les positions non bruitées des projections des cibles de la mire, en respectant parfaitement le modèle sténopé. Cette démarche nous a permis d'effectuer des comparaisons sur des configurations semblables à celles de nos séquences réelles, à la différence que les résultats dérivés des données synthétiques s'épargnent du problème de la modélisation de la caméra.

7.3.2 Évaluation sur des images réelles

Pour l'analyse des méthodes de calibrage, les constatations faites lors des expériences sur des données de synthèse sont limitées par la capacité de modéliser le véritable processus d'acquisition d'images des caméras, la distribution statistique du bruit associé à la détection de primitives, ainsi que d'autres facteurs de distorsion qui pourraient se présenter. Par conséquent, une étude prenant en compte les aspects pratiques du processus de calibrage requiert des observations intégrant les facteurs qui apparaissent dans les situations réelles.

Séquences *Mire* et *Paysage*

Afin de comparer les méthodes de calibrage interne sur des images réelles, nous avons enregistré deux séquences d'images dont les mouvements de la caméra ont été contrôlés à l'aide d'une table robotisée. Nos expériences ont été conduites en observant deux scénarios différents :

- Séquence *Mire* (figure 7.3) : la caméra vise une mire de calibrage sur laquelle on trouve des cibles circulaires dont les positions 3D sont connues (fournies par le fabriquant). La mire (figure 7.4a) est formée par trois surfaces planes dont les deux plans verticaux (le plan #3 à gauche de l'image et le plan #2 à droite) forment un angle de 120°, et sont à la fois perpendiculaires à la base (plan #1).
- Séquence *Paysage* : de manière à obtenir une structure composée par plusieurs surfaces planes texturées (figure 7.4b), nous avons placé des reproductions de paysages (des images imprimées) sur les trois faces de la mire.

Nous avons attribué les mêmes déplacements à la caméra durant l'enregistrement de chaque séquence d'images, de manière à confronter les résultats obtenus avec les deux scénarios.

Le but d'utiliser le même cadre expérimental (les mêmes mouvements de la caméra ainsi que les mêmes positions des plans de référence) est d'obtenir des conditions comparables dans deux contextes d'application différents : en utilisant des marqueurs plans dont les positions des indices sont connues, et des surfaces planes texturées quelconques.

L'ensemble des mouvements a été choisi d'une manière telle que :

- la majeure partie de chacune des trois surfaces de la scène (faces de la mire ou posters) soient observées dans chaque image de la séquence.
- les points de vue de la caméra soient le plus divers possibles (autant que les déplacements de la table micrométrique le permettent) en variant la distance entre la caméra et les plans observés, leur position relative, la distribution sur l'image des points d'intérêt observés, etc. Une description

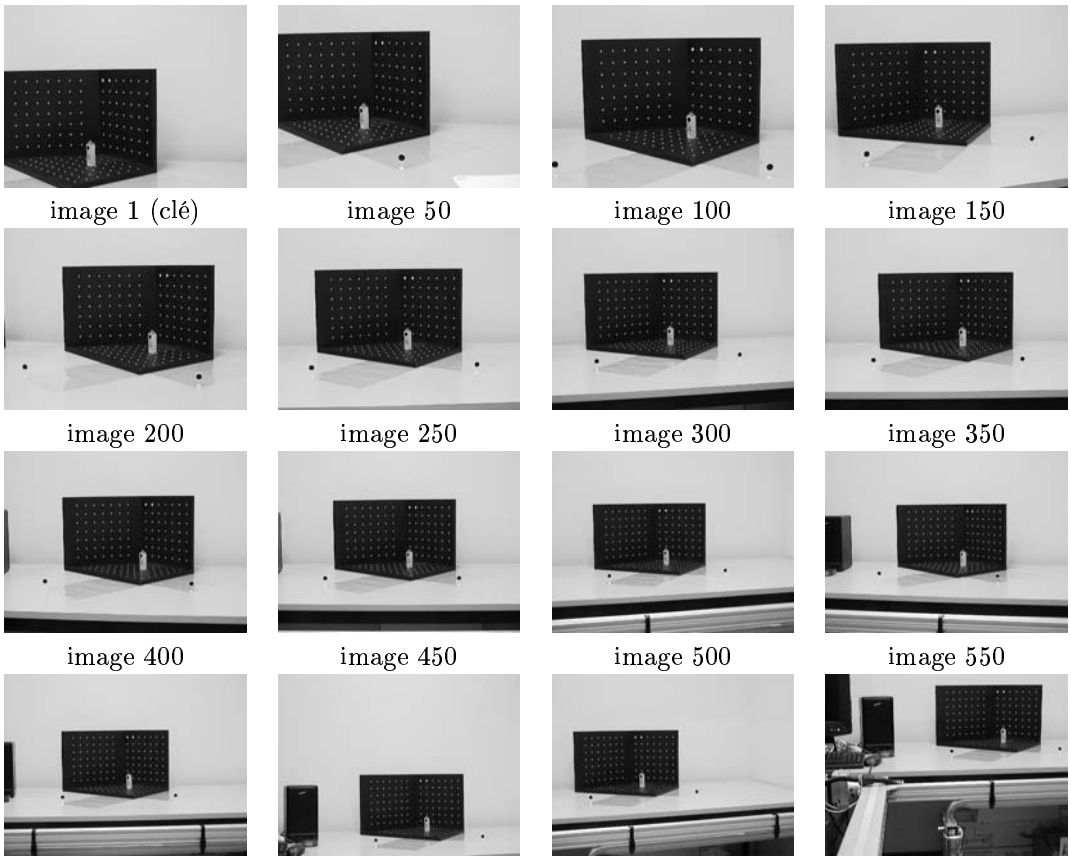
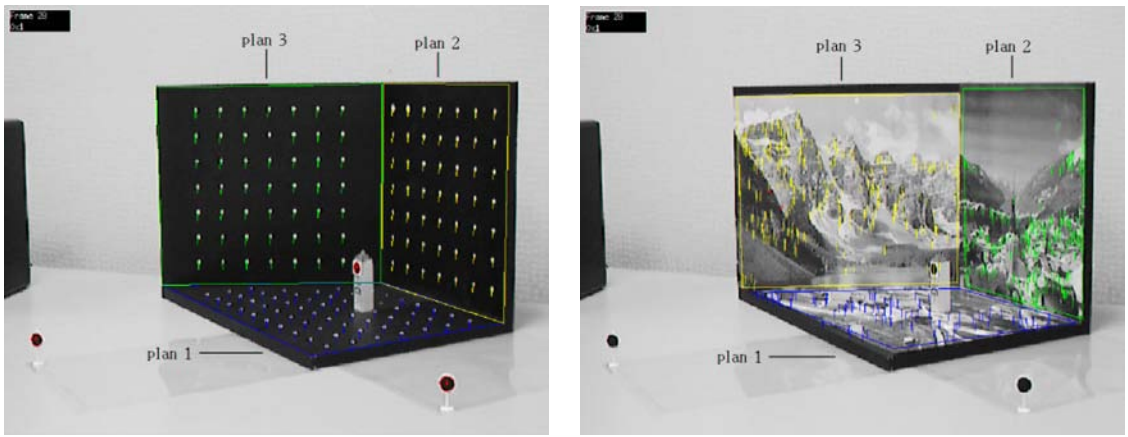


FIG. 7.3 – Quelques images de la séquence de la mire.



(a) (b)

FIG. 7.4 – Exemples des images réelles, utilisées dans le cadre d'évaluation expérimental (a) d'une mire de calibrage, (b) d'une scène avec trois surfaces texturées.

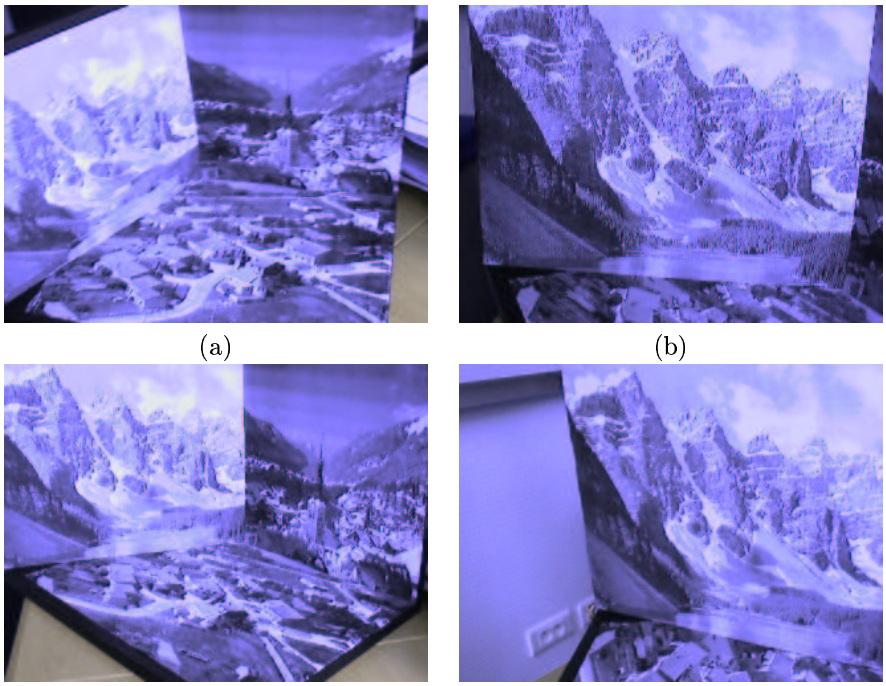


FIG. 7.5 – Séquence du *Paysage2* : Images clés pour les sous-séquences avec (a) trois plans et (b) un seul plan. Exemples de prises de vues pour l'ensemble avec (c) trois plans et (d) un plan.

détaillée du dispositif robotisé ainsi que des mouvements réalisés pendant l'enregistrement de la séquence vidéo a été présentée en section 5.2.

La séquence du *Paysage* a été acquise sans délai à la suite de l'acquisition de la séquence *Mire*. Les images ont été acquises sans changer le zoom de la caméra ni effectuer de mise au point durant l'enregistrement des deux séquences afin de garder les paramètres intrinsèques de la caméra constants pendant les deux séquences d'images.

7.3.3 Extraction et mise en correspondance des points d'intérêt

Nous décrivons à présent les données qui ont été extraites à partir de ces deux séquences d'images, et utilisées comme entrées pour les algorithmes de calibrage planaire.

Pour la séquence de la *Mire*

Nous avons utilisé la méthode de détection fine d'ellipses avec correction des distortions optiques proposée par Brand *et al.* [BMB94] pour effectuer la localisation précise (de l'ordre de 0.1 pixels) des cibles de la mire dans les images.

Les cibles étant localisées sur trois plans différents de la mire, on établit des correspondances 3D-2D entre les points d'intérêt détectés dans les images et les coordonnées des cibles fournies. Ces correspondances sont requises comme entrées par les méthodes de calibrage planaire [SM99, GP01] (voir sections 6.3.2 et 6.3.4).

Des correspondances inter-images ont aussi été obtenues à partir de la détection de cibles de la mire, en appariant les cibles correspondantes entre chaque couple d'images, soit des images consécutives soit entre chaque image et une image clé (la première image, dans nos expériences). Ces correspondances sont employées par les méthodes d'auto-calibrage planaire [NJS03, GS03] (voir sections 6.3.3 et 6.3.5).

Pour la séquence du *Paysage*

Des points d'intérêt ont été extraits à partir des plans texturés de la séquence du *Paysage* (figure 7.4) afin d'estimer les homographies inter-images requises par les méthodes d'auto-calibrage planaire.

Nous avons testé deux *méthodes de détection* des points d'intérêt afin d'examiner comment les procédés utilisés dans l'obtention de ces points et leur précision influent sur la performance des techniques de calibrage interne. Les méthodes de détection évaluées dans nos expérimentations sur la séquence du *Paysage* sont :

- le détecteur de Harris [HS88] ;
- l'obtention des points d'intérêt en utilisant le descripteur SIFT (*Scale Invariant Feature Transformation* [Low99]).

La mise en correspondance en utilisant les points de Harris a été faite uniquement entre des images consécutives de la séquence de manière à que la différence entre les positions des points correspondants entre les deux images ne dépasse pas l'aire de recherche proposée (une fenêtre de 25×25 pixels). L'appariement est effectué en calculant un score de corrélation normalisé entre chaque couple de points candidats.

Les points détectés par la méthode SIFT ont été mis en correspondance en employant l'approche proposée par Skrypnyk et Lowe [SL04] lors de la mise en correspondance de chaque image avec la première image de la séquence.

À la différence de la séquence de la *Mire* où il est simple d'établir des appariements corrects des cibles grâce à la connaissance de leurs positions 3D, dans le cas du *Paysage*, les correspondances requises par les méthodes de calibrage ont été obtenues de manière robuste en utilisant la méthode RANSAC [FB80] afin de supprimer les correspondances aberrantes.

7.3.4 Estimation des homographies et de leurs covariances

Étant donné un ensemble de n appariements $\{(\mathbf{x}_i, \mathbf{y}_i) : i = 1, \cdots, n\}$, nous avons normalisé les coordonnées des points d'intérêt à l'aide d'une transformation T_{norm} qui déplace le centre de l'image (320×240) à l'origine et change l'échelle de façon que la distance maximale d'un point quelconque de l'image à l'origine soit égal à $\sqrt{2}$.

L'homographie correspondant à cet ensemble d'appariements est estimée par la minimisation de la fonction d'erreur $\sum_{i=1}^{n} \|\mathbf{y}'_i - \mathbf{H}\mathbf{x}'_i\|^2$ contrainte à $\|\mathbf{H}\|_F = 1$ (où $\mathbf{x}'_i$ et $\mathbf{y}'_i$ sont les coordonnées homogènes normalisées de $\mathbf{x}_i$ resp. $\mathbf{y}_i$) et sa matrice de covariance $Cov[\mathbf{H}]$ est approchée par l'application de la méthode de propagation inverse d'incertitudes (section 3.5.2) sur la même fonction.

En ce qui concerne le contexte pratique d'application sur une séquence d'images, nous avons confronté deux méthodes d'estimation des homographies entre une image quelconque i de la séquence et une image clé (dans nos expériences, la première image) :

- D'une part, les points d'intérêt de l'image i et de l'image clé sont mis en correspondance directement, de telle manière que les matrices d'homographies $\mathbf{H}_{1i}$ sont calculées au moyen de la méthode décrite ci-avant.
- D'autre part, les matrices d'homographies et leurs covariances correspondantes sont calculées de manière itérative ; c'est à dire, on utilise les correspondances des points d'intérêt entre les images consécutives pour estimer les homographies inter-images $\mathbf{H}_{j,j+1} (j = 1, ..., i-1)$. Les matrices $\mathbf{H}_{1i}$ et leurs covariances sont donc estimées par composition de matrices : $\mathbf{H}_{1i} = \mathbf{H}_{i-1,i} \cdots \mathbf{H}_{1,2}$, et leurs covariances sont approchées par propagation directe d'incertitudes.

Précision de calcul des homographies

De manière à évaluer la précision d'estimation des homographies, nous avons conduit des tests sur les séquences *Mire* et *Paysage* et en employant les deux méthodes de calcul des homographies. Nous avons proposé de calculer cette précision par la somme des variances des coefficients de $\mathbf{H}_{1i}$, c'est à dire, la trace de sa matrice de covariances : $tr(Cov[\mathbf{H}_{1i}])$.

Les courbes dans les graphiques (figure 7.6a) correspondent à $(tr(Cov[\mathbf{H}_{1i}]))^{1/2}$ pour chacune des images de la séquence, pour les différents cas examinés.

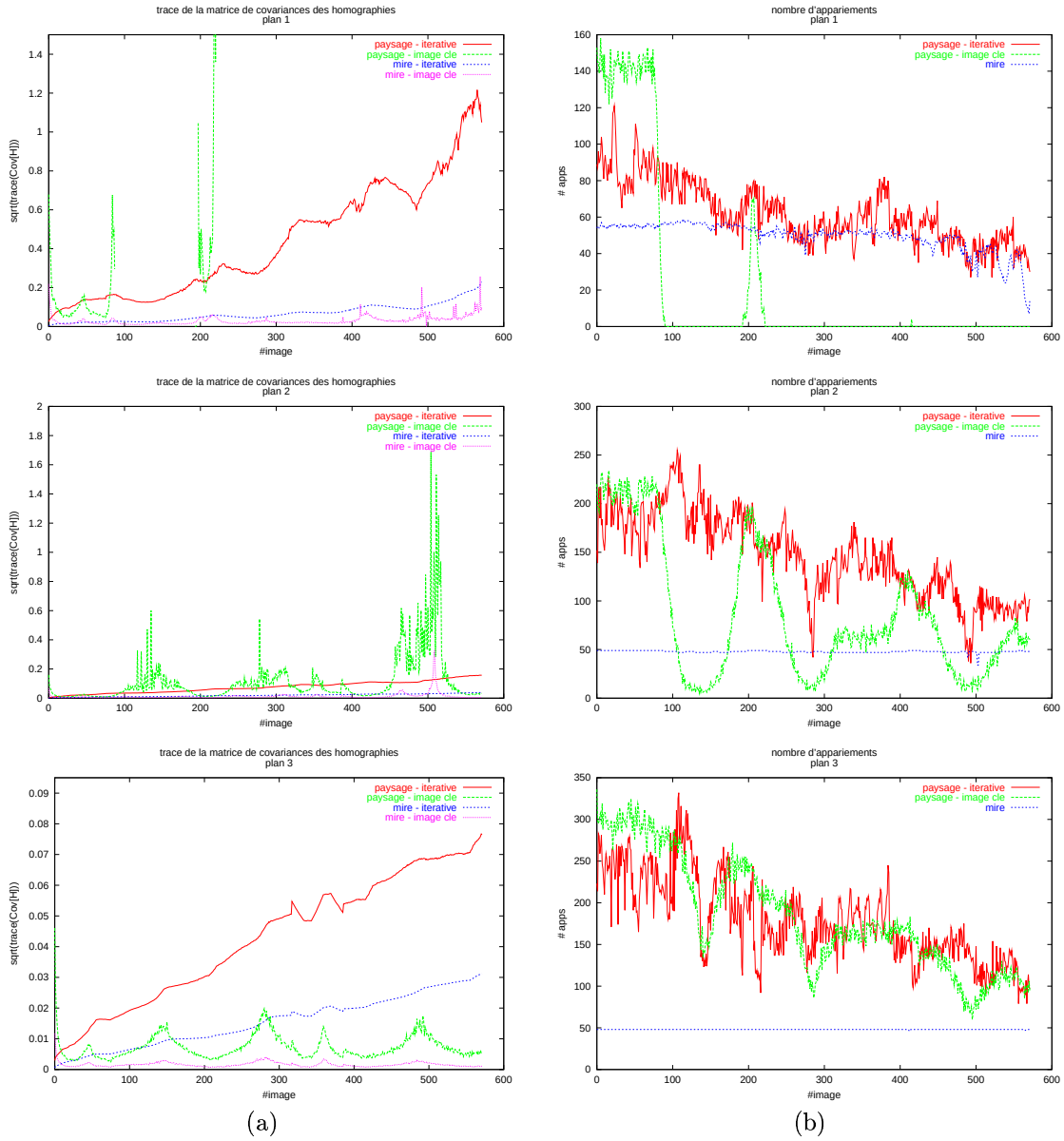


FIG. 7.6 – (a) Racine carré de la trace des matrices de covariance des homographies estimées directement par rapport à l'image clé (première image) et par composition de matrices, pour les séquences *Mire* et *Paysage*. (b) Nombre d'appariements correspondant.

Les résultats montrent qu'une accumulation d'erreurs se produit toujours lorsque les homographies sont calculées de manière itérative. Ce problème ne se manifeste pas quand les homographies sont calculées de manière directe par rapport à l'image clé.

Néanmoins, pour quelques homographies, la précision dans le cas itératif est meilleure que la précision correspondant au cas direct. Ceci a lieu notamment dans les toutes premières images de la séquence ou quand le nombre d'appariements entre l'image correspondante et l'image clé est faible (figure 7.6b) : même si le calcul directe des homographies s'abstient de l'accumulation des erreurs, la mise en correspondance directe entre deux images (spécialement si les points correspondants se trouvent éloignés) repère parfois un nombre réduit d'appariements. Par exemple, pour la séquence du *Paysage*, le nombre d'appariements correspondant au plan #2 (figure 7.6a) est inférieur à 50 pour quelques images, par suite d'un changement

importante dans l'orientation du plan par rapport à l'image clé. Dans le cas du plan #1, le nombre d'appariements entre les images consécutives est toujours réduit à cause de l'inclinaison assez oblique du plan par rapport au plan image (moins d'une centaine de correspondances), et dans le cas direct, la méthode SIFT ne parvient plus à trouver de correspondances au moment où la caméra commence à s'éloigne de la scène.

D'après cette étude, on constate aussi l'incidence des erreurs de détection sur la précision en comparant la mesure d'évaluation pour les deux séquences d'images : les résultats sur la *Mire* sont beaucoup plus précis que sur le *Paysage* grâce au fait que la méthode de détection des cibles circulaires est plus fine.

Les différences d'amplitude dans les graphiques (figure 7.6a) nous permettent d'inférer aussi que la précision de calcul des homographies dépendent des caractéristiques du plan observé :

- le nombre de points d'intérêt détectés par image, et en conséquence, le nombre d'appariements trouvés par rapport à l'image clé ;
- la distribution des points d'intérêt ;
- l'inclinaison du plan par rapport à l'axe focal.

Les deux premiers facteurs expliquent pourquoi il existe des grandes différences de précision sur la séquence *Paysage*, particulièrement entre le plan #3 (dont la mesure est inférieure à 0.1) et les deux autres plans.

Néanmoins, le nombre et la distribution des points d'intérêt sont similaires pour les plans contenants les cibles de la *Mire*, et la seule différence qui permet donc une meilleure précision au plan #3 par rapport aux autres, est l'orientation de ce plan par rapport à la caméra.

7.3.5 Vérité terrain

Pour évaluer les performances des méthodes lors du calibrage avec des images réelles, il est nécessaire d'obtenir des valeurs de référence fiables correspondant aux *'véritables'* paramètres de la caméra utilisés pendant l'acquisition d'images. Cependant, la détermination de cette *vérité terrain* pour le calibrage interne est une tâche complexe, du fait que le modèle sténopé n'est qu'une représentation simplifiée du processus de formation d'images, et donc les paramètres intrinsèques de la caméra ne peuvent pas être directement mesurés par des capteurs.

Afin d'établir la vérité terrain dans nos expérimentations pour l'évaluation des techniques de calibrage planaire, nous utilisons la connaissance *a priori* des positions des cibles de la mire et la constance des paramètres intrinsèques tout au long de la séquence, pour réaliser un calibrage fin de la caméra au moyen de la méthode d'ajustement global, sur la totalité des images de la séquence de la mire. Le tableau 7.2 montre les résultats obtenus. La fiabilité de ce choix a été validée dans le chapitre précédent, satisfaisant les quatre critères de qualité choisis pour évaluer la performance des systèmes de RA (chapitre 2).

	f	τ	u_0	v_0
valeurs estimées	1113.14	1.000025	306.62	253.02

TAB. 7.2 – Paramètres intrinsèques de référence : séquences de la *Mire* et du *Paysage*.

Nous avons comparé les paramètres intrinsèques obtenus dans chacune de nos expériences sur des images réelles, aux valeurs de référence issus de ce procédé et nous avons mesuré pour chacune des méthodes les erreurs sur la distance focale, les coordonnées du point principal et le rapport d'aspect $(\Delta f, \Delta u_0, \Delta v_0, \Delta\tau)$.

7.4 Comparaison des méthodes de calibrage

7.4.1 Méthode des points de fuite

Nous avons conduit une série d'expériences en utilisant les données de synthèse correspondant à la séquence #1 (section 7.3.1). La position du point principal étant connue, la distance focale est calculée

à partir des quatre sommets d'un carré, suivant la procédure basée sur la contrainte des points de fuite décrite dans la section 6.2.

Nous avons d'abord évalué les distances focales obtenues pour des points de vue situés sur une sphère de rayon $r = 40$. Les projections des sommets du carré (c'est à dire, les points de calibrage) ont été générées artificiellement, en considérant le point principal au centre de l'image $(u_0, v_0) = (320, 240)$) et la distance focale $f = 1000$. Pour cet ensemble de points, la surface de chaque carré occupe la plupart de l'image (la distance entre deux sommets opposés peut être jusqu'à environ 400 pixels, dans une image de 640×480).

Les coordonnées des points de calibrage ont été perturbées par un bruit gaussien dont l'écart-type est 0.5 pixels. Les estimations de la distance focale ont été répétées 1000 fois (avec des perturbations aléatoires différentes à chaque fois) pour chacune des positions de la caméra. Les différences entre la distance focale moyenne obtenue à chaque position et la distance focale attendue ont été calculées. La figure 7.7(a) représente cette différence : plus la zone est claire, plus la différence est réduite. La projection des points de vue pour lesquels la valeur absolue de l'erreur sur la distance focale est inférieure à 10 et 100, resp. (c'est à dire, 1 et 10% d'erreur relative) est représentée en figure 7.7(b).

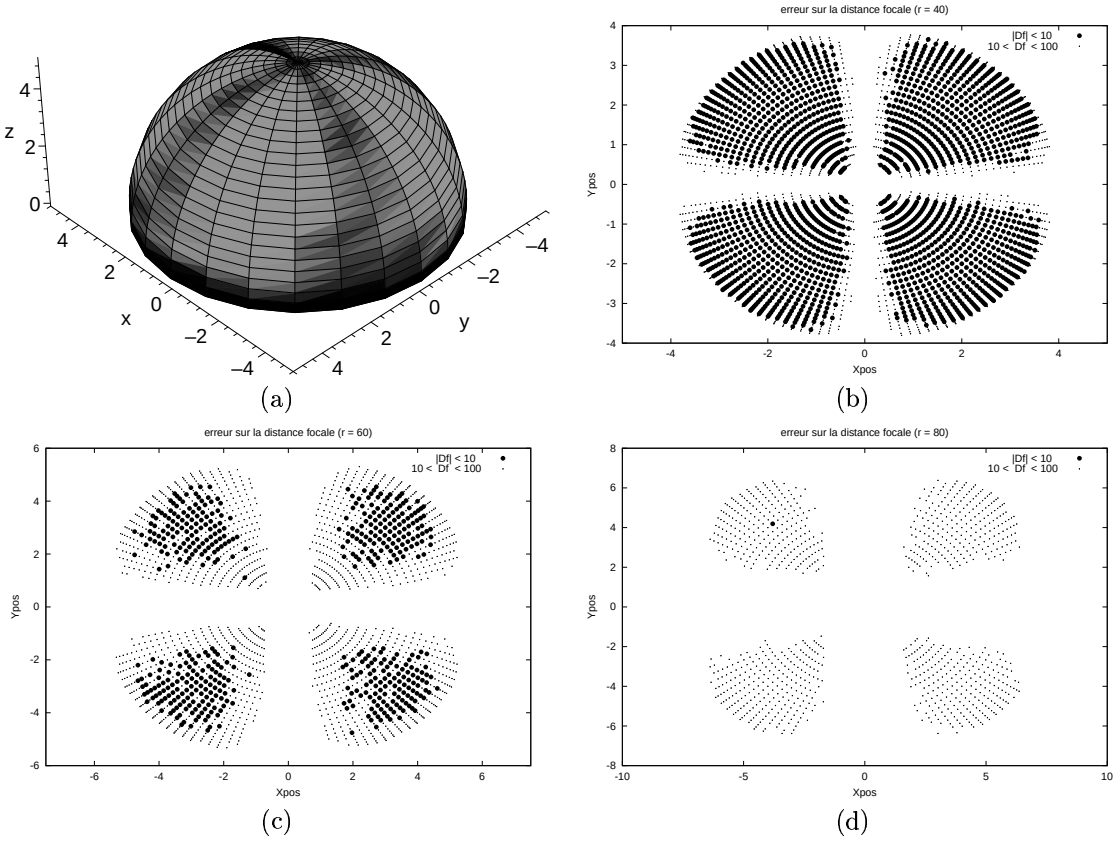


FIG. 7.7 – Erreurs sur la distance focale en fonction de la position de la caméra (données de synthèse). (a) Erreurs de la distance focale : les niveaux de gris les plus clairs correspondent aux meilleures estimations. Positions correspondantes aux erreurs sur la distance focale inférieures à 1% (•) et 10% (·), pour rayons : (b) $r = 40$, (c) 60 et (d) 80.

On retrouve à peu près le même schéma pour des rayons de sphère différents (figure 7.7 (b-d)), mais on constate que les régions correspondant à une erreur relative de f inférieure à 1% (et 10%) se réduit rapidement lorsque la caméra s'éloigne du motif. Une observation importante acquise de ces expériences est le fait que la moyenne des erreurs sur la distance focale est toujours positive, lorsque l'erreur relative est supérieure à 1%.

Du point de vue de la facilité de mise en œuvre du calibrage, la méthode des points de fuite est la

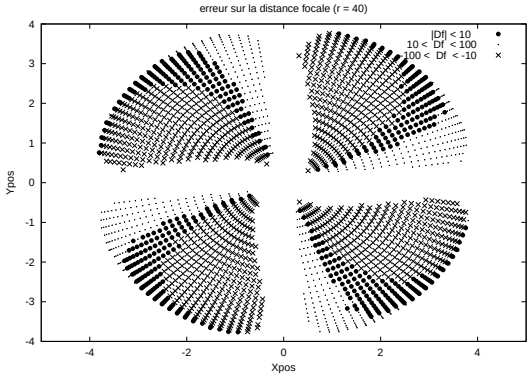
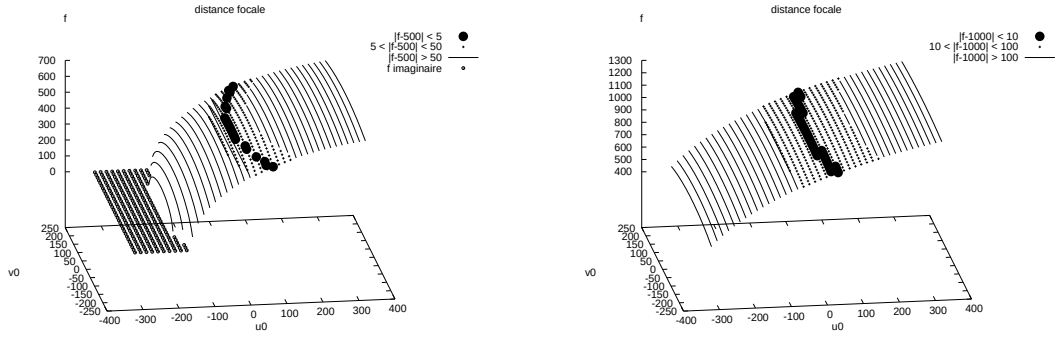


FIG. 7.8 – Erreurs sur la distance focale en fonction de la position de la caméra, lorsque la position du point principal a une erreur de $+20$ pixels sur u_0 et -20 pixels sur v_0. Écarts inférieurs à 1% ($\bullet$), écarts entre 1% et 10% positifs ($\cdot$) et négatifs ($\times$).



(a) (b)

FIG. 7.9 – Estimation de la distance focale en fonction de l'erreur sur le point principal. Écarts inférieurs à 1% ($\bullet$), entre 1% et 10 % ($+$), et lorsque $f^2 < 0$ (distance focale imaginaire, $\circ$).
La position de la caméra par rapport au marqueur correspond à une élévation de 45° et un angle de vue aligné à une des diagonales du motif carré. La distance de la caméra au centre du marqueur est $r = 40$, et la distance focale est : (a) $f = 500$ et (b) $f = 1000$.

technique la plus simple à implémenter lorsqu'un motif simple est connu : il suffit de détecter, dans une seule image, les sommets ou les arêtes d'un contour rectangulaire quelconque, pour estimer la distance focale de la caméra.

Néanmoins, cette méthode présente deux désavantages notables :

- D'une part, la méthode est basée sur le calcul des points de fuite, obtenus par l'intersection d'un ensemble de projections de lignes parallèles (les arêtes opposées du contour). Mais les estimations des points de fuite sont très sensibles à la position du rectangle par rapport à la caméra [BAA98]. Alors, le calcul des positions des points de fuite est aussi très sensible aux variations sur les sommets du rectangle, et de très petites perturbations sur ces points peuvent entraîner des grandes erreurs sur l'estimation de la distance focale, comme le confirment nos résultats.
- D'autre part, il est indispensable de connaître au préalable la position du point principal de la caméra : une erreur d'estimation de ce point peut être à l'origine d'une grande erreur dans l'estimation de la distance focale et même causer des résultats physiquement inconsistants, tels que des valeurs imaginaires de f.
  La figure 7.8 montre que les positions de la caméra par rapport au marqueur correspondant aux

erreurs les plus faibles ont changé à cause des erreurs dans la connaissance du point principal. Dans cette expérience, on a introduit une erreur de +20 pixels sur la coordonnée u_0 et de -20 pixels sur la coordonnée v_0.
Des autres erreurs sur la position du point principal ont été introduits (de -320 à 320 pixels sur u_0 et de -240 à 240 sur v_0) dans le but de mesurer les variations de la distance focale par rapport aux erreurs d'estimation du point principal. Nous avons choisi une disposition caméra-marqueur permettant d'obtenir une estimation très précise de la distance focale (inférieure à 1% en moyenne) lorsque le point principal est connu. Néanmoins, des petites erreurs sur le point principal (voir figure 7.9) entraînent des variations significatives sur l'estimation de la distance focale. Notamment, dans le cas des distances focales petites ($f = 500$ dans l'exemple de la figure 7.9a) les variations sont plus importantes et on peut même trouver des estimées négatives de f^2, c'est à dire, des valeurs imaginaires de la distance focale.

7.4.2 Méthode de Sturm-Maybank

Nous avons testé la méthode de Sturm-Maybank en utilisant les images de la séquence de la mire (séquence #2, voir figure 7.3). Nos résultats ont été obtenus, en effectuant le calibrage planaire interne pour chacun des plans de la mire individuellement, puis en prenant ensemble les trois plans. Les graphiques dans la figure 7.10 indiquent les paramètres intrinsèques (f, u_0, v_0) estimés pour l'ensemble des images à partir de la première image jusqu'à la i-ème image de la séquence. Nous avons comparé les performances de la méthode originale (graphiques 7.10(a)) et de la méthode modifiée en intégrant la stratégie de pondération statistique (graphiques 7.10(b)).

Résultats avec la méthode originale

En observant d'abord les résultats obtenus lors du calibrage pour chaque plan individuellement, nous remarquons notamment que la distance focale est sous-estimée par rapport à la valeur attendue, pour chacun des trois plans (lorsque plus d'une centaine d'images est employée). L'erreur la plus petite est obtenue pour le plan 1 (la base de la mire), dont l'écart varie entre 3 et 5 pixels. Les différences sont plus grandes pour les deux autres plans : de 4 à 6 pixels pour le plan 2, et de 10 à 11 pixels pour le plan 3. Dans tous les cas, l'erreur relative sur la distance focale est inférieure à 1% pour chaque plan, lorsqu'on considère entre les 100 et 500 premières images de la séquence.

Par rapport au point principal, l'erreur sur u_0 devient plus petite que 8 pixels après quelques dizaines d'images prises en compte dans les calculs, cette différence étant inférieure à 3 pixels pour le plan numéro 1. Les estimations de la coordonnée v_0 présentent des erreurs similaires, sauf pour le plan 1 dont les écarts peuvent être jusqu'à 10 pixels, mais en revanche, les résultats semblent moins stables pour v_0 que pour u_0. Une possible explication à ce phénomène est que le type de mouvements effectués par la caméra est prédominant sur l'horizontale, et les ensembles de points d'intérêt présentent moins de variations sur l'axe Y de l'image que sur l'axe X.

Pendant la première centaine d'images, les paramètres estimés oscillent de manière notoire et sont assez écartés de la valeur de référence (jusqu'à 20% sur la distance focale, et plus de 500 pixels sur la coordonnée v_0), étant connu que les déplacements de la caméra effectués pour les images 1-76 sont uniquement des rotations.

Il est possible d'effectuer le calibrage interne en utilisant cette méthode si les positions relatives du plan observé par rapport à la caméra ne sont pas parallèles [SM99, Zha00]. Il est donc faisable d'estimer les paramètres intrinsèques au moyen de rotations pures de la caméra. Cependant, les écarts maximaux pour les rotations (images 1-76) dans nos expériences sont de 10° horizontalement et 8° sur l'élévation. Cela nous indique que les amplitudes de ces rotations n'ont pas été suffisantes pour calibrer efficacement la caméra. Les erreurs sont plus importantes pour le plan de la base de la mire (plan 1) car les rotations horizontales ne modifient pas la position relative du plan par rapport à la caméra.

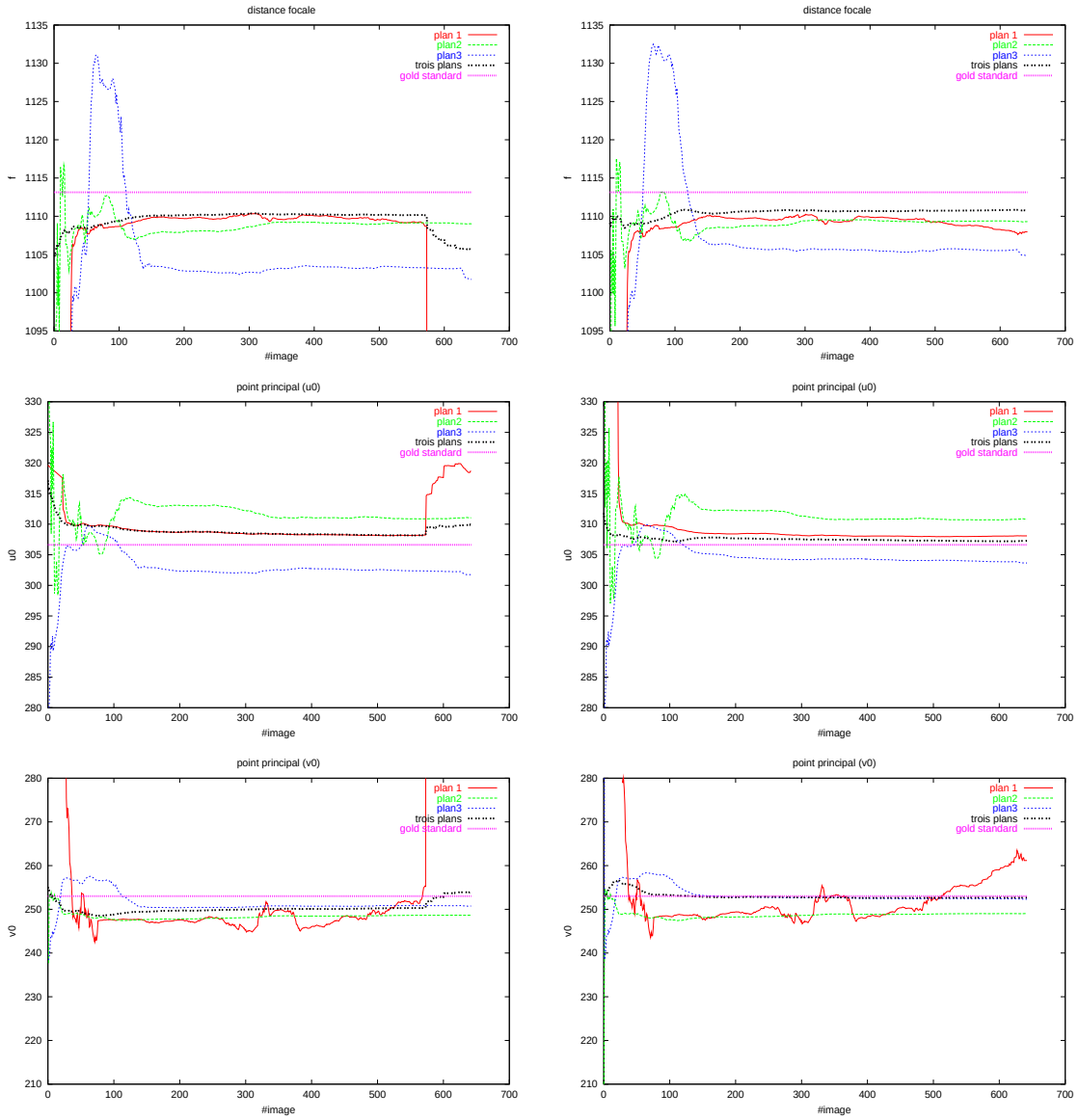


FIG. 7.10 – Méthode de Sturm-Maybank : paramètres intrinsèques calculés (f, u_0, v_0), lorsque τ est connu ; pour les cas planaire et multiplanaire, (a) à partir de la méthode originale et (b) avec l'intégration de la technique de pondération statistique.

Cas multiplanaire

Lorsque les trois surfaces de la mire sont utilisées dans le processus de calibrage (voir trait noir dans la figure 7.10(a)), les différences entre les paramètres estimés et les valeurs de référence sont, en général, plus petites que si on considère les plans 2 et 3 de manière indépendante, et les paramètres estimés sont très proches des valeurs obtenues pour le plan 1 (celui qui a apporté les meilleures estimations pour f et u_0). En général, les résultats obtenus lors de l'utilisation de plusieurs plans de calibrage sont plus stables.

Cependant les estimations de v_0 obtenues avec le plan 3 sont plus près de la référence qu'en utilisant les trois plans ; néanmoins, les écarts entre les coordonnées calculées en observant le plan 3 et les trois plans ensemble sont inférieurs à 0.6 pixels.

Par ailleurs, les erreurs obtenues lors du calibrage avec des rotations pures (images 1-76) sont notable-

(642 images)	méthode originale				avec pondération statistique			
plan	plan 1	plan 2	plan 3	trois plans	plan 1	plan 2	plan 3	trois plans
Δf	-537.1	-4.1	-11.3	-7.4	-5.1	-3.8	-8.3	-2.3
Δu_0	+12.0	+4.4	-4.9	+3.3	+1.5	+4.2	-2.9	+0.7
Δv_0	+806.38	-4.3	-2.3	-1.0	+8.2	-4.0	-0.7	-0.3

TAB. 7.3 – Erreurs des paramètres intrinsèques lors du calibrage avec la méthode de Sturm-Maybank et la version étendue avec pondération statistique.

ment plus petites qu'avec les plans individuels, parce que les orientations relatives des plans de calibrage par rapport à la caméra sont plus variées : dans nos expériences, chaque image fournit des informations des trois plans, où les deux plans latéraux sont orthogonaux au plan de la base, ce qui apporte des contraintes correspondant à trois orientations différentes pour chaque image.

Comparaison avec la méthode pondérée

La distance focale estimée par la méthode originale, en utilisant le plan de calibrage 1, change abruptement à l'image 574 (à 70 pixels d'écart) et dérive ensuite jusqu'à 50% d'erreur relative de la distance focale attendue. La position estimée du point principal est aussi affectée (sautillement des courbes à l'image 574, figure 7.10(a)).

Ceci a lieu à cause des erreurs d'estimation des homographies scène-image du plan 1 : les plans de la mire sont assez éloignés de la caméra pour les toutes dernières images ; malgré cela, la plupart des points de calibrage latéraux sont toujours détectés, mais pour le plan de la base seulement un faible nombre de points d'intérêt est détecté (5 à 20, sur un total de 62 cibles de la mire) dû au fait que l'angle d'inclinaison de la base de la mire par rapport à la caméra empêche que la détection des points soit précise.

Malgré le fait que la plupart des homographies correspondant au plan 1 aient été mieux estimées, les erreurs de calcul dans les toutes dernières homographies nuisent significativement à la précision des résultats, parce qu'aucune mesure de la précision de calcul des homographies n'a été prise en compte dans la méthode de calibrage. L'emploi des trois plans a permis d'atténuer les déviations produites par ces imprécisions, mais pas de les exclure.

Le tableau 7.3 montre les erreurs obtenues sur les paramètres intrinsèques en utilisant les images 1 à 642, avec la méthode originale et avec la technique de pondération statistique. Lorsque la stratégie de pondération statistique est intégrée à la méthode de calibrage (figure 7.10(b)), on remarque que les résultats sont plus stables et plus précis.

Quoique la méthode originale de Sturm-Maybank utilise une approche pour améliorer le conditionnement numérique du système linéaire [Gol89], cette stratégie n'est pas toujours suffisante pour réduire l'influence des erreurs d'estimation des homographies sur les paramètres calculés. En contrepartie, l'emploi de la pondération statistique nous a permis d'atténuer considérablement l'influence des imprécisions des homographies à la fin de la séquence, de telle manière que les nouvelles estimations sont plus robustes à ces erreurs, ce qui est remarquable lors du calibrage avec le plan 1 et dans le cas multiplanaire.

Influence de la position sur l'image des points d'intérêt

Nous avons observé dans nos expériences que les erreurs d'estimation de u_0 avec les plans droit et gauche (2 et 3, respectivement) sont de signes opposés ; c'est à dire, l'erreur d'estimation du point principal est liée à la position sur l'image du plan de calibrage.

Ainsi, quand un plan de calibrage se déplace seulement dans une région limitée de l'image (par exemple, uniquement d'un côté de l'image, près d'un bord ou d'un coin), le point principal calculé a un biais par rapport aux coordonnées attendues. Ceci est le cas des plans latéraux dans notre scène : comme la caméra vise en général la mire au centre de l'image, les points d'intérêt sur un de ces deux plans se trouvent alors groupés d'un seul côté de l'image (figure 7.3).

Pour cette raison, bien que l'estimation des homographies correspondant au plan de la base soit moins fiable que pour les plans latéraux (dû à l'inclinaison assez oblique du plan par rapport à l'axe focal de

la caméra), les résultats de u_0 pour le plan 1 se trouvent plus proches des valeurs attendues parce que la projection de cette surface plane se trouve centrée horizontalement dans l'image dans la plupart des vues.

7.4.3 Méthode de Noirfalise

Paramètres initiaux du filtre de Kalman

Le filtrage de Kalman requiert qu'un ensemble de valeurs initiales soit fourni : pour les paramètres intrinsèques de la caméra, les coefficients (a, b, c, d) de l'homographie scène-image de l'image clé ($\mathbf{H}_{01}$), et pour la matrice de covariances de l'ensemble de paramètres.

Dans nos expériences, nous avons proposé que la matrice de covariance initiale soit diagonale, du fait que les dépendances entre les paramètres sont en principe inconnues, et donc nous avons attribué uniquement des estimées aux variances des paramètres. En essayant différentes valeurs des paramètres initiaux et de leurs variances, nous avons remarqué que la convergence de la solution est assez sensible à cette initialisation.

Dans tous les cas, la position initiale attribuée au point principal a été le centre de l'image $(320, 240)$. Dans nos expériences, le point principal attendu se trouve à la position $(306.6, 253.0)$, c'est à dire, à 18.7 pixels du centre de l'image. Nous avons testé différents écarts-types initiaux pour le point principal, entre 10 et 100 pixels : lorsque l'écart-type initial attribué a été inférieur à 25 pixels, les coordonnées du point principal ont convergé rapidement vers une position incorrecte et proche de la position initiale (les coordonnées estimées peuvent se trouver éloignées de plus d'une dizaine de pixels de la position attendue).

La distance focale initiale attribuée a été de 1000 pixels (la valeur de référence est 1113.1), mais nous avons essayé aussi des valeurs entre 200 et 2000 pixels, en obtenant des résultats similaires, à condition que l'écart-type correspondant ne soit pas trop restrictif. L'écart-type proposé pour la distance focale a été de 1000 pixels ; l'utilisation d'écarts-types initiaux inférieurs à une centaine de pixels nous ont conduit à une convergence prématurée vers des valeurs incorrectes.

Tests sur la séquence *Mire*

Nous avons d'abord conduit des tests avec les images de la séquence *Mire* (fig. 7.4(a)) afin d'obtenir une évaluation sur un ensemble de données réelles dont la technique de détection et de mise en correspondance des points d'intérêt est d'une haute précision.

Les tests consistaient à effectuer le calibrage intrinsèque par la méthode de Noirfalise en utilisant les images 1 à 572 (puisque pour plusieurs vues à la fin de la séquence, le nombre de cibles détectées sur la base de la mire était inférieur à quatre). Les cas suivants ont été évalués :

- par rapport au numéro de plans : monoplanaire (un seul plan de la mire à la fois) ou multiplanaire (les trois plans ensemble) ;
- par rapport à la méthode de calcul des homographies : de manière itérative ou directement par rapport à une image clé (voir section 7.3.4).

Le tableau ci-dessous montre les erreurs sur la distance focale et les coordonnées du point principal obtenues dans ces tests :

	calcul itératif				par rapport à la première image			
	plan 1	plan 2	plan 3	trois plans	plan 1	plan 2	plan 3	trois plans
Δf	- 96.4	-5.2	-9.6	+31.5	- 51.9	-2.4	-7.6	-1.9
Δu_0	+ 7.9	+8.3	-0.1	+ 0.5	- 5.4	+5.6	-0.2	-4.8
Δv_0	+253.5	-4.9	-1.7	+24.2	+155.9	-6.8	-3.0	+2.6

Lors du calibrage avec les deux plan latéraux (2 et 3), les résultats sont similaires pour les deux méthodes de calcul des homographies : la position du point principal est estimée avec une erreur inférieure à 10 pixels dans les deux cas, tandis que la distance focale est obtenue avec une erreur relative inférieure à 1% pour le calcul itératif (0.5% pour le plan 2 et 0.9% pour le plan 3) et plus petite pour la mise en correspondance par rapport à l'image clé (0.2% et 0.7%, respectivement).

Les différences sont plus importantes lors du calibrage avec le plan 1 et du calibrage multiplanaire : les erreurs sur la distance focale et sur le point principal sont plus petites pour la mise en correspondance vers l'image clé. Lors du calibrage multiplanaire, l'erreur sur la distance focale dans le cas itératif est d'environ 3% et de 25 pixels sur le point principal, tandis que pour l'appariement vers l'image clé, l'erreur sur la distance focale est inférieure au 0.2% et d'environ 5 pixels sur le point principal.

Tests sur la séquence *Paysage*

Une deuxième expérience a été conduite en utilisant la séquence d'images *Paysage* (fig. 7.4b) afin de tester la méthode sur un ensemble de surfaces texturées (e.g. posters, marques naturels sur la scène), des éléments qui sont fréquents dans les scénarios courants d'application de la RA.

Les résultats obtenus après d'avoir effectué le calibrage interne par la méthode de Noirfalise sur 572 images de la séquence apparaissent dans les graphiques de la figure 7.11 et le tableau suivant :

	calcul itératif				par rapport à la première image			
	plan 1 $^{(*)}$	plan 2	plan 3	trois plans	plan 1 $^{(*)}$	plan 2	plan 3	deux plans
Δf	-533.8	-808.3	-504.0	-890.5	-457.0	- 9.6	-37.4	-33.7
Δu_0	- 71.4	-416.8	+104.2	-114.6	+ 61.0	-175.2	- 1.1	- 3.5
Δv_0	- 61.9	-167.1	- 42.0	- 57.2	-444.5	- 20.4	-10.3	-11.2

Dans notre exemple, même si le plan 1 est visible sur tout l'ensemble d'images utilisées, la méthode de mise en correspondance éloignée n'a pas toujours obtenu un nombre suffisant d'appariements (au moins quatre) afin que les homographies correspondant à ce plan-ci puissent être calculées. Pour cette raison, le calibrage avec le plan 1 s'arrête à l'image 87 (noté par $^{(*)}$ dans le tableau), et le calibrage multiplanaire a été effectué sur cette séquence en utilisant uniquement les plans 2 et 3 (étiqueté sur les graphiques et le tableau comme *deux plans*).

Quand les homographies ont été obtenues de manière itérative (figure 7.11a), les résultats du calibrage ont été très éloignés des valeurs attendues : jusqu'à plusieurs centaines de pixels pour le point principal de même que pour la distance focale. Dans certains cas, les valeurs estimées semblent être à la dérive (par exemple, la distance focale pour tous les plans, ou le point principal pour le plan 2).

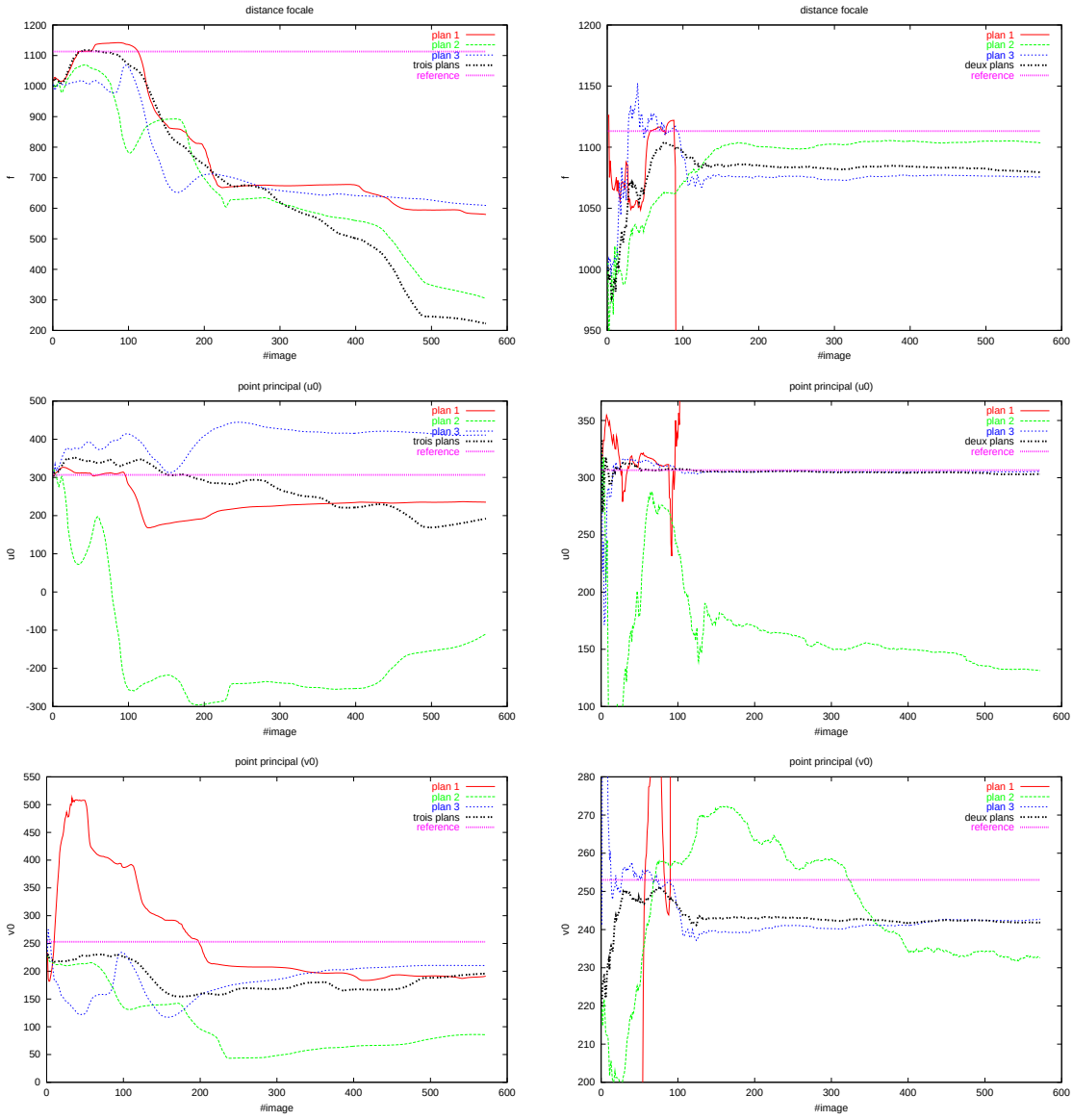
En contrepartie, lorsque la mise en correspondance a été effectuée par rapport à la première image (figure 7.11b), les résultats du calibrage ont été plus proches des valeurs attendues et il n'y a pas de dérive des paramètres intrinsèques, car il n'y a pas d'accumulation d'erreur due à la composition itérative des matrices d'homographie.

Pour le plan 2, nous trouvons que l'erreur sur le point principal est éloignée de 176 pixels bien que la distance focale semble être bien estimée (à moins de 1% d'erreur). La cause de cet écart est que le point principal a dû compenser les erreurs d'estimation des paramètres de rectification (les coefficients $\{a, b, c, d\}$). Pour le plan 3, l'erreur sur la distance focale est de 3.4% et de 10.4 pixels sur le point principal.

Les résultats obtenus avec le calibrage multiplanaire ont été plus stables qu'avec le calibrage à un seul plan. Cependant, les paramètres estimés ne sont pas toujours plus précis qu'avec un seul plan : l'erreur sur le point principal est de 11.7 pixels pour le cas multiplanaire (plus grande que pour le plan 3) et de 3% sur la distance focale (plus élevée que pour le plan 2, mais dans ce cas-là le point principal estimé se trouve assez loin de la position attendue).

Les résultats sont en général plus précis pour la séquence *Mire* que pour la séquence *Paysage*, même si les informations *a priori* pour les deux séquences sont similaires : les paramètres et covariances initiales sont identiques, les positions des plans de calibrage par rapport à la caméra sont les mêmes, et la méthode de calibrage est aussi la même. Les différences dans les résultats sont donc dues uniquement à la qualité de détection et la distribution des points d'intérêt.

Néanmoins, les avantages dans la précision du calcul des paramètres intrinsèques en utilisant la mise en correspondance éloignée au lieu d'une estimation itérative des homographies sont encore plus nettes dans les expériences avec la séquence *Paysage* qu'avec la séquence *Mire*.



(a) (b)

FIG. 7.11 – Paramètres intrinsèques obtenus avec la méthode de Noirfalise pour la séquence *Paysage*, lorsque les homographies inter-images sont calculées (a) de forme itérative ou (b) si la mise en correspondance des points de calibrage a été directement effectuée par rapport à la première image.

7.4.4 Calibrage basé sur la contrainte des lignes centrales

Nous avons testé la méthode de calibrage basée sur la contrainte de la ligne centrale avec les images de la séquence *Mire* (fig. 7.4.a). Plusieurs expériences ont été effectuées pour évaluer les différentes versions de la méthode dérivée de cette contrainte. Les tests de calibrage ont été réalisés sur :

1. plusieurs vues d'un seul plan (d'après la méthode originale proposée dans [GP01]) : les résultats ont été obtenus séparément pour chacun des trois plans de la mire, en prenant plusieurs observations d'un seul plan (figure 7.12) ;
2. une seule vue de plusieurs plans : le point principal a été estimé à partir d'une seule vue, en utilisant les trois plans de la mire et en calculant l'intersection des trois lignes centrales correspondantes (figure 7.14) ;

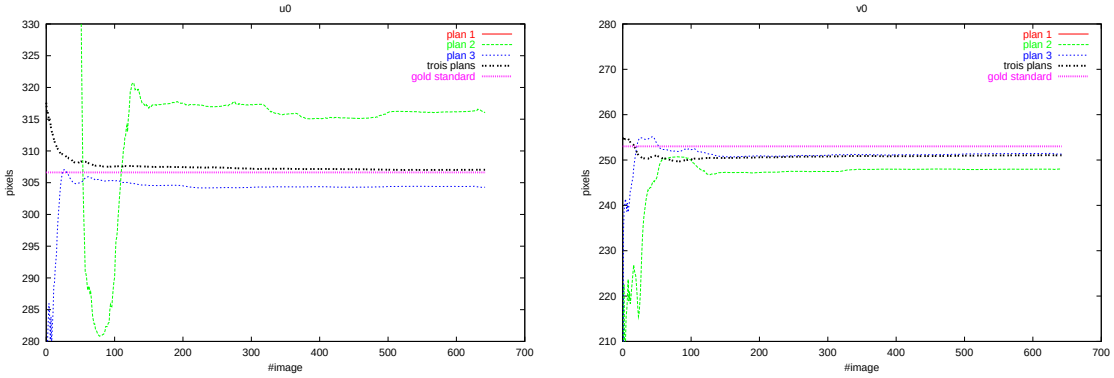


FIG. 7.12 – Point principal estimé par la méthode des lignes centrales, en utilisant un seul plan ou plusieurs.

3. plusieurs vues de plusieurs plans, par deux approches différentes :
   - le point principal est calculé comme l'intersection de toutes les lignes centrales, correspondant aux trois plans par vue, et sur toutes les vues à partir de la première image et jusqu'à l'image i (figure 7.12, courbe des *trois plans*) ;
   - en prenant les résultats des vues indépendantes et en utilisant un filtrage de Kalman, pour raffiner les estimations (figure 7.14, courbe *Kalman*).

Nos contributions concernent les adaptations et l'évaluation de la performance de la méthode lors de l'emploi de plusieurs plans par prise de vue dans le processus de calibrage (les cas 2 et 3 de la liste ci-dessus).

Plusieurs vues d'un seul plan

Sur les graphiques 7.12 et le tableau 7.4, nous montrons les coordonnées du point principal obtenues au moyen de la méthode des lignes centrales (telle qu'elle a été proposée par [GP01], voir l'algorithme dans la section 6.3.4).

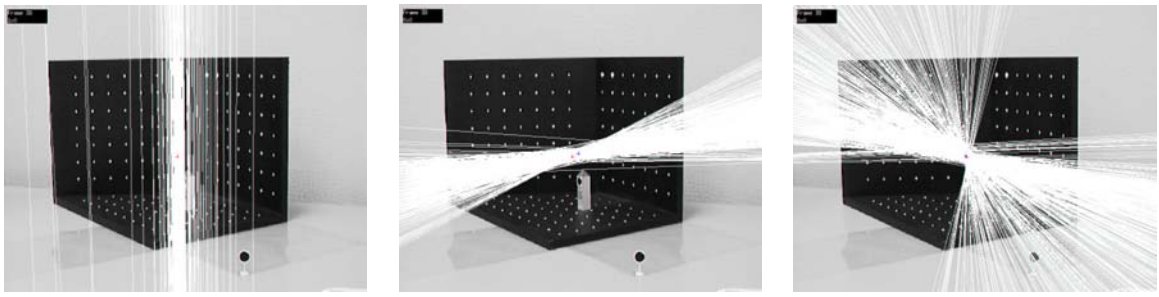
Pour les plans 2 et 3, si un grand nombre d'images est utilisé (plus d'une centaine), les estimées ne varient pas beaucoup lors de l'ajout des nouvelles observations : pour le plan 2, avec une erreur de 10 pixels pour u_0 et de 5 pixels pour v_0, tandis que pour le plan 3, les erreurs sont de 2 pixels sur chacune des coordonnées.

Cependant, pour le plan #1, les estimations sont trop éloignées : les points d'intersection des lignes centrales correspondant à ce plan, se trouvent toujours hors de l'image ; lorsque les images 1 à 642 ont été utilisées (voir tableau 7.4), la position du point principal calculée est $(452.3, 12621.1)$. Le principal problème en utilisant ce plan pour le calibrage est que les lignes centrales du plan pour les différentes vues sont presque parallèles (voir figure 7.13a). Cette difficulté s'explique par le fait qu'une ligne centrale est associée à chaque position d'un plan par rapport à la caméra (non pas à la position des points de calibrage, mais du plan que les contient) ; c'est à dire, un déplacement coplanaire des points de calibrage ne permet pas d'obtenir des lignes centrales différentes. C'est le cas du plan #1 dans notre séquence : le déplacement de la caméra sur la table robotisée (voir section 7.3) a été effectué sur un plan qui est parallèle au plan de la base de la mire ; alors, la position relative du plan de la base par rapport à la caméra est constante tout au long de la séquence.

Toutes les lignes centrales de cette séquence associées au plan de la base de la mire devraient passer par le point principal et former une seule droite ; néanmoins, on observe que les droites estimées sont parallèles, et même si la plupart de lignes sont superposées (le faisceau de droites près du centre de l'image, dans la figure 7.13a), d'autres droites se trouvent décalées. Ce décalage entre les droites est dû aux erreurs de calcul des homographies scène-image. Les lignes centrales les plus éloignées du faisceau central, correspondent aux dernières images de la séquence, dont le nombre de points d'intérêt détectés sur le plan #1 est réduit (moins d'une dizaine), et la distribution de ces points de calibrage n'occupe

	$u_0 \pm \sigma_{u_0}$	$v_0 \pm \sigma_{v_0}$
Plan 1	452.303 ± 31.418	12621.09 ± 2677.9
Plan 2	316.025 ± 11.149	248.058 ± 2.232
Plan 3	304.281 ± 2.008	251.305 ± 1.387
Trois plans	307.033 ± 0.666	251.031 ± 0.774
Filtrage de Kalman	307.067 ± 0.663	250.950 ± 0.771
Valeurs de référence	306.625	253.017

TAB. 7.4 – Estimées du point principal en utilisant la contrainte des lignes centrales : résultats obtenues en utilisant les images 0 à 642 de la séquence *Mire*.



(a) (b) (c)

FIG. 7.13 – Lignes centrales pour les différents plans de la mire de calibrage, obtenues sur la séquence complète : (a) plan de la base, (b) plan à droite, (c) plan à gauche (l'image du fond correspond à la vue #38 de la séquence).

qu'une petite zone de l'image ; donc, l'estimation de l'homographie correspondante n'est pas assez précise ce qui est à l'origine de cette absence de correspondance entre les droites.

Lorsque l'orientation du plan de calibrage par rapport à la caméra est constante, les lignes centrales sont invariantes, et donc l'estimation du point principal n'est pas faisable.

L'inclinaison de la ligne centrale varie plus pour les deux autres plans. Pour le plan #3 la différence d'inclinaison entre les droites est plus grande que pour le plan #2. Ces variations permettent à la méthode de mieux calculer l'intersection entre les droites, et cela se reflète sur la précision du calcul du point principal, qui est d'environ 10.6 pixels pour le plan #2, et 2.9 pixels pour le plan #3.

D'ailleurs, l'affichage de la ligne centrale sur l'image lorsqu'on effectue le processus de calibrage peut être très utile pour montrer les variations dans l'inclinaison des droites issues des images.

Une vue des trois plans

Dans les derniers résultats, nous avons observé l'amélioration dans la performance de la méthode grâce à l'emploi de plusieurs plans de calibrage sur chaque prise de vue. Dans ce cas-là, le processus de calibrage est effectué à l'aide de plusieurs vues de plusieurs plans, et la solution numérique avec la totalité des acquisitions est effectuée à la fin de la séquence.

Chaque vue individuelle d'au moins deux plans de calibrage, nous permet d'estimer les coordonnées du point principal. Dans nos expériences, nous avons estimé pour chaque image le point principal par l'intersection des lignes centrales correspondant aux trois plans de la mire (7.14).

Le tableau 7.5 montre les écarts moyens et maximaux de ces estimations avec et sans pondération, classifiés en trois groupes d'images : le premier groupe où la caméra est tout près de la mire, le deuxième groupe où la caméra se déplace tout au long de la table robotisée, et le troisième groupe où la caméra se trouve dans la position la plus éloignée par rapport à la mire. On observe que les écarts varient en fonction de la distance entre la caméra et la mire. Pour le premier groupe, même si l'erreur moyenne sur le point principal est d'environ 5 pixels, l'écart maximal est de 14.3 pixels.

	Images 1 - 76				Images 77 - 565				Images 566 - 642			
pondération	sans		avec		sans		avec		sans		avec	
	u_0	v_0	u_0	v_0	u_0	v_0	u_0	v_0	u_0	v_0	u_0	v_0
écart moyen	1.9	4.4	1.8	4.0	2.7	2.5	2.7	2.4	55.1	5.9	24.5	4.8
écart max	11.0	9.1	11.0	8.8	23.4	11.2	23.1	11.1	513.1	53.1	346.7	23.5

TAB. 7.5 – Écarts moyens et maximaux lors du calibrage multiplanaire avec une seule image, à l'aide de la méthode de lignes centrales.

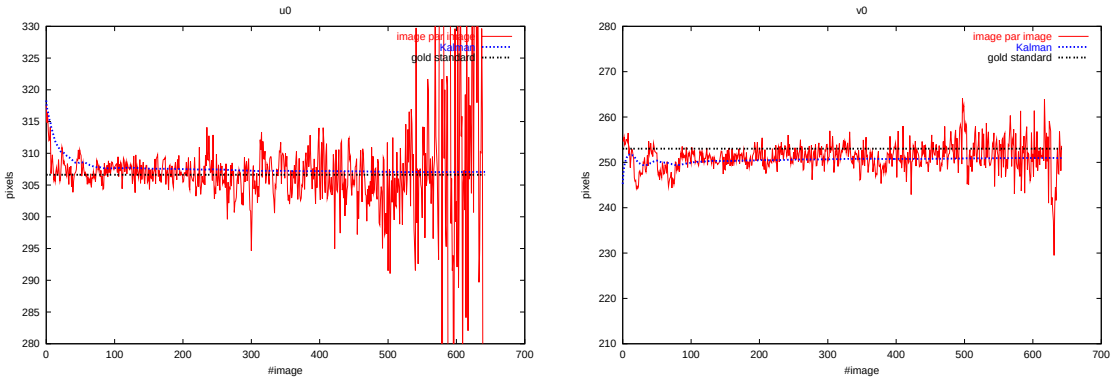


FIG. 7.14 – Point principal estimé par la méthode des lignes centrales, image par image et par filtrage de Kalman.

Lorsque nous utilisons la stratégie de pondération statistique, les écarts entre les points principaux estimés et leurs positions attendues se réduisent. Pour les deux premiers groupes, les différences ne sont pas importantes (moins d'un demi-pixel), mais pour le troisième groupe, les différences peuvent être de plusieurs dizaines et même de centaines de pixel (voir l'écart maximal).

Plusieurs vues des trois plans

En utilisant la même démarche expérimentale sur la séquence de la mire employée pour mesurer la performance des autres méthodes de calibrage planaire, nous avons testé cette méthode en prenant les trois plans de la mire et plusieurs images pour estimer la position du point principal.

Nos résultats (figure 7.12) montrent qu'un nombre inférieur d'images (une dizaine), par rapport au calibrage avec un seul plan, a été nécessaire pour obtenir des estimées proches aux valeurs attendues : l'écart entre le point principal estimé et sa position attendue est 2 pixels (0.4 pour u_0 et 1.9 pour v_0).

Ces résultats sont nettement plus précis que ceux correspondant au calibrage avec un seul plan. En connaissant la signification géométrique de la contrainte, on peut dire que cette amélioration dans la performance est due à la diversité d'orientations des lignes centrales sur les images. Dans ce cas-ci, les différentes orientations des trois plans nous permettent d'obtenir une solution assez précise pour la position du point principal avec très peu de vues.

Raffinement itératif du point principal par filtrage de Kalman

Comme pour certaines configurations, le point principal peut être estimé avec une bonne précision lors du calibrage avec une seule vue de plusieurs plans, et vu que l'estimation avec plusieurs vues est assez précise mais requiert que la solution numérique soit obtenue sur tout l'ensemble de prises de vues, nous proposons une démarche intermédiaire entre ces deux approches basée sur un filtrage itératif :

- La position initiale du point principal est estimée par la méthode des lignes centrales à partir d'une seule vue de plusieurs plans, ainsi que la matrice de covariances associée à cette position. Ceux-ci sont les paramètres initiaux d'un filtre de Kalman.

- Des nouvelles mesures sont obtenues en estimant l'intersection des lignes centrales à partir d'une seule vue de plusieurs plans (et la matrice de covariances correspondant) afin de raffiner le point principal calculé. Les hypothèses du filtre de Kalman pour ce processus sont : le point principal est constant et les mesures correspondantes sont bruitées (dont la distribution initiale est estimée par la méthode de propagation d'erreurs pour la première vue).

Les résultats obtenus lors de l'utilisation de ce processus itératif sont montrés sur les graphiques 7.14 et dans le tableau 7.4. On observe que les estimations faites à l'aide de cette stratégie sont très similaires à celles obtenues quand on calcule le point principal en minimisant la fonction globale décrite par l'équation (6.18), mais l'avantage principal est que le calibrage peut être effectué en ligne de manière itérative, sans que cela nécessite de garder les données (i.e. les points d'intérêt ni les homographies) des vues précédentes.

7.4.5 Calibrage monoplan vs. multiplan : un cas d'étude pratique

Les cas d'étude analysés dans les sections précédentes correspondent à des séquences d'images dont la scène observée contient plusieurs régions planes texturées visibles sur toutes les prises de vues. Dans ces cas d'étude, les approches multiplanaires se sont avérées plus efficaces dans le calcul précis et stable des paramètres intrinsèques par rapport aux approches basées sur l'observation d'un seul plan (dites *monoplanaires*), quand les positions de la caméra par rapport aux plans observés, ainsi que les déplacements de la caméra sur la scène, sont les mêmes.

Par exemple, lorsque l'information d'entrée pour le calcul des paramètres intrinsèques est une séquence de vidéo pre-enregistrée, la stratégie de calibrage doit tirer avantage de la présence de plusieurs surfaces planes si elles sont déjà présentes sur les images.

Un cas de figure différent consiste à filmer une scène dans le but spécifique d'obtenir une séquence d'images pour le calibrage interne si aucun objet de calibrage est disponible. Si la scène est composée de plusieurs surfaces planes, le calibrage interne peut donc être effectué :

- soit en observant un seul plan rapproché de la caméra, de telle manière que les points d'intérêt du plan soient largement distribués sur chaque image ;
- soit en visant plusieurs plans, ce qui permet d'avoir des surfaces avec des orientations et distributions de points d'intérêt assez variées, mais au détriment de la qualité des homographies estimées, puisque l'observation simultanée de plusieurs plans impose la contrainte d'observer chacune des régions dans une portion réduite de l'image et limite aussi le rang de déplacements de la caméra afin de toujours regarder la totalité des plans.

Afin de connaître laquelle de ces deux démarches est la plus convenable, dans le sens de la précision de calcul des paramètres intrinsèques, nous avons utilisé la séquence *Paysage2* (décrite dans la section 7.3.2, voir figure 7.5) prise caméra à la main, qui comprend les deux cas cités ci-dessus à manière de deux ensembles d'images (90 images sur chaque ensemble). La mise en correspondance des points d'intérêt a été effectuée par rapport à l'image clé.

Nous avons testé les méthodes d'auto-calibrage planaire de Noirfalise et de Gurdjos-Sturm sur trois ensembles de données :

- *Trois plans*, les trois surfaces planes sont visibles sur l'ensemble d'images (figure 7.5a et 7.5c) et affectées au processus de calibrage.
- *Plan 3 rapproché*, la caméra a été rapprochée au plan #3 de la scène (plan à gauche, voir figure 7.5b et 7.5d) pendant l'enregistrement de la séquence et ce plan-ci est employé comme le seul plan de calibrage dans ce cas.
- *Plan 3 éloigné*, les images utilisées sont les mêmes que pour l'ensemble des *Trois plans*, mais le plan #3 est le seul à être pris en compte pour le calibrage, cela dans le but de comparer les résultats avec l'ensemble *Plan 3 rapproché* et de montrer l'influence de la distance entre le plan et la caméra sur la précision du calibrage interne.

Les résultats de cette expérience (voir tableau ci-dessous) montrent que la position du point principal a été estimée avec une meilleure précision avec la méthode de Noirfalise qu'avec la méthode de lignes centrales, sur les trois ensembles de données.

	Méthode de Noirfalise		Méthode de lignes centrales	
Trois plans	$f = 958.96$ $u_0 = 307.76$ $v_0 = 226.08$	$\Delta f = +69.94$ $\Delta u_0 = +2.55$ $\Delta v_0 = -20.39$	$u_0 = 279.2$ $v_0 = 256.96$	$\Delta u_0 = -26.01$ $\Delta v_0 = +10.49$
Plan 3 rapproché	$f = 894.96$ $u_0 = 310.94$ $v_0 = 245.51$	$\Delta f = +5.94$ $\Delta u_0 = +5.73$ $\Delta v_0 = -0.96$	$u_0 = 315.32$ $v_0 = 246.21$	$\Delta u_0 = +10.11$ $\Delta v_0 = -0.26$
Plan 3 éloigné	$f = 652.2$ $u_0 = 406.41$ $v_0 = 592.23$	$\Delta f = -236.82$ $\Delta u_0 = +101.2$ $\Delta v_0 = +345.76$	$u_0 = 543.11$ $v_0 = 1348.35$	$\Delta u_0 = +237.9$ $\Delta v_0 = +1101.88$
Paramètres de référence	$f = 889.02$ $\tau = 1.000105$ $(u_0, v_0) = (305.21, 246.47)$			

Cependant, la méthode de Noirfalise possède une difficulté liée à son caractère itératif : comme le filtre de Kalman est restreint à des estimations linéaires du modèle des observations, les résultats obtenues ne sont qu'une approximation de la solution par moindres carrés ajustée à mesure que des nouvelles observations sont ajoutées. Ainsi, si l'ordre d'entrée des homographies est modifié, les résultats peuvent être différents. Dans l'expérience, l'ordre d'ajout des homographies au processus a été inversé (par rapport au résultats du tableau ci-devant) et les résultats obtenus sont les suivants :

- pour le *Plan 3 rapproché* : $f = 897.45, u_0 = 310.28, v_0 = 253.61$
- pour le *Trois plans* : $f = 708.93, u_0 = 406.92, v_0 = 582.47$

Malgré l'avantage de la méthode de Noirfalise de permettre le calibrage interne en ligne de manière itérative, la qualité des résultats n'est pas toujours garantie, à la différence de la méthode de lignes centrales qui nous permet la répétabilité des résultats.

7.5 Discussion

Les chapitres 6 et 7 étaient dédiés au calcul des paramètres intrinsèques de la caméra à partir de l'observation d'objets plans.

Comme les contraintes de calibrage employées sont issues de principes géométriques différents, les méthodes dérivées de ces contraintes présentent des performances différentes lors de leur mise en œuvre. Dans ce chapitre, nous avons comparé expérimentalement quatre techniques de calibrage interne planaire [SM99, GP01, NJS03, GS03] issues de la littérature du domaine de la vision par ordinateur et nous avons proposé des versions modifiées de ces techniques :

- En propageant les erreurs de détection des points d'intérêt sur les images vers les critères de pénalisation des contraintes de calibrage. L'approche proposée pondère les mesures de pénalisation issues des différentes images ou plans. L'emploi de cette stratégie montre que les paramètres intrinsèques estimés sont toujours plus précis que ceux obtenus par les méthodes de calibrage originales. Ces résultats sont aussi plus stables, dans le sens où l'ajout de données imprécises a une influence réduite et toujours inférieure à celles des méthodes originales.
- Plusieurs plans sont employés dans le processus de calibrage. Étant donnée une séquence d'images dont la scène observée est composée par plusieurs surfaces planes, le calcul des paramètres intrinsèques est plus précis lorsqu'on se sert simultanément de tous les plans observés au lieu d'effectuer le calibrage avec un seul de ces plans. Il faut noter que cette assertion est vraie si la pondération statistique des contraintes a eu lieu car, dans le cas contraire, l'ajout d'un plan dont les homographies correspondantes sont très imprécises pourrait nuire à l'estimation des paramètres intrinsèques au lieu de contribuer à leur amélioration.

Cependant, le calibrage multiplanaire impose aux plans observés la restriction de n'occuper que des petites régions dans les images, et la précision de calcul des homographies correspondantes peut être donc inférieure à celle d'une seule de ces surfaces visée en *gros plan*, de telle manière que le calibrage monoplanaire s'avère plus précis dans cette situation.

La précision du calibrage dépend notamment :

- de la précision de calcul des homographies employées dans le processus de calcul, et celle-ci est à la fois liée à la qualité de détection des points d'intérêt, au nombre et distribution des points d'intérêt

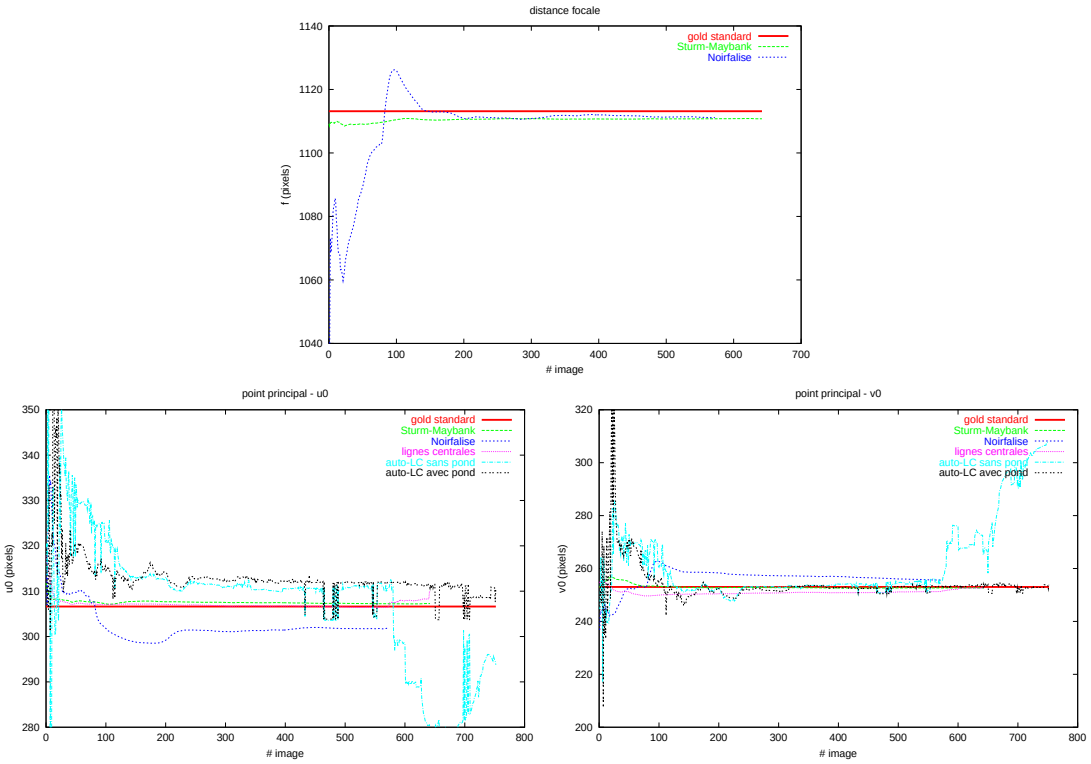


FIG. 7.15 – Comparaison des méthodes de calibrage interne, en utilisant les trois plans de la mire.

sur les images, à la position (distance et inclinaison) du plan par rapport à la caméra, à la méthode de mise en correspondance entre les images, et à la méthode de calcul des homographies (directe ou itérative).

- du type de déplacements de la caméra permettant de contraindre l'espace de solutions des paramètres intrinsèques.

Nous avons examiné deux situations pratiques différentes du problème de calcul des paramètres intrinsèques à partir d'un ensemble de primitives coplanaires :

1. quand on utilise un motif pour le calibrage dont les positions des points d'intérêt contenus dans le motif sont connues (calibrage planaire) ;
2. quand on emploie des surfaces texturées quelconques, de façon qu'uniquement leurs projections et leurs correspondances entre les images sont données (auto-calibrage planaire).

D'après nos expériences, les erreurs sur les paramètres estimés en utilisant le motif de calibrage sont en général inférieures aux erreurs issues du calibrage avec des surfaces texturées. Il existe notamment deux facteurs qui expliquent cette constatation :

- d'une part, parce que les méthodes d'auto-calibrage requièrent le calcul simultané des paramètres intrinsèques et de l'homographie entre le plan de calibrage sur la scène et sa projection sur l'image clé (c'est à dire, la rectification projective du plan).
- d'autre part, parce qu'il existe des méthodes plus précises pour la détection de primitives sur des motifs spéciaux (tels que damiers, grilles, mires avec des cibles circulaires, etc.) que pour la détection de points d'intérêt généraux sur des surfaces quelconques.

Lors du processus d'auto-calibrage, il est nécessaire de proposer des valeurs initiales pour les paramètres intrinsèques et pour les paramètres de rectification du plan (ainsi que pour les matrices initiales des covariances dans la méthode de Noirfalise). Les résultats sont en général assez sensibles à ces estimées initiales.

En particulier, l'initialisation pour la rectification de la surface plane peut être effectuée par l'utili-

sateur en signalant sur l'image (à l'aide d'une interface) quatre points de la surface dont les positions relatives sont connues (e.g. un carré, ou un rectangle si le rapport d'aspect est donné), ce qui permet d'estimer une homographie entre le plan de la scène et l'image. Cependant cette démarche ne peut être introduite que si on trouve un élément sur la scène permettant à l'utilisateur d'indiquer, par exemple, une forme rectangulaire.

Lors de nos expérimentations avec un motif de calibrage, nos résultats montrent que, pour la méthode de Sturm-Maybank, la distance focale est toujours sous-estimée par rapport à sa valeur attendue. Cette même observation a été effectuée pour la méthode d'auto-calibrage planaire de Triggs [Tri98b] mais dans un cadre d'évaluation différent. D'après Kanatani [Kan93a], ce biais systématique est dû au fait que les perturbations sur les points d'intérêt ne se propagent pas de manière linéaire sur les estimées de la distance focale.

Dans la contrainte de la conique absolue, les paramètres internes sont tous regroupés dans les coefficients de la matrice correspondante. La contrainte de la ligne centrale que nous avons utilisé permet par contre de découpler l'estimation du point principal de la distance focale, en évitant que le calcul simultané de tous les paramètres entraînent des erreurs de compensation entre eux.

Cependant, les configurations critiques entre le plan et la caméra sont plus nombreuses pour la méthode des lignes centrales que pour la méthode de la conique absolue :

- Pour la contrainte de la conique absolue, tous les plans parallèles ont les mêmes points circulaires et en conséquence fournissent les mêmes contraintes au processus de calibrage. Ceci est le cas de tous les déplacements de la caméra dont la direction d'incidence de son axe focal par rapport au plan est maintenue constant. Afin d'avoir des nouvelles contraintes, il suffit d'effectuer des rotations de la caméra, autres que autour de l'axe focal.
- Pour la contrainte de la ligne centrale, l'orientation de la ligne de fuite associée au plan de calibrage détermine la position de la ligne centrale (dû à la perpendicularité entre ces deux droites). Comme tous les plans parallèles ont la même ligne de fuite, ils sont aussi reliés à la même ligne centrale. De plus, tous les plans avec des lignes de fuites parallèles génèrent aussi les mêmes contraintes, ce qui rend plus grand le nombre de singularités résultantes de cette contrainte par rapport à celle de la conique absolue.
- Les méthodes basées sur les lignes centrales ne fournissent par contre pas d'estimées de la distance focale et ce paramètre doit donc être calculé par une autre méthode, en supposant par la suite que le point principal n'est plus une variable à déterminer.

La contrainte de la ligne centrale peut offrir une assistance au calibrage à manière d'une représentation graphique de l'espace de solutions du point principal (figure 7.13), de manière à évaluer si les déplacements de la caméra fournissent des nouvelles contraintes au calcul du point principal.

Cette manœuvre ainsi que l'évaluation statistique de la précision des homographies (figure 7.6) et de l'incertitude des paramètres, peuvent être intégrées aux applications pratiques, dans le but d'assister le calibrage interne de caméras, notamment aux utilisateurs quelconques des systèmes de RA.

En conclusion à ce travail, on peut insister sur l'intérêt qu'il y a à mettre en place des interfaces graphiques montrant l'évolution des lignes centrales. Ceci peut servir à guider les mouvements que l'utilisateur fait subir à la caméra afin d'améliorer le calibrage.

Chapitre 8

Conclusion

8.1 Apports de la thèse

Dans cette thèse, nous nous sommes intéressés aux problèmes du calcul du positionnement d'une caméra et de ses paramètres intrinsèques. Nous présentons notre contribution à la recherche de solutions précises et stables à ces deux problèmes et à l'analyse des repercussions des erreurs de calibrage dans un contexte spécifique à la réalité augmentée. Pour effectuer cela, nous avons proposé :

1. Un cadre d'évaluation expérimental nous permettant d'examiner comment le choix du modèle de la caméra et les erreurs de calibrage interne peuvent affecter le calcul du point de vue de la caméra, la projection des objets de synthèse sur la scène, et l'impression visuelle des séquences augmentées, autant dynamique que statique.

   Des études par rapport à ce sujet avaient déjà été effectuées pour d'autres problèmes tels que la navigation robotique [Zha90] ou la reconstruction 3D [LT87, Bou98]. Cependant, nous avons proposé une analyse pour des critères spécifiques à la RA.

   Dans nos expériences, nous avons montré que :
   - le calcul de la totalité des paramètres du modèle de caméra (tel qu'il est fait par certains systèmes de RA [SH00]) entraînent des problèmes de stabilité dans les valeurs estimées, et par conséquent, produisent des effets visuels gênants dans les images augmentées, tels que des oscillations et sauts abrupts.
   - les erreurs de calcul des paramètres intrinsèques et de distorsion radiale entraînent des erreurs sur le positionnement de la caméra, à cause de compensations entre les paramètres. Nous avons examiné analytiquement les relations entre les erreurs des paramètres intrinsèques et les erreurs de projection. Par rapport à la perception visuelle, ces erreurs sont peu importantes lorsque les objets de synthèse sont placés près des primitives de référence employées pour le positionnement de la caméra. En revanche, lorsque les objets incrustés sont placés dans des positions éloignées de ces primitives, les incohérences visuelles sont plus importantes : les erreurs d'alignement sont plus grandes si l'objet de synthèse est placé (i) distant de l'axe focal de la caméra ou (ii) près du plan image de la caméra.
   - certaines hypothèses fréquemment utilisées dans les techniques de calibrage peuvent affecter la performance des systèmes de RA. Deux de ces hypothèses sont : de fixer la position du point principal au milieu de l'image, ou d'affecter le centre de distorsion radiale au point principal, ce qui représente l'emploi de paramètres intrinsèques erronés dans les applications.
2. Un algorithme itératif de calcul robuste du point de vue, basé sur : (i) l'observation de plusieurs plans de référence de la scène (où les plans sont des surfaces texturées quelconques) et (ii) la sélection de modèles de mouvement de la caméra afin de réduire le nombre de paramètres variables dans les estimations.

   Ces stratégies ensemble nous ont permis de réduire deux problèmes importants inhérents aux méthodes itératives : les fluctuations statistiques dans les estimations et l'accumulation des erreurs dans le calcul du point de vue de la caméra. Cela se traduit, dans le cadre des applications de la

RA, comme une diminution des effets visuels gênants de tressaillement et de la dérive des objets incrustés sur la scène réelle.

Notre approche est une extension de la méthode proposée par Kanatani [CM00]. Néanmoins, l'approche multiplanaire que nous proposons : (i) semble beaucoup plus robuste que la méthode monoplanaire originale parce que l'emploi de plusieurs plans fournit une dimension additionnelle qui aide à la stabilité des calculs par rapport à l'utilisation d'un seul plan de référence ; (ii) les possibilités de déplacement de la caméra sont plus grandes lorsque plusieurs plans sont pris en compte.

Une autre contribution à ce sujet a été l'évaluation de plusieurs critères de sélection de modèles de mouvement et la sélection d'un critère adapté au problème du positionnement itératif de la caméra : les critères asymptotiques (comme ceux utilisés par [CM00]) sont souvent utilisés parce qu'ils sont simples à déterminer et facilement aménageables à plusieurs applications, mais ils ne prennent pas en compte les différences de précision des paramètres ni ses corrélations. Après d'avoir évalué les différents critères en situations réelles à l'aide d'une démarche dont les vrais mouvements de la caméra étaient connus, nous avons choisi un critère basé sur la matrice d'information de Fisher des paramètres.

3. Des méthodes de calibrage interne incorporant des techniques de pondération statistique qui permettent d'obtenir des estimations plus stables et précises des paramètres intrinsèques par rapport aux méthodes classiques, en réduisent l'influence des données les plus imprécises dans les calculs.

   Comme partie de nos études sur le calcul des paramètres intrinsèques de la caméra, nous avons mentionné aussi deux stratégies d'assistance au calibrage interne : (i) l'une basée dans la représentation graphique des lignes centrales afin d'indiquer les meilleurs mouvements de la caméra pour obtenir des solutions numériques mieux conditionnées du calcul du point principal ; (ii) l'autre appuyée sur l'estimation théorique des incertitudes des paramètres par la méthode de propagation d'erreurs, de manière à obtenir une mesure de la précision des résultats.

8.2 Perspectives

Les axes de recherche que nous envisageons poursuivre sont : la conception d'interfaces d'assistance pour le calibrage, une étude approfondie sur les méthodes de sélection de modèles de mouvement, et l'extension automatique du modèle multi-planaire.

1. Interfaces pour le calibrage

   Dans les applications de la RA orientées vers le grand public, les utilisateurs doivent effectuer eux-mêmes le calibrage de leur propres caméras. Dans ce but, plusieurs méthodes de calibrage permettent d'effectuer le calibrage à partir de plusieurs vues d'un modèle plan simple [BAA98, Zha00, GP01] (e.g. damier, grille imprimée), d'un poster [GS03] et certains en ligne [NJS03]. Cependant, il existe certains mouvements et positions de la caméra qui sont critiques dans le processus de calibrage et qui peuvent empêcher l'estimation précise des paramètres de la caméra.

   De manière à procurer une assistance pour les utilisateurs quelconques, nous proposons de mettre en place des interfaces qui servent à guider les mouvements de la caméra dans le but d'améliorer les résultats de leur calibrage.

   Comme la ligne centrale [GP01] donne une représentation graphique de l'espace de solutions du point principal, une interface graphique montrant l'évolution de ces lignes aidera à l'utilisateur à choisir les mouvements de la caméra les mieux adaptés afin d'obtenir un calibrage interne précis.

2. Sélection de modèles

   La distinction du modèle de mouvement relatif entre deux vues consécutives peut être difficile pour certains déplacements de la caméra ; par exemple, des très petites translations horizontales peuvent se confondre avec des très petites rotations panoramiques car, dans les deux cas, il semble que les points se déplacent dans la même direction.

   Les critères de sélection de modèles peuvent échouer dans ce type de situations : si la disposition des points d'intérêt sur les images ne fournit pas des informations pour distinguer les deux possibles mouvements, si le nombre de points d'intérêt est faible, si les erreurs de détection des points d'intérêt sont importantes.

De nouveaux critères de sélection basés sur l'apprentissage des modèles [Vap98] se sont montrés plus effectifs que les critères classiques pour le problème de recalage 2D-2D [WDLC04]. Nous envisageons donc d'étudier ces critères de sélection dans le contexte du positionnement 3D de caméras.

Un autre sujet de recherche est l'étude de sélection de modèles non emboîtés. Les critères actuels de sélection concernent le plus souvent des modèles emboités. Par exemple, on peut comparer un modèle de rotation et translation simultanées avec un modèle de rotations pures ou un modèle stationnaire. Néanmoins, comparer la complexité d'un modèle de zoom avec un modèle de translations est une tâche difficile pour les critères de sélection connus. Nos efforts peuvent donc s'orienter vers la recherche de stratégies pour comparer des modèles de mouvements non emboités.

3. Extension du modèle multi-planaire

   L'utilisateur d'un système de RA doit pouvoir se déplacer librement sur l'environnement. Cependant, dans le cas du système planaire utilisé dans nos travaux, les déplacements sont restreints parce qu'au moins l'un des plans de référence repérés initialement doit être toujours visible pendant le suivi de la caméra.

   De manière à augmenter l'autonomie du système de suivi, des nouvelles recherches peuvent s'orienter vers :
   - le prolongement automatique des plans suivis ; c'est à dire, juger si les points d'intérêt apparaissant au fur et a mesure dans les images, appartiennent à l'un des plans de référence connus ;
   - la détection et l'intégration en ligne de nouvelles régions planes : détecter des ensembles de points coplanaires et intégrer le plan correspondant au processus de suivi.

Annexes

Annexe A

Filtre de Kalman étendu

Le filtre de Kalman [KB60] est un estimateur du type récursif. Cela signifie que pour estimer l'état actuel d'un système, seulement l'état précédent et des mesures actuelles des variables du système sont nécessaires. L'ensemble complet des observations et des estimations n'est donc pas requise.

Considérons un système décrit par :

$$\mathbf{x}_k = f_k(\mathbf{x}_{k-1}) + \mu_k \tag{A.1}$$

où $\mathbf{x}_k$ est un vecteur représentant l'état du système à l'instant k, $f(\mathbf{x}_{k-1})$ est une fonction décrivant la transition du système de l'instant $k-1$ à l'instant k, et μ_k es le bruit sur les paramètres du système.

À l'instant k, l'état est mesuré d'accord à l'observation :

$$\mathbf{z}_k = h(\mathbf{x}_k) + \epsilon_k$$

où $h(\mathbf{x}_k)$ est l'opérateur de mesure et ϵ_k est le bruit sur les observations. Les variables aléatoires μ_k et ϵ_k sont supposées indépendantes de moyenne zéro et avec matrices de covariances $\mathbf{Q}_k$ et $\mathbf{R}_k$ respectivement.

L'estimation du vecteur d'état $\mathbf{x}_k$ est séquentielle et effectuée uniquement avec les observations acquises jusqu'au instant k. Dans le cas linéaire, ce problème est résolu par le filtre de Kalman [Jaz70]. Dans le cas non-linéaire, le modèle est en général linéarisé autour du vecteur d'état actuel, donnant origine au filtre de Kalman étendu :

- $f_k(\mathbf{x}_k) \approx \mathbf{A}_k\mathbf{x}_k$, où $\mathbf{A}_k$ est le jacobien de f évalué en $\mathbf{x}_k$;
- $h_k(\mathbf{x}_k) \approx \mathbf{H}_k\mathbf{x}_k$, où $\mathbf{H}_k$ est le jacobien de h évalué en $\mathbf{x}_k$.

Le filtre requiert d'une phase d'initialisation et opère de manière itérative en deux étapes : *prédiction* et *correction.*

Initialisation

Une estimée initiale du vecteur d'état $\mathbf{x}_0$ et de sa matrice de covariances $\mathbf{P}_0$ doivent être fournis au filtre.

Prédiction

L'équation du modèle (A.1) est utilisée pour prédire l'état à l'instant k :

$$\hat{\mathbf{x}}_{k|k-1} = \mathbf{A}_k\hat{\mathbf{x}}_{k-1}$$

La matrice de covariance à cette étape est alors déterminée par :

$$\hat{\mathbf{P}}_{k|k-1} = \mathbf{A}_k\hat{\mathbf{P}}_{k-1}\mathbf{A}_k^T + \mathbf{Q}_k$$

Correction

Une nouvelle observation $\mathbf{z}_k$ est donc utilisée pour corriger la prédiction d'accord à la formule :

$$\hat{\mathbf{x}}_k = \hat{\mathbf{x}}_{k|k-1} + \mathbf{K}_k \tilde{\mathbf{y}}_k$$

où $\tilde{\mathbf{y}}_k$ est l'erreur de prédiction ou *innovation* : $\tilde{\mathbf{y}}_k = \mathbf{z}_k - \mathbf{H}_k \hat{\mathbf{x}}_{k|k-1}$

$\mathbf{K}_k$, appelée le *gain* de Kalman est calculé comme : $\mathbf{K}_k = \hat{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T \mathbf{S}_k^{-1}$

$\mathbf{S}_k$ est l'*innovation de la covariance* : $\mathbf{S}_k = \mathbf{H}_k \hat{\mathbf{P}}_{k|k-1} \mathbf{H}_k^T + \mathbf{R}_k$

L'analyse d'erreur correspondante génère donc la matrice de covariances :

$$\hat{\mathbf{P}}_k = (I - \mathbf{K}_k \mathbf{H}_k) \hat{\mathbf{P}}_{k|k-1}$$

Bibliographie

[ABB+01] R. T. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent Advances in Augmented Reality. *IEEE Computer Graphics and Applications*, pages 34–47, December 2001.

[AF98] Steffen Abraham and Wolfgang Förstner. Calibration errors in structure from motion. In Paul Levi, Rolf-Jürgen Ahlers, Franz May, and Michael Schanz, editors, *DAGM-Symposium*, Informatik Aktuell, pages 117–124. Springer, 1998.

[AHR01] L. Agapito, E. Hayman, and I. Reid. Self-calibration of rotating and zooming cameras. *International Journal of Computer Vision*, 45(2), Nov. 2001.

[Aka74] H. Akaike. A new look at the statistical model identification. *IEEE Trans Aut Ctrl*, 19(6) :716–723, 1974.

[Alo90] J. Aloimonos. Perspective approximations. *Image and Vision Computing*, 8 :179–192, 1990.

[ASB06] Michael Aron, Gilles Simon, and Marie-Odile Berger. Use of inertial sensors to support video tracking. *Computer Animation and Virtual Worlds*, 2006.

[Azu97] R. T. Azuma. A Survey of Augmented Reality. *Presence : Teleoperators and Virtual Environments*, 6(4) :355–385, August 1997.

[BAA98] Jorge Batista, Helder Araujo, and A. T. Almeida. Iterative multi-step explicit camera calibration. In *ICCV*, pages 709–714, 1998.

[Bar00] A. Bartoli. *Reconstruction et alignement en vision 3D : points, droites, plans et caméras.* Thèse de doctorat, Institut National Polytechnique de Grenoble, September 2000.

[Bas94] R. Basri. Paraperspective = affine. Technical report, Weizmann Institute, Jerusalem, Israel, 1994.

[Bas95] A. Basu. Active calibration of cameras : theory and implementation. *IEEE Transactions on Systems, Man, and Cybernetics*, 25(2) :256–265, 1995.

[Bev69] Philip R. Bevington. *Data Reduction and Error Analysis for the Physical Sciences.* McGraw Hill, first edition, 1969.

[BGL05] M. Billinghurst, R. Grasset, and J. Looser. Designing augmented reality interfaces. *computer graphics*, 29(1), 2005.

[BJK+06] G. Bianchi, C. Jung, B. Knoerlein, G. Szekely, and M. Harders. High-fi delity visuo-haptic interaction with virtual objects in multi-modal AR systems. In *Proceedings of International Symposium of Mixed and Augmented Reality, Santa Barbara, USA*, 2006.

[BMB94] P. Brand, R. Mohr, and P. Bobet. Distorsions optiques : correction dans un modèle projectif. In *Actes du 9e Congrès AFCET Reconnaissance des Formes et Intelligence Artificielle, Paris*, volume 1, pages 87–98, January 1994.

[Bou98] S. Bougnoux. From Projective to Euclidiean Space under any Practical Situation, a Criticism of Self-calibration. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 790–796, January 1998.

[Boz87] H. Bozdogan. Model Selection and Akaike's Information Criterion (AIC) ; The General Theory and its Analytical Extensions . *Psychometrika*, 52(3) :345–370, 1987.

[Bro71] D.C. Brown. Close-Range Camera Calibration. *Photogrammetric Engineering*, 8(37) :855–866, 1971.

[BS98] K. Bubna and C.V. Stewart. Model selection and surface merging in reconstruction algorithms. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 895–902, 1998.

[BS01] Adrien Bartoli and Peter Sturm. Three new algorithms for projective bundle adjustment with minimum parameters. Technical Report RR-4236, INRIA, Aug 2001.

[BSN06] Christoph Bichlmeier, Tobias Sielhorst, and Nassir Navab. The tangible virtual mirror : New visualization paradigm for navigated surgery. In *International Workshop on Augmented Reality environments for Medical Imaging and Computer-aided Surgery (AMI-ARCS 2006), Copenhagen, Denmark*, 2006.

[CFJS02] A. Chiuso, P. Favaro, H. Jin, and S. Soatto. Structure from Motion Causally Integrated Over Time. *IEEE Transactions on PAMI*, 24(') :523–535, april 2002.

[CJ03] D. Cobzas and M. Jägersand. A comparison of viewing geometries for augmented reality. In *Image Analysis, 13th Scandinavian Conference, SCIA 2003, Halmstad, Sweden, June 29 - July 2, 2003, Proceedings*, pages 501–508, 2003.

[CM00] K. Kanatani C. Matsunaga. Calibration of a Moving Camera Using a Planar Pattern : Optimal Computation, Reliability Evaluation and Stabilization by Model Selection. In *Proceedings of 6th European Conference on Computer Vision, Trinity College Dublin (Ireland)*, pages 595–609, 2000.

[CP00] L.F. Cheong and C.-H. Peh. Characterizing depth distortion due to calibration uncertainty. In D. Vernon, editor, *European Conference Computer Vision ECCV00*, pages 664–677, Dublin, Ireland, June 2000.

[CRZ97] A. Criminisi, I. Reid, and A. Zisserman. A plane measuring device, 1997.

[CT90] B. Caprile and V. Torre. Using Vanishing Points for Camera Calibration. *International Journal of Computer Vision*, 4 :127–140, 1990.

[DB93] F. Du and M. Brady. Self-Calibration of the Intrinsic Parameters of Cameras for Active Vision Systems. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, NY (USA)*, pages 477–482, June 1993.

[DC99] T. Drummond and R. Cipolla. Real-time tracking of complex structures with on-line camera calibration. In *Proceedings of the British Machine Vision Conference, BMVC 99*, 1999.

[DD92] Daniel DeMenthon and Larry S. Davis. Exact and approximate solutions of the perspective-three-point problem. *IEEE Transactions on PAMI*, 14(11) :1100–1105, 1992.

[DD95] D. Dementhon and L. Davis. Model Based Object Pose in 25 Lines of Code. *International Journal of Computer Vision*, 15 :123–141, 1995.

[DM02] Andrew J. Davison and David W. Murray. Simultaneous localization and map-building using active vision. *IEEE Transactions on PAMI*, 24 :865–880, 2002.

[DMM03] A. Davison, W. Mayol, and D. Murray. Real-Time Localisation and Mapping with Wearable Active Vision. In *Proceedings of International Symposium on Mixed and Augmented Reality, Tokyo*, pages 18–27, June 2003.

[Dro93] L. Dron. Dynamic camera self-calibration from controlled motion sequences. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, New York, NY (USA)*, pages 501–506, June 1993.

[Efr86] B. Efron. The Jackknife, the Bootstrap and other resampling plans. In *Regional conference on resampling methods*, Green State University, 1986.

[Enc95] Reyes Enciso. *Auto-calibration des capteurs visuels actifs. Reconstruction 3D active.* Thèse de doctorat, Université de Paris-Sud U.F.R Scientifique d'Orsay, December 1995.

[EVF93] R. Enciso, T. Viéville, and O. Faugeras. Approximation du Changement de Focale et de Mise au Point par une Transformation Affine à Trois Paramètres. Rapport de recherche 2071, INRIA, 1993.

[FB80] M. A. Fischler and R. C. Bolles. Random Sample Consensus : A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. Technical Note 213, Artificial Intelligence Center, SRI International, Menlo Park, California, March 1980.

[FGR93] A. Fournier, A. S. Gunawan, and C. Romanzin. Common Illumination between Real and Computer Generated Scenes. In *Proceedings of Graphics Interface'93*, pages 254–262, May 1993.

[FLM92] O. D. Faugeras, Q.-T. Luong, and S. J. Maybank. Camera Self-Calibration : Theory and Experiments. In *Proceedings of 2nd European Conference on Computer Vision, Santa Margherita Ligure (Italy)*, pages 321–334, May 1992.

[FM96] G. Florou and Roger Mohr. What accuracy for 3d measurements with cameras ? In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna, Austria*, volume I, pages 354–358. IEEE Computer Society Press, 1996.

[FT87] O. D. Faugeras and G. Toscani. Camera Calibration for 3D Machine Vision. In *Proceedings of International Workshop on Machine Vision and Machine Intelligence*, Tokyo, 1987.

[FTG01] V. Ferrari, T. Tuytelaars, and L. Van Gool. Markerless augmenetd reality with a real time affine tracker. In *International Symposium on Augmented Reality, Los Alamitos*, 2001.

[Gag99] A. Gagalowicz. Use of Image Analysis/Synthesis Techniques for 3D Object Tracking. In *Advanced Research Workshop on Confluence of Computer Vision and Computer Graphics, Ljubljana, Slovenia*, August 1999.

[GCP02] P. Gurdjos, A. Crouzil, and R. Payrissat. Another Way of Looking at Plane-Based Calibration : the Centre Circle Constraint . In *7th European Conference on Computer Vision (ECCV'2002)*, volume IV, pages 252–266, 2002.

[GCS+03] Simon Gibson, Alan Chalmers, Gilles Simon, Flavio Vigueras, Marie-Odile Berger, Didier Stricker, and Wolfram Kresse. Photorealistic augmented reality. In *Second IEEE and ACM International Symposium on Mixed and Augmented Reality - ISMAR'03, Tokyo, Japon.* IEEE, ACM, Oct 2003.

[Gen92] D. Gennery. Visual Tracking of Known Three Dimensional Objects. *International Journal of Computer Vision*, 7(3) :243–270, 1992.

[GKB+05] S. Gorges, E. Kerrien, M.-O. Berger, Y. Trousset, and J. Pescatore. Calibration of the intrinsic parameters of a vascular c-arm for 3d enhanced fluoroscopy. In *Computer Assisted Radiology and Surgery, Berlin*, 2005.

[Gol89] G. Golub. *Matrix computations.* Johns Hopkins University, Johns Hopkins U., 1989.

[GP01] P. Gurdjos and R. Payrissat. Plane-based Calibration of a Camera with Varying Focal Length : the Centre Line Constraint. In *12th British Machine Vision Conference (BMVC 2001)*, volume 2, pages 623–632, 2001.

[GRS+02] Y. Genc, S. Riedel, F. Souvannavong, C. Akinlar, and N. Navab. Marker-less tracking for ar : A learning-based approach. In *Proceedings of International Symposium on Mixed and Augmented Reality, Darmstadt, Germany*, pages 295–304, September 2002.

[GS03] P. Gurdjos and P. Sturm. Methods and Geometry for Plane-Based Self-Calibration . In *International Conference on Computer Vision and Pattern Recognition (CVPR 2003)*, 2003.

[HA96] A. Heyden and K. Aström. Euclidean reconstruction from constant intrinsic parameters. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna (Austria)*, pages 339–343, August 1996.

[Har94a] R. Hartley. Self calibration from multiples views with a rotating camera. In *Proceedings of 3rd European Conference on Computer Vision, Stockholm (Sweden)*, pages 471–478, 1994.

[Har94b] R. I. Hartley. Euclidean reconstruction from uncalibrated views. *Lecture Notes in Computer Science*, 825 :237–256, 1994.

[Har94c] Richard Hartley. An algorithm for self calibration from several views. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA (USA)*, pages 908–912, 1994.

[HCD94] Radu Horaud, Stéphane Christy, and Fadi Dornaika. Object pose : The link between weak perspective,paraperspective, and full perspective. Technical report, INRIA, 1994.

[HCLL89] Radu Horaud, Bernard Conio, Oliver Leboulleux, and Bernard Lacolle. An analytic solution for the perspective 4-point problem. *Comput. Vision Graph. Image Process.*, 47(1) :33–44, 1989.

[HDL97] Radu Horaud, Fadi Dornaika, and Bart Lamiroy. Object pose : The link between weak perspective, paraperspective, and full perspective. *Int. J. Comput. Vision*, 22(2) :173–189, 1997.

[HH64] J.M. Hammersley and D.C. Handscomb. *Monte Carlo Methods.* John Wiley & Sons, Ltd, London :Methuen, 1964.

[HJ01] R. Hu and Q. Ji. Camera Self-Calibration from Ellipse Correspondences. In *Proceedings of the IEEE Computer Society Conference on Robotics and Automation*, 2001.

[HJL+89] R. M. Haralick, H. Joo, C. N. Lee, X. Zhuang, V.G. Vaidya, and M. B. Kim. Pose Estimation from Corresponding Point Data. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6), 1989.

[HK02] R.I. Hartley and R. Kaucic. Sensitivity of Calibration to Principal Point Position. In *Proceedings of 6th European Conference on Computer Vision, Trinity College Dublin (Ireland)*, June 2002.

[HLON91] R. M. Haralick, C. Lee, K. Ottenberg, and M. Nölle. Analysis and Solutions of The Three Point Perspective Pose Estimation Problem. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Maui, HI (USA)*, pages 592–598, 1991.

[HS88] C. Harris and M. Stephens. A Combined Corner and Edge Detector. In *Proceedings of 4th Alvey Conference*, Cambridge, August 1988.

[HS96] J. Heikkilä and O. Silvèn. Calibration procedure for short focal length off-the-shelf CCD cameras. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna (Austria)*, pages 166–170, August 1996.

[HTMS82] E. Hall, J. Tio, C. McPherson, and A. Sadjadi. Measuring curved surfaces for robot motion. *The Computer Journal*, pages 42–54, December 1982.

[Hub81] P. J. Huber. *Robust Statistics.* Wiley, New York, 1981.

[HZ00] R. I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision.* Cambridge University Press, ISBN : 0521623049, 2000.

[iOMS81] Tu ichi Ohta, Kiyoshi Maenobu, and Toshiyuki Sakai. Obtaining surface orientation from texels under perspective projection. In *Proceedings of 7th International Joint Conference on Artificial Intelligence, Vancouver, B.C. (Canada)*, volume 2, pages 746–751, 1981.

[Jaz70] A. M. Jazwinsky. *Stochastic Process and Filtering Theory.* Academic Press, 1970.

[JBL+00] "S. Julier, Y. Baillot, M. Lanzagorta, D. Brown, and Rosenblum". Bars : Battlefield augmented reality system. In *NATO Symposium on Information Processing Techniques for Military Systems, Istanbul, Turkey.*, pages 9–11, October 2000.

[JM01] F. Jurie and M.Dhome. Real time 3d template matching. In *International Conference on Computer Vision and Pattern Recognition, Hawaii, USA*, pages 791–796, 2001.

[JSN04] Frédéric Jurie, Peter Sturm, and Eric Noirfalise. Autocalibration itérative de caméras à partir de scènes planaires. In *Actes du 13e Congrès de Reconnaissance des Formes et Intelligence Artificielle (RFIA'2004), Toulouse*, 2004.

[Kan93a] K. Kanatani. *Geometric Computation for Machine Vision*, volume 37 of *Oxford Engineering Science Series.* Clarendon Press, 1993.

[Kan93b] K. Kanatani. Renormalization for Unbiaised Estimation. In *Proceedings of 4th International Conference on Computer Vision, Berlin (Germany)*, pages 599–606, 1993.

[Kan96] K. Kanatani. *Statistical Optimization for Geometric Computation.* Elsevier Science, Amsterdam, 1996.

[Kan02] K. Kanatani. Model Selection for Geometric Inference. In *Proceedings of 5th Asian Conference on Computer Vision, Melbourne, Australia*, pages 23–25, 2002.

[KB60] R. E. Kalman and R. S. Bucy. New Results in linear filtering and prediction theory. *Trans. ASME, journal of Basic Eng.*, 38 :95–101, 1960.

[KB99] H. Kato and M. Billinghurst. Marker Tracking and HMD Calibration for a Video-based Augmented Reality Conferencing System. In *Proceedings of the 2nd International Workshop on Augmented Reality, San Francisco*, 1999.

[KD02] G. Klein and T. Drummond. Tightly Integrated Sensor Fusion for Robust Vision Tracking. In *Proceedings of the British Machine Vision Conference, Cardiff*, pages 787–796, September 2002.

[KKK02] J.-S. Kim, H.-W. Kim, and I.-S. Kweon. A Camera Calibration Method Using Concentric Circles for Vision Applications. In *Proceedings of 5th Asian Conference on Computer Vision, Melbourne, Australia*, volume 2, pages 515–520, 2002.

[KV96] K. N. Kutulakos and J. R. Vallino. Affine Object Representations for Calibration-Free Augmented Reality. In *Proceedings of 1996 IEEE Virtual Reality Annual International Symposium*, pages 25–36, 1996.

[LF97a] Quang-Tuan Luong and Olivier Faugeras. Scene Motion and Structure recovery from point correspondences and Fundamental matrices. *International Journal of Computer Vision*, 3(22) :261–289, 1997.

[LF97b] Quang-Tuan Luong and Olivier Faugeras. Self-Calibration of a Moving Camera from point correspondences and Fundamental matrices. *International Journal of Computer Vision*, 3(22) :261–289, 1997.

[LH81] H. C. Longuet-Higgins. A computer program for reconstructing a scene from two projections. *Nature*, 293 :133–135, September 1981.

[Lin99] D. Lingrand. Approximations de la projection perspective : modèle combiné et étude des singularités homographiques. Rapport de recherche 3682, INRIA, 1999.

[LL95] Mengxiang Li and Jean-Marc Lavest. Some Aspects of Zoom-Lens Camera Calibration. Technical Report TRITA-NA-9503, CVAP, Royal Institute of Technology, Stockholm, Sweden, 1995.

[Low85] D. G. Lowe. *Perceptual Organization and Visual Recognition.* Kluwer Academic Publishers, Norwell, Massachusetts, 1985.

[Low92] D. Lowe. Robust Model based Motion Tracking Through the Integration of Search and Estimation. *International Journal of Computer Vision*, 8(2) :113–122, 1992.

[Low99] D. G. Lowe. Object recognition from local scale-invariant features. In *Proceedings of 7th International Conference on Computer Vision, Greece*, pages 1150–1157, Los Alamitos, CA, 1999.

[Low04] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2) :91–110, 2004.

[LT87] R. K. Lenz and R. Y. Tsai. Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3D Machine Vision Metrology. In *IEEE International Conference on Robotics and Automation*, pages 68–75, Raleigh, 1987.

[LT88] R. K. Lenz and R. Y. Tsai. Techniques for Calibration of the Scale Factor and Image Center for High Accuracy 3-D Machine Vision Metrology. *IEEE Transactions on PAMI*, 10(5) :713–720, 1988.

[Luo92] Q. T. Luong. *Matrice fondamentale et calibration visuelle sur l'environnement, vers une plus grande autonomie des systèmes robotiques.* Thèse de doctorat, Université de Paris Sud, centre d'Orsay, December 1992.

[LV94] Q. Luong and T. Viéville. Canonic Representations for the Geometries of Multiple Projective Views. In *Proceedings of 3rd European Conference on Computer Vision, Stockholm (Sweden)*, volume 1, pages 589–599, 1994.

[LV96] D. Lingrand and T. Viéville. Dynamic foveal 3d sensing using affine models. In *Proceedings of the 13th International Conference on Pattern Recognition, Vienna (Austria)*, page 810, Washington, DC, USA, 1996. IEEE Computer Society.

[LVD98] J.M. Lavest, M. Viala, and M. Dhome. Do we really need an accurate calibration pattern to achieve a reliable camera calibration? In *Proceedings of 5th European Conference on Computer Vision, University of Freiburg (Germany)*, pages 158–174, 1998.

[LVTF03] V. Lepetit, L. Vachetti, D. Thalmann, and P. Fua. Fully Automated and Stable Registration for Augmented Reality Applications. In *Proceedings of International Symposium on Mixed and Augmented Reality, Tokyo*, pages 93–101, June 2003.

[LZ99] D. Liebowitz and A. Zisserman. Combining scene and auto-calibration constraints. In *Proceedings of 7th International Conference on Computer Vision, Greece*, September 1999.

[MC00a] E. Malis and F. Chaumette. 2 1/2 d visual servoing with respect to unknown objects through a new estimation scheme of camera displacement. *Int. J. Comput. Vision*, 37(1) :79–97, 2000.

[MC00b] E. Malis and R. Cipolla. Multi-view Constraints between Collineations : Application to Self-Calibration from Unknown Planar Structures. In *Proceedings of 6th European Conference on Computer Vision, Trinity College Dublin (Ireland)*, pages 610–624, June 2000.

[MC02] E. Marchand and F. Chaumette. Virtual visual servoing : a framework for real-time augmented reality. In *EUROGRAPHICS 2002 Conference Proceeding, Computer Graphics Forum, Volume 21(3)*, 2002.

[MDB98] J. Mendelsohn, K. Daniilidis, and R. Bajcsy. Constrained Self-Calibration for Augmented Reality Registration. In *First International Workshop on Augmented Reality, San Francisco*, 1998.

[MDG+95] P. Milgram, D. Drascic, J. Grodski, A. restogo, S. Zhai, and C. Zhou. Merging real and virtual worlds. In *Proceedings of IMAGINA 95*, Monte Carlo, February 1995.

[MF92] S. J. Maybank and O. D. Faugeras. A theory of self-calibration of a moving camera. *International Journal of Computer Vision*, 8(2) :123–152, 1992.

[MFD+02] K. Masamune, G. Fichtinger, A. Deguet, D. Matsuka, and R. Taylor. An Image Overlay System with Enhanced Reality for Percutaneous Therapy Performed Inside CT Scanner. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention, Tokyo, Japan,2002*, 2002.

[MVFCM03] F. Mourgues, T. Vieville, V. Falk, and È. Coste-Manière. Interactive guidance by image overlay in robot assisted coronary artery bypass. In *Proceedings of the Medical Image Computing and Computer Assisted Intervention, Montréal, Canada, November 2003.*, 2003.

[MZ92] J. Mundy and A. Zisserman. Projective Geometry for Machine Vision. In J. Mundy and A. Zisserman, editors, *Geometric Invariance in Computer Vision*, pages 463–519. MIT Press, 1992.

[Nav03] Nassir Navab. Industrial Augmented Reality : Challenges in Design and Commercialization of Killer Apps. In *Proceedings of International Symposium on Mixed and Augmented Reality, Tokyo*, 2003.

[Nic04] Stéphane Nicolau. *Un système de réalité augmentée pour guider les opérations du foie en radiologie interventionnelle.* PhD thesis, Université de Nice-Sophia Antipolis, November 2004.

[NJS03] E. Noirfalise, F. Jurie, and P. Sturm. Autocalibration itérative de caméras à partir de structures planes. In *ORASIS'2003 - Journées francophones des jeunes chercheurs en vision par ordinateur*, May 2003.

[NPSA04] Stephane Nicolau, Xavier Pennec, Luc Soler, and Nicholas Ayache. An accuracy certified augmented reality system for therapy guidance. In *Proc. of the 8th European Conference on Computer Vision (ECCV 04)*, volume LNCS, Prague, May 2004.

[Oli97] J. Oliensis. A Critique of Structure from Motion Algorithms. Technical report, NECI, April 1997.

[OSYT99] T. Ohshima, K. Satoh, H. Yamamoto, and H. Tamura. RV-Border Guards : A Multi-player Entertainment in Mixed Reality Space. In *Proceedings of the 2nd International Workshop on Augmented Reality, San Francisco*, 1999.

[PFTV88] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C, The Art of Scientific Computing.* Cambridge University Press, 1988.

[PHMG06] J Platonov, H. Heibel, P. Meier, and B. Grollmann. A mobile markerless AR system for maintenance and repair. In *Proceedings of International Symposium of Mixed and Augmented Reality, Santa Barbara, USA*, 2006.

[PJN99] J. Park, B. Jiang, and U. Neumann. Vision-based Pose Computation : Robust and Accurate Augmented Reality Tracking. In *Proceedings of the 2nd International Workshop on Augmented Reality, San Francisco*, pages 3–12, 1999.

[PK94] C. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. In *Proceedings of 3rd European Conference on Computer Vision, Stockholm (Sweden)*, volume 2, pages 97–108, 1994.

[PMP+05] J. Ponce, K. McHenry, T. Papadopoulo, M.Teillaud, and B. Triggs. On the Absolute Quadratic Complex and its Application to Autocalibration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, San Diego, CA (USA)*, 2005.

[Pon62] J.-V. Poncelet. *Applications d'analyse et de géométrie - Traité des propriétés projectives des figures. Tome I.* Imprimerie de Mallet-Bachelier, Paris, 1862.

[PvG97] M. Pollefeys and L. van Gool. A stratified approach to metric self-calibration. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR (USA)*, pages 407–412, 1997.

[QL98] L. Quan and Z. Lan. Linear $N \geq 4$-Point Pose Determination. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, January 1998.

[RCR99] T. Drummond R. Cipolla and D. Robertson. Camera calibration from vanishing points in images of architectural scenes. In *Proceedings of the British Machine Vision Conference, BMVC 99*, 1999.

[RDB00] J. Rolland, L. Davies, and Y. Baillot. A survey of tracking technologies for virtual environments. In *Fundamentals of Wearable Computers and Augmented Reality*, 2000.

[RDLW96] S. Ravela, B. Draper, J. Lim, and R. Weiss. Tracking Object Motion Across Aspect Changes for Augmented Reality. In *ARPA Image Undertanding Worshop, Palm Spring (USA)*, August 1996.

[RF00] J. Rolland and H. Fuchs. Optical versus video see-through head-mounted displays in medical vizualisation. *Presence*, 9(3) :287–309, 2000.

[RLB+02] M. Ribo, P. Lang, M. Brandner, C. Stock, and A. Pinz. Hybrid Tracking for Outdoor Augmented Reality Applications. *IEEE Computer Graphics and Applications*, 2002.

[RMAR02] J. Rolland, C. Meyer, K. Arthur, and E. Rinalducci. Method of Adjustments versus Method of Constant Stimuli in the Quantification of Accuracy and Precision of Rendered Depth in Head-Mounted Displays. *Presence*, 11(6) :610–625, 2002.

[RSKM98] D. Reiners, S. Stricker, G. Klinker, and S. Müller. Augmented Reality for Construction Tasks : Doorlock Assembly. In *First International Workshop on Augmented Reality, San Francisco*, 1998.

[SAYT01] K. Satoh, M. Anabuki, H. Yamamoto, and H. Tamura. A hybrid registration method for outdoor augmented reality. In *International Symposium on Augmented Reality, Los Alamitos*, pages 67–76, 2001.

[SB98] G. Simon and M.-O. Berger. A Two-stage Robust Statistical Method for Temporal Registration from Features of Various Type. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 261–266, January 1998.

[SB02a] G. Simon and M.-O. Berger. Pose estimation for planar structures. *IEEE Computer Graphics and Applications, special issue on Tracking*, pages 46–53, nov 2002.

[SB02b] Gilles Simon and Marie-Odile Berger. Real Time Registration of Known or Recovered Multi-planar Structures : application to Augmented Reality. In *Proceedings of the British Machine Vision Conference, BMVC 02, Cardiff*, pages 567–576, September 2002.

[SBHN06] T. Sielhorst, C. Bichlmeier, S. Heining, and N. Navab. Depth perception a major issue in medical ar : Evaluation study by twenty surgeons. In *Proceedings of Medical Image Computing and Computer-Assisted Intervention (MICCAI 2006), Copenhagen, Denmark, October 2006*, 2006.

[Sch78] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2) :461–464, 1978.

[SFZ00] G. Simon, A. Fitzgibbon, and A. Zisserman. Markerless tracking using planar structures in the scene. In *Proc. International Symposium on Augmented Reality*, pages 120–128, October 2000.

[SH00] Y. Seo and K. Hong. Calibration-Free Augmented Reality in Perspective. *IEEE Transactions on Vizualisation and Computer Graphics*, 6(4), 2000.

[SHC+96] A. State, G. Hirota, D. Chen, W. Garett, and M. Livingston. Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking. In *Computer Graphics (Proceedings Siggraph New Orleans)*, pages 429–438, 1996.

[SK52] J. G. Semple and G. T. Kneebone. *Algebraic Projective Geometry.* Oxford University Press, 1952.

[SK96] R. Szeliski and S. B. Kang. Shape Ambiguities in Structure from Motion. *Lecture Notes in Computer Science*, 1064 :709– ??, 1996.

[SL04] Iryna Skrypnyk and David G. Lowe. Scene modelling, recognition and tracking with invariant image features. In *Proceedings of International Symposium of Mixed and Augmented Reality, Arlington, VA, USA*, 2004.

[Sla80] C.C. Slama, editor. *Manual of Photogrammetry.* American Society of Photogrammetry, 4 edition, 1980.

[SM99] Peter Sturm and S. Maybank. On Plane-Based Camera Calibration : A General Algorithm, Singularities, Applications. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina (USA)*, pages 432–437, 1999.

[SP95] S. Soatto and P. Perona. Dynamic rigid motion estimation from weak perspective. In *ICCV '95 : Proceedings of the Fifth International Conference on Computer Vision*, page 321, Washington, DC, USA, 1995. IEEE Computer Society.

[STN04] T. Sielhorst, J. Traub, and N. Navab. The ar apprenticeship : Replication and omnidirectional viewing of subtle movements. In *Proceedings of International Symposium of Mixed and Augmented Reality, Arlington, VA, USA*, 2004.

[Stu97a] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR (USA)*, pages 1100–1105, 1997.

[Stu97b] P. Sturm. *Vision 3D non calibrée :contribution à la reconstruction projective et études des mouvements critiques pour l'autocalibrage.* Thèse de doctorat, Institut national polytechnique de Grenoble, February 1997.

[Stu00] Peter Sturm. Algorithms for Plane-Based Pose Estimation. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina (USA)*, pages 706–711, 2000.

[Sut68] I. Sutherland. A Head-Mounted Three Dimensional Display . In *Fall Joint Computer Conference, American Federation of Information Processing (AFIPS), Conf. Proc. 33, Thompson Books*, pages 757–764, 1968.

[SUY04] K. Satoh, S. Uchiyama, and H. Yamamoto. A hear tracking method using bird's eye view camera and gyroscope. In *Proceedings of International Symposium of Mixed and Augmented Reality, Arlington, VA, USA*, pages 202–211, 2004.

[TFZ98] P.H.S. Torr, A.W. Fitzgibbon, and A. Zisserman. Maintaining multiple motion model hypotheses over many views to recover matching and structure. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, pages 485–491, 1998.

[TH98] M. Trajkovic and Mark Hedley. Fast corner detection. *Image and Vision Computing*, (16) :75–87, 1998.

[TMHF00] W. Triggs, P. F. McLauchlan, R. I. Hartley, and A. Fitzgibbon. Bundle Adjustment for Structure from Motion. In *Vision Algorithms : Theory and Practice*. Springer-Verlag, 2000.

[Tor97] P.H.S. Torr. An Assessment of Information Criteria for Motion Model Selection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, PR (USA)*, pages 47–52, June 1997.

[Tri97] B. Triggs. Autocalibration and the absolute quadric, 1997.

[Tri98a] B. Triggs. A new approach to geometric fitting. In *Proceedings of 6th International Conference on Computer Vision, Bombay (India)*, January 1998.

[Tri98b] B. Triggs. Autocalibration from planar scenes. In *European Conference on Computer Vision*, Freiburg, June 1998.

[Tsa87] R. Y. Tsai. A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. *IEEE Journal of Robotics and Automation*, 3(4) :323–331, August 1987.

[Vap98] V. Vapnik. *Statistical learning theory*. Wiley, 1998.

[VKA+01] Vassilios Vlahakis, John Karigiannis, Luis Almeida, Didier Stricker, Tim Gleue, Ioannis T. Christou, Renzo Carlucci, and Nikos Ioannidis. Archeoguide : First results of an augmented reality, mobile computing system in cultural heritage sites. In *International Symposyum on Virtual Reality, Archaeology and Cultural Heritage, Glyfada, Greece, 2001*, 2001.

[VKSN02] S. Vogt, A. Khamene, F. Sauer, and H. Niemann. Single camera tracking of marker clusters : Multiparameter cluster optimization and experimental verification. In *Proceedings of International Symposium on Mixed and Augmented Reality, Darmstadt, Germany*, 2002.

[WB96a] Charles Wiles and Michael Brady. Ground plane motion camera models. In *Proceedings of 4th European Conference on Computer Vision, Cambridge (United Kingdom)*, pages 238–247, London, UK, 1996. Springer-Verlag.

[WB96b] Charles Wiles and Michael Brady. On the appropriateness of camera models. In *Proceedings of 4th European Conference on Computer Vision, Cambridge (United Kingdom)*, pages 228–237, London, UK, 1996. Springer-Verlag.

[WCH92] J. Weng, P. Cohen, and M. Herniou. Camera Calibration with Distorsion Models and Accuracy Evaluation. *IEEE Transactions on PAMI*, 14(10) :965–980, 1992.

[WDLC04] H. Wechsler, Z. Duric, F. Li, and V. Cherkassky. Motion estimation using statistical learning theory. *IEEE Transactions on PAMI*, 26(4) :466–479, 2004.

[WHA89] J. Weng, T. S. Huang, and N. Ahuja. Motion and structure from two perspective views : Algorithms, error analysis, and error estimation. *IEEE Transactions on PAMI*, 11(5) :451–477, 1989.

[Wil94] Reg G. Willson. *Modeling and Calibration of Automated Zoom Lenses*. Thèse de doctorat, Carnegie Mellon University, January 1994.

[YNA99] S. You, U. Neumann, and Ronald Azuma. Hybrid inertial and vision tracking for augmented reality registration. In *Proceedings of IEEE VR '99 (Houston, TX, 13-17)*, pages 260–267, 1999.

[YSH00] C. Yang, F. Sun, and Z. Hu. Planar Conic Based Camera Calibration. In *Proceedings of the 15th International Conference on Pattern Recognition, Barcelona (Spain)*, volume 1, pages 555–558, September 2000.

[ZF96] Cyril Zeller and Olivier Faugeras. Camera Self-Calibration from Video Sequences : the Kruppa Equations Revisited. Rapport de recherche 2793, INRIA, February 1996.

[Zha90] Z. Zhang. *Analyse du mouvement à partir d'une séquence de scènes stéréoscopiques et applications à la robotique mobile.* Thèse de doctorat, Université de Paris Sud, centre d'Orsay, October 1990.

[Zha98] Z. Zhang. Image-Based Geometrically-Correct Photorealistic Scene/Object Modeling (IB-PhM) : A Review. In R. Chin and T.-C. Pong, editors, *Proceedings of 3rd Asian Conference on Computer Vision, Hong Kong, China*, volume 1352 of *Lecture Notes in Computer Science*, January 1998.

[Zha00] Z. Zhang. A flexible new technique for camera calibration. *IEEE Transactions on PAMI*, 22(11) :1130–1334, 2000.

[ZK01] Marco Zucchelli and Jana Kosecka. Motion bias and structure distortion induced by calibration errors. In Timothy F. Cootes and Chris Taylor, editors, *BMVC*. British Machine Vision Association, 2001.

[ZS03] H. Zollner and R. Sablatnig. Comparison of methods for geometric camera calibration using planar calibration targets, 2003.

Résumé

Et un petit résumé en français.

Summary

The same in English.