

Σ

Γ

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
**Ministère de l' Enseignement Supérieur
Et de la Recherche Scientifique**
Université de Mentouri – Constantine.
Faculté des Science de l'Ingénieur.
Département de l'Informatique.

N° d'ordre:

Série:

Soutenu le: 10/04/2005

Mémoire
Présenté pour l'obtention du diplôme de
MAGISTER
En Informatique
Option: Information & Computation

Thème

Estimation du Mouvement
pour
la Réalité Augmentée

Présenté par
FERGANI BAHAA

Dirigé par
Pr. MC. BATOUCHE.

Jury

Présidente:	Pr. Z. BOUFAÏDA	Université de Constantine.
Rapporteur:	Pr. MC. BATOUCHE	Université de Constantine.
Examineur:	Dr. M.BENMOHAMMED	Université de Constantine.
Examinatrice:	Dr. S. MESHOU	Université de Constantine.

σ

γ

Remerciements

Je tiens tout d'abord à exprimer ma profonde reconnaissance pour mon encadreur Pr. MC. BATOUCHE.

Je remercie Mr. FEZANI du département d'informatique de Jijel pour son aide et ses conseils. Je tiens aussi à remercier Mr. BOUZNADA qui m'a toujours aidé.

Je remercie Pr. *Z.BOUFAÏDA* d'avoir accepté de présider ce jury, les rapporteurs, Dr. *M.BENMOHAMMED* et Dr. *S. MESHOU*L, d'avoir accepté la lourde tâche d'être rapporteurs et pour la qualité avec laquelle ils ont analysé mon travail.

Je remercie tous les membres de l'équipe "Infographie et vision" et tous les membres du laboratoire " Lire ".

DEDICACES

Je remercie dieu de m' avoir donné l'aide et la volonté de réaliser ce travail.

A mon père , ma mère et à toute ma famille.

A tous ceux que je porte dans mon cœur et que leur amitié m'est très précieuse, et que ni l'encre ni le papier ne suffisent pour leur exprimer mon affection.

A toutes les personnes ayant participé à ce travail, ne serait-ce que par un petit mot d'encouragement ou même un petit sourire.

A tous ceux que j'aime et qui m'aime et à qui je souhaite la réussite et le bonheur.

Que dieu nous aide pour mieux faire.

Résumé

Dans un système de réalité augmentée, la vue réelle de la scène est augmentée en superposant des objets virtuels sur les images de telle sorte qu'ils soient projetés dans la scène avec le même point de vue que celui adopté par la caméra lors du tournage. Le calcul du point de vue pour chaque image ainsi que certaines caractéristiques intrinsèques, est donc primordial pour une application de réalité augmentée.

Généralement les systèmes de la réalité augmentée nécessitent la calibration de la caméra afin d'avoir un alignement précis. Dans ce mémoire, on présente une synthèse des méthodes d'estimation du mouvement et plus précisément les méthodes estimant le mouvement de la caméra (appelé aussi ego-motion et mouvement propre) tout en présentant nos deux plateformes la première est basée sur notre approche "Augmentation basée calibration", la deuxième est basée sur "augmentation sans calibration".

Abstract

Registration in augmented reality aligns the virtual object with the 3D real scene. Accurate registration and positioning of virtual objects in the real environment requires accurate tracking of the user's head (or camera tracking) and sensing the locations of other objects in the environment. The biggest single obstacle to building effective Augmented Reality systems is the requirement of accurate, sensors and trackers that report the locations of the user and the surrounding objects in the environment.

Augmented Reality systems generally require extensive calibration to produce accurate registration. The basic principles of camera calibration are well established, and many manual AR calibration techniques have been developed. In this thesis we present our two frameworks, the first framework is based on our approach based on calibration process, in the second one we tried to avoid calibration process by tracking only four colored fiducials.

Table des matières

Chapitre1

La Réalité Augmentée.

1. Introduction	4
2. Définition	5
3. Les motivations	6
4. Les applications	6
4.1. La médecine	7
4.2. La robotique et la télé-robotique	8
4.3. La maintenance et l'assemblage	8
4.4. La conception (engineering design)	9
4.5. Les systèmes mobiles	9
4.6. La collaboration	10
4.7. La RA collaborative et mobile	11
4.8. Les développements commerciaux	11
5. Les systèmes de la RA	12
5.1. Les systèmes basés casque	12
5.1.1. Les systèmes optiques	12
5.1.2. Les systèmes vidéos	13
5.1.3. Comparaison	14
5.3. Systèmes basés moniteur	15
6. Les problèmes techniques de la RA	15
6.1. Le recalage	16
6.1.1. Le problème de recalage	16
6.1.2. Les erreurs du recalage	17
6.2. L'occlusion	18
6.2.1. Les techniques basées modèle	18
6.2.2. Les techniques basées profondeur	18
7. Conclusion	18

	19
Chapitre2	20
La RA basée vision.	
1. Introduction	
2. La formation d'image	
3. Les modèles de cameras	22
3.1. Le modèle Sténopé	22
3.2. La projection orthographique	25
3.3. La projection perspective faible	25
4. La calibration de la camera	27
5. La reconstruction	27
5.1. La stéréovision	28
5.1.1. Appariement stéréoscopique	30
5.1.2. La triangulation	31
6. Conclusion	32

Chapitre3	37
Analyse du mouvement.	38

1. Introduction	
2. Définitions	
2.1. Le mouvement apparent	40
2.2. Champ de mouvement	41
2.3. Le flot optique	41
	41
2.4. La discontinuité du flot optique	42
2.5. L'analyse du mouvement basé flot	
2.5.1. Les modèles du mouvement 3D-2D	46

3. Les techniques d'estimation du mouvement	48
3.1. Estimation 3D	48
3. 2. L'estimation 2D	49
4. Analyse stéréo-mouvement	49
5. L'estimation du mouvement de la camera	50
5.1. Les techniques basées capteurs	53
5.1.1. Les capteurs magnétiques	54
5.1.2. Les capteurs inertiels	55
5.2. Les techniques visuelles	55
5.2.1. Le mouvement propre et le flot optique	55
5.2.2. Le mouvement propre et le suivi des primitives	56
5.3. Les techniques hybrides	56
6. Estimation du mouvement et le suivi	57
7. Conclusion	59

Chapitre4 **plates-formes développées pour la RA.**

1. Introduction	60
2. Augmentation basée sur la calibration	63
2.1. Appariement 3D-2D	66
2.2. La matrice de projection	68
2.3. La détermination de la matrice fondamentale	70
2.4. La reconstruction du plan de recalage	71
2.4.1. La triangulation	76
2.5. Le positionnement de l'objet virtuel	77
2.5.1. Positionnement 3D	77
2.5.2. Positionnement 2D	78
2.6. Occlusion	78
3. Augmentation sans calibration	79

3.1.Initialisation	80
3.2. Le suivi de la base affine	81
3.3.La reconstruction affine	84
3.4.La transformation objet-affine	84
3.5.Estimation de la matrice de projection	86
3.6. Insertion de l'objet virtuel	86
4. Discussion des résultats	88
5. Conclusion	89
Conclusion générale	90
Références	91

Introduction générale

Le terme *Réalité Augmentée* (RA) est un terme utilisé pour décrire les systèmes permettant d'augmenter la vision de l'utilisateur par ajout d'objets ou d'informations générées par ordinateur. Contrairement à la *réalité virtuelle* (RV) dans laquelle l'utilisateur est complètement immergé dans un environnement virtuel, la RA augmente la perception de l'utilisateur. Cette augmentation peut prendre la forme d'étiquettes, d'objets 3D de synthèse ou de modifications d'ombrage dans la scène.



Le premier et peut être le plus important problème à résoudre est celui du recalage entre les objets des deux mondes réel et virtuel. En effet, pour une visualisation correcte, l'image de l'objet virtuel doit être calculée avec la position de la caméra utilisée pour l'image considéré. Ce recalage doit être réalisé avec soin car l'œil humain détecte très facilement les erreurs, même minimes, de recalage. Un bon recalage nécessite trois cohérences:

- Ø **La cohérence géométrique:** les objets virtuels doivent être projetés au bon endroit de l'image et selon la bonne orientation.
- Ø **La cohérence du mouvement:** l'effet du changement du point de vue (ou du déplacement de la camera) doit être le même sur les objets virtuels et les objets réels.
- Ø **La cohérence photométrique:** pour accroître le réalisme de la scène il est important de prendre en compte les interactions photométriques entre les objets virtuels et la scène (les ombrages en particulier).

L'objectif de ce mémoire est d'examiner l'effet du mouvement de la caméra sur le réalisme de l'augmentation (la cohérence géométrique et la cohérence du mouvement). Ce mémoire présente deux plates-formes, la première est basée sur une approche inspirée de l'approche de Zonglei et de Boufama. Cette dernière utilise la calibration pour estimer le mouvement de la caméra. La deuxième plate-forme est basée sur une approche inspirée de l'approche de Kutulakos et de Vallino, qui est une méthode essayant d'échapper à la lourdeur de la calibration par le suivi de quatre repères.

Par conséquent, ce mémoire est organisé comme suit:

Le premier chapitre est consacré à une étude théorique de la réalité augmentée, le deuxième chapitre présente les notions de base des techniques basées vision utilisées pour la réalité augmentée. Le troisième chapitre présente une étude sur l'estimation du mouvement surtout l'estimation du mouvement de la caméra. Le quatrième chapitre est consacré à l'explication de quelques détails nécessaires à l'implémentation tout en présentant nos résultats.

Chapitre 1

La Réalité Augmentée

" Ce n'est parce que les choses sont difficiles que nous n'osons pas,
c'est parce que nous n'osons pas qu'elles sont difficiles "
Sénèque.

Chapitre1

La Réalité Augmentée

1. Introduction

Contrairement à notre monde réel, un monde virtuel est un monde entièrement modélisé. Ainsi, chaque objet d'un monde virtuel se représente par un ensemble d'éléments géométriques et photométriques. La réalité virtuelle réfère donc à une représentation mathématique d'un monde modélisé par son créateur, l'ordinateur étant l'outil de création et de visualisation.

Le terme réalité virtuelle a été introduit par "Jaron Lanier" pour définir :

"Un environnement 3D, interactif, généré par ordinateur dans lequel l'utilisateur est complètement immergé".

Cette définition comporte trois points principaux:

- L'environnement virtuel est un environnement tridimensionnel, généré par ordinateur, ce qui nécessite une performance élevée pour assurer le réalisme de la scène.
- On parle d'un environnement interactif par l'utilisation, car l'utilisateur a besoin de réponses en temps réel.
- L'utilisateur est immergé dans un environnement complètement calculé par ordinateur.

L'obtention de mondes virtuels réalistes nécessite de disposer de modèles très précis de l'environnement et se révèle donc très coûteuse, surtout dans le cas d'environnements complexes. A l'opposé, la réalité augmentée (RA) cherche à améliorer ou à compléter la vision de l'utilisateur sans chercher à remplacer ou à synthétiser le monde réel.

Milgram interprète la réalité augmentée dans un continuum linéaire (Fig.1.1.) qui va du réel au virtuel. Il définit le terme de la *réalité mixte* (Mixed Reality : MR) l'intervalle entre le réel et

le virtuel. Cette réalité mixte contient la réalité augmentée mais aussi la virtualité augmentée, qui consiste à intégrer du virtuel dans le monde réel [Milgram94,Sairio01].

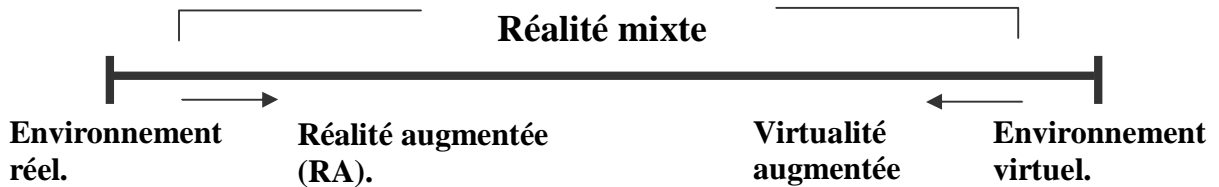


Fig.1.1. Le continuum de la réalité- virtualité [Milgram94].

Dans ce qui suit on commencera par définir le terme *réalité augmentée* et les motivations du développement de cette technologie. Puis on présentera quelques applications de la RA. La quatrième section décrira les systèmes utilisés pour améliorer notre perception. La cinquième section sera consacrée à la présentation des problèmes gênants le réalisme des scènes augmentées.

2.Définition

La réalité augmentée est une variante des environnements virtuels (EV) ou tout simplement *la réalité virtuelle* (RV). Contrairement aux technologies de la réalité virtuelle qui immergent l'utilisateur dans un environnement synthétique (ce qui empêche l'utilisateur de voir l'environnement réel), la RA offre à l'utilisateur la possibilité de voir les objets virtuels en même temps que le monde réel (Fig.1.2.). Idéalement le virtuel et le réel doivent coexister dans le même espace.

Certains chercheurs définissent la RA dans le sens où elle nécessite l'utilisation d'un casque de visualisation ou HMD (Head Mounted Display, littéralement *écran monté sur la tête*). Afin d'éviter de limiter la RA aux technologies spécifiques, AZUMA définit les systèmes de la RA comme étant des systèmes possédant les trois caractéristiques suivantes :

- La combinaison des deux mondes réel et virtuel.
- Interaction en temps réel.
- Alignement tridimensionnel.

Cette définition n'a pas limité les systèmes de la RA à des technologies spécifiques et elle n'est non plus limitée aux sens visuels. La réalité augmentée peut être potentiellement appliquée à tous les sens. Certaines applications demandent non seulement d'ajouter du virtuel au réel mais aussi d'enlever certains objets réels de la scène.



Fig.1.2. Un bureau réel avec une lampe et deux chaises virtuelles [Azuma97].

3. Les motivations

La réponse aux deux questions suivantes, permet d'éclairer les motivations du développement des systèmes de la RA :

- Pourquoi la réalité augmentée est intéressante ?
- Pourquoi la combinaison tridimensionnelle des objets virtuels et du monde réel est utile ?

La réalité augmentée améliore notre perception et notre interaction avec le monde réel. L'incrustation des objets virtuels complète notre perception [Evan01].

4. Les applications

Les applications potentielles de la RA sont multiples : nous trouvons en particulier des applications dans le domaine de la médecine [Fisher04, Schnaider04, Suthan02], de la robotique et la télé-robotique, de l'assemblage et la maintenance des objets complexes, de la conception (engineering design), des systèmes mobiles et des applications d'extérieurs et des développements commerciaux. Cette section éclaire quelques applications de la réalité augmentée (voir Fig.1.3).

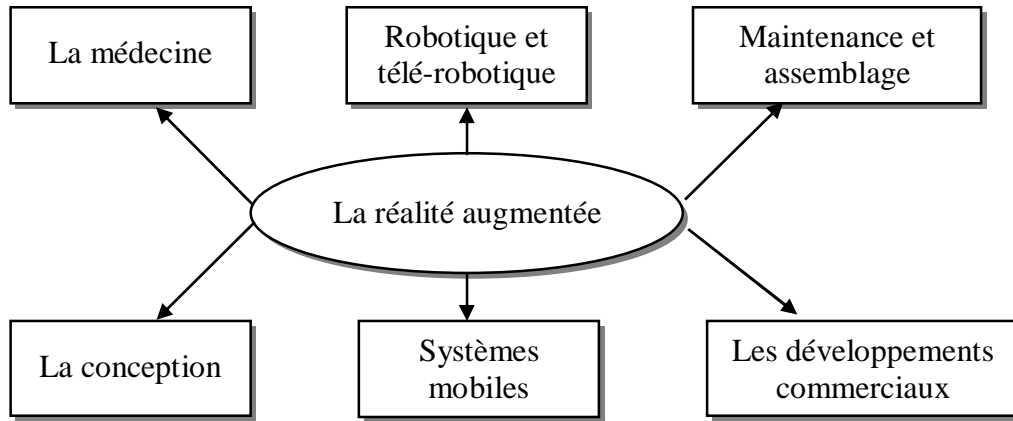


Fig.1.3. Exemples d'applications de la RA.

4.1.La médecine

L'une des applications primaires de la réalité augmentée est dans le domaine médical appelé *chirurgie guidée par l'image* (image guided surgery). La RA peut être utilisée par les médecins pour visualiser les données 3D extraites chez un patient par-dessus le corps du malade (images ultrasonores, tomographie 3D, imagerie par résonance magnétique IRM etc.) [Evan01].

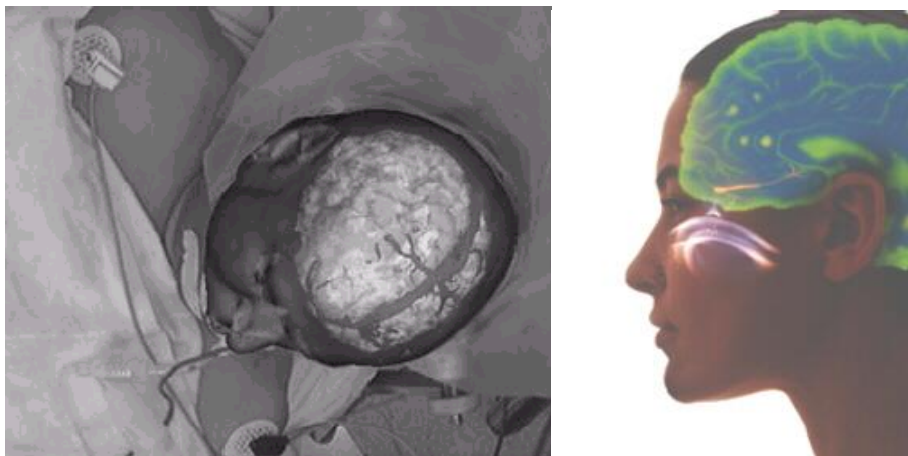


Fig. 1.4. Des images réelles d'un patient augmentées par des visualisations 3D [Evan01, Vallino95].

4.2. La robotique et la télé-robotique

La télé-opération d'un robot est difficile surtout lorsqu'il est éloigné. Les retards de communication causés par la distance nécessitent l'exécution des actions demandées sur une version virtuelle avant de les contrôler directement sur un robot réel. Cette augmentation permet de corriger les erreurs des actions demandées afin d'éviter des perturbations du comportement du robot (voir Fig.1.5) [Azuma97].

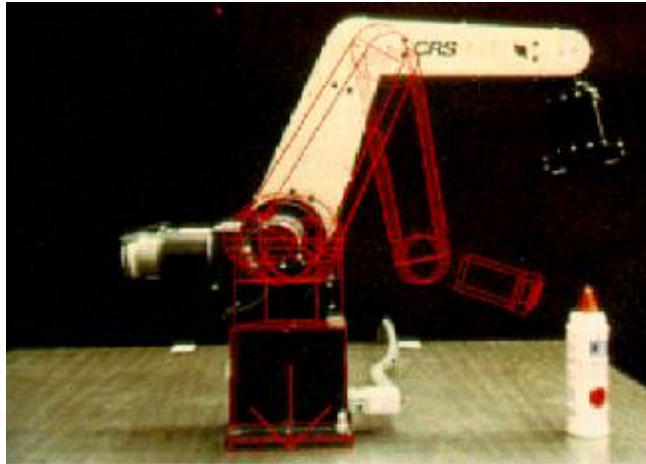


Fig.1.5. Les lignes virtuelles montrent le mouvement planifié du bras du robot [Azuma97].

4.3. La maintenance et l'assemblage

Un autre domaine d'application de la RA est la maintenance et l'assemblage des machines complexes. Des indications virtuelles (des animations tridimensionnelles) peuvent être ajoutées pour désigner certaines pièces d'objets manufacturés, comme les imprimantes (voir Fig.1.6), et faciliter ainsi les opérations de maintenance ou de réparation de ces objets [Azuma97, Vallino95].

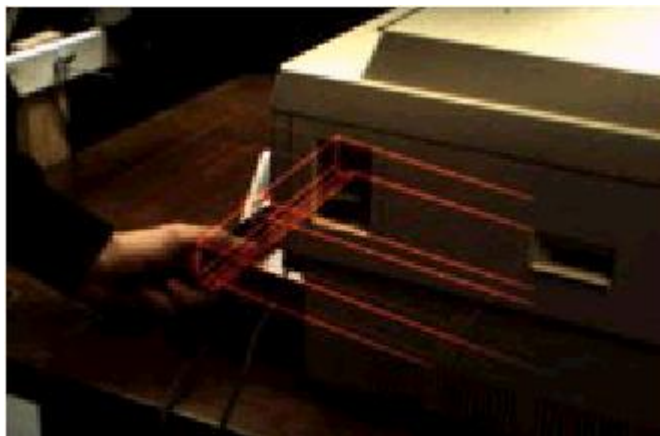


Fig.1.6. Une application de la maintenance d'une imprimante [Azuma97].

Les chercheurs de Boeing ont développé un système permettant de guider les techniciens dans l'assemblage de réseaux électriques pour les avions (voir Fig.1.7.a). Avant que ce système ne soit mis en place, les techniciens plaçaient les fils selon des schémas gravés sur des panneaux. Un 747 comportant plus de 1000 réseaux électriques, et les réseaux étant différents d'un avion à un autre, cela impliquait des coûts considérables pour le stockage, le transport et la construction des panneaux. Avec le nouveau système, tous les réseaux sont stockés en mémoire. Il suffit au technicien de choisir le réseau approprié, qu'il peut alors visualiser par dessus un panneau vierge en s'équipant d'un HMD (dans la Figure 1.7 : l'augmentation résultante du système de la RA utilisé par Neumann et Cho avec la coopération de Mc Donnell- Douglas).

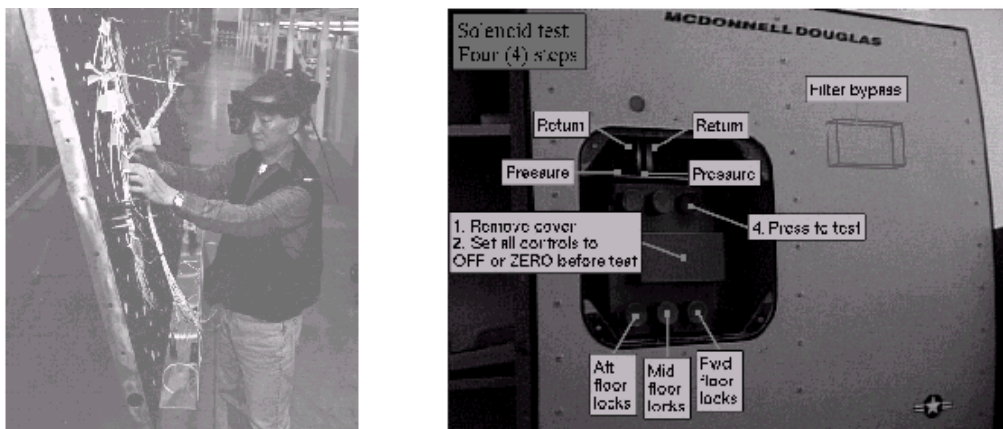


Fig.1.7. L'utilisation de la réalité augmentée pour la construction des avions [Vallino95].

4.4. La conception (engineering design)

Il est important de s'assurer que le modèle construit d'un appareil complexe satisfait les besoins du client. Une augmentation tridimensionnelle et animée offre la possibilité d'illustrer les différents aspects du modèle conçu tout en précisant les interactions du futur appareil avec son environnement.

4.5. Les systèmes mobiles

Le calcul mobile permet d'accéder et de manipuler l'information indépendamment du temps et de la position de l'utilisateur. Grâce aux réseaux sans fils (wireless networking) et à la diminution de la taille des appareils de calcul mobile. L'ordinateur portable est la deuxième génération de ces derniers, il est composé d'un processeur, d'une batterie (gardée dans un sac à dos), d'un casque de visualisation (HMD), des équipements permettant les communications sans fils et d'un organe d'entrée [Sairio01, Zhang03].



Fig.1.8. Exemples de systèmes mobiles [Bensor03].

Les progrès dans le domaine des réseaux sans fils et l'occurrence des ordinateurs portables, ont permis de construire des systèmes (de RA) mobiles. Un exemple important de ces systèmes est la machine de Turing (Turing machine) [Feiner97]. Cette machine permet de localiser les endroits, les immeubles, offrant ainsi la possibilité d'avoir plus d'informations sur les éléments intéressants.

4.6.La collaboration

Plusieurs applications de la RA sont naturellement des activités de collaborations, par exemple la chirurgie assistée par la RA, la maintenance des équipements complexes et de grande taille. D'autre activités de collaboration, surtout celles impliquant la conception et la visualisation des structures 3D, peuvent bénéficier d'avoir plusieurs personnes discutant et interagissant avec les modèles virtuels 3D (cela est vrai même si l'information est bidimensionnelle) [Azuma01].



Fig.1.9. Quatre utilisateurs regardant un modèle commun[Regenbrecht04].

4.7. La RA collaborative et mobile

Le travail de collaboration supporté par ordinateur (CSCW : computer supported collaboratif work) permet d'utiliser l'ordinateur comme intermédiaire pour les communications des êtres humains, la combinaison de la réalité augmentée, du calcul mobile et des CSCW produit une nouvelle technologie appelée : *la réalité augmentée collaborative et mobile* (MCAR : Mobile Collaborative AR), utilisée dans un espace de travail partagé et ouvert (open shared workspace) offrant la possibilité d'une collaboration mobile.

4.8. Les développements commerciaux

Récemment, la réalité augmentée a été utilisée pour les augmentations en temps réel de la diffusion vidéo, soit en augmentant des événements sportifs (voir Fig.1.10.) ou en remplaçant des annonces (voir Fig.1.11) [Azuma01].



Fig.1.10. La RA dans la diffusion sportive [Azuma01, Shahzad02].
Les annotations de la course et la ligne jaune sont ajoutés en temps réel.

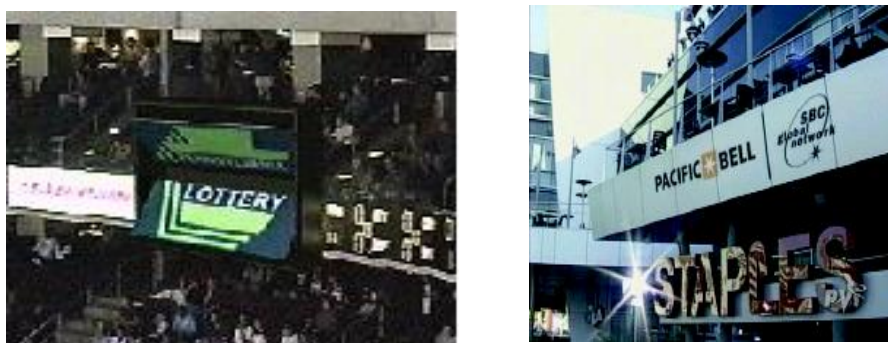


Fig.1.11. Annonce Virtuelle [Azuma01].
"PACIFIC BELL" et "3D LOTTERY" ne font pas partie de la scène.

5. Les systèmes de la RA

Les systèmes de la réalité augmentée améliorent notre perception du monde réel par ajout d'objets virtuels. Les systèmes de la réalité augmentée peuvent être classés, suivant le type des technologies utilisées pour réaliser l'augmentation et le type de la visualisation, en deux catégories : les systèmes basés casque (HMD) et les systèmes basés moniteur.

5.1. Les systèmes basés casque

La perception de la scène augmentée se fait généralement par l'intermédiaire d'un casque de visualisation qui est optique dans le cas d'un système optique et vidéo dans le cas des systèmes vidéos.

5.1.1. Les systèmes optiques

Les casques optiques (voir Fig.1.12) disposent d'un système optique qui est partiellement transparent, c'est-à-dire que la lumière du monde réel le traverse, et qui est partiellement réfléchissant, ce qui permet de visualiser les images virtuelles projetées sur le système optique en même temps que le monde réel (voir Fig.1.13) [Azuma01, Shahzad02, Zonglei02.b].



Fig.1.12. Casques optiques [Azuma97, Azuma01].

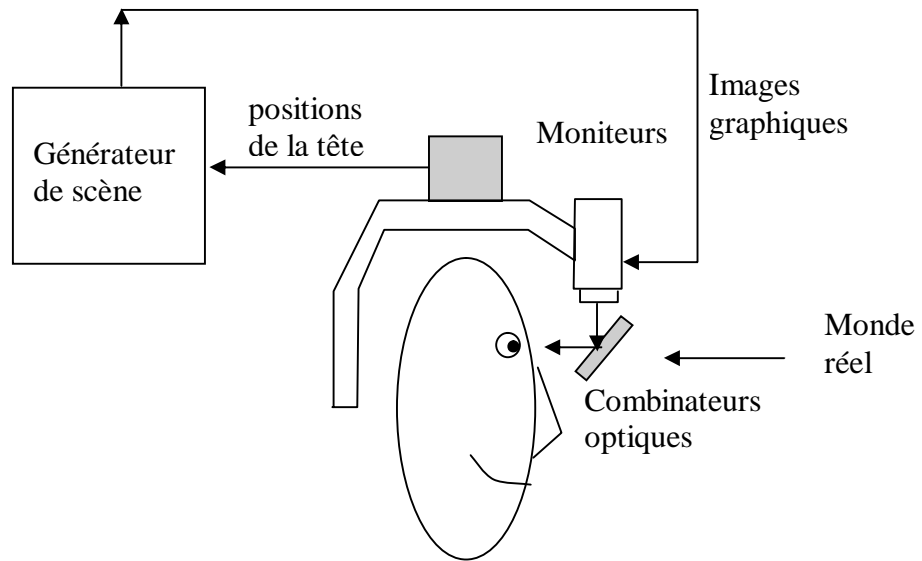


Fig.1.13. Le diagramme conceptuel d'un casque optique.

5.1.2. Les systèmes vidéos

Les casques (HMD) vidéos (voir Fig.1.14) placent deux écrans opaques devant les yeux de l'utilisateur, qui ne reçoit donc plus directement le monde réel. La scène réelle est en fait filmée par deux caméras fixées sur le casque et le film est projeté en même temps que les images virtuelles sur les écrans du casque vidéo (voir Fig.1.15) [Azuma97, Shahzad02, Zonglei02.b].

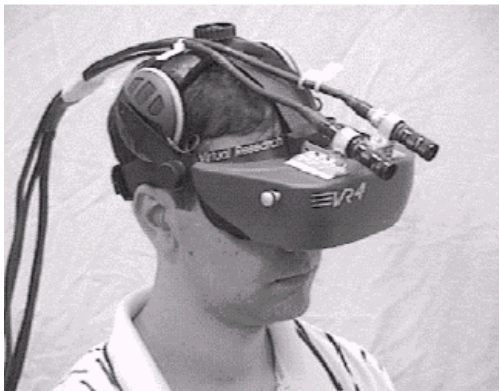


Fig.1.14. Casques vidéos [Azuma97, Vallino95]

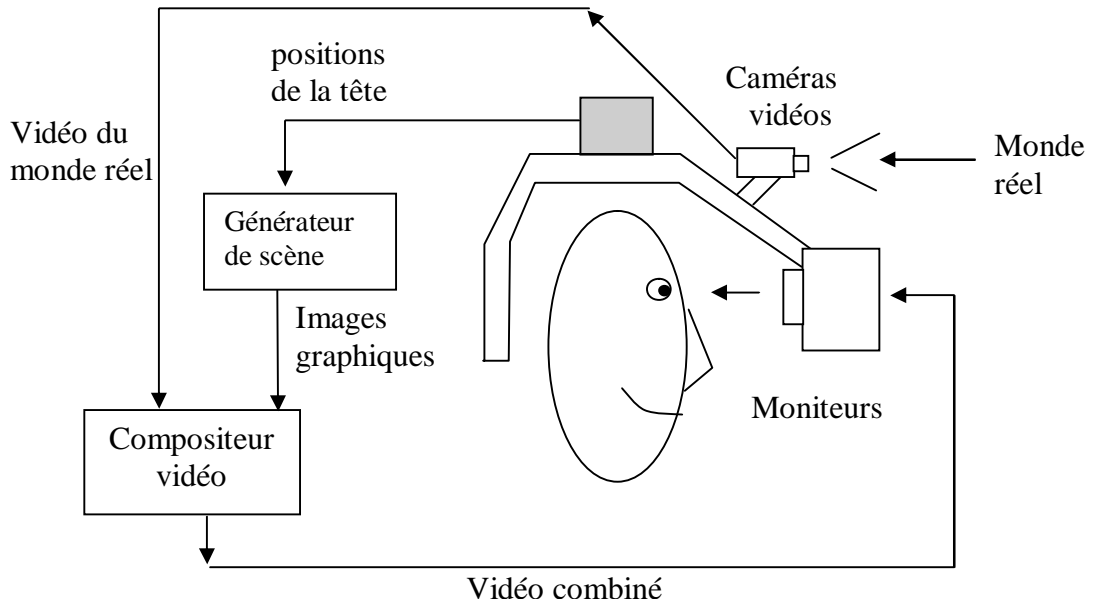


Fig.1.15. Le diagramme conceptuel d'un casque vidéo.

5.1.3. Comparaison

	Les systèmes optiques	Les systèmes vidéos
Simplicité	Traitent un seul flux d'images (images virtuelles).	Traitent deux flux: le flux virtuel et le flux réel.
Commodité (coupure de courant)	Jouent le rôle d'une simple paire de lunettes.	Aucune vue n'est perçue.
Résolution	Le monde réel est directement vu à travers les combinateurs, ce qui implique l'absence de la numérisation.	La numérisation de la scène réelle conduit à une dégradation.
Retard	Le seul retard qui existe est causé par la génération des objets virtuels.	En plus du retard causé par la génération des objets virtuels, un deuxième retard est causé par le traitement des images du monde réel.
Flexibilité	L'occlusion n'est pas parfaite parce que le traitement est réalisé entre les images numériques (virtuelles) et le réel.	Ce type de système est plus flexible parce que le réel et le virtuel sont disponibles sous forme d'images numériques.
Retard de flux des objets virtuels.	Des décalages des objets virtuels par rapport à leurs emplacements prévus peuvent être perçus.	Le retard du flux des objets virtuels est compensé automatiquement et la scène augmentée ne peut être disponible seulement si tout le traitement est achevé.

5.2. Les systèmes basés moniteur

L'approche la plus simple est la visualisation basée moniteur, comme figure dans Fig.1.16. La caméra vidéo capture les trames du monde réel (individuellement) d'une façon continue et les passent au système d'augmentation. Les objets virtuels sont ensuite fusionnés avec les trames, et ce sont les images résultantes de cette fusion qui seront visualisées par un moniteur standard. L'avantage de cette technologie d'affichage est sa simplicité. De plus, en traitant individuellement les trames, le système d'augmentation peut utiliser les approches basées vision pour extraire la position et l'orientation de la caméra afin de réaliser le recalage. Cependant, cette simplicité gêne l'immersion. Clairement, regardant le monde réel à travers l'écran limite la mobilité des augmentations. De plus, le traitement séparé des trames cause un retard potentiel entre l'acquisition des images et la visualisation des augmentations [Azuma97, Shahzad02].

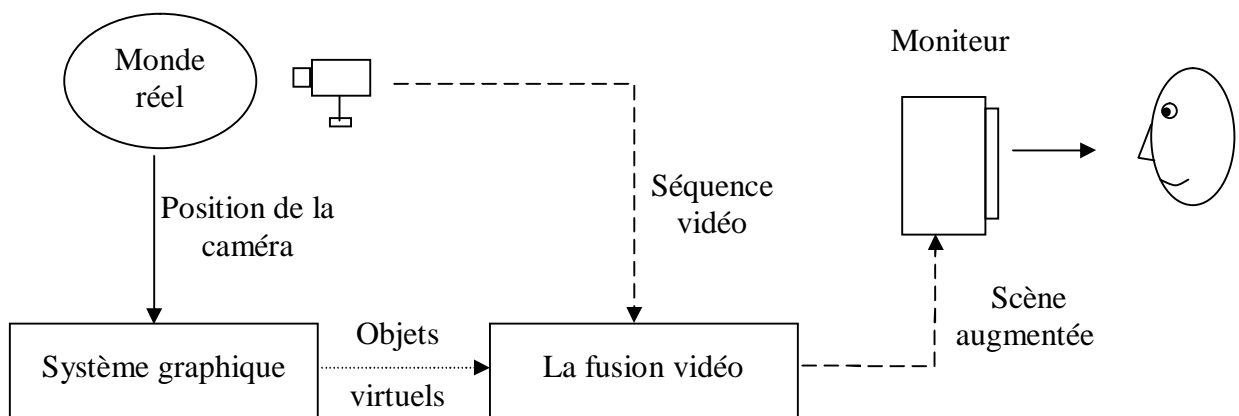


Fig.1.16. La technologie de visualisation basée moniteur.

6. Les problèmes techniques de la RA

Un certain nombre de problèmes techniques doivent être traités afin de produire des systèmes pratiques et convaincants [Koller97] :

Ø Un système de la RA basé vidéo possède essentiellement deux caméras, la première est utilisée pour capturer le monde réel et la deuxième est une caméra virtuelle qui génère les graphiques 3D. Les deux caméras doivent avoir les mêmes paramètres extrinsèques et intrinsèques afin d'aligner correctement les objets des deux mondes virtuel et réel. Pour le réaliser, un processus de calibration (voir chapitre 2) et une mise à jours dynamique des paramètres extrinsèques est nécessaire (c'est à dire le suivi de la caméra, voir chapitre 3).

Ø Afin d'avoir une interaction correcte entre les objets réels et les objets virtuels dans un environnement de la RA, des descriptions précises de la forme et de la position des objets de la scène réelle doivent être obtenues. Ces interactions peuvent comprendre la détection de collision et les occlusions visuelles.

Ø Un éclairage correct est une étape essentielle pour générer des objets virtuels avec un ombrage convaincant. Il est donc important de modéliser correctement l'éclairage de l'environnement réel et le projeter sur les objets virtuels.

Ø Un système de la réalité augmentée fournit interactivement l'information demandée par l'utilisateur. Puisque ce dernier travaille dans un environnement tridimensionnel, le système reçoit les requêtes soit par le suivi du mouvement de l'utilisateur et l'interprétation de ses gestes ou par un système de reconnaissance de la parole.

6.1. Le recalage

6.1.1. Le problème du recalage

Un des problèmes de base couramment limitant les applications de la réalité augmentée est le problème du recalage. Les objets des deux mondes virtuel et réel doivent être alignés correctement. Sans recalage précis, la réalité augmentée est refusée dans plusieurs applications. Par exemple la biopsie, si l'objet virtuel n'est pas au même endroit que la tumeur, le chirurgien n'enlève pas la tumeur et par conséquent la biopsie échoue. Le processus du recalage peut être divisé en trois étapes [Zonglei02.b] :

1. Le positionnement de l'objet virtuel dans la position souhaitée: le positionnement se fait dans le repère scène c'est à dire qu'il faut reconstruire l'environnement.
2. La détermination de la projection 2D de l'objet 3D positionné: cette étape nécessite l'estimation des paramètres de la caméra.
3. La fusion de la projection de l'objet virtuel avec l'image de la scène réelle.

6.1.2. Les erreurs du recalage

Puisque tous les systèmes de la réalité augmentée souffrent des erreurs du recalage de sources différentes, il est important de les décrire. Les erreurs du recalage agissent sur la performance de l'augmentation. Ces erreurs peuvent être classées en deux catégories [Azuma97,Shahzad02], qui sont les erreurs statiques et les erreurs dynamiques.

Ø Les erreurs statiques :

Les erreurs statiques sont indépendantes de la mobilité des objets du monde réel. Par exemple, un objet virtuel statique apparaissant dans des positions différentes lorsque le point de vu change.

Il existe quatre sources principales des erreurs statiques :

- Les déformations optiques : les déformations optiques sont des erreurs inhérentes aux caméras filmant le monde réel et aux systèmes d'affichage.
- Les erreurs du système de suivi: les erreurs du système de suivi sont des erreurs concernant la position de l'utilisateur (position de la caméra ou du casque de la RA) ou la position des objets réels. Ce genre d'erreurs est le plus grave parce qu'il est difficile à mesurer et à éliminer.
- Déviation d'alignement mécanique : Ce genre d'erreurs est causé par la déviation d'alignement des composantes matérielles utilisées par le système de la RA, prenons l'exemple d'un système optique de la RA où l'écran et le combineur (combiner) optique ne sont pas positionnés aux bonnes endroits et ne respectent pas la distance souhaitée entre eux (la distance entre l'écran et le combineur varie d'un moment à l'autre), la déviation d'alignement mécanique cause une fausse projection des images virtuelles.
- Paramètres incorrects de vision (viewing parameters) : les erreurs des paramètres de vision peuvent être considérées comme un cas spécial des erreurs d'alignement où les techniques de la calibration peuvent être appliquées. Les paramètres de vision déterminent comment convertir la position et l'orientation de la camera en matrices de vision utilisées pour dessiner les images graphiques.

Ø *Les erreurs dynamiques*

Les erreurs dynamiques dépendent du mouvement de la caméra et du déplacement des objets réels. Ces erreurs causent un léger déplacement des entités augmentées, autrement dit, les objets virtuels ne sont pas synchronisés avec les objets réels. La plupart des erreurs dynamiques sont causées par les retards du système. Un système de la RA doit tout d'abord capturer une image de la scène, déterminer la pose de la caméra, transformer l'objet vers le repère caméra, projeter l'objet virtuel sur l'image et finalement visualiser l'augmentation.

6.2. L'occlusion

Afin d'assurer une interaction correcte entre les objets des deux mondes réel et virtuel, ils doivent être placés dans le même repère, qui est le repère scène. Permettant ainsi de déterminer les parties occultées. Pour ce faire, les modèles des objets réels doivent être créés.

Les techniques d'acquisition des modèles des objets physiques peuvent être divisées en deux classes: la première classe comprend les techniques basées modèle, ces derniers supposent un modèle et essaient de l'aligner avec les données du monde réel permettant ainsi de déduire l'emplacement de cette objet dans le repère scène. La deuxième classe comprend les techniques basées profondeur, ces dernières reconstruisent la profondeur de l'environnement réel [Breen96].

6.2.1. Les techniques basées modèle

En utilisant l'approche basée modèle, les modèles géométriques des objets réels sont alignés avec leurs correspondants (les objets). En assumant que le modèle représente l'objet réel, cette alignement produit une transformation qui place le modèle géométrique dans la bonne position dans le repère scène. Autrement dit, la projection du modèle géométrique apparaît dans la même position et avec la même orientation que l'image de l'objet lui-même (voir Fig.1.17). Dans la Fig.1.18. figure deux chaises virtuelles placées autour d'une table réelle, la surface de la table occlue correctement les chaises en utilisant l'approche basée modèle.

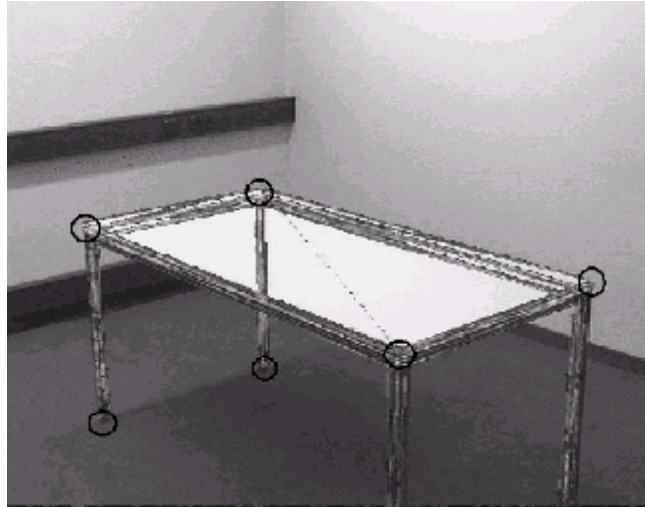


Fig.1.17. Le modèle est superposé sur l'objet réel après le recalage [Breen96].



Fig.1.18. La table réelle occlude deux chaises virtuelles et une partie de sa surface est occultée par la lampe [Breen96].

6.2.2. Les techniques basées profondeur

La deuxième approche utilise la carte de profondeur du monde pour déterminer les occlusions. Si au début de chaque cycle de rendu le Z- buffer est initialisé par les valeurs de la profondeur du monde réel, l'occlusion des objets virtuels est réalisée automatiquement. Au moment du rendu, les pixels éloignés par rapport à la valeur du Z- buffer ne sont pas dessinés. Dans Fig.1.19. figure trois cubes virtuels occultés par le support en bois.



Fig.1.19. Trois cubes virtuels occultés par le support en bois [Breen96].

7. Conclusion

L'objectif de la réalité augmentée (RA) est de superposer les objets virtuels sur l'image de la scène réelle comme s'ils font partie de la scène. Les applications de ce concept concernent principalement la médecine, la robotique et la télé-robotique, la maintenance et l'assemblage, la conception, les systèmes mobiles et les développements commerciaux. Un système de la RA combine les objets virtuels et les images de la scène réelle. Les différentes technologies de visualisation et les différentes façons de combinaison conduisent à deux types de systèmes: les systèmes basés casque et les systèmes basés sur une visualisation utilisant l'écran. Un problème de base rencontré dans le domaine de la RA est le problème du recalage. Les objets des deux mondes virtuel et réel doivent donner l'illusion qu'ils coexistent. Il est donc important d'estimer la position 3D des objets virtuels et celle de la caméra. Le chapitre suivant expliquera les étapes suivies pour extraire l'information 3D à partir des images 2D.

Chapitre 2

La RA basée vision

"Vision: the art of seeing things invisible".
Jonathan Swift

Chapitre2

La RA basée vision.

1. Introduction

L'alignement correct des objets virtuels (tridimensionnels) avec les objets réels nécessite la connaissance des informations tridimensionnelles. Les techniques de la vision par ordinateur permettent l'extraction de l'information tridimensionnelle à partir d'images bidimensionnelles. La stéréovision est une méthode populaire pour extraire la structure 3D d'une scène, elle consiste à déduire le relief en regardant une scène avec deux caméras, disposées comme les yeux humains (ou tout simplement utiliser une seule caméra en mouvement).

Ce chapitre est organisé en quatre sections: la première introduit le processus de formation d'image. La deuxième présente trois modèles de caméras pour faciliter la compréhension des techniques de la calibration présentées dans la troisième section. La quatrième section décrit le problème de la reconstruction tout en présentant la stéréovision.

2. La formation d'image

La formation d'image est un processus qui réduit la représentation tridimensionnelle de la scène en une représentation bidimensionnelle (de la même scène bien sûr) qui est l'image. Considérant seulement l'aspect géométrique, une image est le résultat d'une transformation géométrique. Cette dernière transforme la description tridimensionnelle de la scène en une description 2D [Basri03].

Une camera est attachée à un repère d'origine O appelé *centre de projection*. $o x$ et $o y$ (voir Fig.2.1) sont les axes parallèles au plan image correspondant aux directions des lignes et colonnes des pixels du plan image. L'axe $o z$, appelé *axe optique*, est orientée vers la scène et est

perpendiculaire au plan image. Le *point principale* appelé aussi *centre de l'image* est le résultat de l'intersection de l'axe optique et du plan image. Pour passer des coordonnées définies dans le repère du monde réel aux coordonnées image, et parce que le repère scène est différent du repère caméra, un point 3D doit subir un déplacement rigide 3D (une rotation et une translation) avant sa projection. En plus, les objets sont représentés sous forme de pixels dans le plan image, ce qui nécessite d'appliquer une transformation point-pixel.

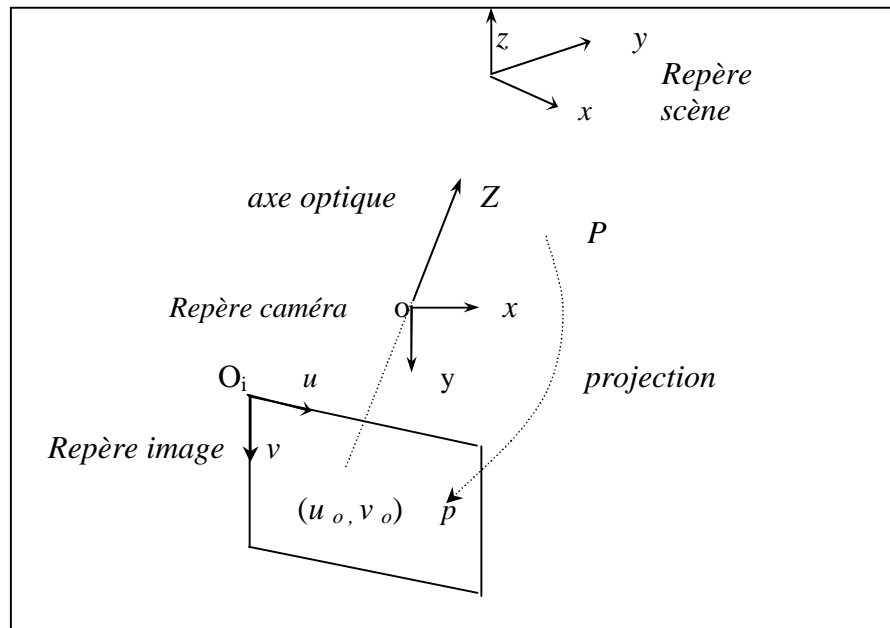


Fig.2.1. La formation d'image.

Les trois étapes précédentes sont appliquées selon un ordre séquentiel comme suit [Basri03, Shahzad02, Zonglei02.a, Zonglei02.b]:

1. *Un déplacement tridimensionnel*: les points 3D exprimés dans le repère de la scène subissent un changement de repère pour passer au repère de la caméra.

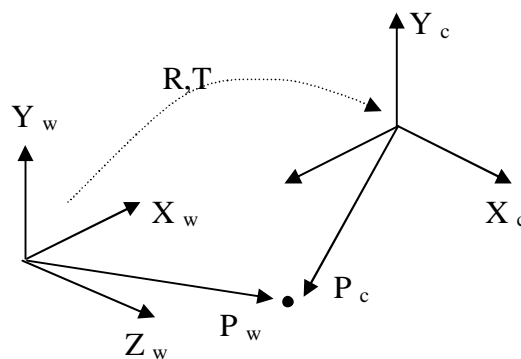


Fig.2.2. Un déplacement 3D [Shahzad02].

Le changement de repère comporte six paramètres correspondant à la rotation 3D et à la translation 3D. Ces paramètres représentent la position et la rotation de la caméra. Ils sont appelés *les paramètres extrinsèques* [Fergani04.b, Zonglei02.a, Zonglei02.b], exprimés par la matrice D (4,4) :

$$D = \begin{pmatrix} r_{11} & r_{12} & r_{13} & t_x \\ r_{21} & r_{22} & r_{23} & t_y \\ r_{31} & r_{32} & r_{33} & t_z \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (2.1)$$

2. *Une projection 3D-2D*: les points tridimensionnels exprimés dans le repère de la caméra sont projetés sur le plan image.

3. *Une transformation 2D-2D* (changement de coordonnées): pour passer aux coordonnées pixels, les points 2D subissent une transformation. Cette dernière comporte cinq paramètres appelés *paramètres intrinsèques*. La transformation 2D-2D s'écrit sous la forme matricielle suivante:

$$A = \begin{pmatrix} \mathbf{a}_u & 0 & u_o \\ 0 & \mathbf{a}_v & u_o \\ 0 & 0 & 1 \end{pmatrix} \quad (2.2)$$

Où:

(u_o, v_o) : est le centre de l'image.

$(\mathbf{a}_u \cdot \mathbf{a}_v)$: la taille effective du pixel horizontalement et verticalement respectivement.

Les trois transformations citées ci dessus peuvent être codées dans une même matrice M (3, 4), appelée *matrice de projection*, tel que:

$$M = AID \quad (2.3)$$

Où:

D : représente le déplacement tridimensionnel.

I : représente la projection 3D-2D.

A : représente le changement de coordonnées.

3. Les modèles de cameras

La caméra est l'outil essentiel autour duquel se développe la vision. Elle possède un rôle central car elle est l'interface entre deux espaces. Le premier est un ensemble de données inconnues évoluant dans un espace inconnu, il s'agit de la scène. Le deuxième est l'espace image, il s'agit des données perçues qui évoluent dans un espace particulier qui est l'image. La caméra se définit comme l'outil géométrique (projectif) qui traduit les relations existantes entre ces deux espaces. On distingue alors autant de modèles que de types de relations. Chacun fournit une interprétation différente de la scène [Heikkila97, Shahzad02].

3.1. Le modèle Sténopé

Le modèle *sténopé*, appelé aussi *trou d'épingle* et *modèle de projection perspective* (en anglais: pinhole model), est le modèle de projection le plus utilisé pour modéliser les caméras réelles. Ce modèle décrit la relation caméra-image c'est à dire la relation entre un point tridimensionnel défini dans le repère caméra et son correspondant (un point bidimensionnel) défini dans le repère image (voir Fig.2.3).

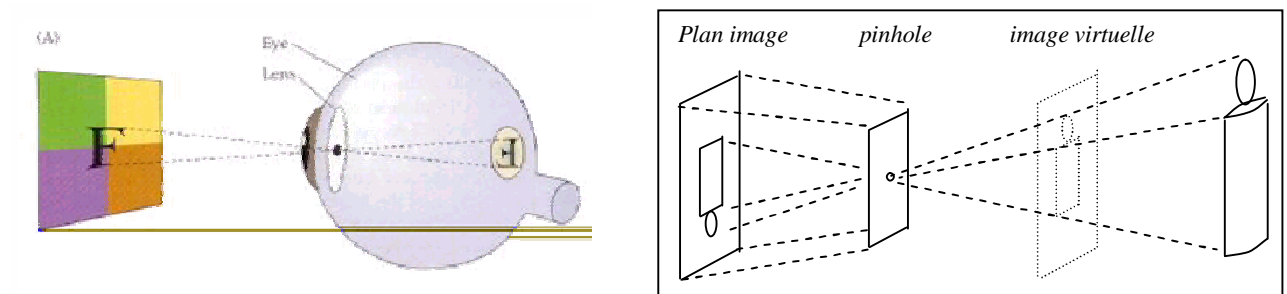


Fig.2.3. Le modèle sténopé [Kriegman03].

Dans ce modèle, un point image p est le résultat de la projection perspective du point physique P (voir Fig.2.4).

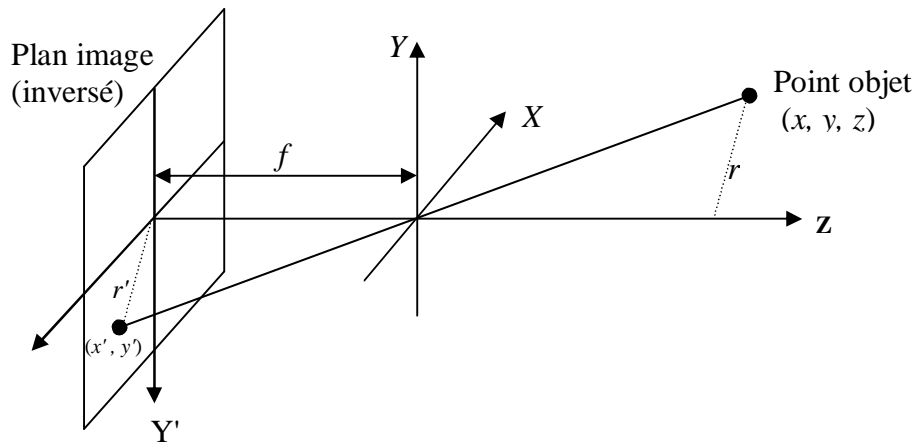


Fig.2.4. Le modèle de la projection perspective.

Afin d'éviter d'avoir une image inversée, il est supposé que le plan image se situe devant le centre de projection (voir Fig. 2.5).

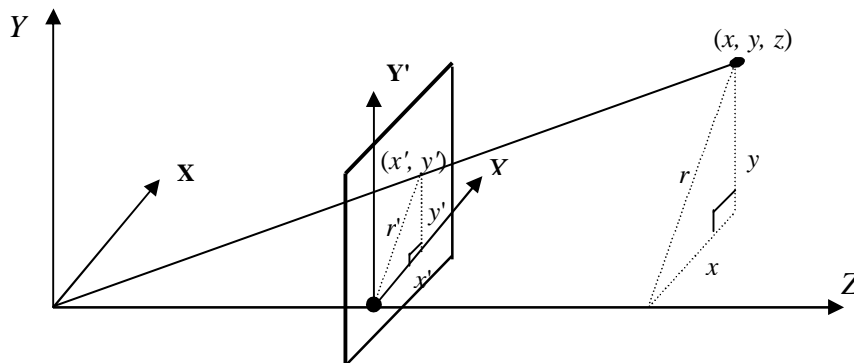


Fig.2.5. Le modèle de projection perspective utilisé.

En utilisant les coordonnées homogènes, la projection perspective d'un point $P(X, Y, Z, T)$ définit dans le repère scène est comme suit [Boufama99]:

$$\begin{pmatrix} x \\ y \\ t \end{pmatrix} = I \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix} \quad (2.4)$$

Où:

I est un facteur d'échelle.

3.2. La projection orthographique

La projection orthographique est l'approximation la plus simple des caméras réelles. Lorsqu'un objet est relativement loin de la caméra ou lorsque la taille de la scène est trop petite par rapport à la distance caméra-scène. La projection d'un objet tridimensionnel sur l'image est réalisée par la projection d'un ensemble de rayons parallèles au plan image [Basri03, Bebis03.a]. On peut écrire:

$$(X, Y, Z) = (x, y) \quad \text{Où: } x = X, \quad y = Y.$$

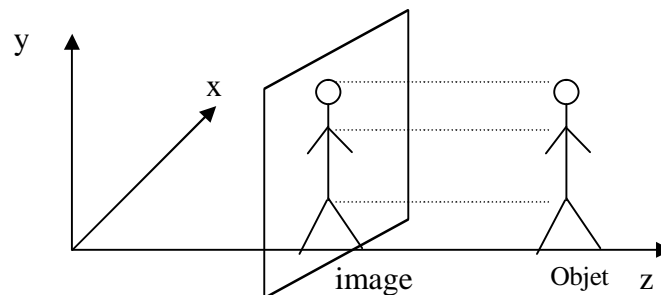


Fig.2.6. La projection orthographique.

∅ Les propriétés de la projection orthographique:

ü La projection d'un ensemble de lignes parallèles est un ensemble de lignes parallèles.

ü La taille est indépendante de la distance caméra-objet.

∅ La matrice de projection orthographique est:

$$\begin{pmatrix} U \\ V \\ W \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix} \quad (2.5)$$

3.3. La projection perspective faible

La projection perspective faible est une amélioration du modèle orthographique car en plus de la projection orthographique, les coordonnées (X, Y, Z) d'un point physique subissent un changement d'échelle, selon un facteur s , et deviennent $(s X, s Y)$. La projection perspective

faible est une bonne approximation de la projection perspective lorsque la taille de l'objet est petite par rapport à la distance objet-caméra [Basri03, Bebis03.a , Gros97].

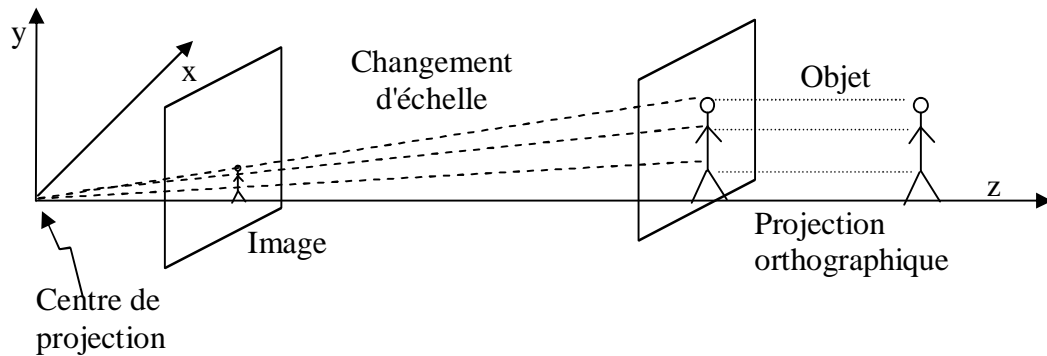


Fig.2.7. La projection perspective faible.

4. La calibration de la camera

La calibration de la camera est un problème fondamental dans la vision par ordinateur. C'est le processus qui permet de déterminer les paramètres extrinsèques et intrinsèques de la caméra ou tout simplement c'est le processus qui permet d'estimer la matrice de projection [Zonglei02.a].

Ø Pourquoi calibrons nous les caméras?

La connaissance du nombre de pixels séparant deux points dans une image (l'unité de mesure au niveau de l'image est bien sûr le pixel) ne signifie pas qu'on connaît la distance séparant leurs correspondants du monde réel, donc il est important d'avoir des équations liant le repère scène avec le repère image. Malheureusement, on ne peut pas lier directement ces deux repères, il faut utiliser un autre repère (un repère intermédiaire) qui est le repère caméra. L'idée de base est de trouver une équation liant le repère scène avec le repère caméra (lien C-I) et une autre liant le repère caméra avec le repère image (lien S-C) pour pouvoir déduire la relation scène-image. La détermination du lien C-I et le lien S-C est équivalente à déterminer les caractéristiques extrinsèques et intrinsèques de la caméra.

Ø Les techniques de la calibration

On dit qu'une caméra est calibrée lorsque ses paramètres extrinsèques et intrinsèques sont connus. Deux techniques sont utilisées pour calibrer une caméra: techniques basées capteur et techniques basées vision.

Malgré la rapidité du calcul du point de vue avec les techniques basées capteurs, elles ne sont pas utilisées parce qu'elles ne sont pas adaptées aux scènes d'extérieur. Elles nécessitent du matériel adéquat et elles sont facilement perturbées par les objets réels (les capteurs magnétiques sont facilement perturbés par les objets métalliques). Les techniques basées vision utilisent les images capturées par le système d'acquisition pour retrouver les paramètres de la caméra. Elles comportent deux méthodes : la calibration forte et la calibration faible.

La calibration forte appelée aussi *calibration basée modèle* (pattern based calibration) repose sur la connaissance des coordonnées tridimensionnelles de n points de référence P_i (au moins six points dans le cas d'un modèle tridimensionnel, voir Fig.2.8, et au moins quatre points dans le cas d'un modèle bidimensionnel, voir Fig.2.9) et leurs projections p_i dans le plan image (détectées par un détecteur automatique de coins, par exemple *HARRIS*). A partir de ces correspondances 3D-2D il est possible de calculer les valeurs de la matrice de projection, ce qui permettra par la suite de déduire les paramètres extrinsèques et intrinsèques. La précision de la calibration est améliorée en réduisant l'erreur des positions mesurées dans le monde réel et au niveau de l'image [Kriegman03, Personnaz02, Zonglei02.b]. Dans [Fergani04.b], [Zonglei02.a] et [Zonglei02.b], le modèle utilisé comporte deux plans orthogonaux, chacun d'eux contient douze carrés noir sur un fond blanc (voir Fig.2.8). Dans Fig.2.9. figure le modèle utilisé dans [Shahzad02].



Fig.2.8. Le modèle 3D.



Fig.2.9. Le modèle 2D.

Faugeras, Luong et Maybank ont introduit l'idée de calibrer la caméra en utilisant seulement l'appariement des images sans aucune connaissance du monde réel (aucun modèle n'est utilisé et

aucune connaissance du mouvement de la caméra), appelée *calibration faible* ou *auto calibration* (self calibration). Le processus de la calibration dans ce cas est réalisé en déplaçant (doucement) la caméra et en filmant en même temps (tout en fixant les paramètres intrinsèques), la mise en correspondance de trois images est suffisante pour calculer les paramètres de la caméra. La calibration faible est une technique flexible mais on ne peut pas encore avoir des résultats fiables (elle est immature) à cause du grand nombre de paramètres à estimer [Zonglei02.b].

5. La reconstruction

Bien que les images soient bidimensionnelles, la perception et la compréhension de l'univers nécessitent les trois dimensions. Le but de la reconstruction tridimensionnelle est de retrouver, à partir d'informations bidimensionnelles, la dimension perdue. La stéréovision offre la possibilité de reconstruire la scène à partir de deux images stéréoscopiques (deux images ayant une partie commune prise par deux caméras en même temps ou par une caméra en mouvement). Afin d'appliquer la stéréovision, il est important de calibrer les caméras utilisées et de trouver, dans les deux images, les primitives homologues, c'est à dire les points correspondants à la même entité physique du monde réel.

Les approches existantes de la reconstruction peuvent être classées en différentes catégories suivant:

- ü *Le type de la reconstruction obtenue:* projective, affine ou euclidienne (voir table 2.1. pour plus d'informations sur les invariants de chaque type). Comme l'information euclidienne est la plus riche, la reconstruction euclidienne est la plus utilisée.
- ü *La calibration de la caméra:* il n'est pas toujours possible de calibrer en temps réel la caméra, ce qui a poussé les chercheurs à développer des méthodes permettant de réaliser la reconstruction euclidienne sans calibration en utilisant des modèles de caméra moins riches (par exemple, le modèle orthographique, le modèle affine). L'utilisation de modèles moins riches apportent plus de limites, ils supposent souvent que la taille de la scène est petite par rapport à la distance scène-caméra [Armstrong96, Basri03, Heikkila97].

La représentation	Les invariants
Euclidienne	Les angles, les distances.
Affine	Le parallélisme, le barycentre.
Projective	Le birapport de quatre points alignés.

Table.2.1. Les invariants des représentations:
euclidienne, affine, projective.

5.1. La stéréovision

La stéréovision est le processus qui permet de construire le modèle tridimensionnel d'une scène à partir de deux images bidimensionnelles. Regarder une scène selon deux points de vue (ou plus) différents, en même temps, permet d'extraire la troisième dimension (appelée *profondeur*) de chaque point. L'importance du modèle tridimensionnel provient de la richesse d'informations qu'il apportent par rapport aux images 2D. Un modèle construit (3D) exprime la géométrie en fonction de coordonnées tridimensionnelles, permettant ainsi de déduire la taille et la forme des objets du monde réel, ce qui n'est pas le cas pour les images 2D (des informations limitées sur la forme et la taille des objets). Acceptez vous d'utiliser une voiture guidée par un ordinateur incapable de déterminer la forme et la taille des objets? Acceptez vous d'être servi par un robot incapable de déterminer la profondeur d'un verre d'eau?

La stéréovision utilise deux caméras disposées comme les yeux humains (voir Fig.2.10) qui vont donner deux images de la même scène (une paire stéréoscopique). Du fait de l'écartement des caméras, les images de cette paire ne sont pas sans relation. Ainsi un objet de la scène verra ses images décalées d'une image de la paire sur l'autre, d'un certain nombre de pixels, ce décalage est appelé *disparité*. Si on est capable d'attribuer une disparité à chaque pixel de l'image, on attribuera par extension une disparité à tous les points d'un objet sur l'image. On sera alors en mesure de replacer tous les points de cet objet dans l'espace, donc de reconstruire l'objet dans la scène.

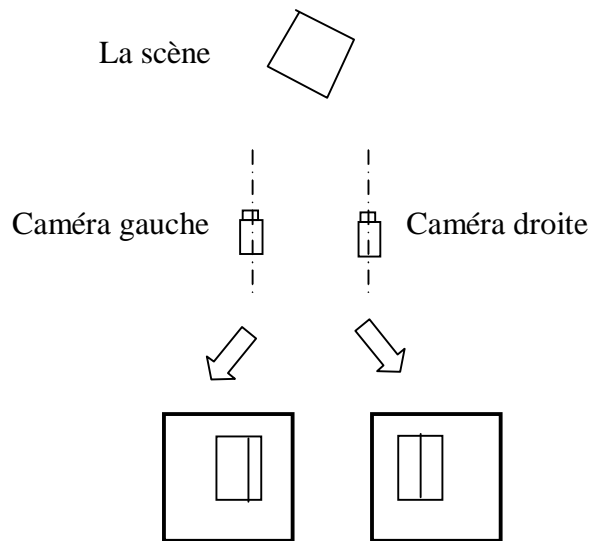


Fig.2.10. La vision stéréoscopique [Fauvel95].

Toute la difficulté réside dans l'appariement des deux images de la paire, appelé aussi *la mise en correspondance stéréoscopique* ou tout simplement *appariement stéréoscopique*. On imagine alors la difficulté du problème lorsqu' on pense qu'une scène de tout les jours peut contenir un grand nombre d'objets, placés à des distances diverses. La perception de ce genre de choses chez les humains est tellement naturelle qu'il est très difficile d'en appréhender le fonctionnement [HS02, JO97, SA02]. Ce qui a été dit permet de diviser le problème de la stéréovision en deux étapes:

- ü *Appariement stéréoscopique*: il consiste à rechercher les éléments d'une image dans l'autre (les éléments représentant la même entité physique).
- ü *La reconstruction du modèle tridimensionnel*: appelé aussi *problème d'estimation de pose* ou tout simplement le problème de la reconstruction stéréoscopique. Une fois les correspondances sont déterminées on fait appel à la triangulation stéréoscopique (une technique qui permet de calculer les coordonnées 3D à partir des appariements).

5.1.1. Appariement stéréoscopique

De nombreux travaux ont été effectués dans le domaine de la stéréovision pendant plus de quatre décennies. Cependant, sa théorie a été défini par le travail de Marr . Marr a identifié

l'ambiguïté du problème d'appariement et il a défini deux contraintes principales permettant de formuler le problème avec précision [Kostkovà02]:

ü *L'unicité*: un point d'une image ne peut pas avoir plus d'un correspondant parce que le résultat de la projection d'un point est au maximum un point.

ü *La continuité*: la disparité est presque la même au niveau de la même image.

Malgré la définition de la théorie de la stéréovision, le problème d'appariement est encore difficile à résoudre à cause de plusieurs facteurs, présentons quelques uns:

ü *L'occlusion*: on parle d'occlusion lorsqu'un point (ou plusieurs) n'a pas de correspondant dans l'autre image. Ce cas figure lorsqu'un objet cache une partie de la scène du champ d'acquisition d'une seule caméra.

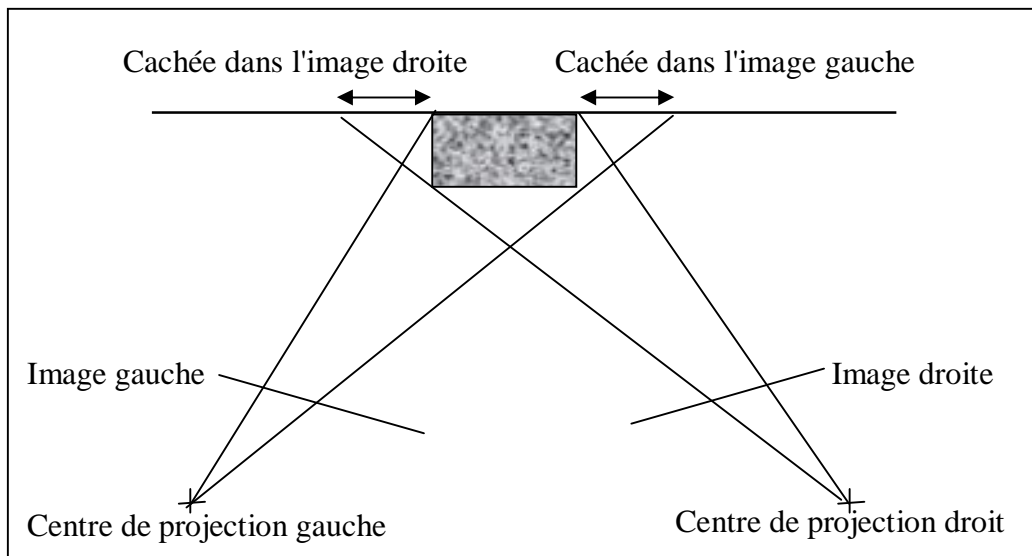


Fig.2.11. Le problème d'occlusion.

ü *Le bruit*: la présence du bruit dans une seule image.

ü *Les distorsions photométriques*: les distorsions photométriques déforment les valeurs d'intensités des pixels.

ü *Les distorsions projectives*: elles déforment la forme des objets, conduisant ainsi à une perte d'informations.

ü *Les régions non-texturées*: il est impossible de distinguer entre les pixels d'une région non texturée parce qu'elle ne présente aucune variation d'intensité.

Afin de déterminer les correspondants, il est important de définir les éléments à mettre en correspondance. Ces derniers peuvent être des primitives ou des intensités de pixels. Les approches utilisant les primitives sont appelées *stéréo basée primitives* (feature based stereo), les autres sont appelées *stéréo basée régions* (area based stereo).

ü *Stéréo basée primitives*

Dans le contexte de la stéréovision, le but des approches basées primitives est d'obtenir des correspondances fiables, même en présence d'une certaine quantité de bruit dans les images. Puisqu'on s'intéresse en stéréo à la reconstruction de la scène 3D, les primitives sont celles comportant une information sur la structure 3D de la scène, en général la structure tridimensionnelle peut être décrite par les contours et les coins (les intersections des contours),

c'est pourquoi la plupart des approches basées primitives se sont intéressées aux contours et aux coins, ce qui nécessite une phase pendant laquelle sont extraits de l'image en utilisant un détecteur automatique.

L'avantage de ces méthodes est qu'elles sont plus rapides grâce à la réduction de l'information contenue dans les images. Cependant, elles sont fortement limitées par le nombre des primitives, et leurs résultats sont des données éparses.

ü *Stéréo basée régions*

Elle consiste en la mise en correspondance des images sans extraction particulière. On traite alors les images au niveau du pixel ou plus souvent au niveau de petits groupes de pixels. Ces méthodes sont massivement parallèles puisque chaque pixel est une entité à part.

Ces méthodes peuvent être divisées en deux classes:

La première classe comporte les méthodes analysant l'intensité autour du pixel étudié, en utilisant des fenêtres de tailles différentes en plus des statistiques de similarités. Les statistiques les plus simples sont ceux comparant seulement les valeurs d'intensités mais elles sont très sensibles au bruit et aux déformations des images. Voici quelques exemples de statistiques utilisés:

- La somme des différences au carré SDC (ou SSD pour Sum of Squared Differences).

- La somme de la valeur absolue des différences au carré (SAD pour Sum of Absolute Differences).
- La somme des différences de l'écart des intensités à leur moyenne calculée sur l'ensemble de la fenêtre de corrélation (ZSSD: Zero mean Sum of Squared Differences).

La deuxième classe comporte les méthodes, dites *globales*, minimisant une fonction appelée *la fonction d'énergie* ou *la fonction objective* [Alghanemai02, Heikkila97, Henele02].

Un outil aidant à la résolution du problème d'appariement stéréoscopique est *la géométrie épipolaire*.

5.1.1.1. La géométrie épipolaire

Etant donnée une paire de caméras, un point tridimensionnel M définit un plan, Π , passant par M et par les centres des deux caméras. Π est appelé *plan épipolaire*, la projection dans une image du centre de projection de l'autre caméra est appelé *épipole*, autrement dit l'épipole est le résultat de l'intersection de la ligne liant les centres de projection avec le plan image (voir Fig.2.12) [Alghanemai02].

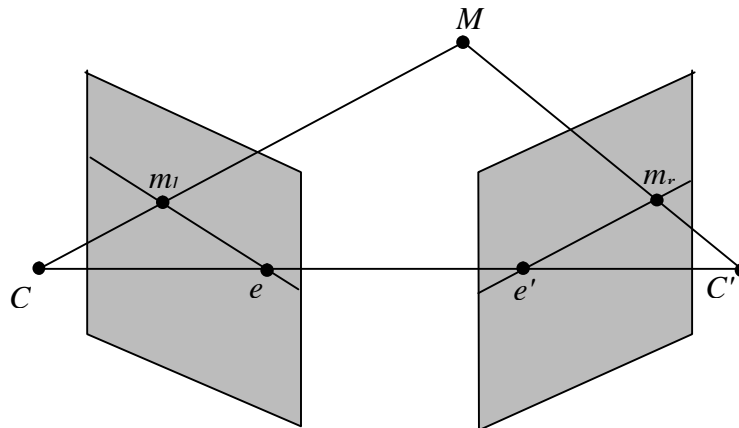


Fig.2.12. La géométrie épipolaire [Loop99].

Dans le domaine de la stéréovision, la géométrie épipolaire fournit le résultat suivant: Etant donné un point m_l de la première image, son stéréo-correspondant m_r appartient nécessairement à une droite de la deuxième image entièrement définie par la donnée de m_l . Cette droite est appelée *droite épipolaire* associée à m_l .

Grâce à cette propriété géométrique importante, la recherche du stéréo correspondant d'un point de l'image 1 se ramène à une recherche 1D dans l'image 2 (le long de la droite épipolaire associée) plutôt qu'une recherche 2D (recherche exhaustive dans toute l'image).

5.1.1.2. La détermination de la géométrie épipolaire

La géométrie épipolaire est fondée sur des propriétés algébriques de la géométrie projective, elle repose sur le calcul de la *matrice fondamentale*, F (c'est la représentation algébrique de la géométrie épipolaire), qui impose des contraintes géométriques entre deux vues, dites *contrainte épipolaire*:

$$m_r^t F m_l = 0 \quad (2.6)$$

Où:

$$m_l = (u, v, 1) \text{ et } m_r = (u', v', 1)$$

Plusieurs techniques sont utilisées pour déterminer la matrice fondamentale, l'algorithme des 8 points est le plus utilisé parce qu'il est le plus rapide et le plus simple à implémenter. La détermination de F nécessite un minimum de huit correspondances de points [Aishy02, Alghanemai02]. On peut réécrire la contrainte épipolaire comme suit:

$$\begin{bmatrix} u'_i & v'_i & 1 \end{bmatrix} \begin{bmatrix} f_{11} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} \begin{bmatrix} u_i \\ v_i \\ 1 \end{bmatrix} = 0 \quad (2.7)$$

L'équation (2.7) s'écrit:

$$u' u f_{11} + u' v f_{12} + u' f_{13} + v' u f_{21} + v' v f_{22} + v' f_{23} + u f_{31} + v f_{32} + f_{33} = 0 \quad (2.8)$$

L'équation (2.8) peut être réarranger comme $Af = 0$, A est une matrice ($n \times 9$). La décomposition en valeurs singulières est une bonne méthode pour résoudre le système:

$$\begin{pmatrix} u'_1 u_1 & u'_1 v_1 & u'_1 & v'_1 u_1 & v'_1 v_1 & v'_1 & u_1 & v_1 & 1 \\ \mathbf{M} & & & \mathbf{M} & & & \mathbf{M} & & \mathbf{M} \\ u'_n u_n & u'_n v_n & u'_n & v'_n u_n & v'_n v_n & v'_n & u_n & v_n & 1 \end{pmatrix} f = Af = 0 \quad (2.9)$$

Pour déterminer le correspondant d'un pixel de la première image, il ne suffit pas bien sûr de comparer les intensités des pixels deux à deux, on mesure la ressemblance entre deux pixels, en calculant un score de corrélation déterminé sur leur voisinage. Etant donné un pixel de la

première image et son voisinage associé, son correspondant est celui minimisant (ou maximisant le score, cela dépend du critère utilisé) le score le long de la ligne.

5.1.2. La triangulation

Une fois les correspondants sont déterminés, la structure tridimensionnelle est facilement reconstruite en utilisant la triangulation (voir Fig.2.13, Fig.2.14).

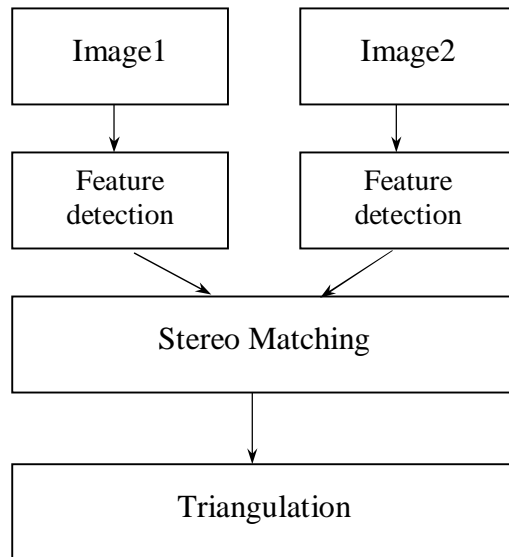


Fig.2.13. Le processus de la reconstruction.

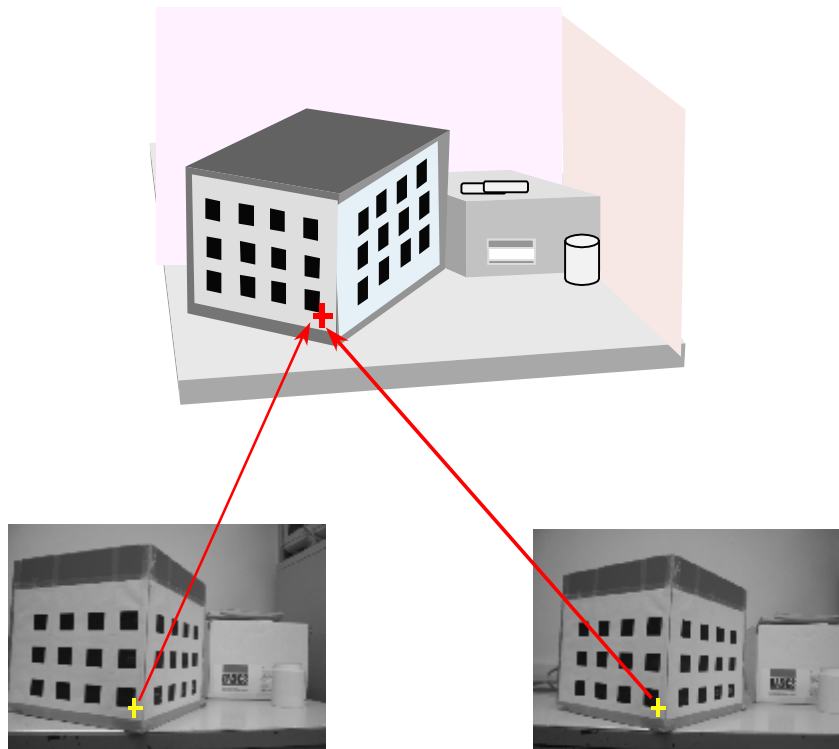


Fig.2.14. La triangulation.

6. Conclusion

Les techniques de la vision par ordinateur ont le potentiel de fournir un alignement précis entre les objets des deux mondes virtuel et réel. En utilisant les techniques basées vision, l'information 3D est facilement extraite à partir des images 2D, permettant ainsi de traiter le problème d'occlusion. Ce chapitre a présenté quelques techniques basés vision qui seront utilisées pour réaliser le recalage. La calibration est le processus permettant d'extraire les paramètres de la caméra. Ces paramètres sont utilisés par la stéréovision pour extraire la structure 3D de la scène, permettant ainsi d'achever la première étape du recalage. Le chapitre suivant présente l'estimation du mouvement de la caméra qui est une des étapes permettant d'augmenter une scène. L'estimation du mouvement des objets permet de traiter le problème d'occlusion dans les scènes dynamiques.

Chapitre 3

Analyse du mouvement

" Il y a diverses sortes de curiosité : l'une d'intérêt, qui nous porte à désirer d'apprendre ce qui nous peut être utile, et l'autre d'orgueil, qui vient du désir de savoir ce que les autres ignorent."

La Rochefoucauld, Maximes, 1678.

Chapitre 3

Analyse du mouvement.

1. Introduction

Un intérêt majeur d'analyse du mouvement est l'estimation du mouvement tridimensionnel (le mouvement de la caméra, le mouvement des objets réels ou la combinaison des deux). Les tâches d'estimation du mouvement peuvent être classées en trois catégories: 3D-3D, 2D-3D et 2D-2D dépendant des correspondances. Le problème 3D-3D consiste à calculer le mouvement tridimensionnel à partir des correspondances 3D. Afin de faciliter le problème d'obtention des données tridimensionnelles on peut supposer des correspondances 2D-3D. Le problème 2D-3D consiste à déterminer le mouvement tridimensionnel à partir des correspondances entre le modèle 3D et les projections 2D sur le plan image. L'analyse 2D-2D suppose seulement les correspondances entre les projections images et vise à calculer le mouvement 3D à partir des correspondances 2D.

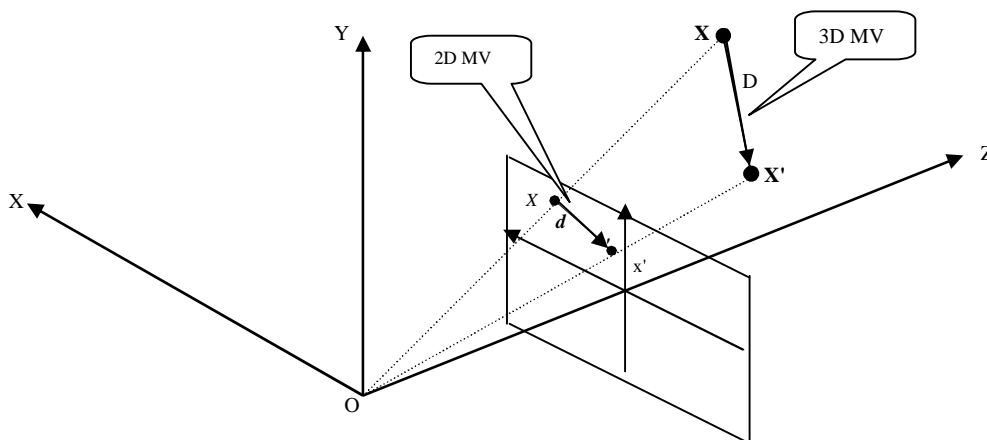


Fig.3.1. Le mouvement 3D et le mouvement 2D [Davidy00].

Ce chapitre est structuré comme suit : la première section présente quelques définitions utiles pour la compréhension de ce qui suit. La deuxième section présente les techniques d'estimation du mouvement suivi par le gain apporté lorsqu'on fusionne la stéréovision et le mouvement dans la troisième section. La quatrième section présente les techniques d'estimation du mouvement de la caméra, appelée mouvement propre ou tout simplement ego-motion, qui est très important pour une augmentation réaliste. La dernière section présente la relation entre l'estimation du mouvement et le suivi d'objets (tracking).

2.Définitions

2.1. Le mouvement apparent

L'image est la projection réalisée par le système d'acquisition du monde réel (3D) sur le plan image (2D). Tout mouvement tridimensionnel causé par le mouvement d'objet filmé par une caméra fixe, ou causé par le mouvement d'une caméra filmant une scène statique, entraîne des variations au niveau de l'image appelé "*mouvement apparent*". On distingue deux classes de variations, la première appelée "variations globales", causée par le mouvement de la caméra ou par un changement global d'illumination. La deuxième classe appelée "variations locales" causée par le développement d'objet ou par un changement local d'illumination ou même causée par le bruit.

2.2. Champ du mouvement

Lorsqu'un objet mobile est placé devant une caméra, on remarque qu'il y a un changement dans l'image ou dans la séquence d'images obtenue. Ainsi, le déplacement d'un point image P_i est causé par le mouvement du point physique P_o à une vitesse V_o (P_i est la projection de P_o sur le plan image), ou à chaque point P_i un vecteur V_i (voir Fig.3.2). L'ensemble de ces vecteurs est appelé *champ de mouvement*.

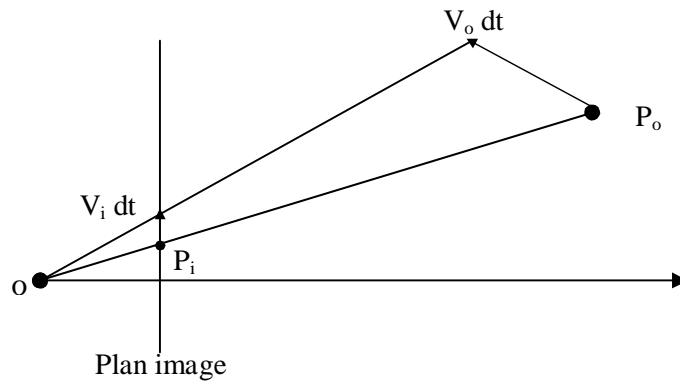


Fig.3.2. Le mouvement de l'objet crée un champ de mouvement sur le plan image [Robyn97].

Lorsqu'on traite seulement les rotations et les translations d'un objet rigide, le champ de mouvement est continu à l'exception des frontières de l'objet (voir Fig.3.3).

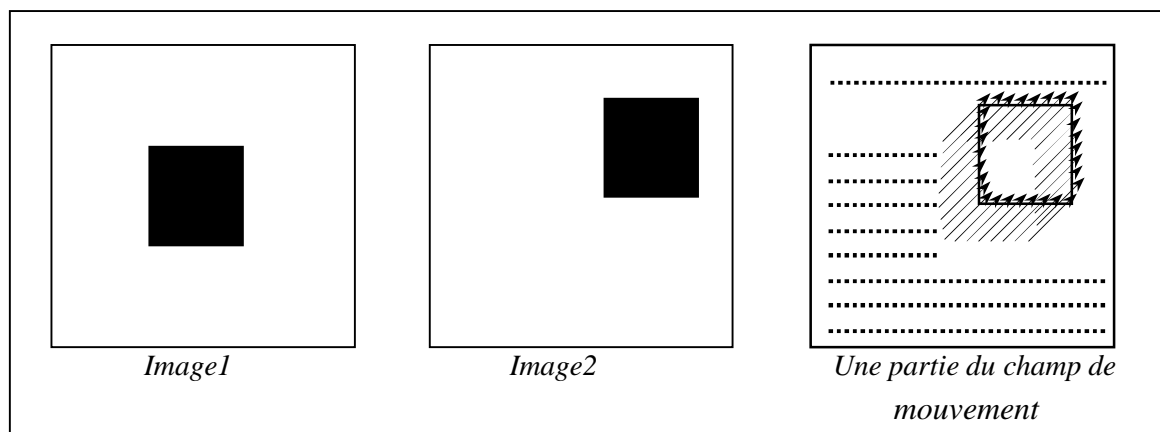


Fig.3.3. le champ de mouvement d'un carré en mouvement [Robyn97].

2.3 . Le flot optique

On distingue trois champs de mouvement différents représentant le mouvement en vision par ordinateur. Le premier est le mouvement tridimensionnel représentant le mouvement réel d'un objet dans le monde réel. Le deuxième est la projection bidimensionnelle sur le plan image du champ de mouvement réel (tridimensionnel). Le troisième est le champ de mouvement créé par le mouvement des gradients de l'intensité ce qui est appelé *le flot optique*.

L'idée du flot optique peut être illustrée comme suit : Premièrement, considérant une région d'intensités d'une image au niveau de gris représentant un ensemble de pixels. On peut remarquer que le gradient augmente de deux unités par pixel vers la droite.

$$\begin{pmatrix} 10 & 12 & 14 & 16 & 18 & 20 \\ 10 & 12 & 14 & 16 & 18 & 20 \\ 10 & 12 & 14 & 16 & 18 & 20 \end{pmatrix} \Rightarrow \text{Augmentation du gradient vers la droite.}$$

Considérant maintenant la même région dans l'image suivante.

$$\begin{pmatrix} 14 & 16 & 18 & 20 & 22 & 24 \\ 14 & 12 & 18 & 20 & 22 & 24 \\ 14 & 16 & 18 & 20 & 22 & 24 \end{pmatrix} \Rightarrow \text{L'image est décalée de deux pixels à droite.}$$

En assumant qu'un pixel maintient sa valeur d'intensité entre deux images consécutives (frames), le mouvement des deux pixels explique l'augmentation des valeurs d'intensités.

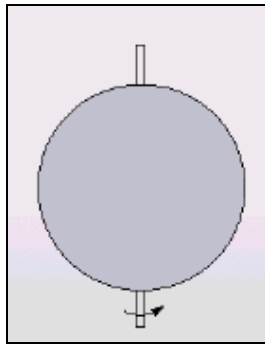
Par définition, le flot optique est le champ du mouvement associé aux variations de l'intensité lumineuse de l'image, c'est un mouvement apparent parce qu'il ne correspond pas nécessairement à la projection du mouvement tridimensionnel, donc le champ du mouvement diffère du flot optique. Pour être plus précis donnons l'exemple suivant :

Soit une scène parfaitement fixe dans laquelle une boule réfléchissante ou non est animée d'une rotation pur autour d'un de ses axes (voir Fig.3.4.a). Dans cette image :

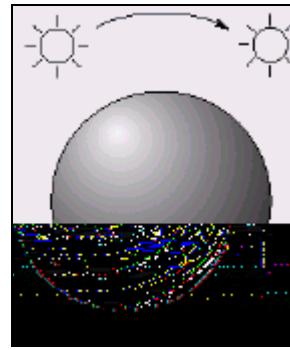
- le flot optique est nul.
- Le champ de mouvement n'est pas nul.

Soit une scène parfaitement fixe (voir Fig.3.4.b). Lorsque la source d'éclairage se déplace :

- Le flot optique n'est pas nul.
- Le champ de vitesse est nul.



a. Une boule en mouvement



b. Une boule immobile

Fig.3.4. Le mouvement apparent [Zaccarin03].

Un point clé est que l'information du mouvement est portée dans la structure de l'image. Par conséquent une estimation du mouvement non ambiguë peut être seulement achevée lorsque la projection d'un objet animé expose un degré suffisant de variation d'intensité [Aishy02, Corbett02, Habed00, Krüger98].

La figure Fig.3.5 présente quelques exemples du flot optique :

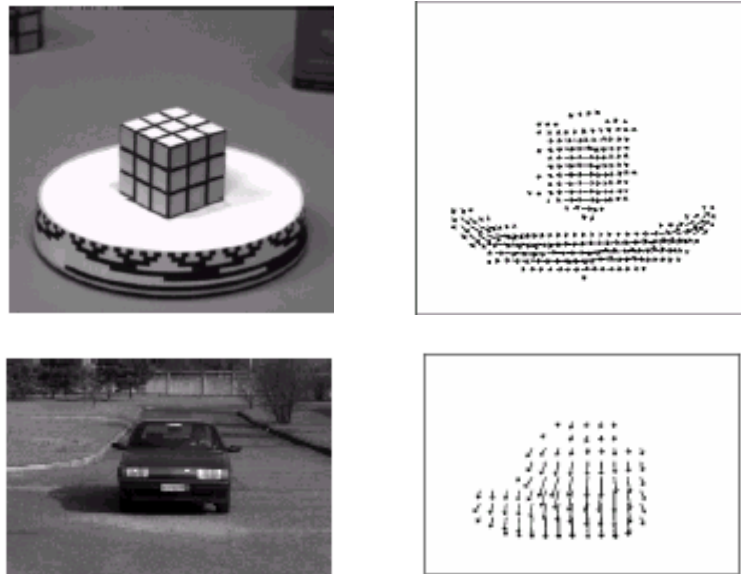


Fig.3.5. Exemples du flot optique.

L'estimation du mouvement basée sur le flot optique nécessite la connaissance des limites rencontrées, les hypothèses qu'on doit respecter, et les problèmes qui peuvent rendre impossible l'estimation du mouvement:

∅ Les trois hypothèses suivantes sont les plus importantes :

1. Champ de vecteurs vitesses localement lisse : les points voisins situés sur un même objet sont animés par un mouvement semblable.

2. Invariance de l'intensité lumineuse : l'intensité lumineuse de chaque point de la scène reste quasiment constante au cours du mouvement, c'est l'hypothèse de base de toutes les techniques d'estimation du mouvement. Le non-respect de cette hypothèse (variation de l'éclairage, bruit) cause le plus grand nombre d'erreurs d'estimation.

3. Les objets doivent être de nature non homogène (des objets texturés) car il est impossible de détecter le mouvement sur les zones homogènes (aucune variation n'est détectée).

Généralement, le flot optique correspond au champ du mouvement mais pas toujours. Prenons l'exemple (Fig.3.6) du poteau de Barber (Barber' s pole) où le champ de mouvement est différent du flot optique (ce genre de cas est inhabituel).

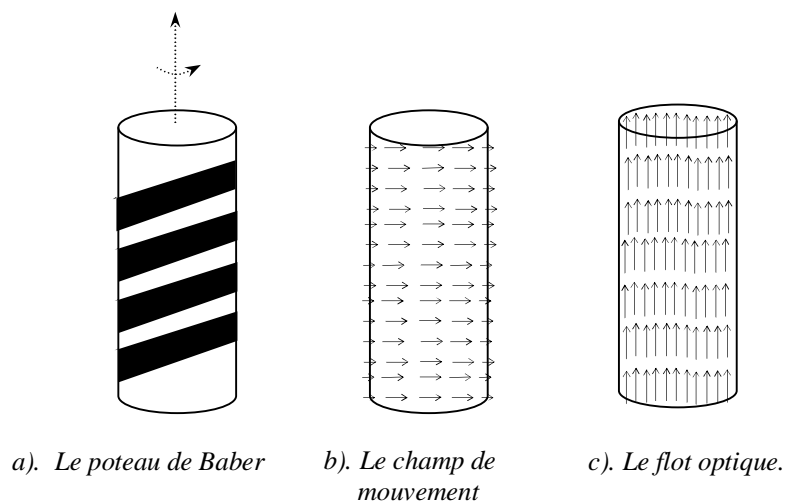
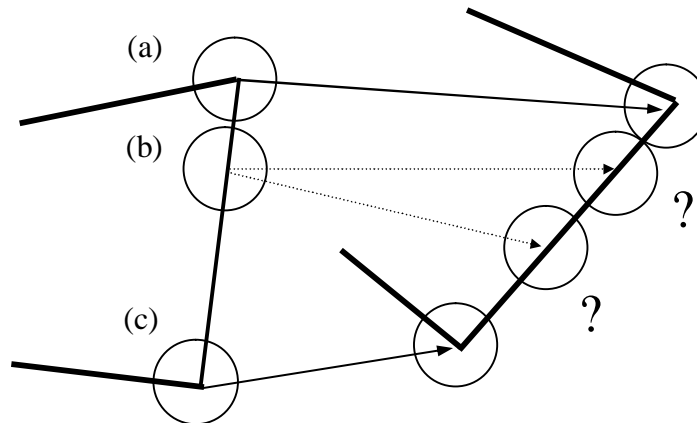


Fig.3.6. Le champ de mouvement et le flot optique du poteau de Barber.

Ø Un problème rencontré est le problème d'ouverture (the aperture problem) :

Ce problème est causé par l'insuffisance de variation de niveaux de gris. Pour un point situé sur un contour rectiligne (en ligne droite) dans une image, on peut considérer comme correspondant tout point de ce contour (voir Fig.3.7. section b). L'estimation du mouvement est donc ambiguë. Dans les deux sections a, c de la figure 3.7, représentant des coins, le flot optique est déterminé facilement.



*Fig.3.7. une partie des extrémités d'un objet en mouvement
illustrant le problème d'ouverture.*

2.4.La discontinuité du flot optique

Une source riche et particulière d'information sur le mouvement visuel est découverte aux extrémités des surfaces où le flot optique est discontinu à cause de l'indépendance du mouvement des objets. Par conséquent, on peut parler des " extrémités du mouvement" (motion boundaries) entre les régions adjacentes possédants des vitesses différentes, elles fournissent une information sur la position et l'orientation des extrémités de la surface dans la scène [Black00]. L'information sur les extrémités de surface et la profondeur est utilisée pour différentes tâches, citons: la reconstruction à partir du mouvement, la reconnaissance d'objets, la segmentation (pour plus de détails sur la segmentation et les extrémités du mouvement, voir [Dengsheng01]) .

La scène présentée dans la Figure 3.8. contient plusieurs objets qui peuvent présenter des champs de mouvement différents par rapport à un observateur actif, par exemple un robot mobile. La concentration sur une petite bande horizontale permet de découvrir trois régions possédant de petites variations de vitesses (voir Fig.3.8). Une estimation optimale du flot optique peut être obtenue en divisant la bande aux régions précédentes et en calculant les vitesses séparément (tout en utilisant le modèle approprié du flot optique). La recherche des extrémités est un problème d'optimisation difficile, il n'était pas pris en compte par les méthodes traditionnelles ce qui conduisait à des résultats non fiables [Deutsch97]. Papademetris et Belhumeur dans [Papademetris96], ont présenté un algorithme basé sur la programmation dynamique permettant de détecter la présence des extrémités du mouvement en même temps qu'on estime le flot optique.

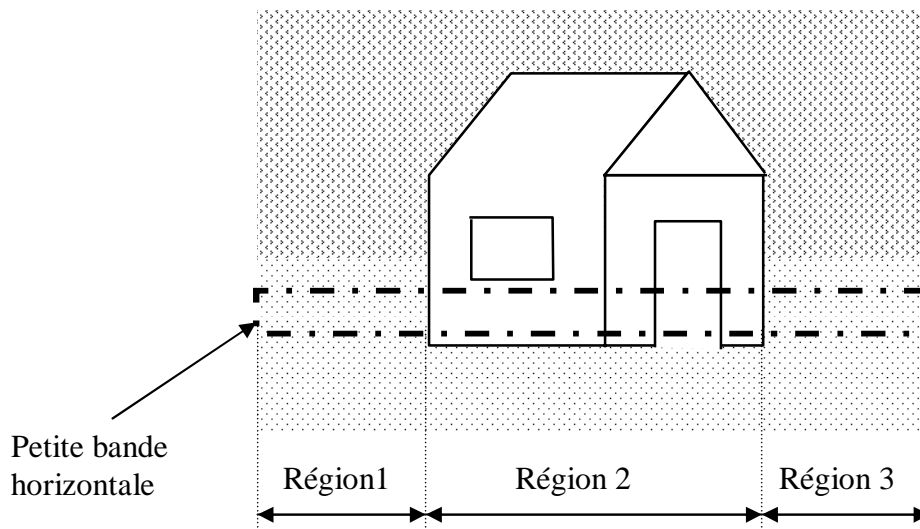


Fig.3.8. Une petite bande horizontale contenant les extrémités du mouvements.

2.5. L'analyse du mouvement basée flot

2.5.1. Les modèles du mouvement 3D-2D

Le vecteur de vitesse 3D d'un objet rigide est écrit comme suit :

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = \begin{bmatrix} 0 & -\Omega_3 & \Omega_2 \\ \Omega_3 & 0 & -\Omega_1 \\ -\Omega_2 & \Omega_1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix} + \begin{bmatrix} T_x \\ T_y \\ T_z \end{bmatrix}$$

Où

$$\dot{X} = \Omega \times X + T$$

$\Omega = |\Omega_1, \Omega_2, \Omega_3|^T$: le vecteur de vitesse angulaire.

$T = |T_x, T_y, T_z|$: le vecteur de vitesse de translation.

Ø Une projection orthographique

Lorsqu'il s'agit d'une projection orthographique, on peut décrire le flot optique par

$$v_x = \dot{x} = \dot{X} = T_x + \Omega_2 Z - \Omega_3 Y.$$

$$v_y = \dot{y} = \dot{Y} = T_y + \Omega_3 X - \Omega_1 Z.$$

Ø Une projection perspective

$$v_x = f \frac{\dot{X}}{Z} - x \frac{\dot{Z}}{Z} = f \left(\frac{T_x}{Z} + \Omega_2 \right) - \frac{T_z}{Z} x - \Omega_3 y - \frac{\Omega_1}{f} xy + \frac{\Omega_2}{f} x^2.$$

$$v_y = f \frac{\dot{Y}}{Z} - y \frac{\dot{Z}}{Z} = f \left(\frac{T_y}{Z} + \Omega_1 \right) - \frac{T_z}{Z} y - \Omega_3 x + \frac{\Omega_2}{f} xy - \frac{\Omega_1}{f} y^2.$$

Lorsque $f = 1$, on obtient :

$$v_x = \frac{-T_x + xT_z}{Z} + \Omega_1 xy - \Omega_2(1 + x^2) + \Omega_3 y.$$

$$v_y = \frac{-T_y + yT_z}{Z} - \Omega_2 xy + \Omega_1(1 + y^2) - \Omega_3 x.$$

En éliminant Z , on aboutit à une équation non linéaire :

$$-v_y e_1 + v_x e_2 - x (\Omega_1 + \Omega_3 e_2) - x y (\Omega_2 e_1 + \Omega_1 e_2) + (x^2 + y^2) \Omega_3 + (1 + y^2) \Omega_1 e_1 + (1 + x^2) \Omega_2 e_2 = y v_x - x v_y$$

Où :

$$e_1 = T_x / T_z \text{ et } e_2 = T_y / T_z.$$

$e_1, e_2, \Omega_1, \Omega_2, \Omega_3$: inconnus.

x, y : les coordonnées image.

v_x, v_y : flot optique.

Puisque ces équations possèdent 5 inconnus $\{e_1, e_2, \Omega_1, \Omega_2, \Omega_3\}$:

- Il existe au plus 10 solutions avec 5 vecteurs du flot optique.
- Le flot optique à 6 points (au plus) détermine souvent le mouvement 3D uniquement.
- Si le mouvement est une rotation pure, donc, il est déterminé par 2 valeurs de flot optique.
- Dans le cas d'une surface tridimensionnelle, plane le flot optique à 4 points donne deux solutions.

3. Les techniques d'estimation du mouvement

Les techniques d'estimation du mouvement peuvent être classées en deux classes :

Les méthodes d'estimation tridimensionnelle, elles essayent de résoudre les équations de projection directement par la mise en correspondance des images consécutives de la même séquence . Les méthodes d'estimation bidimensionnelle estime le flot optique [Evan01] .

3.1. Estimation 3D

L'estimation du mouvement tridimensionnel à partir des projections bidimensionnelles est un problème difficile. Il fait partie de plusieurs applications à savoir : la navigation de robot. Les méthodes d'estimation tridimensionnelle peuvent être classés en deux classes: les méthodes directes et les méthodes basées sur l'appariement des primitives .

Ø Les méthodes directes déduisent le mouvement tridimensionnel à partir du flot optique (2D). Ces méthodes peuvent être considérées comme une extension des méthodes (2D) du flot optique .

Ø Les méthodes basées sur les primitives déduisent le mouvement à partir des correspondances des primitives. Ce problème comprend trois phases : la détection des primitives de chaque image de la séquence, la détermination des correspondances et la déduction du mouvement. Les primitives utilisées par les estimations 3D peuvent être des points, des coins, des contours ou des lignes, le choix d'un de ces types dépend de l'application, de la disponibilité des primitives dans les images et de la fiabilité de leurs mesures. Le problème de la mise en correspondance des primitives peut être vu comme un problème de recherche, les techniques du filtrage temporel tel que le filtrage de Kalman [Crowley01, Jennings96, Ku01],

sont souvent appliquées pour le suivi des primitives dans le temps afin de réduire l'espace de recherche associé à chaque pixel. L'appariement des points a été utilisé par Ullman, qui a formulé une méthode permettant de découvrir le mouvement 3D . Une méthode similaire a été proposée par Roach et Aggarwal, en utilisant 8 points dans deux images consécutives de la séquence tout en supposant que la scène est statique et la caméra est en mouvement . Le résultat de cette méthode est la solution d'un système de 20 équations et 27 inconnus .

L'utilisation des appariements de lignes rend le problème plus difficile, parce qu'on ne peut pas déterminer le mouvement à partir de deux images seulement. L'utilisation des correspondances de ligne conduit à un grand nombre d'équations non linéaires, pour les résoudre il faut appliquer des méthodes itératives.

En résumé, on peut dire que l'utilisation des primitives d'un niveau inférieur, tel que les points, facilite les formulations mais présente un inconvénient gênant qui est la sensibilité au bruit. L'utilisation des primitives stables et d'un niveau élevé, tel que les lignes , augmente la difficulté du problème d'appariement. Les méthodes basées primitives sont utilisées dans les situations où une description simplifiée de la scène est suffisante ou lorsque les correspondances existent déjà. Elles ne sont pas appropriées pour les applications demandant une estimation dense du champ de mouvement parce que la complexité du problème augmente exponentiellement avec le nombre de primitives.

3. 2. L'estimation 2D

La plupart des applications pratiques d'estimation du mouvement utilisées aujourd'hui utilisent une approche bidimensionnelle, dans laquelle l'estimation du mouvement est obtenue à partir du mouvement (2D) apparent (dans le temps) [Irani94]. Ces méthodes s'appuient sur l'hypothèse d'invariance des intensités lumineuses (voir Fig.3.10) c'est à dire qu'un changement de l'intensité d'un point est causé seulement par le mouvement.

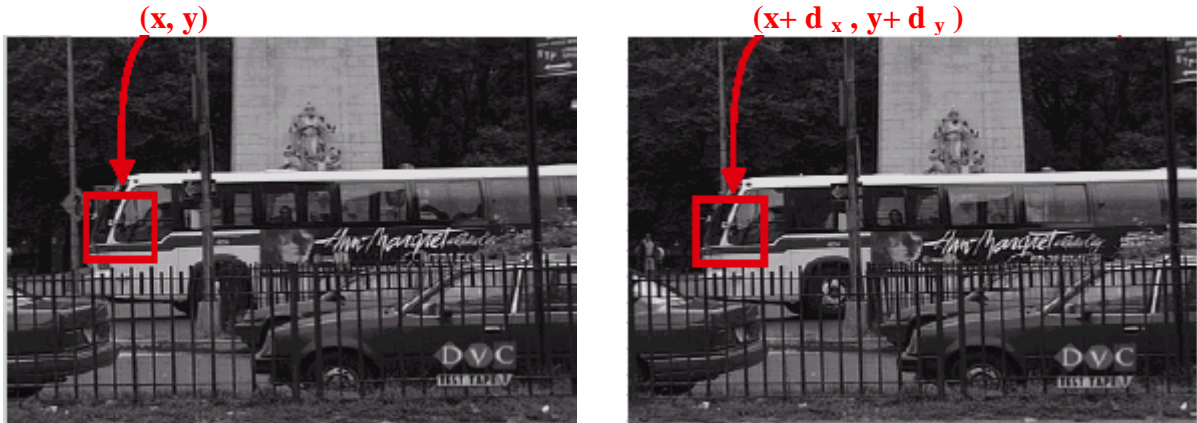


Fig.3 .9. Invariance de l'intensité lumineuse entre les trames [Zaccarin03].

Cette hypothèse est modélisée par une fonction continue dans l'espace et dans le temps [Verplaetse97]:

$$\frac{d I (x , y , t)}{dt} = 0 \quad (3.4)$$

Où:

$I(x, y, t)$: la valeur de l'intensité lumineuse de l'image aux coordonnées x et y et à l'instant t .

On peut facilement dériver et obtenir l'équation de la contrainte du changement d'intensité (Brightness change constraint equation):

$$\frac{d I (x , y , t)}{dt} = \frac{\partial I (x , y , t)}{\partial x} \frac{dx}{dt} + \frac{\partial I (x , y , t)}{\partial y} \frac{dy}{dt} + \frac{\partial I (x , y , t)}{\partial t} = 0 \quad (3.5)$$

On peut écrire l'équation précédente comme suit:

$$I_x (x , y , t) v_x + I_y (x , y , t) v_y + I_t (x , y , t) = 0 \quad (3.6)$$

Avec:

$$v_x = \frac{dx}{dt} \quad \text{et} \quad v_y = \frac{dy}{dt}$$

L'équation (3.6) est similaire à:

$$I_t + u I_x + v I_y = 0 \quad (3.7)$$

Cette équation est appelée "*la contrainte du flot optique*" (optical flow equation: OFE) puisqu'elle exprime une contrainte sur les composantes u et v du flot optique. On a besoin de plus de contraintes parce qu'on a une seule équation avec deux inconnus.

Une autre façon de voir le problème d'estimation du mouvement (2D) est de considérer " la différence du déplacement d'image " (displaced frame difference: DFD), définie par:

$$D = I(x + d_x, y + d_y, t + d_t) - I(x, y, t) \quad (3.8)$$

Il est clair, que la formule DFD diminue lorsque la précision du champ de mouvement estimé augmente, ainsi l'estimation du mouvement peut être formulé comme un problème de minimisation. C'est à dire chercher (d_x, d_y) minimisant (3.8) (on suppose de petits déplacements). En pratique, la valeur de D n'est jamais nulle à cause du bruit.

Ces différents points de vue conduisent à deux différentes philosophies d'estimation du mouvement 2D: l'une utilise OFE directement en mesurant le taux spatio-temporal du changement de l'intensité de l'image, l'autre minimise le DFD sur un ensemble de régions locales en recherchant sur un ensemble des vecteurs de mouvement.

Les algorithmes d'estimation du mouvement 2D peuvent être classés en quatre groupes:

1. Les méthodes basées gradient (les méthodes différentielles): Ces méthodes s'appuient sur l'estimation des gradients spatio-temporaux de l'intensité en chaque pixel. Elles utilisent la formule OFE.

2. Les méthodes d'appariement de régions (region matching methodes): ces méthodes sont introduites pour surmonter le problème de la sensibilité au bruit des méthodes différentielles, elles minimisent la formule DFD (ou un autre critère d'erreur similaire, généralement SSD), l'image considérée est découpée en régions, pour chacune on cherche la région qui lui semble le plus (dans l'image suivante de la séquence).

3. Les méthodes fréquentielles (frequency – domain methods): Elles exploitent les propriétés de la transformé de Fourier pour estimer le mouvement.

4. Les méthodes Bayésiennes (Bayesian methodes):elles sont appelée aussi" techniques stochastiques". L'approche stochastique est une technique très prometteuse grâce à son caractère général, sa robustesse vis à vis du bruit mais son inconvénient réside dans la recherche d'un modèle du mouvement approprié ce qui constitue un problème très difficile [Romanos03, Vaxoncelos01].

4.Analyse stéréo-mouvement:

La stéréovision et l'analyse du mouvement sont deux méthodes très répandues en vision par ordinateur, elles permettent l'extraction de l'information à partir des ensembles d'images. Dans la stéréovision, deux images acquises en même temps mais selon deux points de vu différents, sont utilisées pour calculer la profondeur. L'analyse du mouvement utilisent une séquence d'images acquises à des moments différents, pour localiser et mesurer le mouvement entre elles (les images). Les problèmes de la stéréovision et de l'analyse du mouvement sont similaires, ils sont des problèmes de mise en correspondance. Dans les deux cas, la tâche essentielle est la localisation d'un point physique dans chaque image. La stéréovision et l'analyse du mouvement sont appelées" forme à partir de stéréo (Shape From Stereo) et forme à partir du mouvement (Shape From Motion) respectivement [Milles97]. Semblable à la stéréovision, l'analyse du mouvement est essentiellement un processus d' appariement. Le but est de trouver les points d'une séquence d'images correspondant au même point physique.

Cependant, plusieurs différences existent entre l'appariement stéréoscopique et l'appariement du mouvement . En stéréovision, les deux images sont prises en même temps c'est à dire qu'un objet physique ne peut pas bouger par rapport à une seule image, les différences entre les deux images sont causées par la différence de positions des deux caméras. Les images utilisées par l'analyse du mouvement sont prises à des moments différents, donc l'objet peut bouger. Une autre différence est la connaissance de la géométrie de la caméra, généralement les applications de la stéréovision supposent qu'elle est connue ce qui n'est pas le cas pour l'analyse du mouvement.

Dans [Milles97], Milles a examiné la possibilité de combiner la stéréovision et l'analyse du mouvement. L'association de ces deux méthodes permet d'augmenter la robustesse, d'accéder à la fois à la structure et aux paramètres 3D du mouvement. Cette information est utilisée dans plusieurs applications à savoir : le suivi des objets animés, la reconnaissance des objets basée sur le mouvement et le système visuel d'un robot mobile [Milles97]. Une autre motivation de cette combinaison est l'obtention de contraintes supplémentaires pour les processus d'appariement. Ces contraintes résultent de l'observation que les correspondances stéréoscopiques et celles du mouvement doivent être consistantes, c'est à dire que les appariements stéréoscopiques doivent être conservés par les correspondances du mouvement (Voir Fig.3.10). Dans la première image de Fig.3.10. ($t = 0$), le point b (de l'image droite) est le

correspondant du point a (de l'image gauche) par stéréovision . Donc, au temps $t = 1$, a a bougé vers a' et b vers b' , et parce que les correspondances stéréoscopiques doivent être maintenues a' et b' sont liés. Dans le cas des caméras parallèles (les plans alignés) on a une contrainte supplémentaire, a et b ont la même composante y du mouvement parce qu'ils ont la même composante y , de même pour a' et b' .

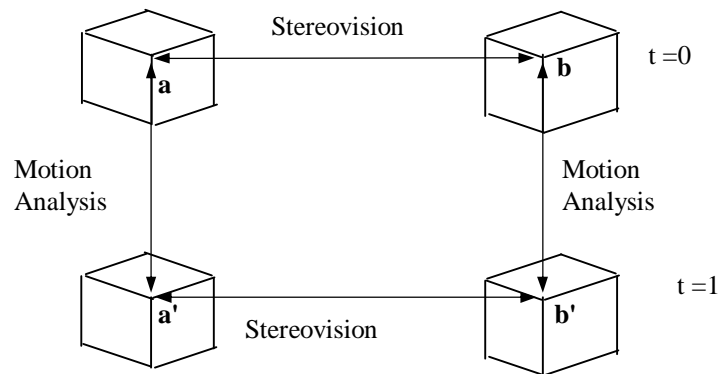


Fig.3.10. Mouvement – stéréo [Milles97].

5. L'estimation du mouvement de la camera

L'estimation du mouvement de la caméra, appelée *mouvement propre* (ego –motion), a pour but d'estimer le mouvement tridimensionnel de la camera à n'importe quel moment . Le mouvement propre peut être décomposé en une rotation R et une translation T [Tian96].

Le mouvement propre peut être estimé en utilisant des techniques basées capteurs (sensors based techniques) ou des techniques basées vision (le *suivi visuel*) [Lang02.b]. Les techniques basées capteurs utilisent des capteurs de position qui permettent d'obtenir directement et rapidement la position et l'orientation de la caméra à chaque instant. Cependant cette approche ne peut être utilisée que dans des environnements de tailles restreintes et elle est de plus sensible à la présence de matériaux perturbant le bon fonctionnement du capteur.

L'intérêt des approches de nature vision est de ne nécessiter aucune instrumentation de la scène. Ces approches cependant dépendent de la détection de primitives pour calculer le point de vue. La plupart des systèmes existants utilisent donc soit des environnements suffisamment simples, soit ils positionnent des modèles dans des environnements plus complexes afin de permettre un calcul facile du mouvement de la caméra.

5.1. Les techniques basées capteurs

La position et l'orientation de la caméra par rapport au système des coordonnées réelles, appelées "*la pose*" de la caméra, déterminent la transformation scène- caméra. A cause de la qualité de la précision demandée par le processus d'alignement des objets réels et des objets virtuels, aucun capteur tout seul n'a pu fournir le résultat souhaité pour la plupart des applications de la réalité augmentée [Shahzad02, Vallino95, Verplaetse97].

5.1.1. Les capteurs magnétiques

Durant les 20 ans précédents, le suivi magnétique était le chemin préféré pour mesurer les mouvements dans les applications d'infographie, parce qu'il facilite le calcul de la position et l'orientation des objets animés et même le mouvement de la caméra.

Le problème principal associé aux capteurs magnétiques est leur sensibilité aux interfaces électromagnétiques et à la présence des objets métalliques dans le champs magnétique. Un autre problème rencontré lorsqu'on utilise les capteurs magnétiques est la région de couverture.

5.1.2. Les capteurs inertiels

Un type de capteur plus effectif est appelé *capteur inertielle*, utilisé dans les avions, les sous marines pendant plusieurs années. Ce type de capteur est attaché directement à l'objet animé. On distingue deux parties composant ce type de capteur [Verplaetse97]:

- L' accéléromètre (accelometer) : mesure l'accélération linéaire de la tête de l'utilisateur (ou de la caméra) pour déterminer les translations.
- Le gyroscope (gyroscope): détermine la rotation.

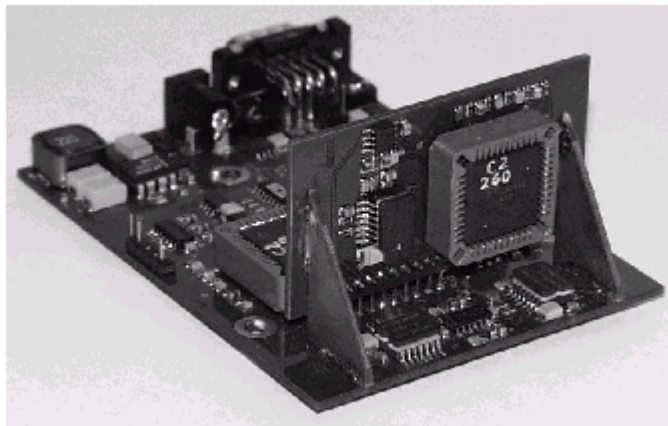


Fig.3.11. Le matériel utilisé [Verplaetse97].

Ce type de capteur souffre de l'accumulation des erreurs, avec le temps, se qui gêne la stabilité de l'alignement.

5.2. Les techniques visuelles

Les techniques visuelles utilisent les images capturées par le système d'acquisition pour retrouver le mouvement de la caméra. Ces techniques peuvent être classées en deux classes, les techniques de la première classe sont basées sur l'analyse du flot optique pour estimer le mouvement propre, elles sont appelées "méthodes différentielles" ou "méthodes continues" [Gonçalves02] ou les algorithmes à temps instantané dans [Tian96] (instantaneous-time algorithms). La deuxième classe contient des techniques appelées "suivi de primitives" ou les méthodes discrètes" (voir [Gonçalves02]), elles permettent de retrouver le mouvement et les propriétés d'un ensemble de primitives visuelles d'une image à l'autre [Verplaetse97].

5.2.1. Le mouvement propre et le flot optique

Le mouvement de la caméra cause trois formes principales du flot optique (différentes combinaisons peuvent exister) [Radthe01]:

- Une translation de la caméra dans le plan image ou une rotation (de petits angles) autour d'un axe perpendiculaire à l'axe optique permet d'obtenir des vecteurs du flot parallèle (possédant la même longueur).
- Une rotation est causée par la rotation de la caméra autour de l'axe optique.
- Le mouvement de la caméra, le long de l'axe optique cause un changement d'échelle.

Ø **La modélisation affine:** étant données les données du flot optique extraites d'une séquence vidéo, le mouvement des objets sur le plan image est lié aux mouvements des objets réels par un des modèles existants. Le modèle affine donné par l'équation (3.8) fournit une transformation linéaire entre les séquences d'images (flot optique) et le mouvement de la caméra [Verplaetse97].

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} b_x & c_x \\ b_y & c_y \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} + \begin{pmatrix} a_x \\ a_y \end{pmatrix} \quad (3.8)$$

Où

u, v : les termes du flot optique.

a_x, a_y : représente la translation.

b_x : un facteur d'échelle pour x dans la direction des x .

c_x : la rotation de x dans la direction des y .

b_y : la rotation de y dans la direction des x .

c_y : un facteur d'échelle pour les y dans la direction des y .

x, y : les coordonnées du monde réel.

Remarque: le modèle affine traite les translations selon z comme un zoom (avec des changements constants dans les termes b_y et c_y) ce qui explique l'absence du terme z .

Ø **La projection perspective:** le modèle de la projection perspective (équation 3.9) est une autre façon pour lier le flot optique avec le mouvement du monde réel [Verplaetse97]. Dans ce modèle non linéaire :

$$\begin{pmatrix} u \\ v \end{pmatrix} = \begin{pmatrix} x \\ y \end{pmatrix} \frac{1}{1 + zb} \quad (3.9)$$

Où :

u, v : les termes du flot optique.

x, y, z : les termes du mouvement du monde réel.

b : la distance focale inversée.

Six algorithmes pour calculer le mouvement propre à partir des vitesses d'image ont été évalués dans [Tian 96] (Bruss& Hurn , Jepson & Heeger, Tomasi & Shi, Prazdny, Kanatani, Kanatani).

5.2.2. Le mouvement propre et le suivi des primitives

Le suivi des primitives est le problème de la mise en correspondance des primitives d'une image à l'autre le long d'une séquence. Les points ont été la source d'information pour les algorithmes du suivi basés image (voir [Henele02]). Récemment, les contours actifs (voir [Deusch97]) ont été proposé comme des primitives suivies surtout lorsqu'il s'agit des applications ou il n'est pas demandé d'avoir une bonne précision tel que la navigation du robot. Le suivi des contours actifs permet de localiser le robot (l'estimation de la position et de l'orientation du robot) [Hirshmüller02]. Une méthode estimant le mouvement propre d'un robot

est présentée dans [Alenyà03]. Cette méthode est basée sur le suivi des contours en temps réel, en utilisant des images acquises par une caméra calibrée.

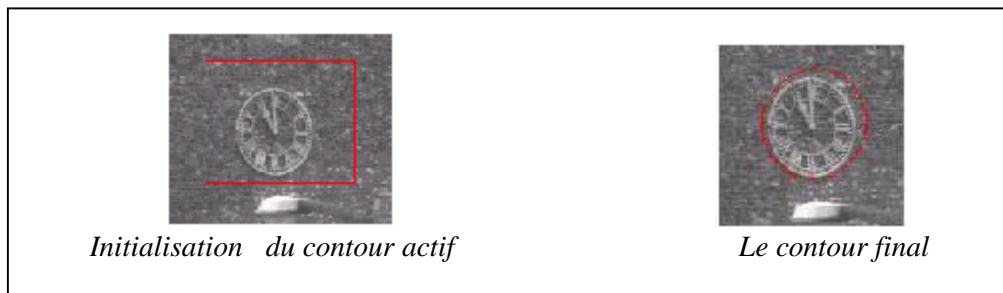


Fig.3.12.un exemple de contour actif [Deusch97].

L'estimation du mouvement propre est utilisée dans plusieurs applications. Habituellement, le suivi des primitives est réalisé via la séquence d'images pour déterminer les correspondances. De plus, des techniques statistiques robustes sont utilisées pour manipuler les points aberrants [Fisher04]. Une technique populaire utilisée pour le suivi des primitives est le filtre de Kalman (utilisée dans [Jennings96]). C'est une procédure récursive estimant la position d'un point comme une incertitude de l'image courante vers la suivante.

L'estimation de la pose d'une caméra est obtenue en suivant des repères artificiels (fiducials) dont leurs coordonnées 3D sont connues. Cette approche donne de bon résultat lorsque les images sont calibrées, les repères sont détectés et identifiés d'une façon fiable. Cependant, la calibration dans des grands environnements est difficile et même si elle est réalisée, l'exploitation du système de suivi est limitée aux régions où un minimum de fiducials est bien détecté. L'occlusion d'une partie du fiducial peut causer l'échec du suivi.

Dans [Koller97], Koller et Klinker ont proposé un repère automatique basé sur la calibration de caméra et initialisent l'estimateur de mouvement qui utilise les techniques du filtre de Kalman étendu pour suivre les repères et pour estimer le mouvement de la caméra à n'importe quel moment.

Lorsqu'une scène est planaire, l'estimation du mouvement basée sur l'homographie est le plus utilisée [Henele02, Yue04]. Puisque les caméras sont des outils projectifs liant le monde tridimensionnel avec le plan image, il est intuitif que l'estimation du mouvement utilisant une séquence d'images doit considérer la nature tridimensionnelle de l'environnement. Cependant, la forme de la scène est généralement inconnue et par conséquent il est préférable d'utiliser des

méthodes qui n'ont pas besoin de cette connaissance explicite. Une classe de technique connue pour l'estimation visuelle du mouvement de la caméra utilisent "les homographies" .Une homographie est une matrice 3 x 3 capable de décrire la transformation projective des points, situés sur un plan, entre les images. L'idée générale de cette classe de techniques est l'utilisation des primitives suivies pour calculer le déplacement entre deux positions de caméra. Ces déplacements peuvent être intégrés pour estimer le mouvement de la caméra [Henele02].

Une méthode permettant de calculer le mouvement 3D d'une caméra dans un environnement statique est introduite dans [Irani94], qui est basée sur le calcul du mouvement image d'une seule région directement à partir des intensités. Ce mouvement calculé est utilisé pour recalibrer les images. Dans cette méthode, le calcul du mouvement propre passe par deux étapes, la première consiste à calculer tout d'abord la translation toute seule, suivie par le calcul de la rotation dans la deuxième étape.

5.3.Les techniques hybrides

5.3.1. Les techniques optique- inertiel

La fusion du suivi basé vision avec le suivi inertiel (inertial tracking)est utilisée pour surmonter les inconvénients inhérentes à chaque classe ([Lang02.a, Strelow02,Verplaetse97, You99]). Dans [Verplaetse97], le système IOME cam (Inertial- optical motion- estimating camera) est conçu, pour combiner les techniques optiques et les techniques utilisant les capteurs inertiels afin de déterminer le mouvement de la caméra durant la séquence vidéo. IOME Cam utilise un estimateur joint au filtre de kalman (Kalman filter joint estimator) pour observer les deux types de données (optiques et inertiel) afin de prédire les paramètres du mouvement de la caméra. L'utilisation de deux types de données permet de corriger les erreurs liées à chacun des deux types de suivi.

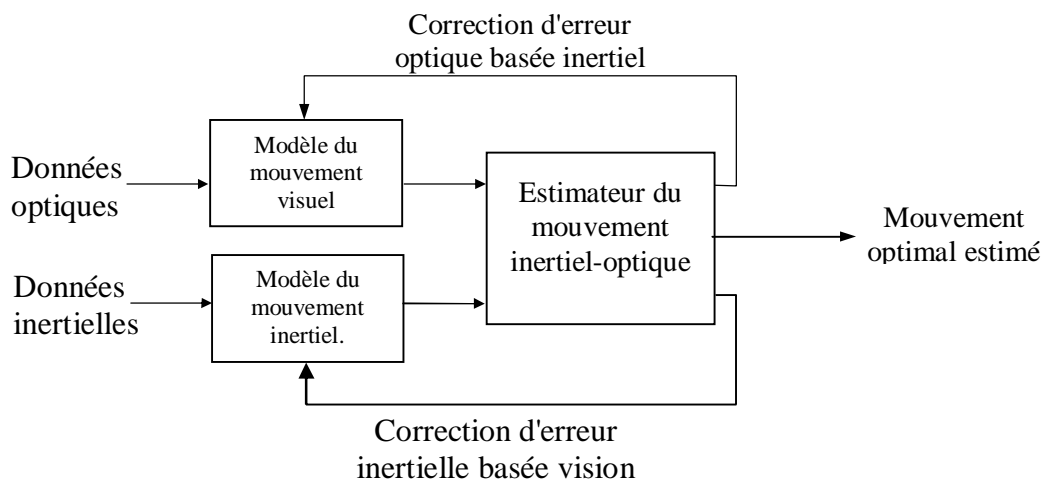


Fig.3.13.Le schéma du système IOME Cam [Verplaetse97].

6. Estimation du mouvement et le suivi d'objets

Le problème typique du mouvement est d'analyser le mouvement au niveau des données d'images (numériques) bidimensionnelles correspondant à une séquence d'échantillons spatio – temporelles, appelé " trame " (frame), d'une scène tridimensionnelle. Les trames sont acquises par une caméra à un taux de trame donné qui est 25HZ (ou 30). Cette fréquence temporelle permet d'avoir un mouvement régulier, réaliste lorsque l'être humain perçoit le vidéo clip [Krüger98].

Puisque la séquence capturée représente une projection 3D –2D, la relation entre les changements observables d'intensités (flot optique) et les mouvements des objets physiques tridimensionnels est complexe. Une raison principale de cette complexité est le fait q'une projection 3D –2D cause une perte d'information qui est la profondeur de la scène, causant ainsi une ambiguïté de projection. Cela signifie que plusieurs points physiques (avec des mouvements différents) seront projetés dans la même position sur le plan image . Fig. 3.15 illustre ce qui a été dit dans le cas d'une projection perspective.

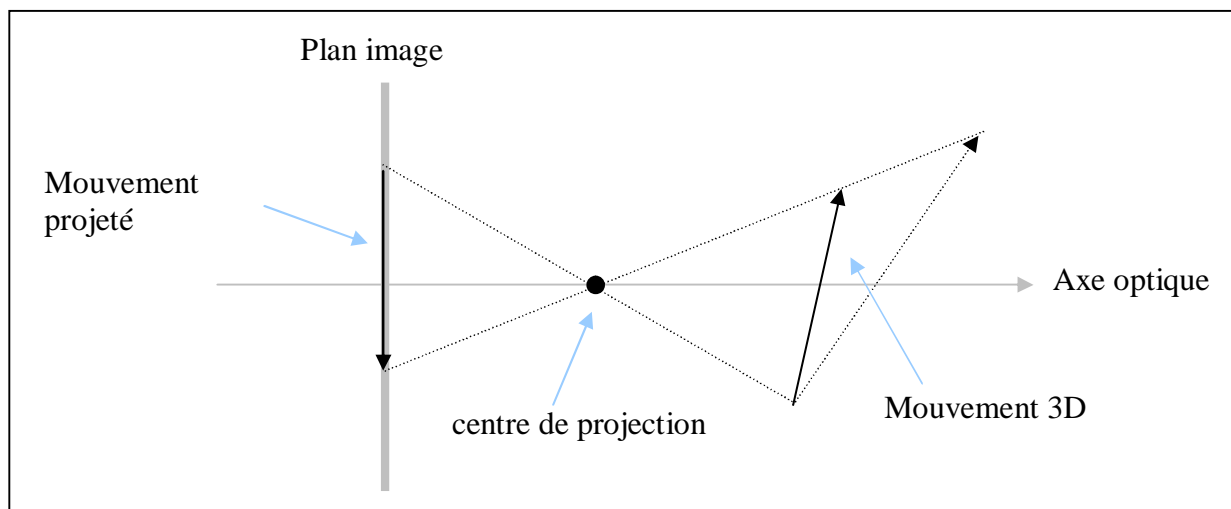


Fig.3.14.La géométrie de la projection perspective

La relation entre le flot optique et la scène dépend aussi du type du mouvement existant, on distingue trois classes: La première classe concerne les scènes statiques capturées par une caméra mobile, la deuxième classe concerne les scènes dynamiques acquises par des caméras statiques. La dernière classe comprend les scènes dynamiques prises par des caméras dynamiques. Tout ce qui a été dit suppose que les objets sont rigides, imaginez- vous la complexité du suivi d'un objet

non rigide qui se déplace dans un environnement dynamique capturé par une caméra mobile [Krüger98].

Une complication supplémentaire est le fait que la projection du mouvement tridimensionnel ne correspond pas nécessairement au flot optique, parce qu'il dépend d'autres facteurs tel que l'illumination globale de la scène. Deux points doivent être pris en considération [Krüger98]:

Ø Le premier point consiste à déterminer l'entité à suivre ou l'entité pour laquelle on estime le mouvement. On peut envisager des formulations en termes d'objets, primitives, pixels, le choix est dicté par le type d'application pour laquelle l'analyse du mouvement est envisagée. L'approche centrée pixel peut être vue comme adopter une approche de vision ancienne (early vision approche) tandis que l'utilisation de modèle d'objet est une tâche de vision d'un niveau plus élevé. L'utilisation des régions et des primitives se situe entre ces deux extrémités.

Ø Le deuxième point est l'utilisation d'un modèle tridimensionnel ou d'un modèle bidimensionnel. La scène est une représentation tridimensionnelle mais la séquence d'images est un ensemble de projections bidimensionnelles. L'application peut nécessiter l'analyse du mouvement 3D en termes d'objets physiques, comme il peut être suffisant d'analyser le mouvement dans le plan de projection (2D). En général, l'estimation du mouvement 3D à partir des projections 2D est un problème difficile, qui peut être considéré, en grande partie, non résoluble à moins que l'environnement est contraint, les applications pratiques des robots autonomes (autonomous) habituellement font des hypothèses sur l'entourage dans lequel ils opèrent, tel que l'existence de repères communs et fixes.

Dans la décennie précédente, le suivi d'objets est devenu très populaire parce qu'il est applicable aux problèmes quotidiens tel que les caméras de surveillances, la détection des plans. La supériorité du suivi des objets par rapport à leurs reconnaissances est devenue claire après les développements du traitement vidéo et de l'estimation du mouvement [Sebe02].

Les objets peuvent être classés en trois catégories: rigide, articulé et non rigide. Le mouvement d'un objet rigide est composé d'une translation et d'une rotation. Un objet articulé comprend des parties rigides liées par joints [Corbett02]. Le mouvement des objets non rigide peut être considéré comme une composante rigide plus une déformation non rigide. Par exemple, la tête d'un être humain peut bouger rigidement (tournant à droite ou à gauche) en même temps qu'il parle (un changement des expressions faciales), pour plus de détails voir [Torresani03].

La plupart des applications vidéo, tel que la surveillance, suppose la rigidité des objets [Corbett02, Mensouri00].

L'utilisation d'une seule caméra conduit à des ambiguïtés inhérentes. Par conséquent, il est avantageux d'utiliser plusieurs caméras au lieu d'une seule. D'un point de vue pratique. Il est plus ennuyeux de : calibrer un système de plusieurs caméras, modifier les données de la calibration pour pouvoir calibrer une caméra . La plupart des approches considèrent un système multi – caméra comme un appareil stéréo ([Delhumeau03, Frédérick01, Svoboda97]).

Le suivi des objets acquises par une caméra en mouvement est un défi à surmonter en vision par ordinateur. Lorsque la caméra bouge, elle génère un mouvement sur la totalité des vecteurs de position de l'image (un mouvement global), le problème du suivi ne peut pas être résolu directement par une simple détection du mouvement comme dans le cas des caméras statiques. Les méthodes du suivi, avec une caméra mobile, peuvent être divisées en deux catégories. Les méthodes de la première catégorie supposent que le mouvement de la caméra est connu ou que le fond de la scène présente des propriétés différentes (sur le plan image), cela conduit à des contraintes qui sont valide seulement pour les régions du fond. Les régions des objets mobiles violent ces contraintes et peuvent être ainsi détecté, les méthodes de la deuxième catégorie, accomplissent généralement le suivi en trois étapes successives [Feghali03] :

1. Elles supposent que le mouvement du fond est représenté par un modèle paramétrique, puis elles estiment les paramètres du mouvement du fond.
2. Elles détectent les objets mobiles en se basent sur le mouvement après la compensation du mouvement de la caméra.
3. Le suivi est réalisé en utilisant un appariement temporel entre les trames.

Les méthodes du suivi peuvent être basées contour ou basées région. Récemment, ces deux types de suivi ont été formulé en utilisant les PDE (level sets partial differential equations), ce formalisme est intéressant pour plusieurs raisons [Feghali03] :

1. Il considère les changement de topologie durant l'évolution du contour.
2. Il peut être implémenté en utilisant des méthodes numériques stables.

Les méthodes existantes basées sur ce formalisme, suivent les objets mobiles trame par trame, à l'exception de [Feghali03]. Elles ont besoin d'une segmentation des objets en utilisant un processus externe [Dengsheng01]. Quelques schémas supposent qu'une bonne estimation du mouvement de l'image est disponible à chaque instant ou que le fond possède des propriétés connues. Les méthodes basées PDE existantes, à l'exception de [Yilmaz04], partagent les défauts suivants : elles ne sont pas valides lorsque la caméra bouge, elles peuvent suivre seulement les objets en mouvement du départ (au début du suivi).

Un effort très important a été effectué sur le suivi des objets, les méthodes du suivi de plusieurs objets peuvent être classées en trois classes [Fisher04] :

1. Le suivi d'objet basé correspondances : les méthodes de cette classe, ont besoin de détecter l'objet à chaque instant. Le suivi est réalisé par l'établissement de correspondance de l'objet dans les trames consécutives. Les objets sont représentés par leurs barycentres ou leurs silhouettes.
2. Le suivi basé transformation : les méthodes de cette classe ont besoin de détecter l'objet une seule fois. Le suivi est réalisé en estimant le mouvement des objets dans des trames consécutives. L'objet est représenté par des surfaces planaires tel que le rectangle ou l'ellipse.
3. Le suivi basé contour : les méthodes basées sur le suivi basé contour ont besoin de détecter l'objet une seule fois. Le suivi est réalisé en cherchant le contour de l'objet à partir du contour de la trame précédente. Les objets sont représentés par leurs contours.

Remarque : la détection d'objet qui est commune à toutes les classes, peut être réalisée par détection de primitives, la soustraction de fond ou par segmentation.

7. Conclusion

Un alignement correct des objets virtuels avec l'image des objets réels nécessite la détermination du mouvement de la caméra (la cohérence du mouvement), les objets virtuels doivent se comporter d'une manière similaire au comportement des objets physiques causé par le changement du point de vue (ou le déplacement de la caméra). L'estimation du mouvement permet de déterminer les relations d'occlusion (dans le cas des scènes dynamiques). Le chapitre suivant est consacré à nos résultats. En premier lieu, nous présenterons notre approche "augmentation basée calibration" qui est inspirée de l'approche de Zonglei et de Boufama. Dans

la deuxième section, les étapes suivies dans notre approche "augmentation sans calibration" avec le résultat de l'augmentation sont présentés. Nous clôturerons le chapitre par une discussion.

Chapitre 4

Plates-formes

"If you can find a path with no obstacles, it probably doesn't lead anywhere"
Frank A. Clark

Chapitre4

Plates-formes développées pour la RA.

1. Introduction

Le but principal des systèmes de la réalité augmentée est d'améliorer la perception et l'interaction de l'utilisateur avec le monde réel, en incrustant des objets virtuels tridimensionnels.

Les systèmes de la réalité augmentée sont des systèmes : capables de combiner le monde réel et les objets virtuels, interactifs, en temps réel, capables de réaliser l'alignement tridimensionnel correct entre les objets des deux mondes réel et virtuel. Nous présentons dans ce chapitre nos deux plates-formes expérimentales permettant d'insérer un cube virtuel dans des images bidimensionnelles en utilisant des repères artificiels. Parler de plates-formes signifie qu'elles peuvent être utilisées comme bases pour insérer des objets virtuels tridimensionnels plus complexes (vases, bouteilles, êtres humains...) parce qu'on peut en-capsuler tout objet 3D par un rectangle.

Insertion d'un cube virtuel dans une image bidimensionnelle suit les deux étapes suivantes:

- Ø Ramener l'objet virtuel de son repère euclidien vers le repère scène (voir Fig.4.1.) en déterminant la transformation objet-scène. La détermination de cette dernière nécessite la reconstruction de la scène parce que le repère scène est inconnu.

- Ø Ramener l'objet virtuel du repère (3D) scène vers le repère (2D) image en déterminant la transformation scène-image (représentée par la matrice de projection).

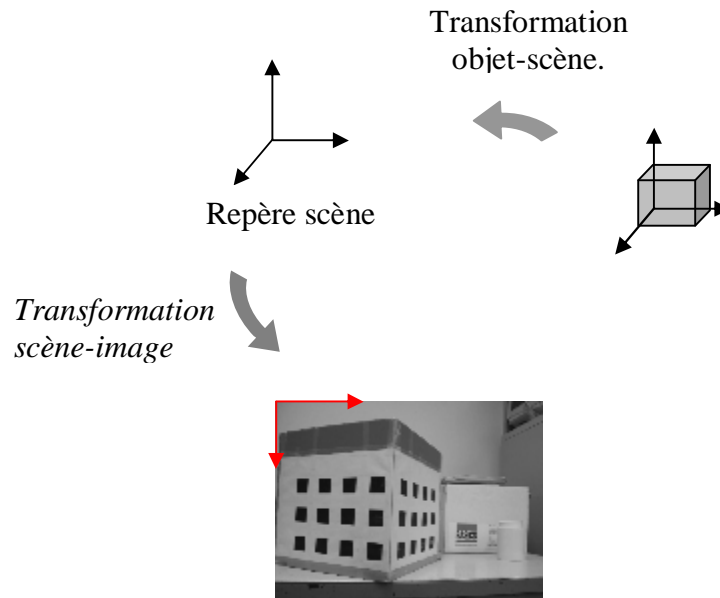


Fig.4.1. Etapes suivies pour insérer les objets virtuels.

La première plate-forme utilise un repère artificiel dont les coordonnées 3D sont connues afin de d'avoir une reconstruction fidèle à la scène réelle, on parle de la reconstruction euclidienne. La transformation objet-scène est ainsi réduite à une translation entre les deux repères euclidiens (le repère scène et le repère objet). La transformation scène-image est déterminée en utilisant le processus de la calibration.

La deuxième plate-forme utilise un repère artificiel dont les coordonnées tridimensionnelles sont inconnues. La transformation objet-scène appelée transformation objet-affine parce qu'il s'agit de déplacer l'objet virtuel de son repère euclidien vers le repère scène qui est un repère affine. La transformation scène-image est déterminée en utilisant la propriété de projection affine.

Ce chapitre est constitué de deux sections, la première section présente notre première plate-forme basée sur notre approche " Augmentation basée calibration" tout en présentant nos résultats. La deuxième section présente les étapes suivies pour augmenter une séquence d'images basée sur notre approche "Augmentation sans calibration". Cette section présente aussi les résultats de notre deuxième plate-forme. Nous clôturons ce chapitre par une discussion des résultats.

2. Augmentation basée calibration

Cette approche s'inspire de l'approche de Zonglei et de Boufama pour insérer un objet virtuel. Elle traite en plus le problème d'occlusion. Dans cette approche, l'utilisateur utilise une paire stéréoscopique d'images calibrées et un objet virtuel tridimensionnel. La figure Fig.4.2. présente les images utilisées dans notre plate forme.

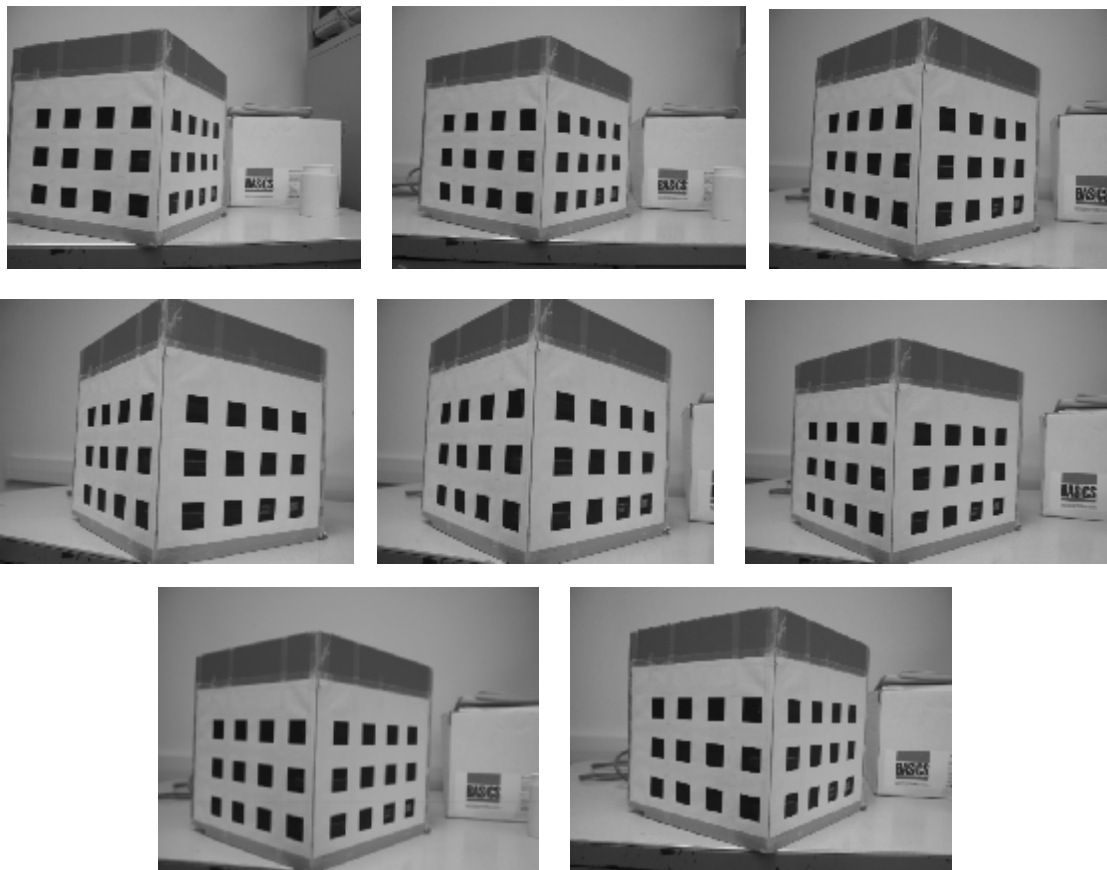


Fig.4.2. Exemples des données de test.

Les étapes suivies par l'approche " augmentation basée calibration" sont :

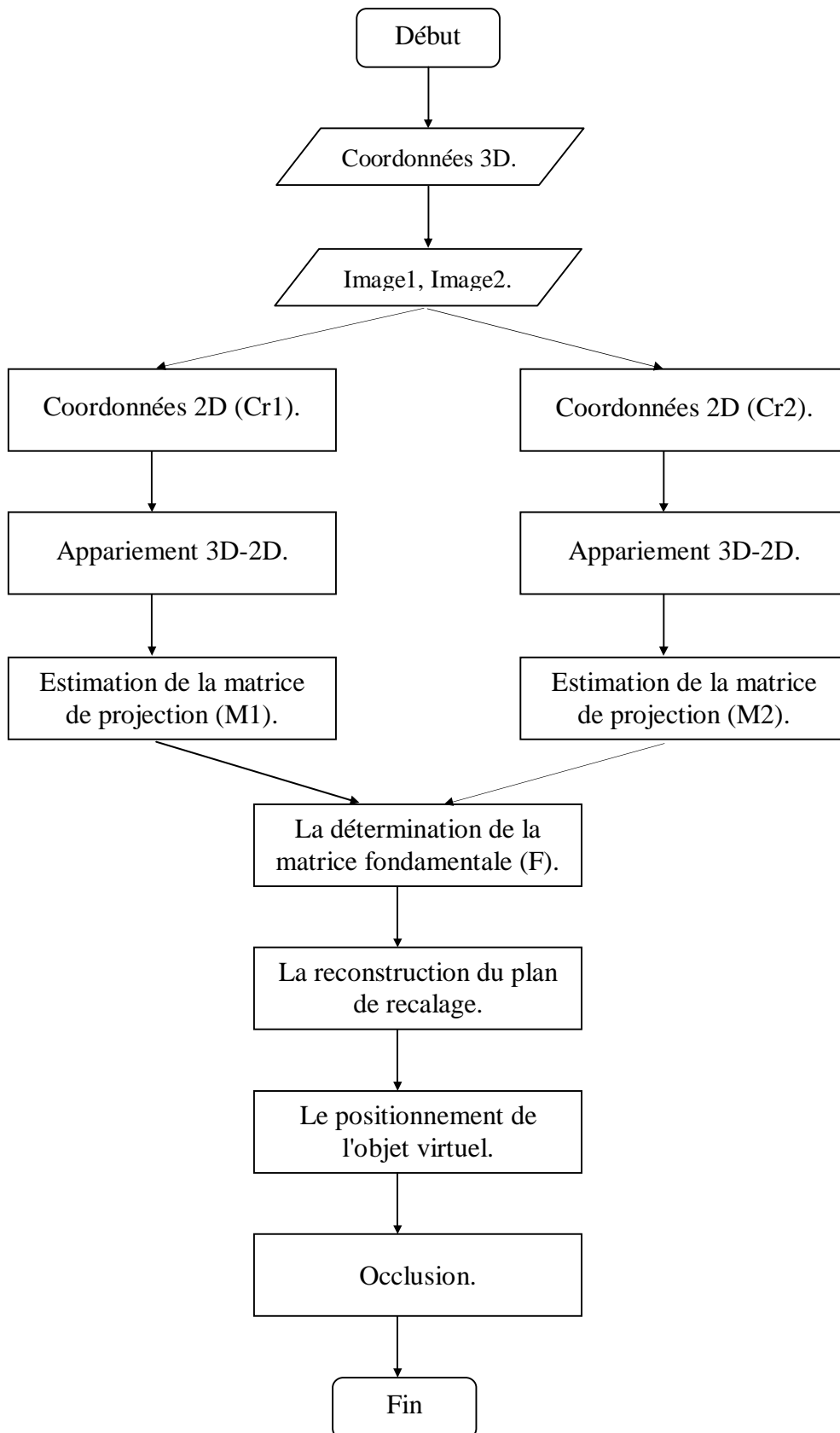


Fig.4.3. Les étapes suivies par l'augmentation basée calibration.

2.1. Appariement 3D-2D

Apparier les coordonnées tridimensionnelles avec leurs correspondants bidimensionnelles, nécessite l'appariement des coordonnées 3D des coins des fenêtres (les carrés noir) de la mire (le modèle 3D utilisé) avec leurs correspondants 2D. Zonglei et Boufama [Zonglei02.a, Zonglei02.b] ont proposé une méthode semi automatique permettant de trouver les appariements 3D-2D, en calculant une homographie pour chacun des deux plans.

Dans cette approche, l'utilisateur doit cliquer sur les quatre coins extrêmes de chacun des deux plans, le plan XY et le plan XZ (voir Fig.4.4). Pour calculer les valeurs des deux homographies.

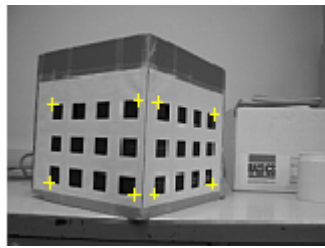


Fig.4.4. Les coins utilisés.

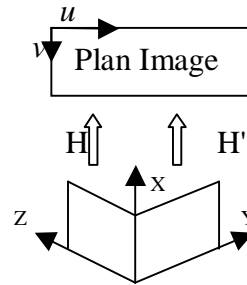


Fig.4.5. L'homographie.

L'équation (4.1) décrit les deux homographies:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = I \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \quad (4.1)$$

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = I \begin{pmatrix} \dot{h}_{11} & \dot{h}_{12} & \dot{h}_{13} \\ \dot{h}_{21} & \dot{h}_{22} & \dot{h}_{23} \\ \dot{h}_{31} & \dot{h}_{32} & \dot{h}_{33} \end{pmatrix} \begin{pmatrix} x \\ z \\ 1 \end{pmatrix}$$

La méthode SVD (Singular Value Decomposition: Décomposition en valeurs singulières) est utilisée pour résoudre le système ci dessous:

$$\begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -u_1 x_1 & -u_1 y_1 \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -v_1 x_1 & -v_1 y_1 \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & & & & \\ x_4 & y_4 & 1 & 0 & 0 & 0 & -u_4 x_4 & -u_4 y_4 \\ 0 & 0 & 0 & x_4 & y_4 & 1 & -v_4 x_4 & -v_4 y_4 \end{pmatrix} \begin{pmatrix} h_{11} \\ h_{12} \\ h_{13} \\ h_{21} \\ h_{22} \\ h_{23} \\ h_{31} \\ h_{32} \end{pmatrix} = \begin{pmatrix} u_1 \\ v_1 \\ \mathbf{M} \\ u_4 \\ v_4 \end{pmatrix} \quad (4.2)$$

Les coordonnées calculées par les deux homographies ne sont pas exactes car l'utilisateur est incapable de cliquer exactement sur le bon point. Pour cette raison le résultat des homographies est utilisé comme référence aux points résultants du détecteur de coins, on garde les coordonnées détectées proches des coordonnées calculées c'est à dire qu'on garde les coordonnées présentant une erreur de deux pixels. Le calcul de l'erreur se fait suivant la formule (4.3)

$$E = \sum_{i=1}^N \|x_i - u_i\|^2 + \|y_i - v_i\|^2 \quad (4.3)$$

Où:

(x_i, y_i) les coordonnées détectées par le détecteur de coins.

(u_i, v_i) les coordonnées calculées par l'homographie.

2.2. La matrice de projection

La relation entre un point de la scène $P(X, Y, Z, 1)$ est sa projection $p(u, v, 1)$ est donnée par la relation suivante:

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \quad (4.4)$$

L'équation (4.4) fournit deux équations fondamentales:

$$u = \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \quad (4.5)$$

$$v = \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}}$$

Où:

m_{ij} sont les éléments de la matrice de projection.

La méthode de décomposition en valeurs singulières est une bonne méthode pour résoudre le système suivant:

$$\begin{pmatrix} X_1 & Y_1 & Z_1 & 1 & 0 & 0 & 0 & 0 & -u_1X_1 & -u_1Y_1 & -u_1Z_1 & -u_1 \\ 0 & 0 & 0 & 0 & X_1 & Y_1 & Z_1 & 1 & -v_1X_1 & -v_1Y_1 & -v_1Z_1 & -v_1 \\ X_2 & Y_2 & Z_2 & 1 & 0 & 0 & 0 & 0 & -u_2X_2 & -u_2Y_2 & -u_2Z_2 & -u_2 \\ 0 & 0 & 0 & 0 & c & Y_2 & Z_2 & 1 & -v_2X_2 & -v_2Y_2 & -v_2Z_2 & -v_2 \\ \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} & \mathbf{M} \\ X_n & Y_n & Z_n & 1 & 0 & 0 & 0 & 0 & -u_nX_n & -u_nY_n & -u_nZ_n & -u_n \\ 0 & 0 & 0 & 0 & X_n & Y_n & Z_n & 1 & -v_nX_n & -v_nY_n & -v_nZ_n & -v_n \end{pmatrix} \begin{pmatrix} m_{11} \\ m_{12} \\ m_{13} \\ m_{14} \\ m_{21} \\ m_{22} \\ m_{23} \\ m_{24} \\ m_{31} \\ m_{32} \\ m_{33} \\ m_{34} \end{pmatrix} = 0 \quad (4.6)$$

La table 4.1 contient les matrices de projection des images utilisées.

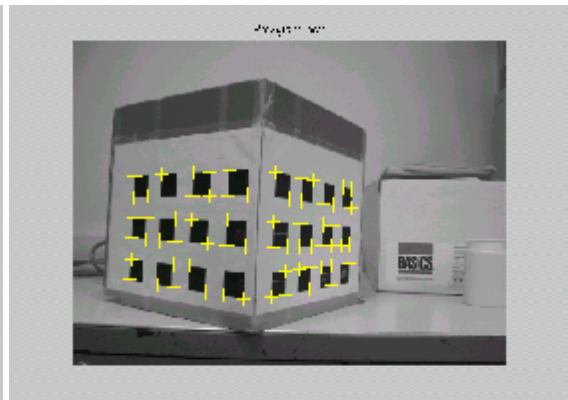
m_{ij}	1	2	3	4
1	0.4665	10.1725	-5.6292	213.1173
2	-10.366	4.0210	2.4108	346.6853
3	0.0003	0.0172	0.0125	1.0000
1	0.5990	9.4401	-7.7436	215.2268
2	-10.9095	4.5161	1.8683	354.2236
3	0.0008	0.0198	0.0103	1.0000
1	0.6993	13.5463	-4.4762	217.7523
2	-12.4950	4.3817	3.7306	376.6168
3	0.0008	0.0183	0.0184	1.0000
1	0.8468	15.8208	-1.6972	221.5089
2	-13.6594	4.0600	4.9324	392.8306
3	0.0009	0.0161	0.0235	1.0000
1	0.8992	16.0335	-4.1648	244.5548
2	-14.2423	4.8166	4.4899	399.4926
3	0.0013	0.0199	0.0219	1.0000
1	0.2025	12.2634	-3.8008	203.3318
2	-11.4630	3.4425	3.0748	390.6496
3	-0.0006	0.0161	0.0167	1.0000
1	0.2464	11.4637	-4.8848	205.8581
2	-11.1894	3.6064	2.7215	387.7937
3	-0.0004	0.0170	0.0150	1.0000
1	-0.164	11.4440	-5.4828	264.3412
2	-11.4670	3.3254	2.1697	362.0337
3	-0.0014	0.0182	0.0140	1.0000

Table.4.1. les matrices de projection.

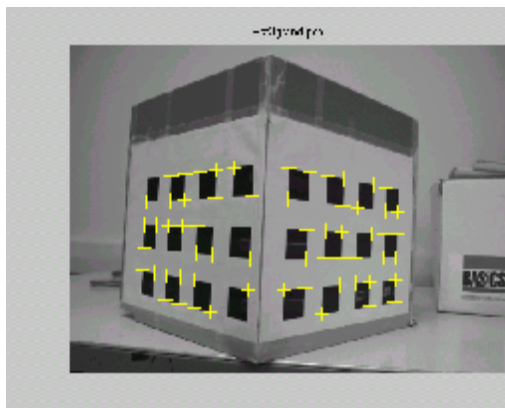
Les figures Fig.4.6. et Fig.4.7. présentent un test visuel des points projetés, c'est un test illustrant la précision de la calibration. La table.4.2. présente un autre test de la précision de la calibration qui est le calcul de l'erreur de re-projection. On remarque que l'erreur est autour de deux pixels.



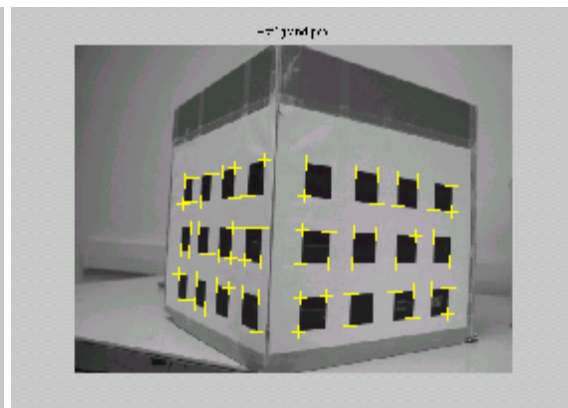
(a)



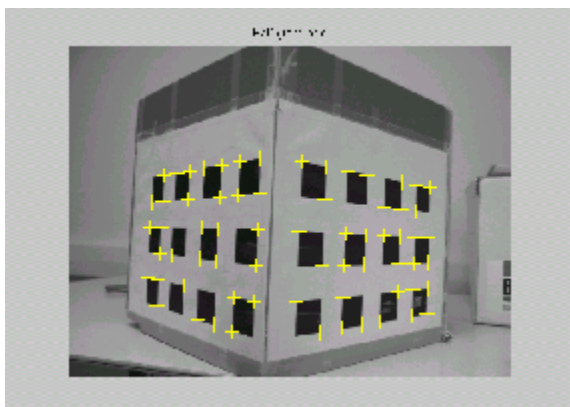
(b)



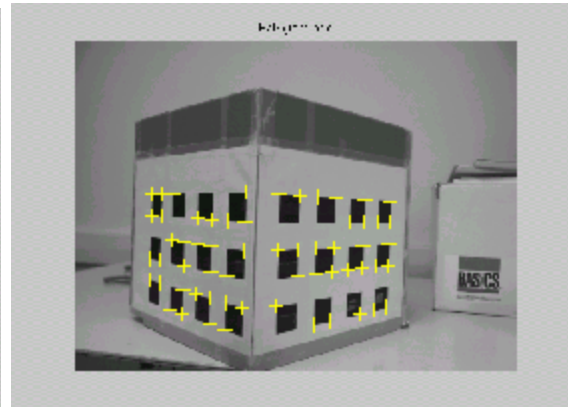
(c)



(d)



(e)



(f)

Fig.4.6. Un test visuel de la précision de la calibration des images (a, b, c, d).

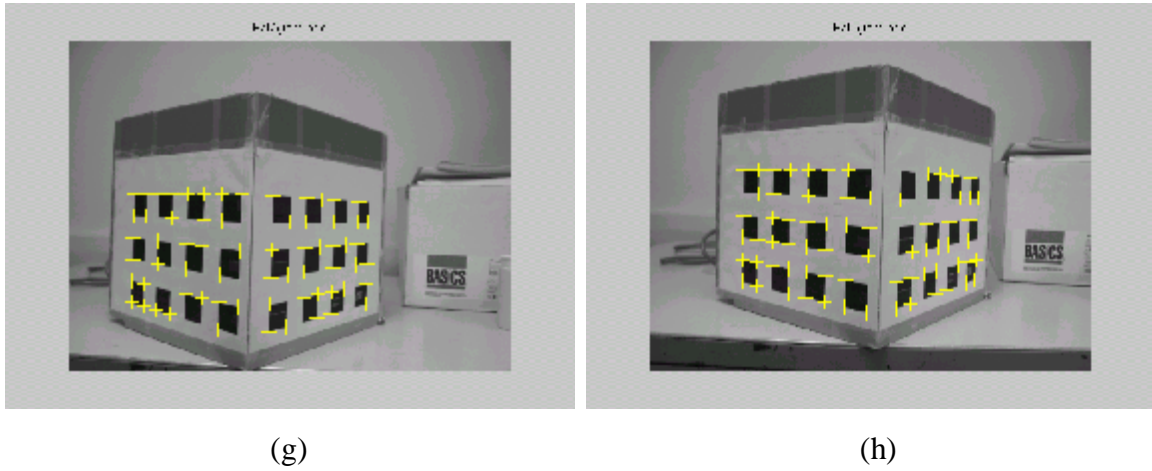


Fig.4.7. Un test visuel de la précision de la calibration des images (g, h).

Les coordonnées des points images		Les coordonnées des points re-projetés		Erreur
u	v	x	y	$\sqrt{(u-x)^2 + (v-y)^2}$
377.0000	278.0000	377.6121	277.4507	0.8225
369.0000	280.0000	367.5876	281.1306	1.8092
356.0000	286.0000	356.7774	285.0988	1.1902
346.0000	289.0000	345.0855	289.3908	0.9945
331.0000	295.0000	332.3993	294.0478	1.6926
319.0000	299.0000	318.5864	299.1183	0.4302
302.0000	305.0000	303.4897	304.6602	1.5280
288.0000	310.0000	286.9215	310.7422	1.3092
378.0000	256.0000	378.7368	254.5067	1.6652
370.0000	258.0000	368.7022	257.3362	1.4577
357.0000	262.0000	357.8768	260.3887	1.8344
347.0000	265.0000	346.1634	263.6916	1.5530
332.0000	269.0000	333.4480	267.2770	2.2507
320.0000	272.0000	319.5963	271.1828	0.9115
303.0000	277.0000	304.4488	275.4540	2.1188
288.0000	281.0000	287.8149	280.1443	0.8755
379.0000	231.0000	379.8730	231.3289	0.9329

Table.4.2. Les erreurs de la re-projection.

2.3. La détermination de la matrice fondamentale

La matrice fondamentale est la représentation géométrique de la géométrie épipolaire.

$$m_r^t F m_l = 0 \quad (4.7)$$

Où:

$$m_l = (u, v, 1) \text{ et } m_r = (u', v', 1)$$

La connaissance de la scène et l'estimation des matrices de projection permet d'avoir 96 correspondances ce qui permet d'échapper au problème des points aberrants et facilite la détermination de la matrice fondamentale. Une fois la matrice fondamentale est déterminée on peut déterminer pour chaque point la droite épipolaire sur laquelle se trouve son correspondant. Dans Fig.4.8. figure deux exemples de deux droites épipolaires. Dans la figure Fig.4.8.b figure la droite épipolaire du point de la figure Fig.4.8.a. Dans la figure Fig.4.8.d. figure la droite épipolaire du point de la figure Fig.4.8.c.

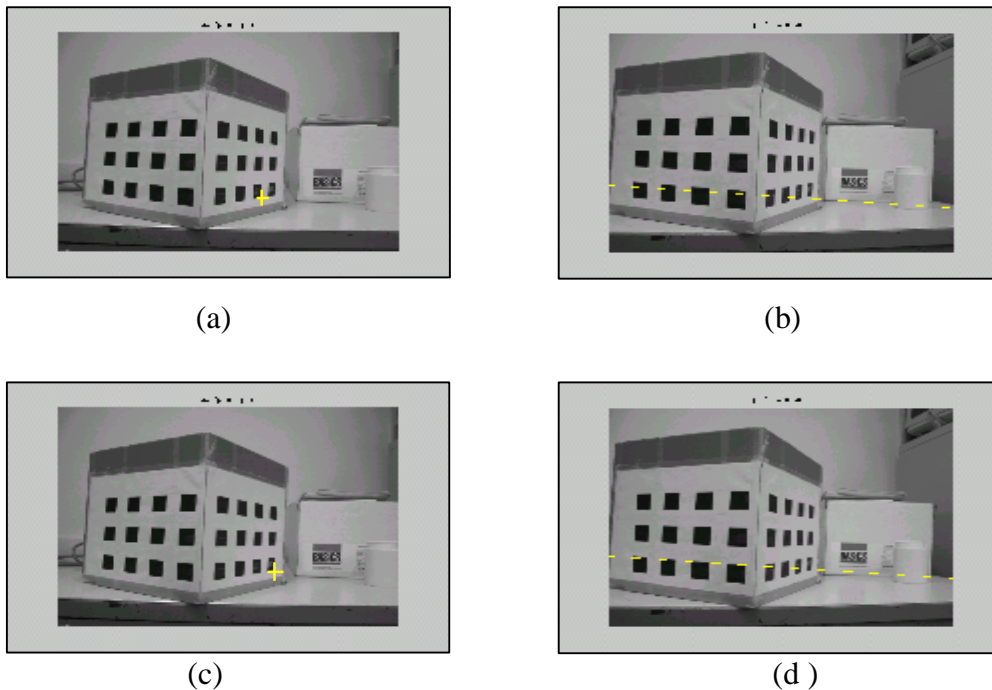


Fig.4.8. Illustration des droites épipolaires.

2.4. La reconstruction du plan de recalage

L'insertion d'un objet tridimensionnel dans une paire d'images bidimensionnelles nécessite la reconstruction de la scène. Un des avantages de cette approche est qu'elle reconstruit seulement " le plan de recalage ", qui est le plan sur lequel l'objet virtuel est placé.

La reconstruction du plan de recalage suit les étapes suivantes:

- ü L'utilisateur spécifie trois points en cliquant sur l'une des deux images.
- ü Les correspondants sont automatiquement déterminés en utilisant la contrainte épipolaire et une mesure de similarité.
- ü Ainsi, on peut déterminer les coordonnées tridimensionnelles des trois points par triangulation.
- ü Le plan du recalage est défini en déterminant les coefficients de l'équation (4.8):

$$0 = a X + b Y + c Z + d \tag{4.8}$$

2.4.1. La triangulation

Rappelez vous qu'un point physique est lié par la matrice de projection avec sa projection bidimensionnelle.

$$\begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = M \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix} \tag{4.9}$$

Si on connaît la matrice de projection et les projections d'un point tridimensionnel dans les deux images, on peut facilement déterminer les coordonnées (X, Y, Z) par la résolution du système d'équations (4.10).

$$\begin{aligned}
 u &= \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\
 v &= \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\
 u' &= \frac{m'_{11}X + m'_{12}Y + m'_{13}Z + m'_{14}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}} \\
 v' &= \frac{m'_{21}X + m'_{22}Y + m'_{23}Z + m'_{24}}{m'_{31}X + m'_{32}Y + m'_{33}Z + m'_{34}}
 \end{aligned}
 \tag{4.10}$$

Le système (4.10) comporte quatre équations et trois inconnus. On peut le réécrire comme suit:

$$\begin{pmatrix} m_{11} - m_{31}u & m_{12} - m_{32}u & m_{13} - m_{33}u \\ m_{21} - m_{31}v & m_{22} - m_{32}v & m_{23} - m_{33}v \\ m'_{11} - m'_{31}u' & m'_{12} - m'_{32}u' & m'_{13} - m'_{33}u' \\ m'_{21} - m'_{31}v' & m'_{22} - m'_{32}v' & m'_{23} - m'_{33}v' \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \begin{pmatrix} m_{34}u - m_{14} \\ m_{34}v - m_{24} \\ m'_{34}u' - m'_{14} \\ m'_{34}v' - m'_{24} \end{pmatrix}
 \tag{4.11}$$

La détermination de (X, Y, Z) est réalisé par SVD. En résumé, les coordonnées d'un point physique peuvent être déterminées si on connaît ses projections sur au moins deux images (pour une scène connue).

2.5. Le positionnement de l'objet virtuel

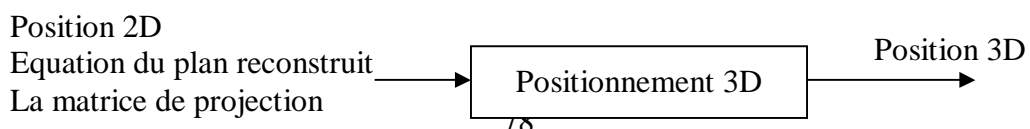
Le positionnement de l'objet virtuel suit deux étapes:

ü Positionnement 3D.

ü Positionnement 2D.

2.5.1. Positionnement 3D

Le but de cette étape est de déterminer la position tridimensionnelle de l'objet virtuel sur le plan tridimensionnel reconstruit à partir de la position bidimensionnelle spécifiée par l'utilisateur en déterminant la transformation objet-scène.



La position de l'objet virtuel sur le plan de recalage est déterminée en suivant:

- ü L'utilisateur doit spécifier où il veut mettre l'objet virtuel (u, v) en cliquant sur la projection bidimensionnelle du plan de recalage reconstruit (sur une image).
- ü La position tridimensionnelle est déterminée par la résolution de l'ensemble des équations:

$$\begin{aligned}u &= \frac{m_{11}X + m_{12}Y + m_{13}Z + m_{14}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}} \\v &= \frac{m_{21}X + m_{22}Y + m_{23}Z + m_{24}}{m_{31}X + m_{32}Y + m_{33}Z + m_{34}}\end{aligned}\tag{4.12}$$

$$0 = a X + b Y + c Z + d.$$

Où

(u, v) : les coordonnées 2D de la position désirée.

(X, Y, Z) : les coordonnées 3D de la position (u, v) .

m_{ij} : sont les éléments de la matrice de projection.

2.5.2. Positionnement 2D

Le but du positionnement 2D consiste à positionner l'objet virtuel sur la projection du plan de recalage, ce qui revient à appliquer la transformation scène-image (la matrice de projection) sur l'objet virtuel.

2.6. Occlusion

L'occlusion entre l'objet rouge et l'objet jaune de la figure Fig.4.9. est déterminée en utilisant une technique basée profondeur, c'est à dire que la région possédant une profondeur plus grande est placée derrière la région possédant une profondeur minimale. Dans les figures 4.9.a et 4.9.b, l'occlusion de deux objets virtuels est présentée. L'objet ayant la couleur jaune occulte correctement le rouge dans la figure 4.9.a. L'inverse figure dans la figure .4.9.b (l'objet rouge occulte correctement le jaune).

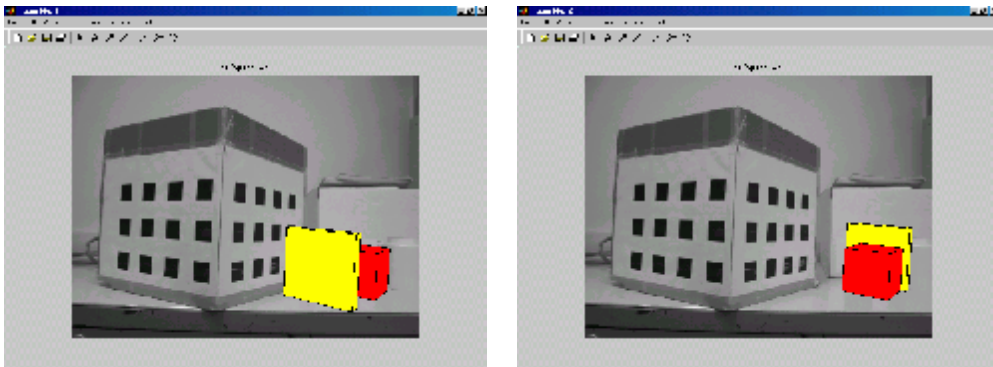


Fig.4.9. Illustration de l'occlusion pour la RA.

3. Augmentation sans calibration

Notre approche "augmentation sans calibration" s'inspire des travaux de Kutulakos et de Vallino. Elle est motivée des approches récentes de la réalité augmentée basées vidéo. Ces approches suivent un petit nombre de repères placés dans la scène réelle pour affaiblir les besoins de la calibration. Dans notre plate forme, le nombre de repères est égal à quatre (voir Fig.4.10).

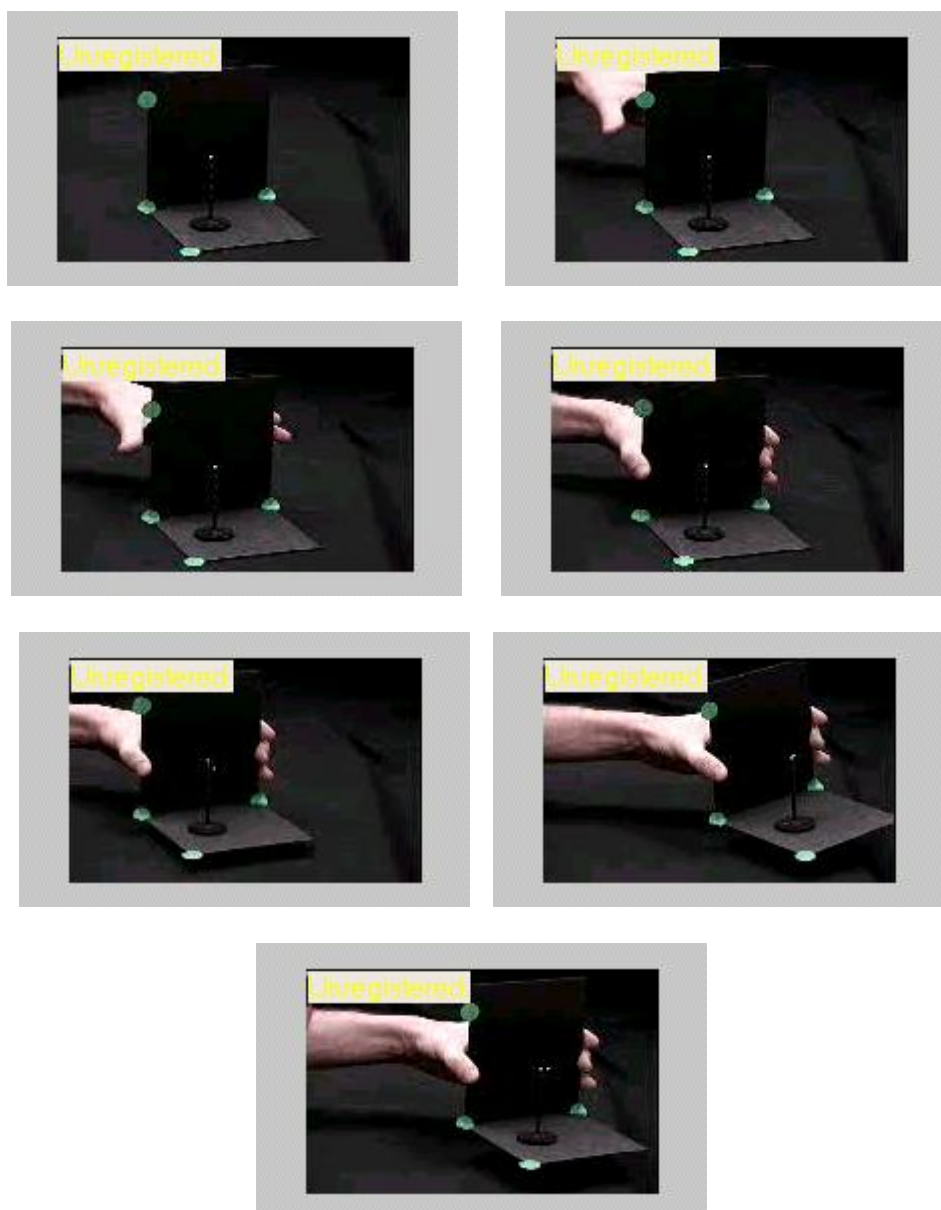


Fig.4.10. Exemples de données de test.

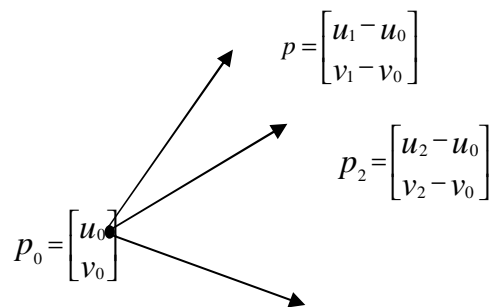
La caractéristique principale du système basé sur cette approche est qu'il permet le positionnement de l'objet virtuel sans la prise en considération d'aucune information sur les paramètres de la caméra et la position tridimensionnelle de l'objet virtuel. Le seul besoin est la capacité de suivre au moins quatre primitives dont les coordonnées tridimensionnelles sont inconnues.

Cette approche est basée sur l'observation suivante (observée par : Koenderink, Van Doorn, Ulman et Basri) :

Etant donné un ensemble de quatre (ou plus) points tridimensionnels non coplanaires, la projection de tous les points de l'ensemble peut être calculée comme une combinaison linéaire de projections de seulement quatre points.

Cette observation peut être exploitée comme suit :

- ü La représentation des objets virtuels afin que leurs projections peuvent être calculées comme une combinaison linéaire de la projection des points.
- ü La représentation affine comprend trois composantes :
 1. L'origine qui est un des points p_1, \dots, p_n .
 2. Les points de la base affine, qui sont trois points de la collection (non coplanaires avec l'origine).
 3. Les coordonnées affines des points p_1, \dots, p_n , exprimant les points $p_i, i=1..n$ en terme de points de base affine et de l'origine.



$$P_3 = \begin{bmatrix} u_3 - u_0 \\ v_3 - v_0 \end{bmatrix}$$

Fig.4.11. la représentation de point affine.

Les étapes suivies par l'approche " augmentation sans calibration" sont :

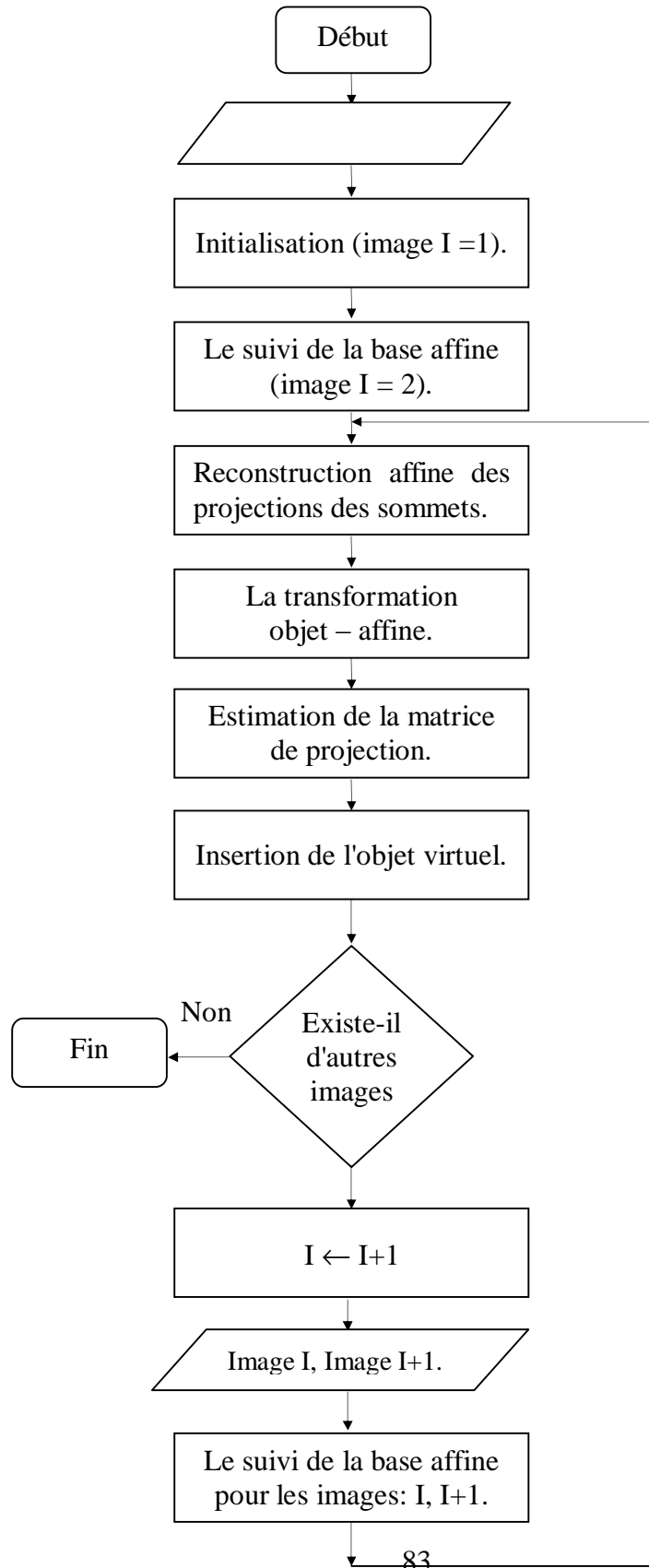


Fig.4.12. Les étapes suivies par l' augmentation sans calibration.

3.1. Initialisation

L'utilisateur sélectionne les quatre repères de la première image pour déterminer les couleurs des repères. Il doit aussi déterminer les projections de quatre sommets non-coplanaires de l'objet virtuel.

3.2. Le suivi de la base affine

Le suivi de la base affine est réalisé par une segmentation des régions basée couleur. Pour ce faire on suit les étapes suivantes:

- Ü Un seuillage basé sur les valeurs RGB est appliqué sur l'image en entrée. Le but est de garder seulement les repères. La figure Fig.4.13.b. représente le résultat du seuillage de la figure Fig.4.13.a.



Fig.4.13. le résultat du seuillage.

ù La recherche des composants connectés suit l'algorithme suivant [Bebis03.b]:

1. Balayer l'image pour trouver un 0 pixel (un pixel noir) non étiqueté et lui assigner une étiquette L.
2. Récursivement assigner l'étiquette à tous les pixels voisins.
3. Arrêter lorsqu'il y a plus de 0 pixels non étiquetés.
4. Retourner à l'étape 1.
5. Balayer l'image de gauche à droite et du haut vers le bas.
6. Si le pixel est noir (0 pixel) donc :
 - 6.1. Si seulement un de ses voisins (supérieur ou celui à gauche) est étiqueté, copier l'étiquette.
 - 6.2. Si les deux voisins possèdent la même étiquette, copier l'étiquette.
 - 6.3. Si les deux voisins possèdent deux étiquette différentes, donc copier l'étiquette du voisins supérieur et entrer les étiquettes dans la table d'équivalence.
 - 6.4. Sinon, assigner une nouvelle étiquette au pixel concerné.
7. S'il n'y plus de pixels à considérer, retourner à l'étape 6.
8. Trouver l'étiquette la plus basse pour chaque ensemble d'équivalences.
9. Balayer l'image. Remplacer chaque étiquette par l'étiquette la plus basse trouvée dans l'étape8.

ü Chaque repère est représenté par son barycentre réduisant ainsi le suivi de la base affine a un suivi des barycentres (voir Fig.4.14).



Fig.4.14. le résultat du suivi. La position des repères est déterminée par les barycentres de chaque ellipse.

3.3. La reconstruction affine

Les coordonnées affines de n'importe quel point peuvent être calculées de l'équation (4.13), si sa projection sur au moins deux vues est connue et les projections des points de base affine sont aussi connues dans ces deux vues :

$$\begin{pmatrix} u_p^1 \\ v_p^1 \\ u_p^2 \\ v_p^2 \end{pmatrix} = \begin{pmatrix} u_{p1}^1 - u_{po}^1 & u_{p2}^1 - u_{po}^1 & u_{p3}^1 - u_{po}^1 & u_{po}^1 \\ v_{p1}^1 - v_{po}^1 & v_{p2}^1 - v_{po}^1 & v_{p3}^1 - v_{po}^1 & v_{po}^1 \\ u_{p1}^2 - u_{po}^2 & u_{p2}^2 - u_{po}^2 & u_{p3}^2 - u_{po}^2 & u_{po}^2 \\ v_{p1}^2 - v_{po}^2 & v_{p2}^2 - v_{po}^2 & v_{p3}^2 - v_{po}^2 & v_{po}^2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (4.13)$$

3.4. La transformation objet-affine

Avant la superposition des objets virtuels sur l'image de l'environnement 3D, la relation entre ces objets et l'environnement doit être établie. L'approche "augmentation sans calibration" emprunte quelques résultats de la stéréovision :

Etant donné un point dans l'espace, sa position tridimensionnelle est uniquement déterminées par la projection du point

dans les deux images prises selon deux points de vue différents (deux positions différentes).

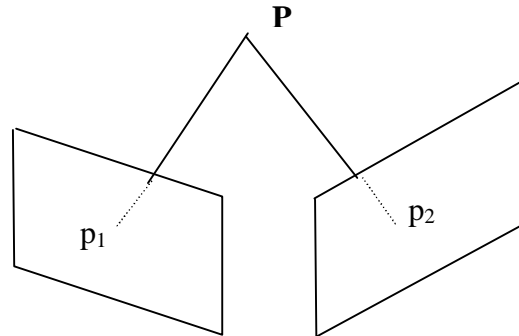


Fig.4.15. La position de P est uniquement déterminée par sa projection sur deux images.

Pratiquement cela entraîne la détermination des points de l'objet virtuel dans deux images dans lesquelles les points de base affine sont visible. Les questions principales sont :

1. Combien de projections de points doivent être déterminées dans les deux images ?
2. Comment fait-on pour déterminer la projection de ces points ?
3. Comment déterminer la représentation affine de l'objet à partir de ces projections ?

Le nombre de correspondances de points nécessaire pour déterminer la position et la forme de l'objet virtuel, est équivalent au nombre de points qui déterminent la transformation objet-affine. Cette transformation est déterminée en spécifiant la position 3D de quatre points non coplanaires de l'objet virtuel. Si p_1, p_2, p_3, p_4 sont les coordonnées de quatre points non coplanaires de l'objet virtuel exprimées dans le repère objet et p'_1, p'_2, p'_3, p'_4 sont leurs correspondants dans le repère affine, les repères sont liés par une transformation irréversible objet-affine A tel que :

$$[p'_1 \ p'_2 \ p'_3 \ p'_4] = A [p_1 \ p_2 \ p_3 \ p_4] \quad (4.14)$$

L'équation (4.15) peut être réécrite comme suit:

$$\begin{pmatrix} x_A \\ y_A \\ z_A \\ 1 \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} \end{pmatrix} \begin{pmatrix} x_0 \\ y_0 \\ z_0 \\ 1 \end{pmatrix} \quad (4.15)$$

Où: $(x_A, y_A, z_A, 1)$: les coordonnées affines 3D.

$(x_0, y_0, z_0, 1)$: les coordonnées 3D dans le repère objet.

La détermination de A revient à résoudre le système:

$$\begin{pmatrix} x_0^1 & y_0^1 & z_0^1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_A^1 x_0^1 & -x_A^1 y_0^1 & -x_A^1 z_0^1 & -x_A^1 \\ 0 & 0 & 0 & 0 & x_0^1 & y_0^1 & z_0^1 & 1 & 0 & 0 & 0 & 0 & -y_A^1 x_0^1 & -y_A^1 y_0^1 & -y_A^1 z_0^1 & -y_A^1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_0^1 & y_0^1 & z_0^1 & 1 & -z_A^1 x_0^1 & -z_A^1 y_0^1 & -z_A^1 z_0^1 & -z_A^1 \\ \mathbf{M} & & & & \mathbf{M} & & & & \mathbf{M} & & & & & & & \mathbf{M} \\ x_0^4 & y_0^4 & z_0^4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -x_A^4 x_0^4 & -x_A^4 y_0^4 & -x_A^4 z_0^4 & -x_A^4 \\ 0 & 0 & 0 & 0 & x_0^4 & y_0^4 & z_0^4 & 1 & 0 & 0 & 0 & 0 & -y_A^4 x_0^4 & -y_A^4 y_0^4 & -y_A^4 z_0^4 & -y_A^4 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & x_0^4 & y_0^4 & z_0^4 & 1 & -z_A^4 x_0^4 & -z_A^4 y_0^4 & -z_A^4 z_0^4 & -z_A^4 \end{pmatrix} \begin{pmatrix} a_{11} \\ a_{12} \\ a_{13} \\ \mathbf{M} \\ a_{42} \\ a_{43} \\ a_{44} \end{pmatrix} = 0$$

Le processus déterminant la position 3D est réalisé par triangulation.

3.5. Estimation de la matrice de projection

La matrice de projection est calculée en utilisant la propriété de projection affine. Lorsque la projection de l'origine et des points de la base sont connues, la projection affine $(u_p, v_p, 1)$ de n'importe quel point représentée par les coordonnées affines $(x, y, z, 1)$ est exprimée par l'équation (4.17) :

$$\begin{pmatrix} u_p \\ v_p \\ 1 \end{pmatrix} = \begin{pmatrix} u_{p1} - u_{p0} & u_{p2} - u_{p0} & u_{p3} - u_{p0} & u_{p0} \\ v_{p1} - v_{p0} & v_{p2} - v_{p0} & v_{p3} - v_{p0} & v_{p0} \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} \quad (4.17)$$

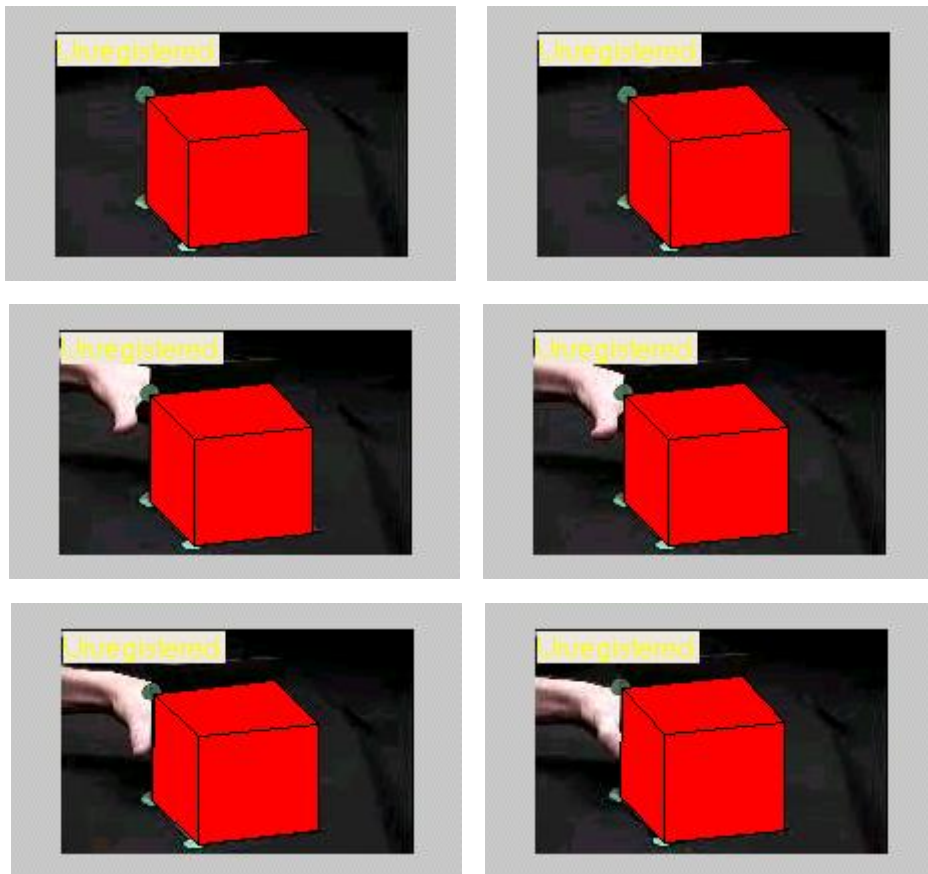
Où :

$(u_{pi}, v_{pi}, 1)^T$ est la projection de la base (quatre points non coplanaires).

L'équation (4.17) définit la projection d'un point 3D dans n'importe quelle nouvelle image, comme une combinaison linéaire des projections des points de la base affine dans cette image. Le calcul de la projection d'un point de l'objet virtuel a été réduit à la connaissance de la position des projections des points de la base dans la nouvelle image et les coordonnées affine homogènes du point virtuel. Les points de la base seront définis par un suivi visuel des primitives dans la scène et déterminant les projections dans chaque nouvelle image.

3.6. Insertion de l'objet virtuel

Le suivi de la base affine permet de mettre à jours dynamiquement la matrice de projection permettant ainsi de réaliser automatiquement la superposition de l'objet virtuel sur l'image de la scène réelle. Dans la figure Fig.4.16 figure quelques résultats de notre plate- forme basée sur "l'augmentation sans calibration".



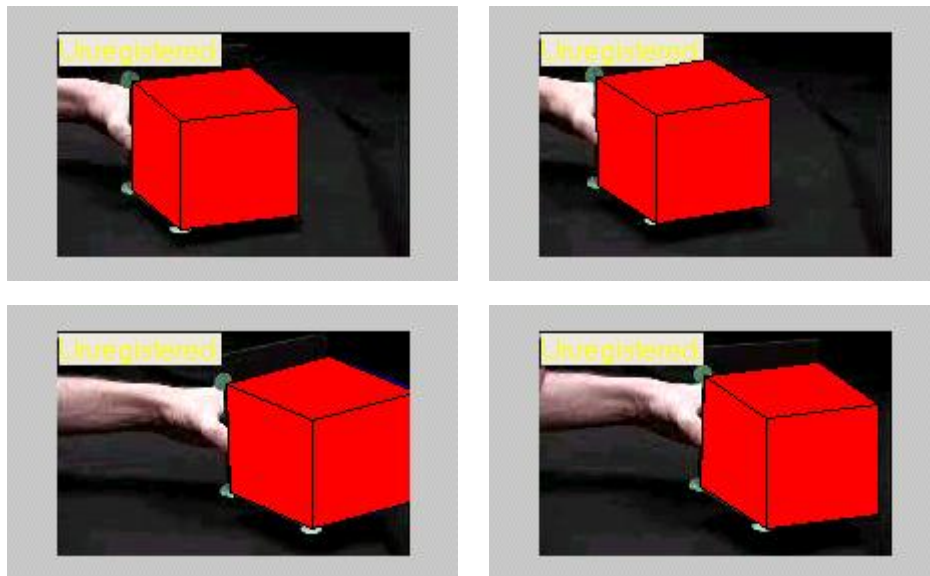


Fig.4.16. les résultats de notre plate-forme.

4. Discussion des résultats

La table table.4.3. présente les points principaux des deux approches précédentes:

Augmentation basée calibration	Augmentation sans calibration
L'approche utilise un modèle 3D (mire) placé dans la scène.	L'approche utilise quatre repères.
La nécessité de connaître les coordonnées tridimensionnelles d'un ensemble de points physiques.	Une connaissance minimale sur la scène, cette approche utilise quatre repères dont les coordonnées tridimensionnelles sont inconnu.
La connaissance des paramètres de la caméra.	Aucune connaissance des paramètres de la caméra.
Le modèle de caméra utilisé est un modèle sténopé utilisant une projection perspective. Ce modèle est fidèle à la réalité.	Le modèle de caméra est un modèle affine, c'est un modèle simplifié utilisant une projection perspective faible.
La calibration est un processus or-ligne, c'est à dire que la détermination des paramètres intrinsèques se fait une seule fois pendant la phase d'initialisation. Par conséquent, les paramètres intrinsèques sont supposés fixes.	La matrice de projection est dynamiquement mise à jours, permettant ainsi de prendre en compte la variation de la distance focale.
L'objet virtuel est représenté dans un repère euclidien.	La détermination de la projection des quatre sommets permet de déplacer l'objet virtuel de son repère euclidien vers le repère affine.

L'objet virtuel est modélisé par la méthode B_Rep, ou il est représenté par six faces et huit sommets. Par conséquent, la projection du cube est réduite à la projection de huit points au lieu de projeter un ensemble de points.	
La position de l'objet virtuel est déterminée par un seul point image.	La position de l'objet virtuel est déterminée par quatre points.
<i>La robustesse</i> : il est possible d'estimer la matrice de projection en utilisant six appariements 3D-2D. La présence du modèle tridimensionnel offre 96 appariements 3D-2D.	La détermination de la matrice de projection nécessite la présence des quatre repères, ce qui explique l'échec du suivi si on n'arrive pas à localiser un repère(voir Fig.4.16.).
La précision du recalage dépend de la précision de la détection des coins.	Les modèles affines de vision sont utiles lorsqu'une grande précision n'est pas nécessaire.

Table.4.3. Points principaux des deux approches.

En résumé, l'augmentation basée sur la calibration est une bonne tentative pour construire un système de la réalité augmentée efficace et portable. Cette approche a échappé au processus de la calibration. Par conséquent, elle est souhaitable dans les situations où la distance focale n'est pas fixe.



Fig.4.17. L'absence d'un repère cause échec.

5. Conclusion

Ce chapitre a été consacré pour présenter les résultats de nos deux plates-formes tout en détaillant les aspects nécessaires à l'implémentation. Notre première approche sur laquelle est basée notre première plate-forme s'inspire des travaux de Zonglei et de Boufama, elle nécessite la calibration des images utilisées et la reconstruction tridimensionnelle du plan de recalage. Notre deuxième plate-forme est basée sur l'approche "Augmentation sans calibration", qui essaye d'échapper à la calibration en suivant un nombre minimal de primitives (quatre repères) dont les coordonnées tridimensionnelles ne sont pas connues.

Conclusion générale

Le domaine de la réalité augmentée (RA) est un nouveau domaine mais actif. La réalité augmentée est considérée comme une variante de la réalité virtuelle (RV). Dans la Réalité virtuelle l'utilisateur est complètement immergé dans une scène virtuelle, tandis que les systèmes de la réalité augmentée complètent la perception de l'être humain en superposant des objets virtuels sur les images de la scène réelle. Un des problèmes de base actuellement limitant les applications de la réalité augmentée est le problème du recalage. Les objets des deux mondes réel et virtuel doivent être correctement alignés donnant l'illusion qu'ils coexistent.

Le problème du recalage est loin d'être résolu. Plusieurs systèmes supposent que le point de vu est statique (une caméra statique) ou que les objets sont statiques (une scène statique) ou même les deux. Même si le mouvement des objets ou du point de vu est autorisé, il n'est pas libre. Un alignement précis, des objets des deux mondes réel et virtuel, nécessite un suivi précis du mouvement de la caméra ou de la tête de l'utilisateur (selon le type du système utilisé) et une localisation précise des objets réels (tout en prenant en compte le mouvement de ces objets). Le grand obstacle rencontré pour construire des systèmes efficace de la réalité augmentée est le besoin de moyens précis permettant de localiser l'utilisateur (ou la caméra) et les objets physiques c'est à dire la détermination du mouvement de la caméra et la détermination du mouvement des objets.

Dans ce mémoire on a présenté nos plates-formes, la première est basée sur notre approche "augmentation basée calibration", qui est basée sur la reconstruction tridimensionnelle du plan de recalage en utilisant des images stéréoscopique calibrées. Notre deuxième approche "augmentation sans calibration" est une approche essayant d'échapper à la calibration. Elle est basée sur le suivi de seulement quatre primitives dont leurs coordonnées tridimensionnelles ne sont pas connues. Augmentation basée calibration est conseillée pour les applications de la réalité augmentée demandant la robustesse et une grande précision où le temps réel n'est pas demandé. Augmentation sans calibration est conseillée pour les application de la réalité augmentée en temps réel où une grande précision n'est pas demandée.

Les systèmes de la réalité augmentée basés sur les deux approches précédentes requièrent un degré de connaissance plus ou moins important sur la scène. Nous souhaitons axer nos recherches futurs sur l'augmentation des scènes *d'extérieurs* en suivant *des primitives naturelles* pour insérer des objets virtuels tridimensionnels *complexes*.

Références

- [Aishy02] A.Aishy, A.Mitiche, E.Dubois. *Real-time Motion Estimation by Object-Matching for High Level Video Representation*. In Proc. IAPR/CIPPRS 15th Int. Conf. on Vision Interface (VI), pp. 31-38, Calgary, Canada, May 2002.
- [Alenyà03] G. Alenyà, E. Martinez, C. Torras. *Fusing Visual and Inertial Sensing to Recover Robot Ego-motion*.
www-iri.upc.es/people/torras/articles/jrs2003.pdf.
- [Alghanemi02] S.Alghanemi, A.Elhussuna, A.Abdul, N.Johansen. *Stereovision*. Four weeks Project in Introduction to multimedia technologies, 11/05/2002.
- [Armstrong96] MN. Armstrong. *Self-Calibration from Images Sequences*. Thesis submitted for the degree of Doctor of Philosophy, 1996.
- [Azuma01] R.Azuma, Y.Baillet, R.Behringer, S.Feiner, S.Julier, B.Macintyre. *Recent Advances in Augmented Reality*. Computer & graphics, November 2001.
- [Azuma97] R.Azuma. *A Survey of Augmented Reality*. Presence: Teleoperators and Virtual Environments, 6(4): 355-385, 1997.
- [Basri03] R.Basri. *Introduction to Computer and Human Vision*. Courses 2003-2004.
- [Bebis03.a] Bebis. *The geometry of Perspective Projection* (13/03/2003).
[www.cs.unr.edu/~bebis/cs791E/Notes/Perspective Projection. CS491E/791E](http://www.cs.unr.edu/~bebis/cs791E/Notes/Perspective%20Projection.CS491E/791E).
- [Bebis03.b] Bebis. *Connected Components* (27/02/03).
www.cse.unr.edu/~bebis/CS791E/Notes/Connectedcomponents.pdf.
- [Bensor03] K.Bensor. *How will Augmented Reality work? How stuff works* (referenced 18/09/2003). [Http://www.howstuffworks.com /Augmented-reality.html](http://www.howstuffworks.com/Augmented-reality.html).
- [Black00] MJ. Black, DJ. Fleet. *Probabilistic Detection and Tracking of Motion Boundaries*. International Journal of Computer Vision 38(3), 231-245, 2000.
- [Boufama99] B. Boufama. *On the recovery of motion and structure when cameras are not calibrated*. In International Journal of Pattern Recognition and Artificial Intelligence,(13) 5 (1999), pages 735--759.
- [Breen96] D. Breen, R. Whitaker, E.Rose, M. Tuceryan. *Interactive Occlusion and Automatic Object Placement for Augmented Reality*. Eurographics '96 Proceedings (Poitiers, France, August 1996) pp. 11-22
- [Corbett02] DR. Corbett. *Multiple Object Tracking in Real-time*. October 20, 2002. Submitted for the degree of Bachelor of engineering (Pass) in the division of electronical and electronics.

- [Cornelis01] K. Cornelis, M. Pollefeys, M. Vergauwen, L.V. Gool. *Augmented Reality using uncalibrated video sequences*. Lecture Notes in Computer Science, 2018,2001.
- [Crowley01] J.L. Crowley. *Suivi des Formes par Filtre de Kalman*. Seance 8 (DEA IVR), 4/12/2001.
- [Davidy00] Davidy. *Introduction to Dynamic Vision; Image Motion*. Computer Vision. Lecture 10.
www.cogs.susx.ac.uk/users/davidy/compvis/slides_lec10.pdf.
- [Delhumeau03] J. Delhumeau, L. Morin, S. Pateux. *Segmentation basée Mouvement 3D pour la Détection d'Objets Indépendants*. Proc. CORESA, Lyon, France, January, 2003.
- [Dengsheng01] Z. Dengsheng, G. Lu. *Segmentation of Moving Objects in Image Sequence: A Review*. Circuits, Systems and Signal Processing (Special Issue on Multimedia Communication Services), 20(2):143-183, 2001.
- [Deutsch97] C. Deutsch. *Suivi et Localisation d'une Cible Complexe Connue par Vision Monoculaire*. Rapport présenté aux membres du CAS pour l'examen pré-doctoral, Juin 1997.
- [Evan01] J.Evan. *A Tool for Augmented Reality*. Interiam Project Report, May 2001.
- [Fauvel95] M.Fauvel. *Contribution à la Stéréovision: Une approche Connexionniste*. Thèse soutenue le mardi 31/01/1995.
- [Feghali03] R.Feghali, A.Mitiche. *Tracking with Simultaneous Camera Motion Substraction by Level Set Spatio-Temporal Surface Evolution*. In proceeding ICIP 2003, Barcelon, spain, sept.14-17, 2003.
- [Feiner97] S. Feiner, BM. Intyre, T. Hollerer, T. Webster. *A Turing Machine: Prototype 3D Mobile Augmented Reality Systems for exploring the Urban Environment*. Proc. 1 st Int'L Symp. Wearable Computers (ISWC'97). Cambridge, MA, 13-14 Oct. 1997, pp.74-81.
- [Fergani04.a] B. Fergani, MC. Batouche, B. Boufama. *Robust Registration of Virtual Objects for Industrial Applications*. IEEE International Conference on Industrial Technology (IEEE ICIT'04).ISBN: 0-7803-8663-9.Yasmin Hammamat, December 8-10, 2004, Tunisia.
- [Fergani04.b] B.Fergani, M.Fezani, MC. Batouche, A. Benhocine. *Calibration, Stéréovision et Réalité augmentée*.CISC'04, The International Conference on Complex Systems. September 6-8, 2004. Jijel, Algeria.
- [Fisher04] J. Fisher, M. Neff, D. Freudenstein, D. Bartz. *Medical Augmented Reality based on Commercial Image Guided Surgery*. Eurographics Symposium on Virtual Environments 2004.
- [Frédérick01] M.Frédérick, H.Radu. *Multiple Camera Tracking of Rigid Objects*. Rapport de recherche N°4268, September 2001.

- [Gonçalves02] N. Gonçalves, H. Araújo. Estimation of 3D Motion from Stereo Images-Differential and Discrete Formulations. ICPR'2002--16th International Conference on Pattern Recognition, Quebec City, Canada, August 11-15, 2002.
- [Gros97] P. Gros, R. Mohr, L. Quan, R. Hartley. *How Useful is Projective Geometry?* www.inrialpes.fr/movi/publi/Publications/19997/GHMQ97.
- [Habed00] A. Habed, B. Boufama. *Novel View Synthesis : A Comparative Analysis study.* <http://www.cim.mcgill.ca/~dudek/vi2000/vifinal.html>.
- [Heikkila97] J.Heikkila, O.Silvèn . *A Four Step Camera Calibration Procedure with Implicit Image Correction.* www.vision.caltech.edu/bouguetj/calib_doc/papers/heikkila97.pdf.
- [Henele02] A. Henele , S.Singh, D. Strelow. *An Empirical Comparaison of Methods for Images -based Motion Estimation.* IEEE/RSJ International Conference on Intelligent Robots and Systems, October 2002.
- [Hirshmüller02] H. Hirshmüller, PR. Innocent, JM. Garibaldi. *Fast, Unconstrained Camera Motion Estimation from Stereo without Tracking and Robust Statics.* Accepted for ICARCV 2002, 2-5 December 2002 in Singapore.
- [Hwan02] JO. Hwan , P. Sankuratir. *Automatic Distinction of Camera and Object Motions in Video Sequences.* Accepted for ICME 2002, Lausanne, Switzerland.
- [Irani94] M. Irani, B. Rousso, S. Peleg. *Recovery of Ego-motion Using Image Stabilization.* www.vision.huji.ac.il/papers/ego-mtn-cvpr94.pdf.
- [Jennings96] C.Jennings. *Structure and Motion from Stereo Image Sequences* (14 May 1996). www.cs.ubc.ca/spider/jennings/paper525.pdf
- [Jian00] B.Jiang, S.You, U.Neumann. *Camera Tracking for Augmented Reality Media.* IEEE International Conference on Multimedia 2000, August 2000.
- [Koller97] D.Koller, G. Klinker, E.Rose, D.Breen, R.Whitaker, M.Tuceryan. *Real-time Vision-based Camera Tracking for Augmented Reality Applications.* Appeared in the Proceeding of the symposium on Virtual Reality Software and Technology (VRS T-97), pp.87-94, Lausanne, Switzerland, September 15-17,1997.
- [Konrad00] J. Konrad. *Motion Detection and Estimation In Image and Video Processing.* Handbook, ch.3.8, Academic Press, 2000.
- [Kostkovà02] J.Kostkovà. *Stereoscopic Matching : Problems and Solutions.* PHD Thesis, September 4, 2002.
- [Kriegman03] Kriegman. *Background : Camera models, Transforms, Radiometry.* Topics in Images – based Modeling and Rendering. CSE 291J00, 2003.

- [Krüger98] S.Krüger. *Motion Analysis and Estimation Using Multiresolution Affine Models*. Thesis submitted for the degree of Doctor of philosophy, July 1998.
- [Ku01] PJ. Ku, LY. Chen, J. Zou. *Features Tracking and Shape & Structure from Motion*. Projet #4, ECSE 6650-Computer Vision. December 3, 2001.
- [Kutulakos96] Kutulakos, J.Vallino. *Affine Object Representations for Calibration-Free Augmented Reality*. To appear in: Proc. 1996 IEEE Virtual Reality Ann. Int. Symp. (VRAIS'96).
- [Lang02.a] P. Lang, A. Kusej, A. Pinz, G. Brasseur. *Inertial Tracking for Augmented Reality*. IEEE instrumentation and measurement technology conference. Anchorage, AK, USA, 21-23 May 2002.
- [Lang02.b] P. Lang, M. Ribo, A. Pinz. *A New Combination of Vision- based and Inertial Tracking for Fully Mobile, Wearable and Real-time Operation*. 26th workshop of the AAPR/ÖAGM, OCG schriftenreihe, Band 160, pp.141-148, Graz, Austria, september 2002.
- [Loop99] C.Loop, Z.Zheng. *Computer Rectifying Homographies for Stereo-Vision*. Technical Report MSR-TR-99-21, April 8, 1999.
- [Mensouri02] AR. Mensouri. *Region Tracking Via Level Set PDEs without Motion Computation*. IEEE transactions on Pattern Analysis Machine Intelligence, vol.24, No7, July 2002.
- [Milgram94] P. Milgram, H.Takemura, A.Utsumi, F.Kishimo. *Augmented Reality: A class of displays on the reality-virtuality continuum*. 282/SPIE Vol.2351, Telemanipulator and Telepresence Technologies (1994).
- [Milles97] S. Milles. *Stereo-Motion Analysis of Image Sequences*. Appeared in the proceedings of DICIA/IVCNZ'97.
- [Papademetris96] X. Papademetris, PN. Belhumeur. *Estimation of Motion Boundary Location and Optical Flow using Dynamic Programming*. Technical Report 9607, June 1996.
- [Personnaz02] M. Personnaz, P.Strum. *Calibration of a Stereo-Vision System by the Non-Linear Optimization of the Motion of a Calibration Object*. Rapport technique N° 0269 (INRIA), september 2002.
- [Radthe01] T.Radthe, V. Zerbe. *Tracking of Dynamic Objects based on Optical Flow*. www.sst.theinf.tuilmnau.de/research/publications/documents/2001/ICIMADE2001Paper.pdf
- [Regenbrecht04] HT. Regenbrecht, MT. Wagner. *Interaction in a Collaborative Augmented Reality Environment*. ACM 1-58113-454-1/02/2004.
- [Robyn97] Robyn .*Motion*. Computer Vision IT412 (29/10/1997). www.cse.iitk.ac.in/~amit/courses/768/vision/robyn/motion/
- [Romanos03] P. Romanos, S. Suguru, N.Masayuki. *Automatic Calibration of Magnetic Motion Trackers Using Bayesian-Neural Cascade*. Vol.2 N°3 pp.105-113 (accepted)

15/06/2003). <http://www.art-science.org/journal/v2n3>.

- [Ruf00] A.Ruf, H.Radu. *Rigid and Articulated Motion*. September 25, 2000.
- [Sairio01] M.Sairio. *Augmented Reality*.
www.tml.hut.fi/studies/Tik-111.590/2001s/papers/Mikko_sairio.pdf.
- [Schnaider04] M.Schnaider, B.Schwald. *Augmented Reality in Medecine: A review to the patient's inside*. CG topics 1/2004.
- [Sebe02] IO. Sebe. *Object Tracking Using Multiple Constraints*. EE392J, Digital Video Processing, Winter 2002.
- [Shahzad02] M.Shahzad. *Robust Registration of Virtual Objects for Real-time Augmented Reality*. Thesis submitted for the degree of master of Computer Science. May 8,2002.
- [Simon99] G. Simon. *Vers Un Système de Réalité Augmentée Autonome*. Thèse de Doctorat soutenue le 14/12/1999.
- [State96] A. State, M. Livingston, W. Garrett, G. Hirota, M. Whitton, E. Piani, H. Fuchs. *Superior Augmented Reality Registration by Integrating Landmark Tracking and Magnetic Tracking*. In Computer Graphics Proceeding, Annual conference series: SIGGRAPH ' 96 (New Orland, LA), pp. 429-438, ACM SIGGRAPH, New York, August 1996.
- [Strelow02] D. Strelow, S. Singh. Optimal motion estimation from visual and inertial measurements. *IEEE Workshop on Applications of Computer Vision*, December 2002, pages 314-319.
- [Suthau02] T.Suthau, M.Vetter, P.Hassenpflug, H.Meinzer, O.Hellwich. *A concept work for Augmented Reality Visualisation based on a medical application in liver surgery*. IAPRS, Corfu, 2002
- [Svoboda97] T. Svoboda, P.Sturm. *Badly Calibration Camera in Ego-motion Estimation, Propagation of Uncertainty*. 7th International Conference Computer Analysis of Images and Patterns, pages 183-190. Springer Verlag in the LNCS series, september 1997.
- [Tian96] TY. Tian, C.Tamasi, DJ. Heeger. *Comparaison of approaches to Ego-motion Computation*. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 315-320, June 1996.
- [Torresani03] L. Torresani, A. Hertzmann, C. Bregler. *Learning Non- Rigid 3D Shape from 2D Motion*. October 30, 2003.
- [Tsai03] R.Tsai, S. Osher. *Level Set Methods in Image Science*. January 2003.
http://users.ntua.gr/karank/topo/PhD_notes/Segmentation/downl/cam03-04.pdf.
- [Vallino95] JR. Vallino. *Interactive Augmented Reality*. Thesis submitted for the degree Doctor of philosophy, 1995.

- [Vaxoncelos01] N. Vaxoncelos, A. Lippman. *Empirical Bayesian Motion Segmentation*. February 2001(vol.23, N.02).
www.svcl.ucsd.edu/publications/journal/2001/pami/pami01.pdf.
- [Verplaetse97] C J.Verplaetse. *Inertial- Optical Motion- Estimating Camera for Electronic Cinem Atography*. Submitted for the program in Media Arts and Sciences, school of architecture and planning, of the requirements for the degree of Master of science in Media Arts and Sciences, on March 15, 1997.
- [Vu03] D.Vu. *Calibration-Free Augmented Reality* (Nov 20, 2003) CSE252C Fall 03 Project Report.
- [Wegger02] J. Wegger. Estimation of Motion In Image Sequences. Master 's thesis, 2002.
- [Yilmaz04] A.Yilmaz, X.Li, M.Shah. *Object Tracking Using Level Sets*.
www.cs.ucf.edu/~vision/papers/yilmaz_accv_2004.pdf
- [You99] S.You, U.Neumann, R.Azuma. *Orientation Tracking for Outdoor Augmented Reality Registration*. IEEE Computer Graphics and applications, december 1999.
- [Yue04] Z.Yue, KZ. Shaohua, Ramachelleppa. *Robust Two-Camera Tracking Using Homography*. www.cfar.umd.edu/~shaohua/papers/yue04icassp.pdf.
- [Zaccarin03] A.Zaccarin. *Estimation du Mouvement*. GIF-19263, vision numérique, Automne 2003.
- [Zehang01] S. Zehang, C.Rayburn. *Camera Calibration* (June 4th , 2001).
www.cs.unr.edu/~rayburn/Reports/report1.pdf.
- [Zhong03] X. Zhong, P.Liu, ND. Georgamas, P. Boulanger. *Designing a Vision-based Collaborative Augmented Reality Application for Industrial Training*. Information technology. 45 (2003) 1.
- [Zonglei02.a] H. Zonglei, B.Boufama. *A Semi Automatic Camera Calibration Method for Augmented Reality*. Systems, Man and cybernetics, 2002 IEEE International conference on.
<http://intl.ieeexplore.ieee.org/xpl/tocresult.jsp?isNumber=262988page=8>.
- [Zonglei02.b] H. Zonglei. *Computer Vision-based Registration for Augmented Reality*. Thesis submitted for the degree of master of science, 2002.