

# *Projet MOVI*

*Modélisation pour la vision par ordinateur*

*Rhône-Alpes*

THÈME 3B

*R* *apport*  
*d'Activité*

2000



## Table des matières

<b>1</b>	<b>Composition de l'équipe</b>	<b>2</b>
<b>2</b>	<b>Présentation et objectifs généraux</b>	<b>3</b>
<b>3</b>	<b>Domaines d'applications</b>	<b>4</b>
3.1	Panorama . . . . .	4
3.2	Vision, robots et leur couplage . . . . .	4
3.3	Indexation de bases d'images . . . . .	5
3.4	Vision et défense nationale . . . . .	6
3.5	Virtualisation d'objets . . . . .	6
3.6	Synthèse d'images à partir d'images . . . . .	7
3.7	Analyse vidéo . . . . .	7
<b>4</b>	<b>Résultats nouveaux</b>	<b>7</b>
4.1	Géométrie multi-images et multi-caméras . . . . .	7
4.2	Couplage vision/robotique . . . . .	11
4.3	Indexation d'images et reconnaissance d'objets . . . . .	14
4.4	Structuration de vidéos . . . . .	18
<b>5</b>	<b>Contrats industriels (nationaux, européens et internationaux)</b>	<b>19</b>
5.1	Reconnaissance et indexation de vues aériennes . . . . .	19
5.2	Indexation de vidéo . . . . .	19
5.3	Projet RNRT "AGIR" . . . . .	19
5.4	Projet PRIAMM "Visteo" . . . . .	20
<b>6</b>	<b>Actions régionales, nationales et internationales</b>	<b>20</b>
6.1	Actions régionales . . . . .	20
6.2	Actions nationales . . . . .	20
6.3	Actions financées par la Commission Européenne . . . . .	21
6.4	Relations bilatérales internationales . . . . .	22
6.4.1	Amérique . . . . .	22
6.4.2	Asie . . . . .	22
<b>7</b>	<b>Diffusion de résultats</b>	<b>23</b>
7.1	Animation de la communauté scientifique . . . . .	23
7.2	Enseignement universitaire . . . . .	23
7.3	Participation à des colloques, séminaires, invitations . . . . .	24
<b>8</b>	<b>Bibliographie</b>	<b>24</b>

## 1 Composition de l'équipe

### Responsable scientifique

Radu Horaud [directeur de recherche]

### Assistante de projet

Véronique Roux [à partir du 1er janvier 2000]

### Personnel INRIA

Cordelia Schmid [chargée de recherche]

Peter Sturm [chargé de recherche]

### Personnel CNRS

Long Quan [chargé de recherche]

William Triggs [chargé de recherche depuis le 1er octobre 2000]

### Personnel universitaire

Edmond Boyer [maître de conférences à l'université Joseph Fourier]

### Collaborateur extérieur

Roger Mohr [Xerox European Research Center]

### Ingénieurs experts

Bart Lamiroy [jusqu'au 30 septembre 2000]

Mathieu Personnaz [à partir du 1er avril 2000]

**Chercheurs doctorants**

Marc-André Ameller [allocataire ENS Cachan]  
Adrien Bartoli [allocataire MENESR, depuis le 1er octobre 2000]  
David Demirdjian [allocataire MENESR, jusqu'au 30 août 2000]  
Yves Dufournaud [boursier CIFRE avec la Société Aérospatiale]  
Fabien Evrard [allocataire ENS Cachan, depuis le 1er octobre 2000]  
Ryad Hammoud [boursier INRIA, sur contrat avec la Société Alcatel]  
Maxime Lhuillier [allocataire MENESR]  
Frédéric Martin [allocataire MENESR]  
Krystian Mikolajczyk [allocataire CIES]  
Andreas Ruf [boursier TMR (bourse Marie-Curie)]  
Cristian Sminchisescu [allocataire CIES]  
Marta Wilckowiak [boursière INRIA, depuis le 1er octobre 2000]

**Chercheurs post-doctorants**

Ragini Choudhury [contrat PRIAMM, depuis le 1er avril 2000]

**Professeur invité**

Andrew Zisserman [Oxford University, du 1er avril au 30 juin 2000]

## 2 Présentation et objectifs généraux

Le projet MOVI est un projet commun entre le CNRS, l'INPG, l'UJF et l'INRIA, localisé à l'INRIA Rhône-Alpes et appartenant au laboratoire GRAVIR de la fédération IMAG.

Comprendre l'espace tridimensionnel perçu par une ou plusieurs caméras, identifier les objets qu'il contient, se localiser et agir forment un premier ensemble d'activités qui peut se regrouper schématiquement sous le vocable de "géométrie de la vision 3D". Un second groupe d'activités, plus récent s'intéresse à la recherche d'objets ou d'images dans une base de référence très large par des "techniques d'indexation". Un troisième groupe d'activités s'appuie sur le savoir faire méthodologique développé autour de la géométrie et de l'indexation et a comme objectif le "continuum réel-virtuel". Plus précisément, sur ces trois thèmes, nous développons les aspects suivants :

- modélisation d'une scène 3-D à partir d'une seule image, l'étalonnage étant implicite ;

- couplage de la vision avec le contrôle : asservissement visuel, manipulation guidée par la vision, étalonnage du couple caméra-robot, caméras actives et à paramètres internes variables ;
- mise en correspondance de deux ou plusieurs images ;
- construction interactive de modèles d'objets et de scène complexes à partir de plusieurs vues ;
- indexation de large base d'images ;
- traitement de séquences vidéo : segmentation, suivi et analyse d'objets en mouvement ;
- rendu réaliste de scènes complexes à partir de quelques images.

Outre les approfondissements des aspects mentionnés ci-dessus, notre projet vise à développer des démonstrateurs intégrant différents aspects de ce savoir-faire. Plus que la juxtaposition de différentes techniques, cette intégration amène à reposer les problèmes fondamentaux dans des cadres nouveaux, comme par exemple celui de la commande par la vision non-métrique de mouvements d'un mécanisme articulé ou la création de bases d'images interrogeables par leur contenu.

## 3 Domaines d'applications

### 3.1 Panorama

Les domaines d'application habituels de la vision tridimensionnelle ont été liés à la robotique et à la défense. Notre projet continue à être bien présent dans ces domaines. Il est cependant de nouveaux domaines en émergence et liés à l'utilisation de l'information visuelle numérique dans des domaines très variés allant de la visite virtuelle d'un musée jusqu'à la production assisté par ordinateur de vidéos.

Dans ces créneaux nous nous positionnons dans la problématique de la synthèse d'images à partir d'images, de la virtualisation d'objets d'art et de l'analyse dynamique de séquences d'images.

### 3.2 Vision, robots et leur couplage

**Résumé :** *Nous abordons plusieurs aspects du couplage de la vision et de la robotique : étalonnage et auto-étalonnage d'ensembles de caméras et de robots, qu'ils soient liés rigidement ou non, le guidage visuel de robots manipulateurs, la modélisation d'un robot dans un espace visuel calibré ou non calibré.*

Les domaines d'applications de l'intégration vision-robotique sont les suivants :

- l'auto-étalonnage d'une tête stéréoscopique à deux degrés de liberté (deux rotations) ; ce problème a été abordé dans le cadre du contrat européen Reactive LTR VIGOR 26247 (1998-2001) ;

- le guidage visuel d’un robot à plusieurs degrés de liberté à l’aide d’une ou plusieurs caméras non étalonnées ; ce problème est abordé dans le cadre du projet VIGOR en collaboration avec Odense Steel Shipyard (chantiers navals). L’application visée est le positionnement d’une torche de soudure par rapport à une pièce de bateau avec une précision inférieure à 1mm ;
- l’auto-étalonnage de la chaîne cinématique d’un mécanisme articulé “in-situ” ;
- la commande projective d’un mécanisme articulé.

### 3.3 Indexation de bases d’images

**Résumé :** *Les secteurs de la presse et de l’audiovisuel, ceux de l’industrie (imagerie scientifique), de la médecine ou encore ceux de la propriété industrielle collectent des quantités impressionnantes d’images qu’il faut pouvoir gérer. Souvent le processus d’acquisition est plus rapide et simple que celui de l’indexation, ce qui fait naître un besoin urgent d’indexation automatique par le contenu. Une problématique similaire émerge à partir du web pour les moteurs de recherche et les portails de données.*

Les applications des bases d’images peuvent être divisées comme suit :

- agence de presse et de l’audiovisuel : ce segment se caractérise par un énorme volume de données d’images, de plusieurs millions pour les images fixes à des centaines de milliers de vidéos pour les images animées ; les besoins en consultation sont complexes, les requêtes sont souvent de haut niveau et elles requièrent une forte interactivité avec l’utilisateur ;
- les secteurs médicaux, scientifiques et industriels ont des besoins plus spécifiques, très liés aux différents domaines concernés ; les volumes dépassent parfois les centaines de milliers d’images et les types d’interrogation sont très variables ;
- les demandes de la propriété industrielle correspondent à une interrogation bien plus précise : « quelles sont les archives similaires à tout ou partie d’un motif présenté? » Pour cette classe d’application, la réponse exacte, si elle existe, ne doit pas être manquée. Le volume d’images traité peut atteindre celui des agences de presse.
- Les besoins pour les moteurs de recherche et les portails de données sont d’identifier les données (images et vidéos) correspondant à la recherche d’un utilisateur ou au domaine concerné par le portail.

A la différence de l’indexation automatique des textes, les images n’apportent pas directement d’information conceptuelle de haut niveau sémantique. Il faut donc développer, pour les différentes classes d’application, des index qui soient pertinents pour permettre une recherche efficace et une interaction proche des concepts de l’utilisateur.

### 3.4 Vision et défense nationale

Notre projet a développé depuis plusieurs années un partenariat de recherche avec la Division Missiles de la Société Aérospatiale. Après avoir collaboré dans le domaine de la reconstruction tri-dimensionnelle nous abordons depuis 1998 le problème de la comparaison de deux images prises dans des conditions de vue très différentes. L'application envisagée est la mise en correspondance de données visuelles obtenues avec un "drone"<sup>1</sup> de reconnaissance d'une part et un missile opérationnel d'autre part (voir section "contrats industriels").

Ce travail se réalise sous la forme d'une convention CIFRE et d'un contrat de collaboration entre l'Aérospatiale et l'INRIA.

### 3.5 Virtualisation d'objets

#### Résumé :

*Par virtualisation nous entendons la création de modèles tridimensionnels photoréalistes. L'intérêt majeur de l'utilisation de techniques de vision par ordinateur est le gain, potentiellement énorme, en temps d'acquisition, en coût de matériel, en qualité des résultats, par rapport à d'autres approches (qui sont souvent manuelles, nécessitent un équipement spécialisé ou un utilisateur expert ou dont le champ d'action est limité).*

Les domaines d'applications sont :

- tourisme : la visite virtuelle d'un site peut déjà en soi être une expérience enrichissante. Par exemple, un modèle complet permet au touriste virtuel d'adopter des points de vue qui lui ne seraient pas possibles physiquement. Les visites virtuelles pourront avoir ont grand impact au niveau de la publicité pour des destinations de vacances. Des musées qui sont partiellement fermés, par exemple pour effectuer des travaux de rénovation, pourront néanmoins offrir des visites complètes, à travers de kiosques de visite virtuelle.
- archivage du patrimoine architectural : la virtualisation permet un archivage compact de l'aspect visuel et des mesures d'ensembles architecturaux. La navigation virtuelle permettra un accès rapide à des structures d'intérêt.
- marché de l'immobilier, commerce : la virtualisation d'un produit (e.g. d'une maison à vendre) permet une présentation plus attractive car dynamique et interactive. La visite à distance donne à la personne cherchant à acquérir une maison la possibilité de faire une pré-sélection et ainsi de réduire le nombre de déplacements à effectuer.
- jeux/télévision/cinéma : il est de plus en plus important pour ces secteurs de créer des environnements virtuels réalistes, par exemple pour la création d'effets spéciaux ou pour des studios virtuels. On peut aussi s'imaginer que tout le monde puisse créer lui-même des scénarios pour son jeu vidéo préféré, à l'aide d'une caméra digitale.

---

1. un engin volant sans pilote



- enseignement : des visites virtuelles peuvent enrichir l’enseignement dans des domaines telle l’histoire ou la géographie. En plus de l’aspect purement informatif, la présentation attractive d’un sujet ne peut que stimuler la curiosité et la créativité des élèves.

### 3.6 Synthèse d’images à partir d’images

#### Résumé :

*L’approche de la synthèse de nouvelles images à partir d’images (Image-Based Rendering) utilise une collection d’images existantes comme la représentation de la scène 3D. Comparée à la technique classique de l’infographie, elle crée des images photo-réalistes. La modélisation hors ligne et le rendu en ligne sont indépendants de la complexité géométrique et photométrique de la scène.*

Cette approche de synthèse d’images a de nombreuses applications potentielles dans les domaines de multimédia, par exemple,

1. simulation de la caméra virtuelle en créant la nouvelle image dans la position soit pré-spécifiée soit fixée interactivement ; le projet européen IST EVENTS que nous venons de démarrer a pour objectif d’appliquer ce principe pour pouvoir transmettre les grandes scènes 3D par la télévision en temps réel.
2. les effets spéciaux comme ralentissement du temps en interpolant de nouvelles images dans une séquence existante ou encore comme un effet de temps mort tout en tournant autour d’un objet 3D en interpolant les images prises dans les positions différentes.

### 3.7 Analyse vidéo

La vidéo et le multimédia ont de plus en plus d’ubiquité dans notre « société d’information », et leur traitement digitale a pris une importance économique et technologique de première ordre. MOVI a travaillé dans ce domaine depuis plusieurs années, notamment sur les aspects structuration et indexation pour les bases de données vidéos. Cette année nous redoublons nos efforts sur cette axe, avec l’ouverture d’une thème de travail sur le suivi et la reconstruction du mouvement humain à partir de vidéo monoculaire, et le début en fin d’année d’un projet européen majeur VIBES sur l’indexation et la reconstruction « niveau objet » de vidéo...

## 4 Résultats nouveaux

### 4.1 Géométrie multi-images et multi-caméras

**Participants :** Marc-André Ameller, Adrien Bartoli, Edmond Boyer, David Demirdjian, Radu Horaud, Maxime Lhuillier, Long Quan, Andreas Ruf, Peter Sturm, William Triggs, Marta Wilckowiak.

**Mots clés :** géométrie, étalonnage, reconstruction 3D, vision 3D, séquence d’images.

**Résumé :** *Nous avons été présents depuis plusieurs années dans les études sur l'utilisation de la géométrie projective pour la vision tri-dimensionnelle. L'avantage essentiel de cette approche est de permettre de s'affranchir de l'étalonnage des systèmes de vision, offrant ainsi un cadre de calcul rigoureux et exact lorsque les paramètres des systèmes de vision ne sont que partiellement connus ou pas connus du tout.*

*Le projet Esprit Reactive LTR 21914 CUMULI (Computational Understanding of Multiple Images) dont nous avons été contractant principal nous a permis par ailleurs une collaboration étroite avec d'autres équipes européennes universitaires et industrielles. Les travaux entrepris dans CUMULI sont repris dans le cadre du projet IST VISIRE.*

**Synthèse d'images par interpolation** Nous avons déjà proposé la notion de la triangulation conjointe dans deux images pour l'interpolation de nouvelles images. Cette triangulation s'avère être une représentation d'images multiples efficace pour traiter la visibilité. Mais la triangulation conjointe existante est construite sur une division régulière du plan image, elle crée souvent des artifacts indésirables dans les images synthétisées. Pour résoudre ces problèmes, nous avons développé une nouvelle triangulation conjointe qui est contrainte par les points de contour d'images qui intègre aussi les structures rectilinéaires présentes dans beaucoup d'images urbaines [47, 22]. Les exemples d'images re-synthétisées sont donnés dans figure 1.

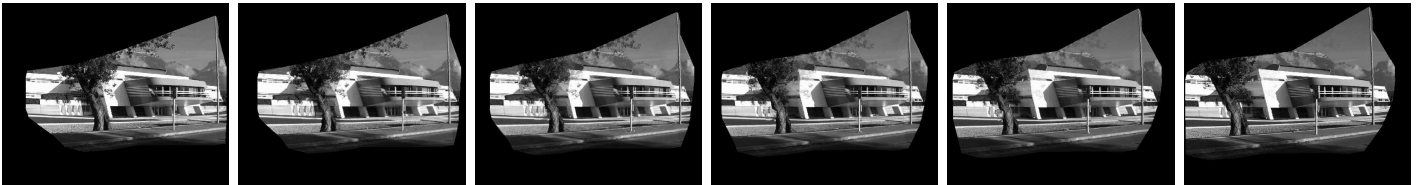


FIG. 1 – *Les images re-synthétisées du bâtiment INRIA.*

### **Appariement dense robuste à l'aide de contraintes géométriques locales et globales**

Dans ce travail [46, 48], nous proposons un nouvel algorithme d'appariement dense. Des points remarquables sont d'abord appariés puis servent de germes pour des croissances de régions qui tendent à densifier les correspondances entre pixels selon une stratégie meilleur d'abord. Enfin la carte des déplacements obtenue est régularisée d'abord à l'aide de contraintes géométriques locales (application affines) puis à l'aide de contraintes globales (matrice fondamentale). L'originalité de l'approche réside dans la stratégie de croissance de région, de régularisation à l'aide de contraintes locales et globales. De plus, la méthode proposée effectue l'étape d'appariement dense avant l'estimation de la matrice fondamentale contrairement aux autres méthodes. L'algorithme est efficace, robuste et peut traiter les grandes disparités. Il est appliqué à l'interpolation d'images et à la reconstruction de surfaces. Ces travaux complètent les travaux plus géométriques [24, 25, 49, 28].

**Approche plan+parallax à la vision tensorielle :** L'étude des relations algébriques « d'appariement » qui lient les différentes images d'une même scène représente un thème majeur dans notre travail depuis plusieurs années. Cette année nous avons serré les liens entre cette approche dite « tensorielle » et les méthodes « plan + parallax » qui encodent les primitives 3D par leurs positions relatives à un plan de référence [53]. Au plan théorique, le travail a mené à une meilleure compréhension de la structure de ces contraintes, et au plan pratique à une nouvelle méthode de reconstruction 3-D par factorisation d'une matrice qui encode les profondeurs relatives au plan et les centres optiques des caméras.

**L'ajustement des faisceaux** est le raffinement d'une reconstruction visuelle approximative par optimisation numérique. Cette étape se montre incontournable pour toute méthode de reconstruction visuelle de haute précision. Elle est bien développée chez les photogrammètres, mais à présent moins pratiquée en vision. Nous avons déjà abordé le sujet il y a quelques années. Cette année nous avons publié une longue synthèse (74 pages) [33], qui essaie à la fois de présenter le sujet à la communauté vision, de résumer sa littérature, et de clarifier ses relations avec la statistique robuste et l'optimisation numérique moderne.

**Le suivi et la reconstruction du mouvement humain** à partir des séquences vidéo monoculaires représentent un nouveau thème abordé pour la première fois cette année. Nous avons pris une approche structurale, avec un modèle 3-D articulaire explicite. Ce modèle paramétrique est recalé aux images par moyen d'une méthode de recherche hybride discrète / continue, qui agit sur une fonction de coût soigneusement robustifiée qui quantifie la qualité d'appariement entre les images et le modèle projeté. Au présent une toute première version de la méthode est opérationnelle, mais il en reste beaucoup plus de validation et d'amélioration à faire.

**Calcul de pose avec une approche matricielle** Le calcul de la pose en vision est un sujet très classique et de nombreux travaux ont déjà été menés dans le projet. Dans ce travail, nous passons d'abord en revue les méthodes existantes et ensuite nous proposons quatre nouvelles méthodes qui ont toutes les points communs d'utiliser les outils de calcul de l'algèbre linéaire pour les résoudre. Les trois premières méthodes donnent une solution unique à partir de quatre points de correspondances par la méthode de SVD. Ces méthodes ont été obtenues en calculant les différentes matrices de résultant de taille  $9 \times 9$ ,  $12 \times 12$  et  $24 \times 24$ . La dernière méthode résout le problème de pose à partir de 3 points de correspondance en cherchant les valeurs propres d'une matrice de taille  $5 \times 5$ .

Au plan pratique et dans le cadre du projet EU CUMULI, nous avons étudié et implanté en MATLAB plusieurs méthodes de pose. Pour le problème important d'orientation relative entre deux images (cas minimal avec 5 points), nous avons développé une bibliothèque plus au point en C [58], qui est à présent en évaluation industrielle.

**Calcul de pose à partir de plusieurs plans vus dans plusieurs images.** Les structures planes ont récemment suscité beaucoup d'intérêt, surtout en vue d'approches de calibrage et de reconstruction 3-D. Il y a plusieurs raisons pour ceci, par exemple : une mire de calibrage

plane peut être créée à l'aide d'une imprimante laser, tandis que l'assemblage précis d'une mire 3-D est assez coûteux ; des structures planes, identifiées dans des images, fournissent des contraintes de calibrage et de calcul de pose simples mais puissantes. En 1999, nous avons développé un algorithme de calibrage qui prend en entrée une ou plusieurs images d'une ou plusieurs structures planes. Nous avons étendu cette approche par des algorithmes de calcul de pose multi-caméras-multi-plans [52]. Tandis qu'un calcul de pose purement séquentiel (pour des paires d'images successives) peut facilement diverger pour une séquence d'images longue, nos algorithmes prennent en compte des redondances, e.g. la visibilité d'une des structures planes dans plusieurs images. Le calcul de pose ne nécessite que la solution de systèmes d'équations linéaires et permet donc d'initialiser de manière efficace un processus d'optimisation non linéaire.

**Utilité de la contrainte de gravité pour le calibrage et le calcul de pose** Nous avons examiné ce que l'observation d'un objet dont la trajectoire est dictée par la gravité (e.g. un objet lancé ou tombant) peut apporter au calibrage et au calcul de pose relative de caméras [51]. Il a été démontré que tout d'abord la reconstruction affine de la scène est possible, même si les caméras ne sont pas calibrées. Ayant obtenu la reconstruction affine, on dispose de plus de contraintes sur le calibrage et la pose que c'est le cas dans des scénarios non contraints. Il existent par ailleurs moins de configurations singulières pour l'exploitation de ces contraintes supplémentaires. On peut alors s'imaginer de calibrer des systèmes multi-caméras à partir de séquences d'images d'objets, lancés à cet effet dans le champ d'opération.

**Mouvements critiques pour l'auto-calibrage** Ce sujet est abordé dans le projet depuis 1997 [18]. Les premières études concernaient les mouvements critiques *génériques*, c'est-à-dire qui constituent des singularités pour tout algorithme d'auto-calibrage. Nous nous sommes ensuite intéressés à des singularités supplémentaires, induites par différents types d'algorithmes. Entre autre, nous avons montré que l'approche classique, basée sur les "équations de Kruppa", induit des singularités non génériques, et qu'elle subit plus de singularités que même quelques méthodes linéaires [32].

**Reconstruction optimale d'une scène plane par morceau à partir de deux images** L'information de coplanarité apporte une contrainte forte pour la reconstruction 3-D d'une scène. Beaucoup de travaux dans le domaine ont exploité cette contrainte, mais typiquement de manière sous-optimale (e.g. la coplanarité n'est typiquement imposée qu'après avoir effectué une reconstruction non contrainte). Nous avons développé une paramétrisation minimale et consistante d'une scène plane par morceau et de la géométrie de deux caméras [55]. Celle-ci nous a permis de créer un estimateur au maximum de vraisemblance qui estime simultanément le mouvement entre deux images et la structure 3-D de la scène. Les résultats expérimentaux, aussi bien sur des données simulées que réelles, montrent que la prise en compte de la (multi-) coplanarité mène à une réduction de l'erreur dans des reconstructions de jusqu'à 50%.

**Auto-étalonnage stéréoscopique** Ce thème est abordé dans le projet depuis 1997. En 1998 nous avons obtenu d'importants résultats théoriques et expérimentaux. Nous avons déve-

loppé un nouvel outil mathématique permettant de caractériser d'un point de vue algébrique et géométrique un système formé de deux caméras rigidement liées l'une à l'autre et se déplaçant dans l'espace. Nous avons montré que dans le cas d'un mouvement général la solution est linéaire et il faut alors deux mouvements pour calibrer le système complètement [27]. Nous avons également étudié plusieurs cas particuliers comme les mouvements planaires et les mouvements de translation. En 1999 nous avons abordé le cas particulier des rotations pures et nous avons proposé une solution complète pour l'auto-étalonnage projectif et/ou euclidien d'une paire de caméras fixée sur un dispositif rotatif à deux degrés de liberté.

En 2000 nous avons abordé le problème de calibration dans le cas particulier où le capteur observe une scène plane. En collaboration avec Andrew Zisserman (professeur invité) nous avons développé une méthode permettant d'étendre notre approche à ce cas particulier [35].

**Stéréo dynamique** Un des problèmes clé concernant l'auto-étalonnage stéréoscopique est l'estimation de l'homographie reliant deux reconstructions projectives obtenues avec le système stéréo à partir de deux positions différentes. Le problème est difficile car, ne connaissant pas a priori le type de mouvement, il n'est pas possible de paramétrer l'homographie. Il faut alors mettre au point des outils statistiques afin d'obtenir la meilleure estimation possible. Nous avons proposé dans un premier temps des méthodes linéaires ainsi qu'une méthode non-linéaire pour estimer cette transformation alors que le système stéréoscopique observe un seul mouvement. Ensuite nous avons étendu cette approche au cas où le système stéréoscopique effectue un mouvement alors qu'il observe plusieurs objets en mouvement [23]. Ceci est illustré sur la figure 2.

Ces travaux sont décrits en détail dans la thèse de doctorat de David Demirdjian [21].

## 4.2 Couplage vision/robotique

**Participants :** Nicolas Andreff, Bernard Espiau, Radu Horaud, Bart Lamiroy, Frédéric Martin, Andreas Ruf.

**Mots clés :** asservissement visuel, auto-étalonnage, étalonnage caméra-pince, modélisation de robots, suivi multi-caméras..

**Résumé :** *Les approches classiques en asservissement visuel considèrent le cas d'une caméra étalonnée intervenant dans la boucle d'asservissement d'un robot [26]. Nos travaux, menés en collaboration avec le projet BIP (projet européen VIGOR) et avec les projets BIP, VISTA et ICARE (Action de recherche coopérative ARC/AVEC), s'intéressent au cas de caméras non-étalonnées en posant la question suivante : peut-on faire de l'asservissement visuel sans un étalonnage préalable des caméras ? On étudie l'élargissement du paradigme "asservissement visuel" au cas de deux caméras liées rigidement (couple stéréoscopique). Ces travaux sont intimement liés au problème d'auto-étalonnage d'une paire de caméras. Les travaux théoriques que nous développons montrent qu'on peut représenter la cinématique d'un*



FIG. 2 – Un capteur stéréoscopique observe plusieurs objets en mouvement alors qu’il bouge lui-même. Sur cette figure on peut voir deux paires d’images extraites d’une séquence de 10 paires d’images, les points extraits de la scène statique (en bas à gauche) ainsi que les deux objets en mouvement (en bas à droite).

*robot avec des transformations projectives. Il apparaît donc possible de modéliser un robot et ses actions aussi bien dans l'espace projectif qu'eulidien et le passage du projectif à l'eulidien n'est, ni plus ni moins, une forme d'auto-étalonnage de l'ensemble robot-caméras.*

*Les expérimentations sont menées en utilisant un robot portique à cinq degrés de liberté sur lequel est fixé une tête stéréoscopique. On dispose également d'un robot manipulateur à six degrés de liberté (cf. figures 3 et 4).*

**Auto-étalonnage caméra-pince.** Les techniques classiques d'étalonnage caméra-pince utilisent une mire d'étalonnage et des mouvements spécifiques de la pince afin que les rotations associées à ces mouvements soient bien conditionnées. Lorsque l'on veut intégrer cet étalonnage dans le processus boucle-fermée de l'asservissement, on ne dispose ni d'une mire ni de la possibilité de choisir les mouvements de la pince. Nous avons intégré la reconstruction 3-D dans le processus d'auto-étalonnage et nous avons développé une méthode qui accepte de faibles déplacements. L'implémentation de cette méthode sur la plate-forme expérimentale du projet donne des résultats comparables à ceux obtenus avec un étalonnage classique.

**Asservissement visuel avec des droites** Dans le cadre de l'asservissement visuel, nous nous sommes penchés sur la modélisation des droites de l'espace. Nous avons montré que les modélisations présentes dans la littérature n'étaient pas idéales. Nous avons dérivé, des coordonnées euclidiennes de Plücker d'une droite, une notion de coordonnées euclidiennes de Plücker normées. Elle permet de définir une nouvelle notion d'alignement des droites aussi bien dans l'image, comme dans les approches précédentes, qu'en orientation dans l'espace. À partir de cet alignement mixte (2D-3D), nous avons pu bâtir une loi de commande explicite et partiellement découplée en rotation et translation. Nous avons ainsi pu faire une analyse globale de la convergence de cette commande et en exhiber les singularités [34].

Cette loi de commande a été appliquée, dans le cadre du projet VIGOR, au positionnement d'une caméra par rapport à un trièdre orthogonal. Cette configuration de droites étant invariante par translation le long de son centre, nous avons proposé une méthode originale d'observation de la profondeur. Elle se fait, par similarité avec la stéréovision, par adjonction d'un pointeur laser non étalonné. Ce pointeur étant fixe par rapport à la caméra, la projection du point laser dans l'image est contraint à se déplacer le long d'une droite épipolaire. L'étude de ces déplacements permet de définir la loi de commande en profondeur.

**Asservissement visuel stéréoscopique** Dans le cadre du projet européen VIGOR nous avons étudié plusieurs types de configurations pour la commande visuelle [43]. L'approche particulière que nous avons étudié et expérimenté consiste à observer le robot avec une paire stéréoscopique calibrée ou non-calibré. Dans le cas calibré, il est possible de réaliser la commande avec deux caméras autonomes. Nous avons comparé cette approche à l'utilisation des contraintes stéréo et nous avons montré que l'approche "2×mono" était plus robuste et plus précise que l'approche "stéréo" [44], [45].

**Géométrie projective et asservissement visuel.** On étudie en détail les propriétés algébriques associées à l'observation d'un mouvement rigide (tel que le mouvement de la pince d'un robot) par un système stéréo non-étalonné. Ces propriétés algébriques permettent de paramétrer l'homographie reliant deux reconstructions projectives obtenues avant et après un mouvement rigide et d'extraire des propriétés euclidiennes dans le cas d'un mouvement général ou planaire et des propriétés affines dans le cas d'un mouvement de translation. Ces travaux sont fortement liés aux travaux fondamentaux dans le domaine de la géométrie pour la vision. En particulier nous avons montré cette année qu'il est possible de modéliser un robot manipulateur couplé à une paire stéréo dans un espace non métrique. Ceci permet de représenter des mécanismes tels que le robot dans l'espace visuel projectif associé avec une paire de caméras faiblement calibrées (seule la géométrie épipolaire est connue). Nous avons complètement caractérisé les propriétés algébriques et différentielles des transformations projectives agissant sur cet espace et nous avons mis en évidence la structure de groupe de Lie de ces transformations [29], [50].

**Suivi d'objets en mouvement avec plusieurs caméras** Toujours dans le cadre du projet européen VIGOR nous avons développé un système capable de suivre un objet rigide en mouvement. L'originalité de ce travail consiste d'une part dans l'approche méthodologique et d'autre part dans l'utilisation judicieuse de plusieurs caméras. L'approche méthodologique s'appuie sur l'asservissement visuel. Cependant la commande est virtuelle car il s'agit de commander la position d'un modèle pour que celui-ci s'aligne en temps réel avec une ou plusieurs images de l'objet à suivre. Nous avons développé un système qui utilise trois caméras synchronisées pouvant travailler soit indépendamment soit isolément. Ces travaux ont débuté cette année et un rapport de recherche est en cour de rédaction. La finalité pratique est le démonstrateur final du projet VIGOR.

### 4.3 Indexation d'images et reconnaissance d'objets

**Participants :** Yves Dufournaud, Riad Hammoud, Radu Horaud, Krystian Mikolajczyk, Roger Mohr, Cordelia Schmid.

**Mots clés :** mise en correspondance, indexation d'images, reconnaissance d'objets, invariant géométrique et photométrique, classes d'objets, classification.

**Résumé :** *L'appariement et l'indexation des images font parti des axes de recherche du projet. Cette activité s'est développée selon plusieurs directions : appariement entre images prises de points de vue différents, robustification de l'indexation et reconnaissance de classes d'objets.*

**Appariement entre images prises de points de vue différents** L'appariement entre deux images en présence de changements d'échelle importants et de changements perspectifs est un problème difficile [31, 30]. L'approche adoptée consiste à utiliser une représentation multi-échelle pour l'ensemble des descripteurs ainsi que pour l'extraction de caractéristiques, endroits où sont calculés les descripteurs. L'utilisation de contraintes géométriques entre images et



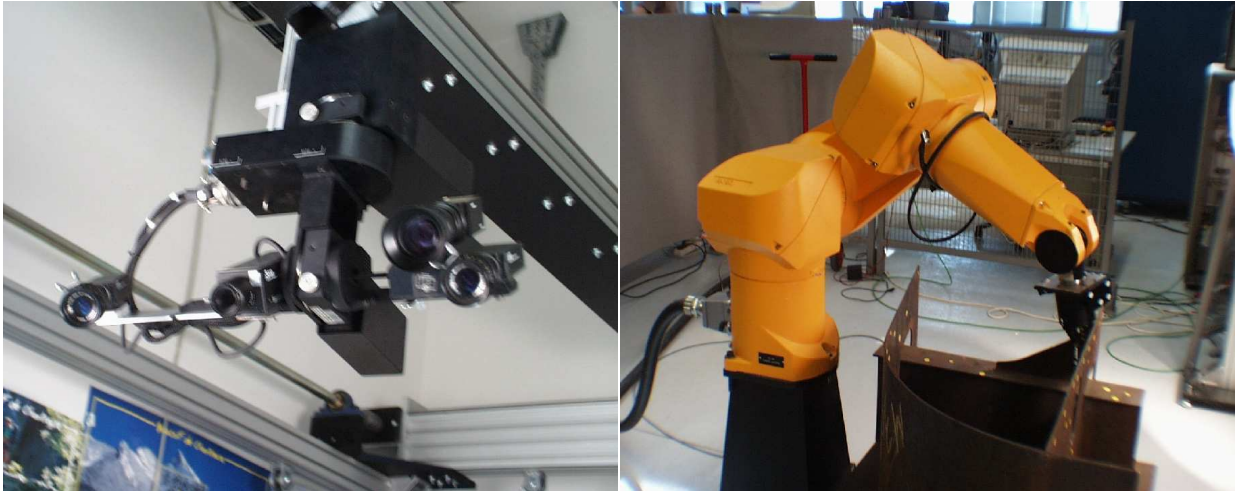


FIG. 3 – La plate-forme expérimentale du couplage vision-robotique comportant une tête stéréo montée sur un robot portique à 5 degrés de liberté (gauche) et un robot manipulateur à 6 degrés de liberté (droite). La pièce métallique visible dans l'image de droite est une pièce de bateau fabriquée par les chantiers navals d'Odense.

l'estimation robuste de ces contraintes permet un rejet important des faux appariements, [37], [36]. L'approche développée permet d'estimer le rapport d'échelle entre deux images sans avoir recours à une initialisation manuelle. D'excellents résultats ont été obtenus lors d'une étude expérimentale sur des données fournies par l'Aérospatiale. La Figure 5 montre des résultats en présence d'un facteur d'échelle de 5 et d'une rotation de 34 degrés.

**Robustification de l'indexation** Dans le cadre de l'indexation, l'approche développée permet l'amélioration de la caractérisation et la robustification de l'indexation par une modélisation d'un objet à partir de plusieurs vues. Une telle approche est particulièrement adaptée dans un contexte de vidéo où plusieurs images d'un même objet sont accessibles (à l'intérieur d'un plan vidéo). La caractérisation ainsi obtenue est ensuite utilisée pour retrouver l'objet ailleurs dans la vidéo (dans un autre plan). Ceci nécessite de modéliser la variabilité de la caractérisation, ce qui est réalisé par l'utilisation d'un mélange de Gaussiennes. L'estimation du nombre de Gaussiennes nécessaires s'avère un problème difficile, [57, 40, 41, 42, 39, 38].

Le problème d'une indexation robuste à des changements d'échelle a également été traité. Une approche multi-échelle s'avère mal adaptée, car elle augmente la complexité de la mise en correspondance et la probabilité de faux appariements. Il a été choisi une méthode qui sélectionne de façon automatique l'échelle optimale pour chaque caractéristique. Une telle échelle est obtenue par maximisation d'un critère dans l'espace d'échelle. Les positions dans l'image où cette maximisation est réalisée sont sélectionnées par un détecteur de points d'intérêt. Des très bons résultats ont montré la possibilité de retrouver une image dans une base contenant 1000 images, alors que le changement d'échelle entre l'image recherchée et l'image de la base peut varier jusqu'à un facteur 4.

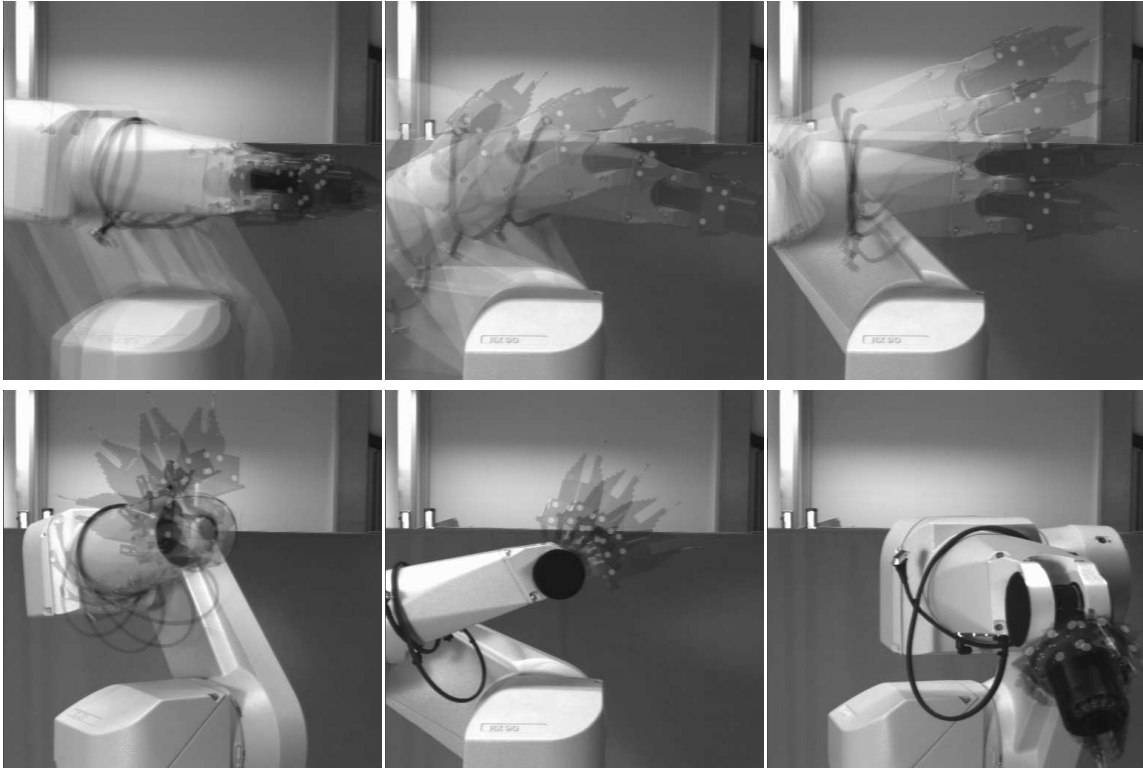


FIG. 4 – Cette figure montre une séquence d’auto-étalonnage du robot. Celui-ci effectue des mouvements articulaires (une articulation à la fois) devant le capteur stéréoscopique. On obtient ainsi un modèle de “cinématique projective” du robot à partir duquel on peut extraire les paramètres euclidiens du robot ainsi que les paramètres du capteur.

**Recherche de classes d’objets** Dans le contexte de la recherche de classes d’objets il ne s’agit pas de retrouver le même objet, mais des images similaires, c’est-à-dire appartenant à la même classe d’objets. Ceci s’applique à des objets comme par exemple des visages, des chevaux, des chaises etc. Ceci est important pour des applications telles qu’il en existe par exemple dans les agences de presse. Ces agences souhaitent pouvoir répondre à des requêtes comme trouver des images contenant des enfants qui jouent.

Dans ce but, un détecteur de visage a développé. Ce détecteur utilise des descripteurs génériques obtenus à partir d’invariants locaux. Ces descripteurs représentent entre autre la bouche, le nez et les yeux. Pour tenir compte de la variabilité et de la diversité d’apparence de telles données, des mélanges de Gaussiennes s’avèrent bien adaptés. Ces mélanges ont été calculés à partir d’un ensemble représentatif d’images de visages. La détection de visages à partir de ces descripteurs permet d’obtenir de bons résultats [54].

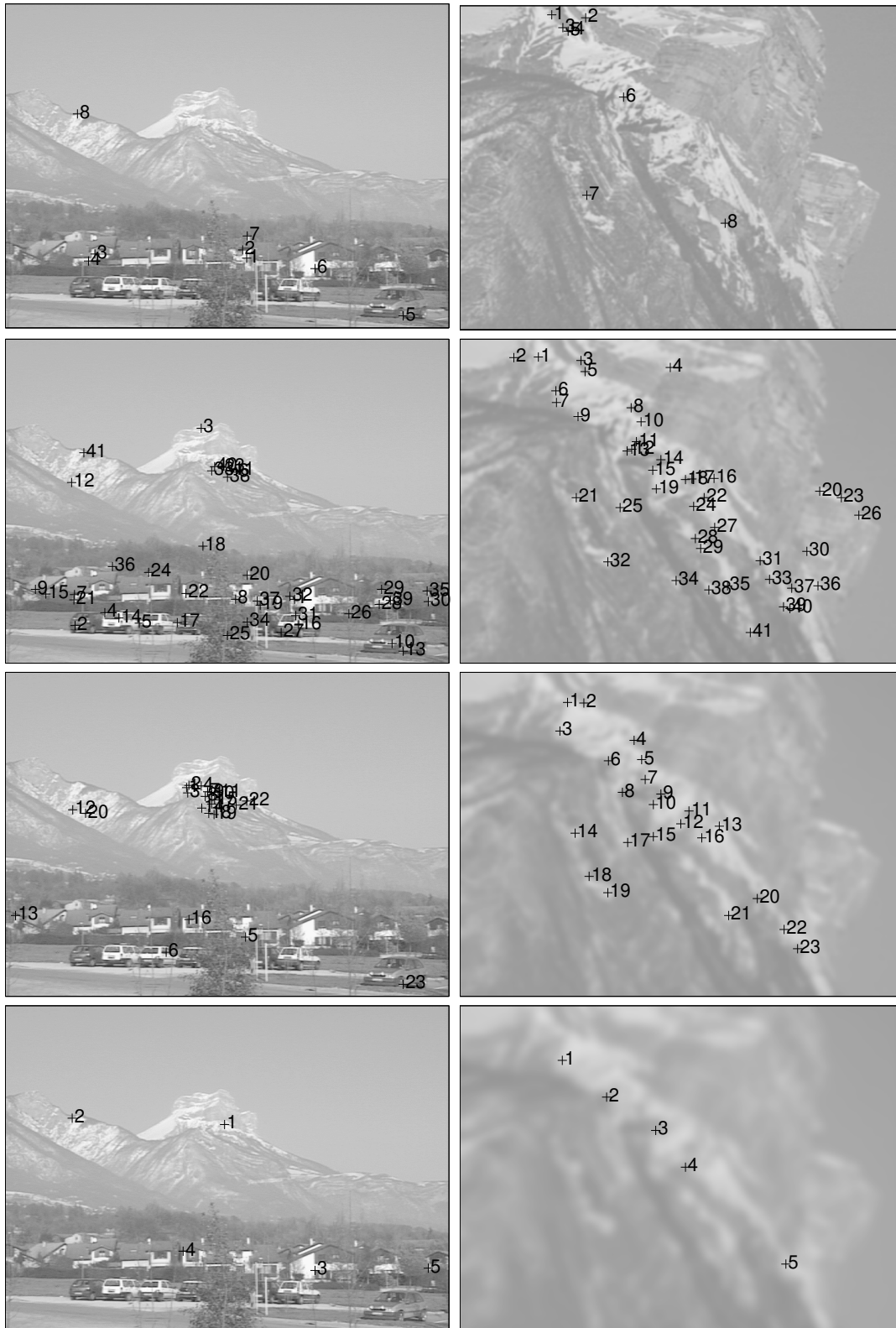


FIG. 5 – Cette figure montre les points d'intérêt détectés à plusieurs niveaux de résolution de l'image de droite et les points correspondants dans l'image basse résolution (gauche).

## 4.4 Structuration de vidéos

**Participants :** Riad Hammoud, Krystian Mikolajczyk, Roger Mohr, Cordelia Schmid.

**Résumé :** *Pour pouvoir manipuler l'information du contenu d'une vidéo, il faut la structurer. Ces structures permettent l'interactivité de l'utilisateur, la recherche d'informations pertinentes, ou plus simplement un parcours adapté au besoin de l'utilisateur. Si structurer automatiquement les vidéos en leur contenu sémantique est actuellement hors de portée de nos programmes, il reste en revanche possible d'établir des liens entre les objets détectés, de découper les objets de la vidéo de façon semi-automatique, et à partir de ces éléments d'autoriser l'éditeur à créer la structure recherchée.*

**Structuration de la vidéo** Actuellement, il est impossible d'achever une indexation de haut niveau de la vidéo par le contenu comme suggéré par exemple par le groupe MPEG-7. Cette indexation se place dans le coeur des applications sur la vidéo et en particulier sur "la création et la présentation de la vidéo hyperliée". Pour créer un film interactif et créer des liens entre des objets "identiques" il faut être capable de repérer des occurrences de ces objets à travers la vidéo. L'identification des objets est une tâche très difficile à cause de l'apparence variable de ces objets à travers le temps.

Dans un premier temps nous avons proposé une technique de caractérisation de bas-niveau bien adapté à nos données visuelles (objets non-rigides suivis), qui permet d'une part de *capturer* les différentes apparences intra-plan de l'objet suivi, et d'autre part elle *réduit* considérablement le nombre des descripteurs de bas niveaux à indexer par objet suivi. La théorie du mélange gaussien a été adoptée à ce stade.

Dans un deuxième temps nous avons proposé deux approches supervisée et automatique de classification des objets suivis. Dans l'approche supervisée l'utilisateur expert de la vidéo interactive désigne les "objets suivis modèles" pour une séquence vidéo traitée et ensuite le système classe tous les objets de cette vidéo dans ces "objets suivis modèles". Deux techniques de classements de requêtes sont testées : classement par maximum à posteriori et classement robuste par vote majoritaire. Les expérimentations sur la séquence vidéo *Avengers* ont montrées que cette dernière technique fournit les meilleurs résultats de classement d'apparences d'objets.

Dans l'approche automatique, la classification est effectuée sur les densités du mélange gaussien et non pas dans l'espace de descripteurs classiques de bas niveaux. La distance entre deux objets suivis est calculée par le minimum de la distance de Kullback entre les composantes gaussiennes des deux densités de mélanges correspondant. Sur la matrice de distances ainsi obtenue, une classification ascendante hiérarchique est effectuée. Une méthode interactive a été proposée également pour choisir le nombre de classes d'objets suivis. Aussi, des outils graphiques ont été développés et mis à la disposition de l'utilisateur expert de la vidéo interactive afin de corriger les résultats de la classification automatique du système.

En se basant sur les résultats de la modélisation des objets suivis dans les plans nous avons proposé une technique automatique pour sélectionner les images- apparences-clés des objets suivis. Celles-ci permettent d'accélérer la navigation dans la vidéo.

Aussi et pour le même objectif, nous avons proposé une approche de macro-segmentation de la vidéo en scènes. L'idée de base de cette approche consiste à caractériser la plan vidéo par plusieurs de descripteurs : histogramme de couleurs et le contenu du plan. Une technique de fusion des clusters de différents descripteurs à été proposée.

Enfin, un système de création et de présentation de la vidéo hyperliée qui intègre les approches développées durant cette thèse est implanté [57, 40, 41, 42, 39, 38].

## 5 Contrats industriels (nationaux, européens et internationaux)

### 5.1 Reconnaissance et indexation de vues aériennes

**Participants :** Radu Horaud, Cordelia Schmid, Yves Dufournaud.

Nous collaborons avec la Société Aérospatiale depuis 1993. Depuis 1998 et pour une durée de trois ans nous avons un contrat de collaboration dans le domaine de l'interprétation d'images aériennes.

Yves Dufournaud effectue une thèse de doctorat dans le cadre d'une convention CIFRE. L'objectif de ce travail est d'étudier, expérimenter et réaliser un logiciel permettant de comparer des images prises dans des conditions de vue très différentes en termes de la focale utilisée.

Les travaux effectués à ce jour sont très encourageants et nous avons mis en place un transfert de savoir-faire méthodologique.

### 5.2 Indexation de vidéo

**Participants :** Riad Hammoud, Roger Mohr, Cordelia Schmid.

La société Alcatel, à travers sa division Alcatel-CRC, a lancé une action de partenariat avec l'INRIA sur le sujet de l'indexation vidéo (autres partenaires de l'institut : le projet VISTA). L'objet de cette action est la réalisation d'outils de base permettant d'indexer et structurer de façon automatique des bandes de vidéo, en se limitant aux images.

Dans cette opération, notre projet est principalement acteur sur le calcul de signatures d'indexation, et de la réalisation de liens entre plans de la vidéo. Une démonstration est en cours de réalisation.

Par ailleurs, Alcatel finance la bourse CIES de R. Hammoud, thèse qui sera soutenue en 2000.

Divers contacts ont été noués avec des PME dans le domaine de l'indexation des images fixes : Alma, Caldera et SGBI.

### 5.3 Projet RNRT "AGIR"

**Participants :** Riad Hammoud, Krystian Mikolajczyk, Cordelia Schmid.

Le projet AGIR est un projet RNRT précompétitif dont le début a été fixé au 1er septembre 1999. Il regroupe le centre de recherche d'Alactel, une PME (Arts Video Interactive), l'INA, et des équipes de recherches (IRISA, INRIA Rhône-Alpes, IRIT, CLIPS-IMAG, INT, LIP6). Ce projet est financé pour 18 mois, avec un renouvellement possible de 12 mois.

Le but du projet est de développer un système complet permettant d'indexer des vidéos, éventuellement accompagnée de textes, puis d'utiliser cette indexation pour rechercher des documents par leur contenu, ceci en mélangeant les divers médias. Ce projet prend donc naturellement la suite du travail fait dans le cadre du projet européen DIVAN, avec un accent plus fort mis sur l'aspect multimédia des données, des index et des requêtes, sur l'aspect architecture globale du projet et sur les liens avec les efforts de normalisation actuels (MPEG7).

Les travaux développés par MOVI concerneront la fourniture de descripteurs d'images. Ces descripteurs, calculés à partir de diverses primitives (points, régions, visages) doivent permettre l'identification d'objets ou de scènes. Un effort spécifique sera fourni pour adapter les techniques utilisées au cas de la vidéo où la redondance des images permet une robustification des descripteurs. Cela sera fait en reprenant les travaux de R. Hammoud sur le calcul de modèles de descripteurs par mélange de gaussiennes. Sera aussi étudié le problème de la composition spatiale des descripteurs élémentaires pour obtenir un langage de requête plus riche.

#### 5.4 Projet PRIAMM "Vistéo"

**Participants :** Edmond Boyer, Ragini Choudhury.

Le projet VISTEO implique deux projets de recherche INRIA (MOVI, SHARP) ainsi que la société Getris-Image. Initié en janvier 2000 pour une durée de 2 ans, il concerne les studios virtuels en télé-programmation. Il s'agit d'ajouter, dans les images d'une émission, les éléments correspondant à un décor virtuel situé autour du présentateur. Le but est de voir le présentateur dans un environnement virtuel issu d'un processus de synthèse d'images et de permettre au présentateur d'interagir avec cet environnement. La réalisation de ces opérations repose sur plusieurs processus de vision par ordinateur : l'étalonnage de la caméra et la détermination de sa position et de son orientation pour chaque image traitée ; le suivi du présentateur dans le flot d'images ; la détermination des positions respectives du présentateur et de la caméra. Le projet MOVI intervient sur ces différents points. Un chercheur post-doctoral Indien : Ragini Choudhury, a développé les outils nécessaires à la réalisation du premier point : le calibrage de la caméra en temps réel [56]. Les autres points sont actuellement en cours d'études.

## 6 Actions régionales, nationales et internationales

### 6.1 Actions régionales

Le projet est impliqué dans une action régionale ACTIV qui fédère onze groupes de recherche régionaux en vision par ordinateur et qui aborde les thèmes de catégorisation et segmentation d'images, de mesure et rendu de la couleur, et d'indexation d'images. R. Mohr est coordinateur de ce dernier thème.

### 6.2 Actions nationales

- L. Quan participe à l'animation du GT5 dans le GDR ISIS.

- Le projet participe également à l'ARC (Action de recherche coopérative) AVEC (Asservissement visuel en environnements complexes). Cette action a débuté en 1998 et se prolonge jusqu'en 2000. Les autres partenaires sont les projets BIP, VISTA (Rennes) et ICARE (Sophia-Antipolis). Voir (<http://www.irisa.fr/AVEC/>)

### 6.3 Actions financées par la Commission Européenne

CUMULI. Le projet LTR 21914 « CUMULI – Computational Understanding of Multiple Images<sup>2</sup> » a pris fin en février 2000 après 3,5 ans. Son but était d'approfondir nos compétences théoriques et pratiques en reconstruction visuelle photogrammétrique à partir d'images multiples, avec pour emphases l'étape d'initialisation, l'autocalibrage, et l'optimisation de la structure. Les partenaires étaient : MOVI (coordinateur); l'équipe ROBOTVIS de l'INRIA Sophia-Antipolis ; le Mathematical Imaging Group à l'Université de Lund, Suède ; les PME photogrammétrique Imetric S.A, Porrentruy, Suisse, et Image Systems A.B., Linköping, Suède ; et l'équipe IGD A4 (réalité virtuelle et augmenté) de l'Institut Fraunhofer, Darmstadt, Allemagne.

VIGOR. Nous sommes coordinateur du projet européen Reactive LTR VIGOR. (<http://www.inrialpes.fr/VIGOR/>). Ce projet auquel participe également le projet BIP s'exécute en partenariat avec deux équipes universitaires (University of Cambridge et Hebrew University of Jerusalem), un institut de recherche (IITB) et deux partenaires industriels (Sinters SA et Odense Steel Shipyard). D'une durée de trois ans et ayant commencé le 1 février 1998, ce projet a comme objectifs d'étendre le concept de caméras non-calibrées au guidage visuel de robot et de proposer un nouveau concept où les caméras et les robots sont considérés ensemble pour des applications navales et aéronautiques.

**Bourses TMR.** MOVI accueille :

- Adreas Ruf (bourse TMR de type 20) pour une durée de trois ans (1997–2000).

TELEDIMOS. Ce projet est réalisé en partenariat avec le projet BIP. La contribution du projet MOVI est le développement logiciel d'une plate-forme de calibration d'une ou plusieurs caméras ainsi qu'un logiciel permettant de valider la qualité des paramètres de calibration. Nous avons également réalisé un logiciel de triangulation stéréoscopique. Une interface graphique utilisateur permet au non-spécialiste de se servir de ce logiciel.

VISIRE. Nous participons depuis le mois de mai au projet européen IST VISIRE dont l'objectif est d'exploiter la limite de la technologie de la vision par ordinateur 3D pour pouvoir obtenir des reconstructions 3D de scénarios à partir de séquences vidéos d'une caméra grand public. Ce projet va durer 3 ans et s'exécute en liaison avec nos partenaires académiques de l'université de Lund (Suède) et de l'université polytechnique de Madrid (Espagne); et de deux partenaires industriels, Eptron S.A. (Espagne) et Giunti Multimedia (Italie).

---

2. <http://www.inrialpes.fr/CUMULI/>

EVENTS. Nous participons depuis le mois d'octobre au projet européen IST EVENTS dont l'objectif est de rechercher de la nouvelle technologie en vision par ordinateur qui permet d'avoir un système en temps réel de l'interpolation d'images inovant pour la transmission par télévision des grands scenarios en 3D. Ce projet va durer 3 ans et s'exécute en liaison avec nos partenaires de l'université d'Oxford (Angleterre), de Siemens IC C-Lab (Allemagne), de Eptron S.A. (Espagne) et de Via Digital (Espagne).

VIBES. Le nouveau projet 5e cadre FET-Open IST-2000-26001 « VIBES, Video Browsing, Exploration and Structuring » a officiellement débuté 1 décembre 2000. Ce projet de 3 ans a pour but de développer des représentations et des techniques de manipulation haut-niveau (niveau objets / personnages) pour la vidéo, en particulière des méthodes d'indexation et des méthodes de reconstruction / modification / resynthèse. Les partenaires sont KTH Stockholm (coordinateur) (Suède), MOVI (France), l'Université d'Oxford (UK), le Katholieke Universiteit Leuven (Belge), l'Ecole Polytechnique Fédérale de Lausanne (Suisse), et le Weizmann Institute of Science (Israel). VIBES représentera 32 personne-ans d'effort, pour un budget totale de 2,3 MEu (15 MF), avec support Européen de 1,7 Meu (11 MF). MOVI travaillera en particulière sur les aspects indexation, et suivi et reconstruction humaine.

## 6.4 Relations bilatérales internationales

### 6.4.1 Amérique

Dans le cadre du programme de la collaboration CNRS/UIUC (The Univ. of Illinois, U.S.A.), nous menons un projet avec l'équipe de J. Ponce sur le rendu d'images à base d'images. Le projet est renouvelé pour l'année 2000.

### 6.4.2 Asie

P. Sturm a effectué un séjour de 10 jours à l'Académie des Sciences à Pékin.

Dans le cadre du laboratoire Franco-chinois, nous avons un projet de collaboration avec l'académie de science de Chine sur la sélection de modèles pour la reconstruction à partir d'images.

Une collaboration de plusieurs années se poursuit avec l'université Xidian de Xi'an (professeur Wu Chengke). Elle porte sur l'utilisation des outils géométriques pour la compression d'images et pour la synthèse d'images à partir d'images.

L. Quan a aussi séjourné pendant l'été 2000 à Microsoft Research, Beijing. Par ailleurs, L. Quan est oversea assessor (élu pour 3 ans) pour l'académie des Sciences de Chine et aussi professeur invité de l'institut de l'académie des Sciences de Chine.



## 7 Diffusion de résultats

### 7.1 Animation de la communauté scientifique

Les membres du projet font partie des comités de rédaction de revues suivantes :

- *Machine Vision and Applications* (R. Mohr),
- *International Journal of Robotics Research* (R. Horaud),
- *Computacion y Sistemas* (R. Horaud),

Les membres du projet font partie des comités de programme des conférences suivantes :

- ECCV'00 (W. Triggs, L. Quan)
- CVPR'00 (C. Schmid, area-chair, L. Quan, W. Triggs)
- ICPR'00 (W. Triggs, P. Sturm)
- ICRA'00 (R. Horaud)
- 9e IMA Workshop on Mathematics of Surfaces (Cambridge, 4–6 septembre 2000), W. Triggs
- 2e Workshop on 3D Structure from Multiple Images of Large-scale Environments and applications to Virtual and Augmented Reality (SMILE'00, Dublin, 1–2 juillet 2000), W. Triggs
- IAPR Workshop Robot Vision 2001 (16–18 février 2001, Auckland, New Zealand), W. Triggs

### 7.2 Enseignement universitaire

- Outils mathématiques pour la vision et la robotique, DEA IVR, INPG, 12h, L. Quan D. Demirdjian,
- Vision 3D, DEA IVR, INPG, 18h, R. Horaud
- Applications de la vision à la robotique, DEA SIP, INPG, 30h, R. Horaud.
- Vision 3D INPG - ENSIEG, 20h, L. Quan
- Introduction à la vision par ordinateur – MAGISTÈRE, UJF, 18h, P. Sturm.

### 7.3 Participation à des colloques, séminaires, invitations

Les membres du projet ont été invités à faire des présentations aux manifestations suivantes :

- P. Sturm fait partie du jury du Prix de Thèse SPECIF 2000.
- W. Triggs est éditeur du volume Springer-Verlag LNCS 1883 « Vision Algorithms: Theory and Practice », (ISBN 3-540-67973-1, septembre 2000), qui contient – avec plusieurs papiers invités – les proceedings post-workshop du workshop INRIA-IEEE du même nom qui a eu lieu à ICCV'99 en Corfou, Grèce en septembre 1999.
- R. Horaud, C. Schmid et A. Ruf ont participé à l'École franco-mexicaine d'imagerie et robotique, juillet 2000.
- R. Horaud et A. Ruf on présenté un tutoriel lors de la conférence ECCV'00, Dublin, juillet 2000.
- R. Horaud a prononcé deux séminaires à l'Université de British Columbia et à Microsoft Research Center, décembre 2000.
- Cordelia Schmid a fait une présentation invitée au Beckman Institute vision workshop, Urbana-Champaign, février 2000.
- Cordelia Schmid a fait une présentation invitée au "Vision Seminar" à l'université de Berkeley, janvier 2000.
- M. Lhuillier a reçu le prix du meilleur article étudiant lors la conférence ICPR'2000.
- Long Quan a fait deux séminaires à l'institut de l'académie des Sciences de Chine et à Microsoft Research, Beijing, août 2000.

## 8 Bibliographie

### Ouvrages et articles de référence de l'équipe

- [1] S. CHRISTY, R. HORAUD, « Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations », *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 11, November 1996, p. 1098–1104, <ftp://ftp.inrialpes.fr/pub/movi/publications/rec-affiter-long.ps.gz>.
- [2] P. GROS, O. BOURNEZ, E. BOYER, « Using Local Planar Geometric Invariants to Match and Model Images of Line Segments », *Computer Vision and Image Understanding* 69, 2, 1998, p. 135–155.
- [3] R. HARTLEY, P. STURM, « Triangulation », *Computer Vision and Image Understanding* 68, 2, 1997, p. 146–157.
- [4] L. HÉRAULT, R. HORAUD, « Figure-ground discrimination: a combinatorial optimization approach », *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15, 9, September 1993, p. 899–914.

- 
- [5] R. HORAUD, F. DORNAIKA, B. ESPIAU, «Visually Guided Object Grasping», *IEEE Transactions on Robotics and Automation* 14, 4, August 1998, p. 525–532.
- [6] R. HORAUD, F. DORNAIKA, B. LAMIROY, S. CHRISTY, «Object Pose: The Link between Weak Perspective, Paraperspective, and Full Perspective», *International Journal of Computer Vision* 22, 2, March 1997, p. 173–189.
- [7] R. HORAUD, F. DORNAIKA, «Hand-Eye Calibration», *International Journal of Robotics Research* 14, 3, June 1995, p. 195–210.
- [8] R. HORAUD, O. MONGA, *Vision par ordinateur: outils fondamentaux*, Editions Hermès, Paris, 1995, *Deuxième édition revue et augmentée*.
- [9] R. MOHR, B. BOUFAMA, P. BRAND, «Understanding Positioning from Multiple Images», *Artificial Intelligence* 78, 1995, p. 213–238.
- [10] L. QUAN, T. KANADE, «Affine Structure from Line Correspondences with Uncalibrated Affine Cameras», *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 8, août 1997, p. 834–845.
- [11] L. QUAN, «Invariants of Six Points and Projective Reconstruction from Three Uncalibrated Images», *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17, 1, janvier 1995, p. 34–46.
- [12] L. QUAN, «Conic Reconstruction and Correspondence from Two Views», *IEEE Transactions on Pattern Analysis and Machine Intelligence* 18, 2, février 1996, p. 151–160.
- [13] L. QUAN, «Self-Calibration of An Affine Camera from Multiple Views», *International Journal of Computer Vision* 19, 1, mai 1996, p. 93–105.
- [14] A. RUF, R. HORAUD, «Visual Servoing of Robot Manipulators, Part I: Projective Kinematics», *International Journal of Robotics Research* 18, 11, November 1999, p. 1101–1118, [ftp://ftp.inria.fr/INRIA/publication/RR/RR-3670.ps.gz](http://ftp.inria.fr/INRIA/publication/RR/RR-3670.ps.gz).
- [15] C. SCHMID, R. MOHR, «Object Recognition Using Local Characterization and Semi-Local Constraints», *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19, 5, mai 1997, p. 530–534.
- [16] P. STURM, S. MAYBANK, «On Plane-Based Camera Calibration: A General Algorithm, Singularities, Applications», *in: Proceedings of the Conference on Computer Vision and Pattern Recognition, Fort Collins, Colorado, USA*, p. 432–437, June 1999.
- [17] P. STURM, B. TRIGGS, «A Factorization Based Algorithm for Multi-Image Projective Structure and Motion», *in: Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, p. 709–720, Avril 1996.
- [18] P. STURM, «Critical Motion Sequences for Monocular Self-Calibration and Uncalibrated Euclidean Reconstruction», *in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico*, p. 1100–1105, Juin 1997.
- [19] B. TRIGGS, «Matching Constraints and the Joint Image», *in: IEEE Int. Conf. Computer Vision*, E. Grimson (éditeur), p. 338–43, Cambridge, MA, juin 1995.
- [20] B. TRIGGS, «Autocalibration and the Absolute Quadric», *in: IEEE Conf. Computer Vision & Pattern Recognition, Puerto Rico*, 1997.

## Thèses et habilitations à diriger des recherches

- [21] D. DEMIRDJIAN, *Le Mouvement Projectif. Théorie et Applications pour l'Autocalibrage et la Segmentation du Mouvement*, thèse de doctorat, Institut National Polytechnique de Grenoble, Juillet 2000.
- [22] M. LHUILLIER, *Modélisation pour la synthèse d'images à partir d'images*, thèse de doctorat, Institut National Polytechnique de Grenoble, Décembre 2000.

## Articles et chapitres de livre

- [23] D. DEMIRDJIAN, R. HORAUD, «Motion-Egomotion Discrimination and Motion Segmentation from Image-Pair Streams», *Computer Vision and Image Understanding* 78, 1, Avril 2000, p. 53–68, <ftp://ftp.inrialpes.fr/pub/movi/publications/DavidRadu-cviu2000.ps.gz>.
- [24] O. FAUGERAS, L. QUAN, P. STURM, «Self-calibration of a 1D Projective Camera and its Application to the Self-calibration of a 2D Projective Camera», *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 10, 2000, p. 1179–1185.
- [25] A. HEYDEN, F. KAHL, L. QUAN, «Minimal Projective Reconstruction Including Missing Data», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
- [26] R. HORAUD, F. CHAUMETTE, «Guest Editorial, Special Issue on Image-Based Servoing», *International Journal of Computer Vision* 37, 1, Juin 2000, p. 5–6.
- [27] R. HORAUD, G. CSURKA, D. DEMIRDJIAN, «Stereo Calibration from Rigid Motions», *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 12, Décembre 2000, à paraître, <ftp://ftp.inrialpes.fr/pub/movi/publications/HoraudCsurkaDemirdjian-pami2000.ps.gz>.
- [28] L. QUAN, «Two-Way Ambiguity in 2D Projective Reconstruction from Three Uncalibrated 1D Images», *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2000.
- [29] A. RUF, F. MARTIN, B. LAMIROY, R. HORAUD, «Visual Control Using Projective Kinematics», *in: Proceedings of the Ninth International Symposium on Robotics Research*, J. Hollerbach et D. Koditschek (éditeurs), Springer-Verlag, 2000, p. 95–104, <ftp://ftp.inrialpes.fr/pub/movi/publications/ruf99g.ps.gz>.
- [30] C. SCHMID, A. ZISSERMAN, «The geometry and matching of lines and curves over multiple views», *International Journal of Computer Vision*, 2000, à paraître.
- [31] C. SCHMID, R. MOHR, C. BAUCKHAGE, «Evaluation of Interest Point Detectors», *International Journal of Computer Vision* 37, 2, 2000, p. 151–172.
- [32] P. STURM, «A Case Against Kruppa's Equations for Camera Self-Calibration», *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22, 10, Septembre 2000, p. 1199–1204.
- [33] B. TRIGGS, P. MCLAUCHLAN, R. HARTLEY, A. FITZGIBBON, «Bundle Adjustment — A Modern Synthesis», *in: Vision Algorithms: Theory and Practice*, B. Triggs, A. Zisserman, et R. Szeliski (éditeurs), *Lecture Notes in Computer Science, 1883*, Springer-Verlag, 2000, p. 298–372.

**Communications à des congrès, colloques, etc.**

- [34] N. ANDREFF, B. ESPIAU, R. HORAUD, «Visual Servoing from Lines», *in: Proceedings of the International Conference on Robotics and Automation, San Francisco, USA*, p. 2070–2075, Avril 2000, <http://www.inrialpes.fr/bip/people/Andreff/docs/icra00.ps.gz>.
- [35] D. DEMIRDJIAN, A. ZISSERMAN, R. HORAUD, «Stereo Autocalibration from One Plane», *in: Proceedings of the 6th European Conference on Computer Vision, Dublin, Irlande, II*, Springer-Verlag, p. 625–639, Juin/Juillet 2000.
- [36] Y. DUFOURNAUD, C. SCHMID, R. HORAUD, «Appariement d'images à des échelles différentes», *in: Actes du 12ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, 2*, p. 327–336, Février 2000.
- [37] Y. DUFOURNAUD, C. SCHMID, R. HORAUD, «Matching Images with Different Resolutions», *in: Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, USA*, IEEE Computer Society Press, p. 612–618, Juin 2000.
- [38] R. HAMMOUD, D. KOUAM, «A mixed classification approach of shots for constructing scene structure for movie films», *in: Proceedings of the Irish Machine Vision and Image Processing Conference, Belfast, Royaume-Uni*, p. 223–230, August/September 2000.
- [39] R. HAMMOUD, R. MOHR, «Interactive Tools for Constructing and Browsing Structures for Movie Films», *in: Proceedings of ACM Multimedia, Los Angeles, USA*, p. 497–498, Octobre/Novembre 2000.
- [40] R. HAMMOUD, R. MOHR, «Mixture Densities for Video Objects Recognition», *in: Proceedings of the 15th International Conference on Pattern Recognition, Barcelone, Espagne, 2*, p. 71–75, Septembre 2000.
- [41] R. HAMMOUD, R. MOHR, «A probabilistic framework of selecting effective key frames for video browsing and indexing», *in: Proceedings of the International workshop on Real-Time Image Sequence Analysis, Oulu, Finland*, p. 79–88, Août/Septembre 2000.
- [42] R. HAMMOUD, R. MOHR, «Probabilistic Hierarchical Framework for Clustering of Tracked Objects in Video Streams», *in: Proceedings of the Irish Machine Vision and Image Processing Conference, Belfast, Royaume-Uni*, p. 133–140, Août/Septembre 2000.
- [43] B. LAMIROY, T. DRUMMOND, R. HORAUD, O. KNUDSEN-NECKELMAN, «Visually Guided Robots for Ship Building», *in: Proceedings of the 1st International Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT'2000), Potsdam/Berlin, Allemagne*, Mars/Avril 2000, <ftp://ftp.inrialpes.fr/pub/VIGOR/DOCUMENTS/LamiroyDrummondCOMPIT.ps.gz>.
- [44] B. LAMIROY, B. ESPIAU, N. ANDREFF, R. HORAUD, «Controlling Robots with Two Cameras: How to Do it Properly», *in: Proceedings of the International Conference on Robotics and Automation, San Francisco, USA*, p. 2100–2105, Avril 2000, [ftp://ftp.inrialpes.fr/pub/movi/publications/LamiroyEspiauAndreffHoraud\\_ICRA2000.ps.gz](ftp://ftp.inrialpes.fr/pub/movi/publications/LamiroyEspiauAndreffHoraud_ICRA2000.ps.gz).
- [45] B. LAMIROY, C. PUGET, R. HORAUD, «What Metric Stereo Can Do for Visual Servoing», *in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Takamatsu, Japon*, p. 251–256, Septembre 2000, <ftp://ftp.inrialpes.fr/pub/VIGOR/DOCUMENTS/LamiroyIROS2000.ps.gz>.

- [46] M. LHUILLIER, L. QUAN, « Appariement dense robuste à l'aide de contraintes géométriques locales et globales », in : *Actes du 12ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle*, 3, p. 215–223, Février 2000, <http://www.inrialpes.fr/movi/people/Lhuillie/demo2.html>.
- [47] M. LHUILLIER, L. QUAN, « Edge-Constrained Joint View Triangulation for Image Interpolation », in : *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, USA, 2*, p. 218–224, Juin 2000, <http://www.inrialpes.fr/movi/people/Lhuillie/demo3.html>.
- [48] M. LHUILLIER, L. QUAN, « Robust Dense Matching Using Local and Global Geometric Constraints », in : *Proceedings of the 15th International Conference on Pattern Recognition, Barcelone, Espagne, 1*, p. 968–972, 2000. ICPR'2000 Piero Zamperoni Best Student Paper Award, <http://www.inrialpes.fr/movi/people/Lhuillie/demo3.html>.
- [49] L. QUAN, B. TRIGGS, « A Unification of Autocalibration Methods », in : *Proceedings of the Fourth Asian Conference on Computer Vision*, p. 917–922, 2000.
- [50] A. RUF, R. HORAUD, « Vision-based guidance and control of robots in projective space », in : *Proceedings of the 6th European Conference on Computer Vision, Dublin, Irlande, II*, Springer-Verlag, p. 50–66, Juin/Juillet 2000, <http://www.inrialpes.fr/movi/people/Ruf/rufECCV00.ps.gz>.
- [51] P. STURM, L. QUAN, « Camera Calibration and Relative Pose Estimation from Gravity », in : *Proceedings of the 15th International Conference on Pattern Recognition, Barcelone, Espagne*, A. Sanfeliu, J. Villanueva, M. Vanrell, R. Alquézar, J.-O. Eklundh, Y. Aloimonos (éditeurs), 1, p. 72–75, Septembre 2000.
- [52] P. STURM, « Algorithms for Plane-Based Pose Estimation », in : *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, USA*, p. 1010–1017, Juin 2000.
- [53] B. TRIGGS, « Plane + Parallax, Tensors and Factorization », in : *Proceedings of the 6th European Conference on Computer Vision, Dublin, Irlande*, Springer-Verlag, p. 522–538, 2000.
- [54] V. VOGELHUBER, C. SCHMID, « Face detection based on generic local descriptors and spatial constraints », in : *Proceedings of the 15th International Conference on Pattern Recognition, Barcelone, Espagne, 1*, p. 1084–1087, 2000.

## Rapports de recherche et publications internes

- [55] A. BARTOLI, P. STURM, R. HORAUD, « A Projective Framework for Structure and Motion Recovery from Two Views of a Piecewise Planar Scene », *Research Report n° 4070*, INRIA, OCTOBRE 2000.
- [56] R. CHOUDHURY, « REAL TIME CAMERA CALIBRATION », *Rapport technique*, INRIA, JUIN 2000.
- [57] R. HAMMOUD, R. MOHR, « GAUSSIAN MIXTURE DENSITIES FOR INDEXING OF LOCALIZED OBJECTS IN A VIDEO SEQUENCE », *rapport de recherche*, INRIA, MARS 2000, <HTTP://WWW.INRIA.FR/RRRT/RR-3905.HTML>.
- [58] B. TRIGGS, « ROUTINES FOR RELATIVE POSE OF TWO CALIBRATED CAMERAS FROM 5 POINTS », *rapport de recherche*, INRIA, 2000.