

Projet MOVI

Modélisation pour la vision par ordinateur

Rhône-Alpes

THÈME 3B

R *apport*
d'Activité

2001

Table des matières

1	Composition de l'équipe	3
2	Présentation et objectifs généraux	4
3	Fondements scientifiques	5
3.1	Optimisation numérique et apprentissage	5
3.2	Commande visuelle de robots	6
3.3	Description d'images	6
3.4	Reconstruction de surfaces	7
3.5	Synthèse d'images à partir d'images	8
3.6	Géométrie	8
4	Domaines d'applications	8
4.1	Panorama	8
4.2	Vision, robots et leur couplage	8
4.3	Indexation de bases d'images	9
4.4	Vision et défense nationale	9
4.5	Virtualisation d'objets	10
4.6	Synthèse d'images à partir d'images	10
4.7	Analyse vidéo	11
5	Logiciels	11
5.1	Trivision	11
5.2	Teledimos	12
5.3	Modélisation interactive de scènes	12
6	Résultats nouveaux	12
6.1	Géométrie multi-images et multi-caméras	12
6.2	Couplage vision/robotique	17
6.3	Indexation d'images et reconnaissance d'objets	19
6.4	Structuration de vidéos	22
6.5	Autres	23
7	Contrats industriels (nationaux, européens et internationaux)	23
7.1	Reconnaissance et indexation de vues aériennes	23
7.2	Projet RNRT « AGIR »	23
8	Actions régionales, nationales et internationales	24
8.1	Actions nationales	24
8.2	Actions financées par la Commission Européenne	24
8.3	Relations bilatérales internationales	25
8.3.1	Europe	25
8.3.2	Amérique	26

8.3.3	Asie	26
9	Diffusion de résultats	26
9.1	Animation de la communauté scientifique	26
9.2	Enseignement universitaire	27
9.3	Participation à des colloques, séminaires, invitations	27
10	Bibliographie	27

1 Composition de l'équipe

Responsable scientifique

Radu Horaud [directeur de recherche]

Assistante de projet

Véronique Roux

Personnel INRIA

Cordelia Schmid [chargée de recherche]

Peter Sturm [chargé de recherche]

Personnel CNRS

Long Quan [chargé de recherche, en détachement à la « Hong Kong University of Science and Technology » à partir du 1 août 2001]

William Triggs [chargé de recherche]

Personnel universitaire

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

Collaborateur extérieur

Roger Mohr [Xerox European Research Center]

Ingénieurs experts

Matthieu Personnaz [contrat européen VISIRE]

Rémi Ronfard [contrat européen VIBES, à partir du 8 septembre 2001]

Chercheurs doctorants

Marc-André Ameller [allocation couplée, ENS Rennes]

Adrien Bartoli [allocataire MENESR et moniteur]

Guillaume Dewaele [allocation couplée, ENS Lyon à partir du 1er septembre 2001]

Yves Dufournaud [boursier CIFRE, jusqu'au 1er juin 2001]

Fabien Evrard [allocation couplée, ENS Rennes]

Riad Hammoud [boursier INRIA, jusqu'au 1er mars 2001]

Frédéric Martin [allocataire MENESR et moniteur]

Krystian Mikolajczyk [boursier INRIA]

Andreas Ruf [boursier TMR (bourse Marie-Curie), jusqu'au 1er mars 2001]

Cristian Sminchisescu [bourse Eiffel, Egide]

Marta Wilczkowiak [boursière INRIA]

Chercheurs post-doctorants

Ragini Choudhury [contrat PRIAMM (depuis le 1er avril 2000), contrat européen VIBES à partir du 1 avril 2001]

Maxime Lhuillier [contrat européen EVENTS]

Ingénieur invité

Markus Michaelis [Société Plettac-Electronics, Allemagne]

Stagiaires longue durée

Magdalena Urbanek

Navneet Dalal



FIG. 1 – Les membres du projet MOVI. Les visages ont été détectés automatiquement avec la méthode [38].

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 3D 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 ;
- tracking en temps réel avec plusieurs caméras ;
- construction interactive de modèles d’objets et de scènes complexes à partir de plusieurs vues ;
- indexation de larges bases 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 Fondements scientifiques

3.1 Optimisation numérique et apprentissage

Le thème de recherche principal de MOVI est la modélisation du monde réel à partir d’images, y compris les aspects géométriques (reconstruction 3D, calibrage des caméras), statistiques (reconnaissance des formes, de classes d’objets, d’actions) ainsi que photométriques (mise en correspondance, suivi, synthèse d’images à partir d’images). En pratique, la modélisation peut souvent être réduite au choix d’un modèle et l’estimation de ses paramètres à partir de données image (la « reconstruction » d’un modèle géométrique, l’« apprentissage » d’un modèle statistique). Pour cette raison, les *techniques d’estimation de paramètres* et de choix de modèle sont une base importante pour nous, surtout celles qui sont adaptées aux problèmes de grande taille (beaucoup de données et/ou de paramètres) et complexes (caractère géométrique/statistique/photométrique mixte, paramétrisations géométriques de topologie non-triviale, contraintes de domaine et invariances géométriques/photométriques à respecter...). En particulier, les différentes phases de la reconstruction 3D (estimation de contraintes d’appariement, auto-calibrage, ajustement de faisceaux...) aboutissent souvent sur des problèmes d’*optimisation numérique continue* à résoudre, avec ou sans contraintes. Pour l’initialisation des modèles géométriques (initialisation des poses des caméras, auto-calibrage...), et aussi pour leur analyse théorique, nous faisons souvent appel aux techniques de la *géométrie algébrique* et de la *résolution de systèmes de polynômes*. Et bien entendu, la *géométrie projective* sert de base théorique à presque tout notre travail en vision géométrique.

Pour nos travaux en classification nous faisons appel aux techniques de la *reconnaissance des formes* et de l’*analyse statistique exploratoire*, et plus récemment, aux techniques de l’*apprentissage statistique / apprentissage automatique*. L’implantation de ces techniques se base encore une fois sur l’*optimisation numérique / la programmation mathématique*.

3.2 Commande visuelle de robots

La commande de robots avec retour visuel fut une des premières applications de la vision par ordinateur. Le terme « asservissement visuel » (visual servoing, en anglais) a été introduit par Gerald Agin en 1977 [Agi77] décrivant un système composé d'un module de vision et d'un bras manipulateur capable de localiser, saisir et insérer des petites pièces. Aujourd'hui un tel système semble naïf pour un certain nombre de raisons. Les pièces étaient plates, les images codées avec 1 bit par pixel et la position de la caméra était contrainte de façon que son plan image soit aligné avec le plan sur lequel étaient posées les pièces. Même avec ces contraintes, la puissance de calcul disponible il y 25 ans n'était pas suffisante pour traiter les images dans le temps limite imposé par le contrôleur du robot. Ce travail a révélé les ingrédients de base de la commande visuelle : le contrôle du robot, la géométrie de la caméra et le suivi.

Depuis l'approche « look-and-move » qu'on vient de présenter, le paradigme de l'asservissement visuel a continuellement évolué en engendrant un grand nombre de théories, méthodes et applications. Un certain nombre de capteurs furent envisagés pour fermer la boucle (lasers, sonars, capteurs proximétriques, etc.) mais petit à petit les caméras et la vision par ordinateur se sont imposés pour un certain nombre de raisons : la quantité d'information pouvant être extraite d'une image est potentiellement très riche et la capacité de doter un robot d'une fonction visuelle le rapproche des aptitudes humaines. Cependant, même avec la puissance de calcul disponible aujourd'hui, la quantité de calcul « visuel » est limitée par la nature temps-réel de la commande d'un robot. Il faut donc étudier les aspects théoriques et méthodologiques de l'asservissement visuel et décider quels types d'algorithmes peuvent être exécutés à l'intérieur de la boucle de commande.

Parmi les aspects théoriques de l'asservissement visuel (commande, géométrie et tracking) le premier a été le plus étudié et un certain nombre de travaux et publications sont disponibles. Des problèmes tels que la commande dans l'espace-tâche, la latence des système de vision, la dynamique des robots ont reçu beaucoup d'attention. Récemment, un des sujets phare de la vision a été l'étude de la géométrie multi-images et multi-caméras. Il est intéressant de noter que ces travaux n'ont pas influencé, avec quelques notables exceptions, la façon dont les chercheurs conçoivent les boucles de commande visuelle. Le tracking fut également un sujet de recherche intense avec prise en compte de longues séquence qui génèrent inmanquablement des occlusions.

Nous croyons que la géométrie des images multiples et le tracking vont continuer d'être des sujets de recherche dans le domaine de la commande visuelle des robots. Le couplage vision-commande permet également un continuum entre mécanismes réels et virtuels, continuum qu'il reste à étudier et expérimenter.

3.3 Description d'images

Des problèmes aussi variés que l'appariement d'images, la recherche d'images ou alors la définition de classes d'objets nécessitent une description d'image appropriée. Obtenir une telle description repose sur la théorie des invariants, la théorie du signal ainsi que sur des techniques

[Agi77] G. AGIN, « Servoing with visual feedback », in : *Proceedings of the Seventh International Symposium on Industrial Robots, Tokyo, Japan*, octobre 1977.

statistiques de sélection. Une première direction de travail consiste à obtenir une description invariante aux transformations image, comme par exemple des changements d'échelle importants. Ceci est basé sur une étude du comportement du signal, notamment la sélection de l'échelle caractéristique. De façon similaire il est possible d'estimer la transformation affine locale du signal. En ce qui concerne l'invariance aux changements de luminosité, différents modèles peuvent être adoptés et on peut ensuite définir l'invariance par rapport au modèle choisi. Ensuite, la définition d'une hiérarchie de description permet d'obtenir des descriptions plus discriminantes. Des exemples sont l'utilisation de contraintes spatiales de voisinage ou de contraintes statistiques de voisinage. Les contraintes doivent elles-mêmes être invariantes au groupe de transformations considérées et appropriées aux objets choisis. Par exemple des contraintes géométriques de voisinage ne permettent pas de représenter des objets déformables. Enfin, pour obtenir une description appropriée de l'image, il est nécessaire de sélectionner les caractéristiques (descripteurs) et relations de voisinage les plus adaptées en utilisant des techniques d'apprentissage.

3.4 Reconstruction de surfaces

Déterminer une approximation d'une courbe ou d'une surface à partir d'un ensemble de points est un problème qui concerne plusieurs domaines dont la vision par ordinateur, le graphisme et la géométrie algorithmique. Les applications sont multiples, de la reconstruction de courbes en analyse d'images à la construction de modèles tridimensionnels à partir de données issues de processus de vision par ordinateur, de données laser, ou de données médicales. Le problème considéré dans ses applications est le suivant : à partir d'un ensemble discret de points reconstruits d'un objet (une courbe ou une surface), comment construire une approximation de l'objet d'origine qui permettent, en particulier, de visualiser et de manipuler la reconstruction effectuée.

Ce problème a reçu beaucoup d'attention des différentes communautés citées où deux classes principales d'approches se distinguent. Une première classe concerne les approches d'approximation de l'ensemble des points de données où, classiquement, un modèle connu *a priori* est déformé de manière à correspondre aux données. Une deuxième classe concerne les approches d'interpolation où une structure polygonale est construite sur la base des points de données. Les récents résultats de cette deuxième classe portent notamment sur les fondements théoriques de la reconstruction à partir de points échantillons, et mettent en lumière le lien qui existe entre la densité de points qui est nécessaire pour une reconstruction *correcte* et la forme de l'objet.

À partir du problème général de reconstruction mentionné, il est possible de dériver plusieurs sous-problèmes spécifiques à un groupe d'applications. C'est le cas notamment en vision par ordinateur où il s'agit de construire un modèle tridimensionnel à partir de points et de segments reconstruits à l'aide d'images. Les images constituent en effet une source d'informations importante sur l'objet à reconstruire, qui inclut des primitives, telles que les points ou les segments, mais aussi une information photométrique et éventuellement une information sur les normales lorsque l'objet est une surface lisse. La prise en compte de ces informations supplémentaires modifie le problème de reconstruction même si les principes et fondements théoriques évoqués précédemment restent valides.

3.5 Synthèse d'images à partir d'images

Contrairement à la synthèse d'images traditionnelle, utilisant un modèle interactif de maillage, la synthèse d'images à partir d'images crée de nouvelles images directement à partir d'images existantes. Ceci a l'avantage d'être plus économe et de pouvoir créer de nouvelles images photo-réalistes. Dans la communauté de l'infographie, l'approche est souvent basée sur un échantillonnage dense de la scène avec une caméra contrôlée par un robot, et dans la communauté de la vision, des systèmes de stéréo synchronisés ont été proposés pour des objectifs similaires. Notre approche se situe plutôt dans un cadre de vision non calibrée, c'est-à-dire de pouvoir synthétiser de nouvelles images à partir d'images prises à la main. Les problèmes fondamentaux sont les mêmes que ceux de la vision, mais le traitement des parties occultées doit être particulièrement considéré, aussi il exige le développement de nouveaux algorithmes pour le rendu, la synthèse de la nouvelle vue.

3.6 Géométrie

Une grande partie de nos travaux concerne la modélisation géométrique pour la vision par ordinateur. En particulier, nous avons besoin de modèles géométriques pour exprimer les projections effectuées par des caméras, pour représenter des objets à reconstruire, mais aussi pour modéliser des mouvements (rigides ou articulés).

D'un côté, nous utilisons la géométrie pour analyser des problèmes de manière théorique (faisabilité, singularités), surtout des problèmes de calibrage, d'auto-calibrage, de calcul de pose et de reconstruction (souvent en utilisant la géométrie projective). De l'autre côté, la formulation géométrie des problèmes et leurs contreparties algébriques, sont à la base de nos algorithmes numériques.

4 Domaines d'applications

4.1 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.

4.2 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 modé-*

lisation d'un robot dans un espace visuel calibré ou non calibré, le tracking d'objets complexes.

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) ;
- le guidage visuel d'un robot à plusieurs degrés de liberté à l'aide d'une ou plusieurs caméras non étalonnées ; ce problème fut 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 de mécanismes articulés dans un espace non-métrique.

4.3 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.

4.4 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 recons-

truction 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 « drone¹ » de reconnaissance d'une part et un missile opérationnel d'autre part (voir section « contrats industriels »).

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

4.5 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.
- 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.

4.6 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*

¹un engin volant sans pilote

des images photoréalistes. 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.

4.7 Analyse vidéo

Les flots d'images vidéos ont de plus en plus d'ubiquité dans notre « société d'information », et leur traitement digital est un enjeu économique et technologique de premier ordre. MOVI a travaillé dans ce domaine depuis plusieurs années, notamment sur les aspects structuration et indexation des bases de données vidéos. L'année dernière nous avons abordé un nouveau thème – le suivi et la reconstruction du mouvement humain à partir de vidéos monoculaires – que nous avons poursuivi cette année avec des travaux sur la reconstruction des mouvements articulaires (notamment les aspects mise en correspondance robuste modèle-image et initialisation et optimisation du modèle articulaire) [47, 59, 46] et des travaux initiaux sur la reconstruction des mouvements dans les événements sportifs de haut niveau. Cette année nous avons également commencé à travailler sur une approche intégrée de détection et de suivi de visages qui permet de localiser les visages dans une séquence vidéo. Cette intégration permet de rendre plus robuste les résultats de détection. En décembre 2000 nous avons débuté le projet européen VIBES sur l'indexation et la reconstruction « niveau objet » de vidéos.

5 Logiciels

5.1 Trivision

Participants : Radu Horaud [correspondant], Bart Lamiroy, Frédérick Martin.

Mots clés : asservissement, robot.

Logiciel complet pour l'asservissement visuel d'un robot à l'aide d'une paire de caméras stéréoscopiques. Logiciel développé dans le cadre du projet européen VIGOR avec Bart Lamiroy (ingénieur expert) et dans le cadre du doctorat de Frédérick Martin (24 mois).

Logiciel déposé à l'APP sous la référence : IDDN.FR.001.080011.00.R.P.2000.000.00000

5.2 Teledimos

Participants : Radu Horaud [correspondant], Matthieu Personnaz.

Mots clés : calibrage.

Calibrage d'une caméra ainsi que d'un couple stéréoscopique à l'aide d'une mire de calibrage. Il s'agit d'un logiciel complet allant de l'extraction de cibles de calibrage, leur localisation dans l'image avec une précision d'un vingtième de pixel, leur mise en correspondance semi-automatique, l'estimation des paramètres internes et externes, ainsi que la vérification expérimentale de la précision obtenue.

Ce logiciel a été développé avec Matthieu Personnaz dans le cadre d'un contrat d'ingénieur expert. Entièrement en Java il sera bientôt distribué en ligne.

5.3 Modélisation interactive de scènes

Participants : Marta Wilczkowiak [correspondant], Edmond Boyer, Peter Sturm.

Mots clés : calibrage, modélisation, reconstruction.

Un logiciel pour la modélisation interactive de scènes à partir d'une seule image a été développé. Il permet de calibrer la caméra et de reconstruire la scène à partir de contraintes géométriques introduites par l'utilisateur. L'interface graphique est basée sur OpenGL et le logiciel permet d'exporter des modèles texturés en VRML.

6 Résultats nouveaux

6.1 Géométrie multi-images et multi-caméras

Participants : Marc-André Ameller, Adrien Bartoli, Edmond Boyer, Radu Horaud, Maxime Lhuillier, Long Quan, Andreas Ruf, Peter Sturm, William Triggs, Magdalena Urbanek, Marta Wilczkowiak.

Mots clés : géométrie, calibrage, 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.

Modèle « distribution conjointe de primitives » de mise en correspondance multi-images La mise en correspondance des primitives homologues entre images est une opération fondamentale en vision. Dans le cas idéal, les positions des différentes images d'une primitive 3D sont liées par des « contraintes d'appariement » algébro-géométriques strictes, comme la contrainte épipolaire. Mais en pratique il est souvent difficile d'estimer ces contraintes avec précision. Dans ce travail nous avons repensé dès le début le problème de mise en correspondance géométrique, pour donner une approche fondée sur la statistique – la distribution conjointe des positions de primitives homologues entre plusieurs images – plutôt que sur la géométrie explicite. Notre approche donne des mises en correspondance plus précises, plus faciles à interpréter, et plus stables à estimer que l'approche classique [51]. Notamment, elle s'adapte parfaitement aux objets quasi-plans (de basse profondeur), qui sont fréquents en pratique, et difficiles à traiter avec les contraintes d'appariement classiques.

Le suivi et la reconstruction du mouvement humain L'année dernière nous avons débuté une étude sur la reconstruction du mouvement articulaire humain à partir de séquences vidéos monoculaires. Il y a des applications potentielles dans la biométrie, la surveillance, les sports, la production des films. . . Mais le problème est délicat en raison des difficultés de la modélisation humaine et de la mise en correspondance modèle-image, mais aussi parce que chaque image monoculaire laisse environ 33% des articulations (les degrés de liberté fronto-parallèles) quasi-indéterminés. Notre approche est basée sur une modélisation 3D articulaire explicite et une mise en correspondance modèle-image multi-primitives et robuste. Cette année nous avons complété l'implantation de notre méthode de base, et développé des méthodes d'optimisation qui sont mieux adaptées aux exigences de ce domaine, où les minima locaux et le mauvais conditionnement du système sont endémiques [47, 59, 46].

Optimisation de modèles déformables Dans ce travail collaboratif entre MOVI et les équipes de D. Metaxas (Université de Pennsylvania) et de S. Dickinson (Université de Toronto), nous avons étudié comment améliorer le suivi d'un modèle 3D déformable dans plusieurs caméras, en incorporant dynamiquement les droites (par exemple les tangentes aux contours occultants) dans le modèle. La méthode [45, 58] prend en compte l'incertitude des positions des droites 3D par rapport au modèle, et maintient les contraintes de consistance géométriques qui lient les images de chaque droite.

Mosaïques vidéo pour l'analyse du mouvement La capture du mouvement humain est un des thèmes centraux des recherches en vision par ordinateur. Une nouvelle technologie, la vidéogrammétrie, est en train de naître. Il s'agit d'extraire des mouvements euclidiens et de mesurer leurs paramètres à partir d'une séquence vidéo. Une des difficultés réside dans le fait que la caméra effectue elle-même un mouvement pour couvrir un champ de vue suffisamment large. Une technique couramment employée consiste à obtenir une vue panoramique (ou une mosaïque) à partir d'une séquence vidéo. Cette technique est bien maîtrisée lorsque la scène est fixe. En présence d'un mouvement, comme le mouvement d'une personne par exemple, les choses se compliquent. En 2001 nous avons commencé des travaux dans ce domaine en mettant l'accent sur la création de mosaïques comportant des scènes non rigides. Nous avons

mis au point une méthode robuste de détermination du mouvement de la caméra qui nous permet ensuite de segmenter chaque image de la vidéo en « background » et « foreground ». Les résultats préliminaires de ces travaux n'ont pas encore fait l'objet de publication et peuvent être consultés à : <http://www.inrialpes.fr/movi/people/Horaud/videomosaic/videomosaic.html>

Calibrage d'appareils photo et vidéo numériques Actuellement on dispose d'une grande gamme d'appareils photo et vidéo numériques munis généralement d'un zoom. Les techniques de calibrage utilisant une mire peuvent être utilisées à condition de connaître les réglages de l'appareil et de reproduire ces réglages facilement. Dans le cas d'appareils grand public ceci n'est pas le cas. Les seuls réglages connus sont lorsque la mise au point est à l'infini et lorsque la position du zoom est soit au minimum soit au maximum. Par ailleurs, des techniques d'auto-calibrage existent permettant l'estimation des paramètres sans mire. Ces méthodes sont difficiles à mettre en œuvre en pratique pour de multiples raisons. Les travaux que nous avons effectués [52, 60] combinent les avantages respectifs du calibrage avec mire et de l'auto-calibrage. En pratique, une forme simple des équations de Kruppa nous permet de proposer une solution ayant très peu de mouvements critiques.

Auto-calibrage de la distance focale à partir de deux images L'un des scénarios d'auto-calibrage les plus basiques, mais aussi les plus importants en pratiques, concerne l'estimation de la distance focale d'une caméra, à partir de deux images d'une scène inconnue. Des travaux antérieurs ont surtout considéré l'estimation de *deux* valeurs de la distance focale, une pour chacune des images, ce qui rend possible l'utilisation du zoom. Bien qu'en général, ce problème ait une solution, il existe une singularité qui le rend quasiment inapplicable en pratique : quand les axes optiques des deux images se coupent dans l'espace, il n'y a pas de solution (et s'ils se coupent presque, la solution est instable). Ceci nous a motivé pour regarder un problème plus simple, mais qui a plus de chances d'être appliqué en pratique : l'auto-calibrage de la distance focale, en sachant (ou en supposant), qu'elle est la même pour les deux images. Nous avons décrit toutes les singularités pour ce problème, et il se trouve que celle décrite ci-dessus, pour le cas plus général, disparaît en général [49]. Aussi, parmi différentes solutions algébriques directes, nous avons décrit les singularités qui leurs sont propres, pour enfin proposer une solution qui ne souffre que des singularités inhérentes au problème traité.

Mouvements critiques pour l'auto-calibrage de caméras à focale variable Nos premiers travaux sur les mouvements critiques pour l'auto-calibrage considéraient des caméras à focale fixe. Dans [30], nous étudions le cas de focales variables, et pour une seule caméra, et pour des systèmes stéréo. La principale conclusion de ce travail est peut-être que, en dehors des translations pures, il n'y a pas de mouvement critiques pour les systèmes stéréo en pratique.

Modélisation 3D (calibrage et reconstruction) en utilisant des contraintes L'utilisation de contraintes géométriques telles que le parallélisme, l'orthogonalité ou la coplanarité rend les processus de modélisation à partir d'image plus fiables. Une seule image peut, en particulier, suffire pour une reconstruction. Ces contraintes sont fréquemment présentes dans

les environnements humains (les bâtiments par exemple). Nous avons proposé une approche dans ce cadre qui permet de calibrer une image à partir de la projection connue d'un parallélépipède. Nous avons en effet montré que les caractéristiques intrinsèques d'une caméra étaient complètement déterminées par celles d'un parallélépipède lorsqu'une image de ce dernier est disponible, et inversement. L'ajout de contraintes de coplanarité permet alors de construire très facilement des modèles de scènes urbaines à partir d'une seule image [54, 53] (voir des exemples dans la Figure 2). Un logiciel implémentant cette approche a été développé au sein de l'équipe MOVI.



FIG. 2 – Exemples de modèles obtenus à partir d'une seule image (images originales sur la gauche de chaque ligne).

Reconstruction optimale d'une scène plane par morceau à partir de n images L'information de coplanarité apporte une contrainte forte pour la reconstruction 3D 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 n caméras (voir [32, 31] pour le cas

de deux vues et [35, 33] pour le cas de n vues). Celle-ci nous a permis de créer un estimateur au maximum de vraisemblance qui estime simultanément le mouvement entre les images et la structure 3D 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%.

Estimation de mouvement à partir de correspondances de droites 3D La droite est une primitive utilisée pour le suivi 2-D (dans les images) mais peu en revanche en 3D. En effet, représenter algébriquement une droite 3D n'est pas trivial et il n'existe pas de mesure de distance entre deux droites 3D ayant un sens physique et étant universellement reconnue. De plus, transférer une droite entre deux bases de l'espace (e.g. d'une reconstruction à une autre) peut être difficile selon la représentation choisie. Nous avons mis en évidence la *matrice de mouvement* pour droites 3D permettant de transférer aisément une droite d'une base de l'espace à une autre [34]. Ce transfert est linéaire en terme des coordonnées de Plücker de la droite. Nous avons établi le lien entre cette nouvelle représentation du mouvement et la représentation usuelle (pour les points ou les plans : une matrice 4×4). Ceci permet de concevoir des estimateurs de mouvement entre deux reconstructions 3D de droites. L'évaluation de ces algorithmes sur des données simulées et réelles illustre leurs performances, comparables aux estimateurs traditionnels basés sur des points.

Reconstruction de surface à partir de points tridimensionnels Le projet MOVI collabore avec le projet ISA du LORIA sur le problème de la détermination d'une approximation polygonale de surface à partir de points tridimensionnels (voir plus haut). Une approche basée sur un critère local de densité des points par rapport à la forme de la surface a été proposée [28]. À la différence des approches existantes, le critère proposé s'applique à des surfaces discrètes et non continues, et s'avère plus réaliste dans la pratique. En effet, la reconstruction consiste ici à retrouver les connections entre des points de la surface et repose donc sur un modèle discret de la surface : les points et leurs connections, et non sur un modèle continu. Les résultats obtenus viennent confirmer ce fait.

Topologie des enveloppes visuelles À la suite d'un séjour à l'université de l'Illinois à Urbana Champaign (USA), une collaboration a été mise en place sur le problème de la reconstruction de surface d'objets courbes à partir de projections images de l'objet. Cette collaboration a débouché sur des résultats théoriques nouveaux sur l'enveloppe visuelle. Cette enveloppe constitue l'approximation de la surface de l'objet qui peut-être reconstruite à partir des contours de l'objet dans les images. Les résultats obtenus portent sur la topologie de l'enveloppe visuelle et permettent, en pratique, de déterminer cette topologie à partir des contours images [36]. Cette description permet ensuite de construire des modèles de l'enveloppe visuelle à tous niveaux de précision.

Une analyse des relations entre la mosaïque co-centrique, le mouvement planaire et la caméra 1D La méthode générale de structure à partir de mouvement donne des résultats plutôt médiocres pour les images acquises par les mouvements contraints tel que le mouvement

planaire de mosaïque co-centrique. Nous avons proposé une nouvelle méthode de structure à partir de mouvement pour les images acquises par la mosaïque co-centrique et l'image composite mosaïque à partir de la mosaïque co-centrique [41]. Nous introduisons d'abord un modèle de caméra 1D affine pour compléter les modèles de caméra. Ensuite, nous démontrons que les images 2D de mosaïque co-centrique peuvent être décomposées en deux images de dimension un : une projective de dimension un et une affine de dimension un. Une image mosaïque composite peut être ré-échantillonnée en une image 1D projective et panoramique. Finalement, nous donnons une méthode de reconstruction de sous-espaces pour démontrer l'avantage théorique et pratique de la méthode de décomposition qui intègre le mouvement contraint le plus tôt possible dans l'analyse de mouvement.

Mosaïque de relief par la triangulation conjointe Nous avons proposé une approche sur la mosaïque de relief d'images par la triangulation conjointe [37]. La mosaïque de relief est une collection d'images enregistrées qui étend la mosaïque traditionnelle en supportant la parallaxe du mouvement. Une méthode simple d'interpolation de parallaxe permet une mosaïque de haute qualité et sans « blur » en utilisant les images acquises à la main, qui auraient faire échouer les méthodes traditionnelles de création de mosaïque. L'algorithme de rendu permet aussi d'afficher les changements des parallaxes locaux, tout en donnant une illusion de profondeur. Aussi cette méthode permet d'approximer la modélisation plénoptique à partir d'une caméra à la main avec un taux d'échantillonnage beaucoup plus bas que la méthode de *lightfield*.

6.2 Couplage vision/robotique

Participants : Radu Horaud, Frédérick Martin, Andreas Ruf, Bernard Espiau.

Mots clés : asservissement visuel, 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. Nos travaux, menés en collaboration avec le projet BIP (projet européen VIGOR qui a pris fin en 2001) 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 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. figure 3).

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 3D dans le processus d'auto-étalonnage et nous avons développé une méthode qui accepte de faibles déplacements [25]. 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 [55].

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.

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 [23].

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 [57]. La finalité pratique fut le démonstrateur final du projet VIGOR, figure 4.

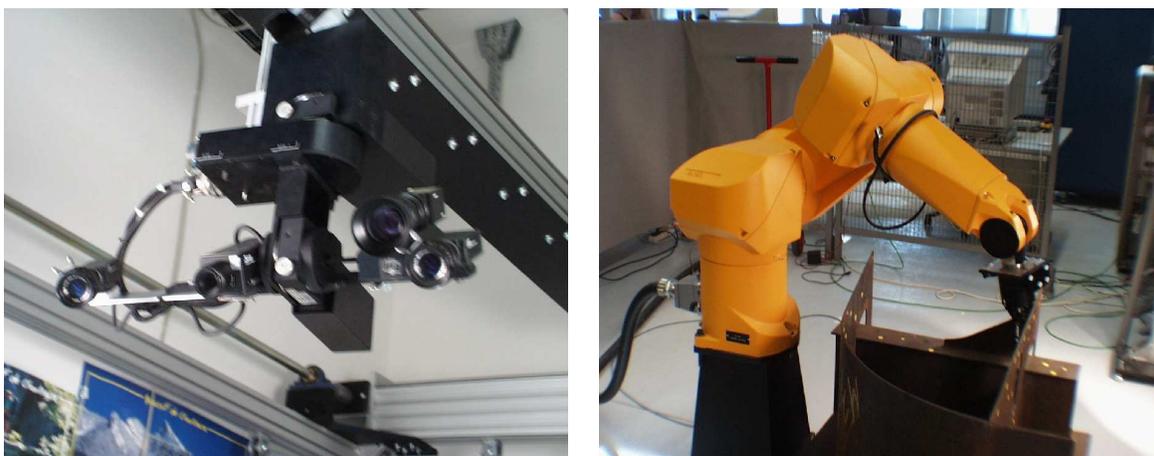


FIG. 3 – La plate-forme expérimentale du couplage vision-robotique comportant une tête stéréo montée sur un robot portique à 5 degrés de liberté (gauche) et un robot manipulateur à 6 degrés de liberté (droite). La pièce métallique visible dans l'image de droite est une pièce de bateau fabriquée par les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

6.3 Indexation d'images et reconnaissance d'objets

Participants : Yves Dufournaud, Riad Hammoud, Radu Horaud, Krystian Mikolajczyk, Roger Mohr, Rémi Ronfard, 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 partie 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 d'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. 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

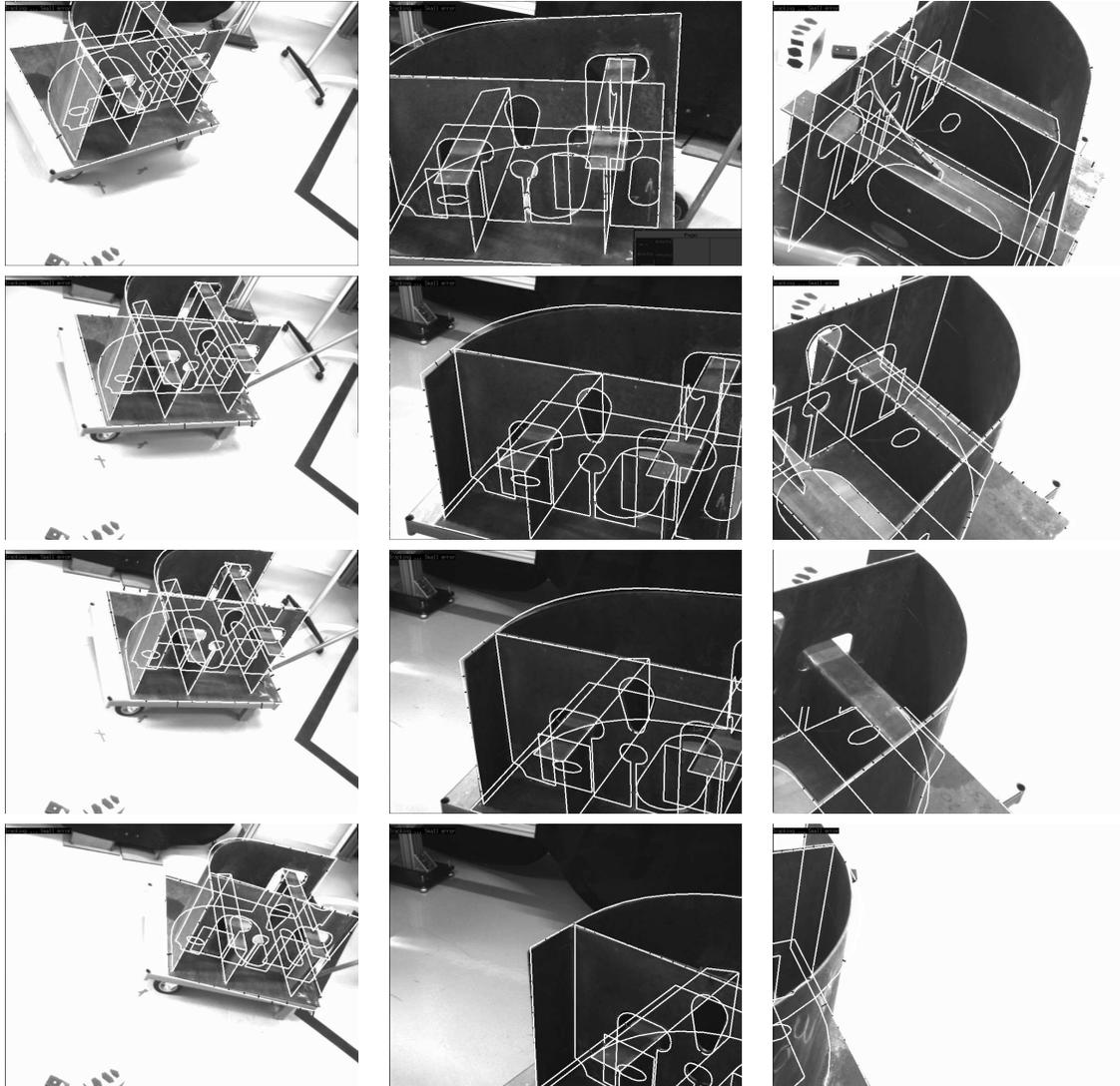


FIG. 4 – Une séquence de tracking avec trois caméras. La première caméra (colonne gauche) a un large champ de vue ce qui permet de maintenir une information continue concernant la position de l'objet. Les deux autres caméras peuvent ainsi tolérer des occlusions partielles, voire totales.

calculés les descripteurs. L'utilisation de contraintes géométriques entre images et l'estimation robuste de ces contraintes permet un rejet important des faux appariements [21]. 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.

Le problème d'une indexation invariante 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. Ceci permet d'obtenir des points d'intérêt invariants à un changement d'échelle [40, 39]. De très bons résultats ont montré la possibilité de retrouver une image dans une base contenant 5000 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. 5 – Recherche d'images en présence de rotations et de changements d'échelle jusqu'à un facteur 4.9. La base d'images contient plus de 5000 images. Sur la gauche, l'image de requête, et sur la droite, quelques images de la base. L'image correspondant à l'image de requête, a été correctement retrouvée (image au milieu de la deuxième ligne).

Définition et détection 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, nous avons commencé par développer un détecteur de visage. 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. Toutefois

une telle approche nécessite la sélection manuelle de caractéristiques.

Pour éviter une telle sélection manuelle nous avons développé une approche non-supervisée [44, 42, 43] qui permet de construire des modèles à partir d'un ensemble d'images positives et négatives. Notre modèle repose sur une représentation originale à deux niveaux. Le premier niveau est constitué de descripteurs « génériques » qui représentent des ensembles de vecteurs de caractéristiques invariantes en rotation. Le second utilise la probabilité jointe des fréquences des descripteurs « génériques » calculées sur un voisinage. Cette représentation permet de capturer efficacement les structures visuelles de type texture ; son invariance à la rotation rend la méthode robuste aux déformations image. La sélection des structures pertinentes permet de déterminer les éléments caractéristiques, et ainsi d'accroître les performances. Les modèles, une fois appris, peuvent être reconnus et localisés en utilisant un score probabiliste. Les résultats expérimentaux sur des animaux « texturés » et sur des visages montrent de très bonnes performances.

6.4 Structuration de vidéos

Participants : Ragini Choudhury, 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.*

Détecteur temporel de visages Actuellement, il est impossible d'obtenir 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.

Afin de pouvoir identifier les objets à travers une séquence, il faut tout d'abord les détecter et les localiser. Dans ce but nous avons développé un détecteur temporel de visages qui permet de localiser des visages dans une séquence vidéo [38]. Cette détection temporelle repose sur les trois principes suivants : 1) Accumulation des probabilités de détection dans la séquence. Ceci permet d'obtenir une détection cohérente au cours du temps et une indépendance du seuil de classification. 2) La prédiction des paramètres de détection, dont la position, l'échelle (taille) et la posture d'un visage. La prédiction garantit l'exactitude de l'accumulation ainsi que la continuité de détection au cours du temps. 3) La représentation de la pose. Cette représentation est basée sur la combinaison de deux détecteurs, un pour la vue de face et un pour la vue de

profil. Les résultats expérimentaux montrent une amélioration importante par rapport à la détection simple « image par image ».

6.5 Autres

Interpolation sous-pixélique optimale Pour l’affichage de texture, la mise en correspondance, et bien d’autres applications, il est souvent nécessaire de re-échantillonner les images à des positions sous-pixéliques. À cette fin, on dispose d’un grand choix de « filtres d’interpolation ». Le filtre « sinc » de largeur infinie est souvent considéré l’idéal théorique, mais en pratique on préfère des filtres bilinéaires et bi-cubiques (dont plusieurs possibles, chacun avec ses partisans...). Dans ce travail, nous avons étudié ce problème de façon empirique : optimiser numériquement le filtre pour minimiser une fonction donnée d’erreur d’interpolation sur un ensemble donné d’images d’apprentissage. Les conclusions [50] : pour minimiser l’erreur moindres carrés ou robuste sur des images réalistes, le « sinc » est sous-optimal (trop d’effets de résonance) et aussi les filtres bi-cubiques et les fonctions de « fenêtrage » du sinc les plus répandues (trop de lissage). Les filtres optimaux sont intermédiaires. Mais la chose la plus importante est d’éviter d’introduire trop d’aliasage dans l’image d’entrée, par exemple en utilisant une mise au point très légèrement floue.

7 Contrats industriels (nationaux, européens et internationaux)

7.1 Reconnaissance et indexation de vues aériennes

Participants : Radu Horaud, Cordelia Schmid, Yves Dufournaud.

Nous collaborons avec la Société Aérospatiale depuis 1993. De 1998 à 2001 nous avons eu un contrat de collaboration dans le domaine de l’interprétation d’images aériennes.

Yves Dufournaud a effectué une thèse de doctorat dans le cadre d’une convention CIFRE [21]. L’objectif de ce travail fut 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 ont donné lieu à un transfert technologique puisque le doctorant a passé 2/3 de son temps à l’INRIA et 1/3 de son temps à l’Aérospatiale. Ceci a permis de valider nombre de nos méthodes d’appariement et de tracking.

7.2 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 janvier 2000. Il regroupe l’INA, une PME (Arts Video Interactive), et des équipes de recherches (IRISA, INRIA Rhône-Alpes, IRIT, CLIPS-IMAG, INT, LIP6). Ce projet a été financé pour 18 mois.

Le but du projet est de développer un système complet permettant d’indexer des vidéos, éventuellement accompagnées 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 concernent la fourniture de descripteurs d'images. Ces descripteurs, calculés à partir de diverses primitives (points, régions, visages) permettent l'identification d'objets ou de scènes. Un effort spécifique est fourni pour adapter les techniques utilisées au cas de la vidéo où la redondance des images permet une robustification des descripteurs. Cela a été fait en reprenant les travaux de R. Hammoud sur le calcul de modèles de descripteurs par mélange de gaussiennes. Il a été aussi étudié le problème de la composition spatiale des descripteurs élémentaires pour obtenir un langage de requête plus riche.

8 Actions régionales, nationales et internationales

8.1 Actions nationales

- L. Quan participe à l'animation du GT5 dans le GDR ISIS.
- Dans le cadre du programme PNRH (Programme National de Recherche en Hydrologie), MOVI collabore avec l'INRA (Centre de Recherches d'Avignon – Science du Sol). Thème du projet : suivi dynamique de la cartographie de la détention hydrique à la surface d'un sol lors du déclenchement du ruissellement à l'aide d'une méthode photogrammétrique.
- Le projet MOVI collabore avec le projet ISA du LORIA sur le problème de la détermination d'une approximation polygonale de surface à partir de points tridimensionnels.
- L'équipe MOVI collabore avec le projet Imagis sur le projet ACI jeune chercheur CYBER. Ce projet initié par Imagis à l'automne 2001 a pour objectif l'incrustation, en temps réel, d'une personne dans un décor virtuel. Les applications visées concernent, en particulier, les émissions télévisuelles, les jeux vidéos et les parcs à thèmes. La difficulté est d'assurer une cohérence maximale entre la personne et le décor virtuel (éclairage, ombre au sol, *etc.*). Les problèmes à résoudre pour cela sont de différents ordres : connaître la position des caméras impliquées en temps réel, extraire le personnage du décor dans les images, avoir une idée de la géométrie du personnage pour effectuer des calculs d'ombres ou de ré-éclairage ou fusionner les données réelles et virtuelles. Leurs résolutions nécessitent la mise en œuvre de différentes techniques dont plusieurs de vision par ordinateur : le calibrage et la reconstruction par exemple. Le projet MOVI interviendra sur ces aspects. Un ingénieur expert est en cours de recrutement sur ce projet et plusieurs stagiaires de DEA devraient y participer.

8.2 Actions financées par la Commission Européenne

VIGOR. Nous avons été coordinateur du projet européen Reactive LTR VIGOR. (<http://www.inrialpes.fr/VIGOR/>). Ce projet auquel a participé également le projet BIP s'est exécuté 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, 1998-2001, ce projet a eu 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.

VISIRE. Le projet IST-1999-10756 « VISIRE, Virtual Image-Processing System for Intelligent Reconstruction of 3D Environments » a débuté en mai 2000. Son objectif est d'exploiter la limite de la technologie de la vision par ordinateur 3D pour pouvoir obtenir des reconstructions 3D de scènes à partir de séquences vidéo acquises avec une caméra grand public. Ce projet de 3 ans s'exécute en liaison avec nos partenaires académiques de l'Université de Lund (Suède) et de l'Université Polytechnique de Madrid (Espagne) ainsi que des deux partenaires industriels, Eptron S.A. (Espagne) et Giunti Multimedia (Italie).

EVENTS. Nous participons depuis le mois d'octobre 2000 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 temps réel de l'interpolation d'images innovant pour la transmission par télévision des grands scénarios en 3D. Ce projet s'exécute en liaison avec nos partenaires de l'Université d'Oxford (Angleterre), de Siemens IC C-Lab (Allemagne), d'Eptron S.A. (Espagne) et de Via Digital (Espagne). Nous travaillons essentiellement sur les méthodes et le prototype d'interpolation de deux images statiques en vue de la transmission de toutes les vues intermédiaires.

VIBES. Le projet 5e cadre FET-Open IST-2000-26001 « VIBES, Video Browsing, Exploration and Structuring » a débuté en décembre 2000. Ce projet de 3 ans a pour but de développer des représentations et des techniques de manipulation haut-niveau (niveau objets / personnages) de la vidéo, en particulier 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), la Katholieke Universiteit Leuven (Belge), l'Ecole Polytechnique Fédérale de Lausanne (Suisse), et le Weizmann Institute of Science (Israël). VIBES représentera 32 personne-ans d'effort, pour un budget total de 2,3 MEu (15 MF), avec un support Européen de 1,7 MEu (11 MF). MOVI travaillera en particulier sur les aspects indexation et suivi et reconstruction humaine.

8.3 Relations bilatérales internationales

8.3.1 Europe

PAI ALLIANCE. Dans le cadre du programme PAI (Programme d'Actions Intégrées) Alliance, nous avons un projet de collaboration (sigle : 03066ZC) avec l'université de Kingston upon Thames, Royaume-Uni, qui porte sur l'interprétation automatique d'événements sportifs à partir de séquences vidéo. P. Sturm a effectué deux séjours d'une semaine à Kingston, et P. Remagnino et E. Corvée (de Kingston) ont visité MOVI pendant une semaine chacun.

8.3.2 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. E. Boyer et C. Schmid ont visité l'Université d'Illinois pendant deux semaines chacun.

K. Mikolajczyk a visité l'Université de Vancouver (Canada) pendant un mois.

C. Sminchisescu a poursuivi sa collaboration avec D. Metaxas (Université de Pennsylvania) et de S. Dickinson (Université de Toronto) sur l'optimisation des modèles déformables [45, 58].

8.3.3 Asie

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.

PRA. Dans le cadre du PRA (Programme de Recherches Avancées Franco-chinois), nous avons un projet de collaboration (sigle : SI00-04) avec l'université Xidian de Xi'an (professeur Wu Chengke), qui porte sur la reconstruction 3D et la visualisation virtuelle à partir de séquences d'images non-calibrées. Ce projet est la continuation d'une coopération datant de plusieurs années. Le professeur Wu Chengke séjournera à Montbonnot pour une durée de deux mois, fin 2001 ou début 2002.

LIAMA. 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.

9 Diffusion de résultats

9.1 Animation de la communauté scientifique

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

- *International Journal of Robotics Research* (R. Horaud est membre de l'*editorial board*)
- *Computer Vision and Image Understanding* (R. Horaud est *area editor*)
- *IEEE Transactions on Pattern Analysis and Machine Intelligence* (C. Schmid et L. Quan sont *associated editors*)
- *Machine Vision and Applications* (R. Mohr)
- *Computacion y Sistemas* (R. Horaud)

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

- ICCV'01 (R. Horaud, program co-chair, L. Quan, W. Triggs)
- CVPR'01 (L. Quan, C. Schmid, P. Sturm, W. Triggs)
- ICRA'01 (R. Horaud)
- IAPR Workshop Robot Vision 2001 (W. Triggs)

- ORASIS'01 – Journées Francophones des Jeunes Chercheurs en Analyse d'Images et Perception Visuelle (C. Schmid).
- 2nd European Workshop on Advanced Video-based Surveillance Systems 2001 (C. Schmid)

Autres : P. Sturm fait partie du comité du Prix de Thèse SPECIF 2001.

9.2 Enseignement universitaire

- Géométrie projective, DEA IVR, INPG, 6h, L. Quan.
- Optimisation, DEA IVR, INPG, 6h, P. Sturm.
- Vision 3D, DEA IVR, INPG, 12h, P. Sturm.
- Mise en correspondance et reconnaissance, DEA IVR, INPG, 12h, C. Schmid.
- Introduction à la vision par ordinateur, MAGISTÈRE INFORMATIQUE, UJF, 15h, R. Horaud.

9.3 Participation à des colloques, séminaires, invitations

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

- P. Sturm a fait une présentation invitée aux Journées Nationales de la Recherche en Robotique à Hyères (octobre 2001) [48].
- P. Sturm a prononcé des séminaires à l'ETH de Zürich (avril 2001) et au LASMEA de Clermont-Ferrand (mai 2001).
- C. Schmid a fait une présentation invitée aux Journées de Travail du GDR-GT10 « Indexation Multimedia », intitulée « recherche d'images en présence de changements d'échelle importants et construction automatique de modèles visuels pour la recherche d'images » (mai 2001).
- R. Horaud a fait des présentations invitée à Microsoft Research (mars) et University of Victoria (juillet).

10 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>.
- [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] Y. DUFOURNAUD, *Navigation aérienne et guidage terminal à partir de données bidimensionnelles*, Thèse de doctorat, Institut National Polytechnique de Grenoble, juin 2001.

- [22] R. HAMMOUD, *Construction et Présentation des Vidéos Interactives*, Thèse de doctorat, Institut National Polytechnique de Grenoble, mars 2001.
- [23] A. RUF, *Closing the loop between articulated motion and stereo vision : a projective approach*, Thèse de doctorat, Institut National Polytechnique de Grenoble, 2001, <http://www.inrialpes.fr/movi/people/Ruf/thesisRUF.ps.gz>.
- [24] C. SCHMID, *De l'appariement d'images à l'apprentissage de modèles visuels*, Habilitation à diriger les recherches, Institut National Polytechnique de Grenoble, 2001.

Articles et chapitres de livre

- [25] N. ANDREFF, R. HORAUD, B. ESPIAU, « Robot Hand-Eye Calibration Using Structure from Motion », *International Journal of Robotics Research* 20, 3, mars 2001, p. 228–248, http://www.inrialpes.fr/bip/people/andreff/docs/ijrr_hand_eye.ps.gz.
- [26] E. BOYER, P. STURM, « Modélisation à partir d'images », in : *Visualisation, F. Sillion (éditeur), Information Géographique et Aménagement du Territoire*, Éditions Hermès, 2001, à paraître.
- [27] A. HEYDEN, F. KAHL, L. QUAN, « Minimal Projective Reconstruction Including Missing Data », *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 4, 2001, p. 418–423.
- [28] S. PETITJEAN, E. BOYER, « Regular and Non-Regular Point Sets : Properties and Reconstruction », *Computational Geometry – Theory and Application* 19, 2-3, 2001, p. 101–126.
- [29] L. QUAN, « Two-Way Ambiguity in 2D Projective Reconstruction from Three Uncalibrated 1D Images », *IEEE Transactions on Pattern Analysis and Machine Intelligence* 23, 2, 2001, p. 212–216.
- [30] P. STURM, « Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length », *Image and Vision Computing*, 2002, à paraître.

Communications à des congrès, colloques, etc.

- [31] A. BARTOLI, P. STURM, R. HORAUD, « Projective Structure and Motion from Two Views of a Piecewise Planar Scene », in : *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, 1*, p. 593–598, juillet 2001, ftp://ftp.inrialpes.fr/pub/movi/publications/Bartoli_Sturm_Horaud_ICCV01.ps.gz.
- [32] A. BARTOLI, P. STURM, R. HORAUD, « Structure and Motion from Two Uncalibrated Views Using Points on Planes », in : *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling, 3DIM'01, Québec City, Canada*, p. 83–90, mai 2001, ftp://ftp.inrialpes.fr/pub/movi/publications/Bartoli_Sturm_Horaud_3DIM01.ps.gz.
- [33] A. BARTOLI, P. STURM, R. HORAUD, « Triangulation projective contrainte par multicoplanarité », in : *Actes des Journées ORASIS 2001, Cahors, France*, p. 47–56, juin 2001, ftp://ftp.inrialpes.fr/pub/movi/publications/Bartoli_Sturm_ORASIS01.ps.gz.
- [34] A. BARTOLI, P. STURM, « The 3D Line Motion Matrix and the Alignment of Line Reconstructions », in : *Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, IEEE Computer Society Press, décembre 2001.
- [35] A. BARTOLI, P. STURM, « Constrained Structure and Motion from N Views of a Piecewise Planar Scene », in : *Proceedings of the First International Symposium on Virtual and Augmented Architecture, VAA'01, Dublin, Ireland*, p. 195–206, juin 2001, ftp://ftp.inrialpes.fr/pub/movi/publications/Bartoli_Sturm_VAA01.ps.gz.
- [36] S. LAZEBNIK, E. BOYER, J. PONCE, « On How to Compute Exact Visual Hulls of Object Bounded by Smooth Surfaces », in : *Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, IEEE Computer Society Press, décembre 2001.

- [37] M. LHUILLIER, L. QUAN, H. SHUM, H. TSUI, « Relief Mosaic by Joint View Triangulation », *in : Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, décembre 2001.
- [38] K. MIKOLAJCZYK, R. CHOUDHURY, C. SCHMID, « Face detection in a video sequence - a temporal approach », *in : Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, décembre 2001.
- [39] K. MIKOLAJCZYK, C. SCHMID, « Indexation à l'aide de points d'intérêt invariants à l'échelle », *in : Actes des Journées ORASIS 2001, Cahors, France*, p. 77–86, mai 2001.
- [40] K. MIKOLAJCZYK, C. SCHMID, « Indexing based on scale invariant interest points », *in : Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, p. 525–531, 2001.
- [41] L. QUAN, L. LU, H. SHUM, M. LHUILLIER, « Concentric Mosaics, Planar Motions and 1D Cameras », *in : Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, II*, p. 193–200, juillet 2001.
- [42] C. SCHMID, K. MIKOLAJCZYK, « Image retrieval in the presence of important scale changes and with automatically constructed models », *in : Multimedia Content-based Indexing and Retrieval Workshop, 2001*.
- [43] C. SCHMID, « Automatic construction of visual models », *in : The Learning Workshop, Snowbird, Utah, 2001*.
- [44] C. SCHMID, « Constructing models for content-based image retrieval », *in : Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA, 2001*.
- [45] C. SMINCHISESCU, D. METAXAS, S. DICKINSON, « Improving the Scope of Deformable Model Shape and Motion Estimation », *in : Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, IEEE Computer Society Press, décembre 2001.
- [46] C. SMINCHISESCU, A. TELEA, « A Framework for Generic State Estimation in Computer Vision Applications », *in : Proceedings of the Second International Workshop ICVS 2001, Vancouver, Canada, Lecture Notes in Computer Science, 2095*, Springer-Verlag, p. 21–34, juillet 2001.
- [47] C. SMINCHISESCU, B. TRIGGS, « Covariance Scaled Sampling for Monocular 3D Body Tracking », *in : Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, IEEE Computer Society Press, décembre 2001.
- [48] P. STURM, « Auto-calibrage de caméras et reconstruction 3D à partir d'images », *in : Actes des Journées Nationales de la Recherche en Robotique, Hyères, France*, octobre 2001.
- [49] P. STURM, « On Focal Length Calibration from Two Views », *in : Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, IEEE Computer Society Press, décembre 2001, <ftp://ftp.inrialpes.fr/pub/movi/publications/Sturm-cvpr01.ps.gz>.
- [50] B. TRIGGS, « Empirical Filter Estimation for Subpixel Interpolation and Matching », *in : Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, II*, p. 550–557, 2001.
- [51] B. TRIGGS, « Joint Feature Distributions for Image Correspondence », *in : Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, II*, p. 201–208, 2001.
- [52] M. URBANEK, R. HORAUD, P. STURM, « Combining Off- and On-line Calibration of a Digital Camera », *in : Proceedings of the Third International Conference on 3D Digital Imaging and Modeling, Quebec City, Canada*, p. 99–106, mai 2001, <ftp://ftp.inrialpes.fr/pub/movi/publications/UrbanekHoraudSturm-3dim01.ps.gz>.

- [53] M. WILCZKOWIAK, E. BOYER, P. STURM, « Calibrage de caméra et reconstruction 3D à l'aide de parallélépipèdes », *in : Actes des Journées ORASIS 2001, Cahors, France*, p. 117–126, juin 2001.
- [54] M. WILCZKOWIAK, E. BOYER, P. STURM, « Camera Calibration and 3D Reconstruction from Single Images Using Parallelepipeds », *in : Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada, 1*, IEEE Computer Society Press, p. 142–148, juillet 2001, <ftp://ftp.inrialpes.fr/pub/movi/publications/WilczkowiakBoyerSturm-iccv01.ps.gz>.

Rapports de recherche et publications internes

- [55] N. ANDREFF, B. ESPIAU, R. HORAUD, « Visual Servoing from Lines », *rapport de recherche n°RR-4226*, INRIA, juillet 2001, <http://www.inria.fr/rrrt/rr-4226.html>.
- [56] A. BARTOLI, P. STURM, « Three New Algorithms for Projective Bundle Adjustment with Minimum Parameters », *rapport de recherche n°RR-4236*, INRIA, août 2001, <http://www.inria.fr/rrrt/rr-4236.html>.
- [57] F. MARTIN, R. HORAUD, « Multiple-camera Tracking of Rigid Objects », *rapport de recherche n°RR-4268*, INRIA, septembre 2001, <http://www.inria.fr/rrrt/rr-4268.html>.
- [58] C. SMINCHISESCU, D. METAXAS, S. DICKINSON, « Incremental Model-Based Estimation Using Geometric Consistency Constraints », *rapport de recherche n°RR-4209*, INRIA, juin 2001, <http://www.inria.fr/rrrt/rr-4209.html>.
- [59] C. SMINCHISESCU, B. TRIGGS, « A Robust Multiple Hypothesis Approach to Monocular Human Motion Tracking », *rapport de recherche n°RR-4208*, INRIA, juin 2001, <http://www.inria.fr/rrrt/rr-4208.html>.
- [60] M. URBANEK, R. HORAUD, P. STURM, « Calibration of Digital Amateur Cameras », *rapport de recherche n°RR-4214*, INRIA, juillet 2001, <http://www.inria.fr/rrrt/rr-4214.html>.