INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Projet movi

Modélisation, localisation, identification et reconnaissance pour la vision par ordinateur

Rhône-Alpes

THÈME 3B

Table des matières

1. Composition de l'équipe **1**
1.1. Nouveautés 2002 1
2. Présentation et objectifs généraux **2**
3. Fondements scientifiques **2**
3.1. Optimisation numérique et apprentissage 2
3.2. Commande visuelle de robots 3
3.3. Description d'images 3
3.4. Reconstruction de surfaces 4
3.5. Apprentissage statistique 4
3.6. Reconnaissance structurelle 4
3.7. Analyse de gestes humains 5
3.8. Synthèse d'images à partir d'images 5
3.9. Géométrie 5
4. Domaines d'application **6**
4.1. Panorama 6
4.2. Vision, robots et leur couplage 6
4.3. Indexation de bases d'images 6
4.4. Indexation de films de cinéma 7
4.4.1. Indexation acteur-lieu-action 7
4.4.2. Alignement film et scénario 7
4.5. Reconnaissance des gestes des mains 7
4.6. Reconnaissance de gestes athlétiques 9
4.7. Virtualisation d'objets 9
4.8. Synthèse d'images à partir d'images 10
4.9. Analyse vidéo 10
5. Logiciels **10**
5.1. Tele2 10
5.2. Modélisation interactive de scènes 10
6. Résultats nouveaux **11**
6.1. Géométrie multi-images et multi-caméras 11
6.1.1. Le suivi et la reconstruction du mouvement humain. 11
6.1.2. Applications de la géométrie algébrique effective à la vision par ordinateur. 11
6.1.3. Mosaïques vidéo pour l'analyse du mouvement 11
6.1.4. Structure et mouvement pour des scènes dynamiques 12
6.1.5. Combiner des caméras omnidirectionnelles avec des caméras perspectives 13
6.1.6. Ajustement de faisceaux 13
6.1.7. Reconstruction de surfaces réfléchissantes 13
6.1.8. Modélisation 3D (calibrage et reconstruction) en utilisant des contraintes 13
6.1.9. Estimation de mouvement à partir de correspondances de droites 3D 15
6.1.10. Acquisition de modèles en temps réel 15
6.2. Couplage vision/robotique 15
6.2.1. Suivi d'objets en mouvement avec plusieurs caméras 15
6.2.2. Asservissement visuel avec plusieurs caméras 15
6.2.3. Asservissement visuel avec une caméra catadioptrique 17
6.3. Indexation d'images et reconnaissance d'objets 17
6.3.1. Appariement d'images prises de points de vue différents 17
6.3.2. Détection des humains 17

6.3.3. Définition et détection de classes d'objets 20
6.4. Structuration de vidéos 20
6.4.1. Mouvements de caméra 20
6.4.2. Reconnaissance des visages 20
6.4.3. Mouvements déformables 21
7. Contrats industriels 21
7.1. Surveillance vidéo 21
8. Actions régionales, nationales et internationales 21
8.1. Actions nationales 21
8.2. Actions financées par la Commission Européenne 22
8.2.1. Visire. 22
8.2.2. Events. 22
8.2.3. Vibes. 22
8.2.4. Lava. 22
8.3. Relations bilatérales internationales 22
8.3.1. Europe 22
8.3.1.1. Pai Alliance. 22
8.3.2. Amérique 23
8.3.2.1. CNRS/UIUC. 23
8.3.2.2. NSF/INRIA. 23
8.3.3. Asie 23
8.3.3.1. Pra. 23
8.3.4. Océanie 23
9. Diffusion des résultats 23
9.1. Animation de la communauté scientifique 23
9.1.1. Les membres du projet font partie des comités de rédaction de revues suivantes : 23
9.1.2. Les membres du projet font partie des comités de programme des conférences suivantes : 23
9.1.3. Autres : 24
9.2. Enseignement universitaire 24
9.3. Participation à des colloques, séminaires, invitations 24
10. Bibliographie 24

1. Composition de l'équipe

Responsable scientifique
Radu Horaud [directeur de recherche]

Assistante de projet
Véronique Roux

Personnel INRIA
Frédéric Devernay [chargé de recherche]
Rémi Ronfard [chargé de recherche]
Cordelia Schmid [chargée de recherche]
Peter Sturm [chargé de recherche]

Personnel CNRS
William Triggs [chargé de recherche]

Personnel universitaire
Edmond Boyer [maître de conférences à l'université Joseph Fourier]
Roger Mohr [professeur à l'Institut National Polytechnique de Grenoble]

Ingénieur expert
Matthieu Personnaz [contrat européen VISIRE]

Chercheurs doctorants
Marc-André Ameller [allocation couplée, ENS Rennes, jusqu'au 1 août]
Ouideh Bentrah [bourse IGN, à partir du 1 octobre]
Adrien Bartoli [allocataire MENESR et moniteur]
Thomas Bonfort [boursier INRIA, à partir du 1 octobre]
Guillaume Dewaele [allocation couplée, ENS Lyon]
Gyorgy Dorko [boursier INRIA, à partir du 1 octobre]
Jean-Sébastien Franco [allocataire MENESR, à partir du 1 octobre]
Frédérick Martin [allocataire MENESR et moniteur, jusqu'au 1 octobre]
Krystian Mikolajczyk [boursier INRIA]
Cristian Sminchisescu [bourse Eiffel, Egide]
Marta Wilczkowiak [boursière INRIA]

Ingénieur invité
Markus Michaelis [Société Plettac-Electronics, Allemagne]

Stagiaires longue durée
Ankur Agarwal [À partir du 1 novembre]
Joao Barreto [Du 1 janvier au 30 juin]
Navneet Dalal

Professeurs invités
Richard Hartley [professeur au National Australian University]
Andrew Zisserman [professeur a Oxford University]

1.1. Nouveautés 2002

Cette année il y a eu d'importants changements en termes de ;a composition de l'équipe.

Rémi Ronfard, qui était ingénieur expert sur le projet VIBES depuis septembre 2001, a été recruté comme CR1 en septembre 2002. Son activité se développe autour de trois thèmes - analyse de films de cinéma, apprentissage statistique et détection des personnes, et plus récemment l'analyse des mouvements déformables. Ces activités trouvent leur champs d'application dans le projet européen VIBES.

Frédéric Devernay, qui était CR2 dans le projet CHIR de Sophia Antipolis nous a rejoint à partir du 1 juillet 2002. L'activité scientifique de Frédéric Devernay va s'articuler autour des thèmes suivants : acquisition de

flux vidéo synchronisés sur grappe de PC, reconstruction 3D à partir de plus de deux caméras, suivi d'éléments de surface texturés sur de multiples caméras, suivi des mouvements de la main.

Roger Mohr, qui était en délégation chez Xerox pour une période de 3 ans nous a rejoint au 1 septembre 2002. Roger Mohr est professeur à l'ENSIMAG-INPG et vient d'être nommé directeur du laboratoire GRAVIR.

En 2002 le projet a accueilli deux professeurs invités, Richard Hartley, Australie et Andrew Zisserman, Grande Bretagne qui ont fait un séjour du 1 juin au 31 juillet.

Le projet accueille également deux étudiants indiens, Navneet Dalal et Ankur Agarwal qui sont étudiants DEA en vue d'études doctorales au sein du projet.

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 ». Dans ce cadre on s'intéresse à la perception de gestes humains, de reconnaissance de visages, etc. 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 une ou 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,
- apprentissage visuel.

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 le tracking de gestes humains dans une séquence d'images, la synthèse d'images à partir d'images, l'apprentissage de modèles à partir d'exemples visuels.

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 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 immanquablement 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 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. Le 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. Apprentissage statistique

Nous avons continué cette année à investiguer les méthodes de machines à vecteurs support (SVM), pour leurs applications à la détection des personnes et des voitures dans les images. Dans certains cas, un classifieur SVM linéaire suffit à distinguer une classe d'objets rigides et la détection est effectuée par simple *template matching*. Nous nous intéressons alors à la mise en oeuvre de pyramides de vecteurs support afin d'améliorer les performances de détection. Pour le cas de la détection des parties du corps (jambes, bras), nous avons également mis en oeuvre des classifieurs orientés selon les axes de symétrie des objets, ce qui nous a amené à proposer une architecture en pyramide échelle+orientation [47].

3.6. Reconnaissance structurelle

Nous avons abordé cette année le problème de la détection de classes d'objets non rigides, en particulier le corps humain. Nous avons abordé ce problème sous l'angle de la reconnaissance de *structures d'images*. La

classe d'objets à reconnaitre est décrite par le graphe de ses parties rigides et de leurs relations spatiales. Si ce graphe est un arbre, nous pouvons détecter l'objet au *maximum de vraisemblance* par l'algorithme de *Viterbi* à partir des détections de ses parties. Cette approche a donné de bons résultats sur la base de données des piétons du MIT [47] et nous travaillons dans le cadre du projet VIBES à la généraliser pour la détection et la reconnaissance d'acteurs dans des films de cinéma.

3.7. Analyse de gestes humains

L'analyse de gestes humains à partir d'images ou de séquences d'images est une discipline en pleine évolution. Auparavant on se contentait de poser des marquages sur les différentes parties du corps, d'identifier et de suivre ces marquages dans les images. Des techniques géométriques tenant compte des liens cinématiques entre les différentes parties du corps permettaient de reconstruire des propriétés tri-dimensionnelles comme des trajectoires, analyser la dynamique du mouvement, etc.

Dans le cas le plus général il est préférable de ne pas utiliser des marqueurs. Par ailleurs on doit distinguer deux cas : analyse avec une seule caméra et analyse avec plusieurs caméras.

Dans le cas de l'analyse mono-caméra il s'agit d'explorer des informations 2D en les combinant avec des connaissances a priori de modèles cinématiques et dynamiques ainsi qu'avec les modèles de caméras. Une des problématiques fondamentales à résoudre est celle de l'alignement de données 2D (zones et points d'intérêt, textures, flots optiques) avec des modèles 3D comportant jusqu'à 30 degrés de liberté. On peut s'intéresser à l'alignement avec une seule image ou avec une séquence temporelle d'images (vidéo). L'absence de marqueurs lance des défis importants afin d'être capable de reconnaître des éléments du corps humain indépendamment des textures (peau, habillages divers, éclairages, ...) dans des circonstances complexes. L'apprentissage statisitique (aussi bien pour la reconnaissance que pour le tracking) et l'optimisation non-linéaires sont deux parmi les outils qu'il faut maîtriser. Dans de nombreux cas la caméra bouge et il faut alors analyser une scène dynamique dont le mouvement apparent est la combinaison du geste humain et du mouvement de la caméra. Le cas des vidéos d'archives (analyse de gestes, reconnaissance de situations, etc.) est l'exemple le plus illustratif quant à l'utilisation massive de ces méthodes.

Dans certains cas on dispose du « luxe » de pouvoir distribuer un grand nombre de caméras autour de la personne dont on veut analyser les gestes. Le réseau de caméras peut être calibré et, moyennant des techniques connues de triangulation, on dispose de données 3D. La tâche de reconstruction et de reconnaissance de gestes humains est alors simplifié, il reste néanmoins beaucoup de problèmes à résoudre, notamment lorsque l'on désire analyser des scènes non contraintes. Des méthodes permettant de déterminer le volume occupé par des personnes ont vu le jour récemment. Le projet MOVI développe quant à lui des méthodes basées sur les enveloppes visuelles [10].

3.8. 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.

3.9. 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'application

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 assistée 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

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, 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

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. Indexation de films de cinéma

4.4.1. Indexation acteur-lieu-action

Dans le cadre du projet européen VIBES, nous nous sommes consacré cette année à constituer une base de données de films de cinéma indexés plan par plan. Ce travail a fait l'objet d'une collaboration étroite entre l'INRIA (pour la détection des visages, la constitution de la base de données et les interfaces de requête) et les universités d'Oxford (pour la reconnaissance des acteurs et des décors) et Leuwen (pour le découpage en plans).

Cette année, nous avons ainsi indexé les 150,000 images du film *Cours, Lola, cours* qui sont accessible en intranet et permettent d'accéder aux plans du film par requêtes sur les acteurs et les décors reconnus dans le film. Ce travail se poursuivra en 2003 en direction de la reconnaissance des actions.

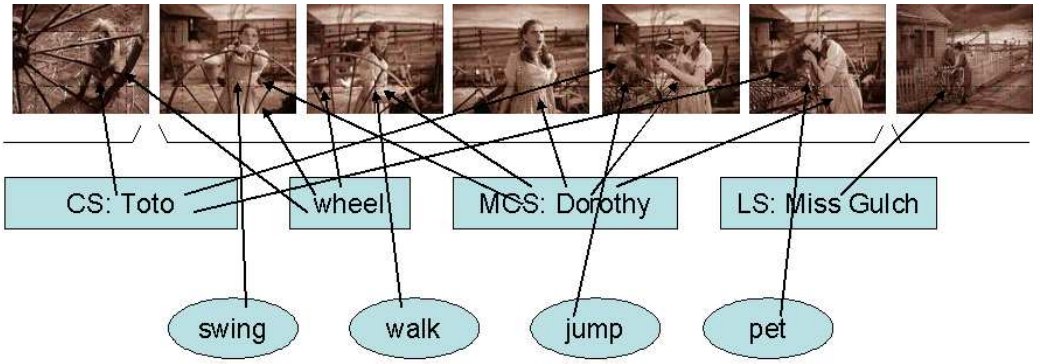
4.4.2. Alignement film et scénario

En parallèle des travaux menés dans VIBES, nous nous avons engagé cette année une collaboration avec le projet OPERA sur la possibilité d'indexer un film de cinéma par les termes de son scénario. Deux stages de maîtrise UFR-IMA ont été consacrés à ce thème, afin de structurer le scénario du film *Le magicien d'Oz* au format XML et de l'aligner temporellement sur un découpage en plans obtenu par détection de ruptures de l'image dans une base d'ondelettes.

Ce travail exploratoire a permis de constituer une base de données des 800 plans du film 'Le magicien d'Oz', accessible en intranet.

4.5. Reconnaissance des gestes des mains

La synthèse de formes 3D articulés et déformables est un processus complexe si l'on utilise des modeleurs graphiques. L'objectif est de concevoir un système permettant de sculpter de telles formes de façon intuitive, en reproduisant les mouvements des mains d'un sculpteur travaillant une argile devant un système de caméras. Les gestes mimés par l'utilisateur permettent la sculpture d'une argile virtuelle à l'intérieur de l'ordinateur. La partie concernant la simulation et le rendu de l'objet sculpté, réalisée conjointement avec l'équipe Evasion [44], a nécessité l'introduction d'un nouveau modèle pour décrire le comportement d'un objet déformable sous l'action de contraintes extérieures, permettant d'obtenir à la fois des déplacements locaux de la matière au voisinage de l'outil (bourrelets) et des déformations à grande échelle de l'objet (pliages), de même que tous les comportements que l'on attend d'une argile (section, fusion, etc). Des méthodes spécifiques de rendu ont été examinées afin de garantir au système une réponse suffisamment rapide pour permettre un travail interactif. L'utilisation des périphériques d'entrée usuels (souris 3D, phantom...) n'est pas suffisamment riche pour permettre le travail d'un tel objet, et l'utilisation de gants est contraignante pour le sculpteur. C'est la raison pour laquelle une interface de type vision a été envisagée pour permettre l'interaction de l'utilisateur avec le logiciel. La prise de vue est assurée par un groupe de trois caméras, fournissant trois séquences de prises de vues synchrones des mouvements de la main de l'utilisateur. On peut extraire de chaque série d'images un ensemble de points de l'espace, correspondant à des points situés sur la peau de la main. D'une image à



Close Shot: Toto behind wheel, listening to song.

Medium Close Shot: Dorothy singing, swings on wheel of rake , then walks forward around wheel. Toto jumps up onto seat of rake. Dorothy pets him, sits on front of rake . Dorothy finishes song. CAMERA PULLS back.
DOROTHY sings
Someday I'll wish upon a star, and wake up where the clouds are far behind…

Long Shot: *Miss Gulch rides forward to front of Gale's home, stops and gets off her bicycle as Uncle Henry comes forward.*

Figure 1. Exemple d'alignement du scenario avec les images du film Le magicien d'Oz. Les changements de plans indiqués par le scénario sont alignés avec le découpage de la bande image. Les descriptions de plans font apparaitre les noms d'acteurs, d'actions et de lieux permettant l'indexation de chaque plan.

l'autre, certains de ces points pourront apparaître ou disparaître, mais ceux visibles à deux instants consécutifs nous permettent d'estimer le mouvement des différentes phalanges. On cherche à estimer le déplacement de chacune d'elles qui va permettre d'amener le nuage de points extrait à un instant donné à coïncider avec celui issu de la série d'images suivante. Les mouvements relatifs des différentes parties de la main sont contraints conformément aux liaisons existantes, selon un modèle usuel comprenant 27 degrés de liberté. Le mouvement de chacune des phalanges est donc estimé de façon à ce qu'il n'y ait pas d'éclatement au niveau des jointures. La géométrie est construite grossièrement à partir de la première image où la main est présentée à la caméra. Pour l'instant, la localisation des joints est faite de façon interactive. Si au cours du mouvement il apparaît que la géométrie est imprécise, le modèle est progressivement affiné par la rectification de la position des joints.

4.6. Reconnaissance de gestes athlétiques

Un des domaines d'application de l'analyse de gestes humains est celui de la compréhension de gestes sportifs. En particulier on s'intéresse aux disciplines sportives pour lesquelles la réalisation d'un geste particulier est la condition d'une excellente performance. Dans ce cadre on s'intéresse aux sports individuels comportant une gestuelle complexe. Parmi toutes les épreuves candidates nous avons choisi le saut en hauteur. Tout d'abord il s'agit d'un geste très complet et très difficile à réaliser. Par ailleurs des nombreuses études biomécaniques sont disponibles ce qui permet de comparer nos résultats à la vérité terrain. Nous avons mis au point une méthode permettant d'extraire la trajectoire d'un saut à partir d'une vidéo. Partant de l'hypothèse que deux athlètes de niveaux équivalents réalisent des trajectoires similaires, nous avons développé une méthode permettant de comparere deux trajectoires, même si les vidéos ont été prises de points de vue différents et avec des paramètres inconnus quant aux caméras [43].

4.7. Virtualisation d'objets

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 peux 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.

4.8. Synthèse d'images à partir d'images

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 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.

4.9. 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. Depuis deux ans nous avons abordé un nouveau thème - le suivi et la reconstruction du mouvement humain dans les flots vidéo. Cette année nous avons poursuivi ce thème avec des nouveaux travaux sur la reconstruction des mouvements articulaires humaines à partir de vidéo monoculaire [51][52][50][28][49]. Ces travaux font aussi partie de notre projet européen VIBES, sur l'indexation et la reconstruction « niveau objet » de vidéos.

5. Logiciels

5.1. Tele2

Participants : Radu Horaud [correspondant], Matthieu Personnaz, Peter Sturm.

Mots clés : *calibrage*.

Calibrage d'une caméra ainsi que d'un couple stéréoscopique à l'aide d'une mire de calibrage [59][60]. Il s'agit d'un logiciel complet - TELE2 - 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 est distribué en ligne à l'adresse http://www.inrialpes.fr/movi/soft/calibration/index.html et référencé sur un site spécialisé sur la calibration de caméras, A few links related to camera calibration : *Tele2 : A complete Calibration Toolkit with additional Computional Vision tools - A great camera calibration software (with complete documentation) for Linux that allows calibrating of single cameras, and stereo systems. This software also contains various other computational vision tools such as stereo matching, and 3D shape computation. Written by Matthieu Personnaz at INRIA.*

5.2. Modélisation interactive de scènes

Participants : Marta Wilczkowiak, Edmond Boyer, Peter Sturm.

Mots clés : *calibrage*, *modélisation*, *reconstruction*.

Un logiciel pour la modélisation interactive de scènes à partir d'une, ou de quelques images, a été développé. Il permet de construire un modèle photo-réaliste d'une scène à l'aide de contraintes géométriques simples

(coplanarité, orthogonalité, parallélisme) 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 : Adrien Bartoli, Thomas Bonfort, Edmond Boyer, Navneet Dalal, Jean-Sébastien Franco, Radu Horaud, Peter Sturm, William Triggs, Marta Wilczkowiak.

Mots clés : *géométrie*, *calibrage*, *reconstruction 3D*, *vision 3D*, *séquence d'images*.

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.

6.1.1. Le suivi et la reconstruction du mouvement humain.

Depuis 2 ans nous étudions 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...Notre approche est basée sur une modélisation 3D articulaire explicite et une mise en correspondance modèle-image multi-primitives et robuste. Le problème est délicat en raison des difficultés de la modélisation humaine et de la mise en correspondance modèle-image, mais surtout parce que chaque image monoculaire laisse environ un tiers des 30-35 degrés de liberté articulaires du corps très mal déterminés. Ce manque d'observabilité engendre aussi quelques milliers de minima locaux (solutions 3D cinématiques) pour chaque configuration image du modèle, ce qui rend difficile l'étape d'ajustement modèle-image. Cette an nous avons amélioré notre approche de base avec le développement de deux méthodes d'optimisation non-locale pour faire face aux minima locaux [51][52][28]. Chacune des deux méthodes permet de créer une « carte de route » des minima en trouvant les « cols » (points de selle de la fonction du coût d'appariement modèle-image) qui mènent d'un minimum donné à ses minima voisins. La première méthode (dont il y a deux variantes, « balayage de surface » et « suivie de vecteur propre ») est basée sur l'optimisation Newton modifiée, la deuxième, l'approche « hyperdynamique », sur une optimisation chaîne de Markov Monté-Carlo modifiée pour converger vers des points de selle. Ces deux méthodes sont d'intérêt beaucoup plus large, et peut s'adapter à bien d'autres problèmes de l'optimisation globale.

Toujours dans le cadre du suivie humaine, nous avons aussi conçu des fonctions d'appariement modèle-image « de plus haut niveau » (qui prennent en compte le support local des appariements primitifs), pour réduire encore l'effet des ambiguïtés et des minima locaux [49][50].

6.1.2. Applications de la géométrie algébrique effective à la vision par ordinateur.

Cette étude était menée avec l'aide de B. Mourrain du projet GALAAD de l'INRIA Sophia-Antipolis. Elle représente la suite des travaux commencés dans notre projet EU Cumuli. Le but était d'examiner l'apport des techniques de la géométrie algébrique computationnelle moderne aux problèmes géométriques de la vision. Cet an, nous avons étudiés plusieurs problèmes de pose de caméra pour mieux comprendre les apports pratiques de plusieurs méthodes de résolution de systèmes de polynômes, y inclut les bases de Gröbner, les résultants de Macaulay, les résultants creux, et la homotopie [36][26].

6.1.3. 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 maitrisée 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 de ces travaux ont fait l'objet de 2 publications [41][43] et peuvent être consultés à : http://www.inrialpes.fr/movi/people/Horaud/videomosaic/videomosaic.html.

Figure 2. Cette figure montre cing images issues d'une vidéo qui en comporte 200 (haut), le panorama composé des objets statiques de la scène permettant de calculer les paramètres de mouvement de la caméra (milieu) et le panorama du mouvement humain (bas) Ici le mouvement de la caméra a été compensée de sorte que le mouvement humain apparaît tel qu'il aurait été filmé par une caméra fixe.

6.1.4. Structure et mouvement pour des scènes dynamiques

Les deux dernières années, le paradigme de la géométrie multi-images, traditionnellement formulé pour des scènes statiques, a connu des extensions à des scénarios dynamiques, où tout bouge : chaque primitive de la scène et les caméras. Le cas général, avec des mouvements complètement indépendants d'une primitive à l'autre, ne peut être traité. Donc, ces travaux récents ont portés sur des scénarios plus contraints. Dans ce cadre, nous avons étudié deux scénarios intéressants :

- une scène composée de primitives (des points) qui bougent indépendemment les unes des autres, à l'exception que chaque primitive bouge dans un plan, et que ces plans de mouvement soient liés : ils forment une faisceau (généralisation du cas de plans parallèles) [54]. Ce type de modèle peut par exemple décrire toute scène composée d'objets rigides, évoluant indépendemment les uns des autres, sur un sol ;
- une scène composée d'une partie tri-dimensionnelle rigide, et d'un plan, sur lequel se trouvent des primitives mobiles [40]. On peut penser à un scénario urbain, où des voitures roulent sur la route (le plan), au milieu de bâtiments et d'autres objets (la partie rigide de la scène).

Nos travaux sur ces scénarios concernent l'estimation du mouvement (de la caméra et des primitives mobiles), la reconstruction 3-D et l'auto-calibrage.

6.1.5. Combiner des caméras omnidirectionnelles avec des caméras perspectives

Les systèmes (multi-) stéréo classiques sont composés de caméras perspectives. Beaucoup de travaux sur la géométrie de tels systèmes ont été effectués, avec des applications en calibrage, auto-calibrage, estimation de mouvement, reconstruction 3-D, etc. Depuis quelques années, des caméras dites *omnidirectionnelles*, ayant un champ de vue hémisphérique voire plus, sont utilisées de plus en plus, surtout pour la navigation et la détection d'obstacles en robotique, et la vidéo-surveillance. Quelques équipes dans le monde étudient la géométrie de systèmes de telles caméras. Pour certaines applications, il peut être intéressant de combiner des caméras omnidirectionnelles avec des caméras classiques, par exemple pour la surveillance : une caméra omnidirectionnelle peut, grâce à son champ de vue étendue, détecter des « événements » qui se produisent n'importe où dans la scène. Une caméra classique pivotable pourra ensuite être commandée telle qu'elle puisse fournir des gros plans de l'événement, afin de procéder à une analyse plus détaillée. Nous avons donc étudié la géométrie de tels systèmes (multi-) stéréo hybrides [53]. Des notions de géométrie épipolaire ou de tenseur trifocal ont été décelées, et des applications en calibrage et auto-calibrage ont été montrées.

6.1.6. Ajustement de faisceaux

L'ajustement de faisceaux est une technique permettant d'obtenir une reconstruction précise de caméras et de primitives (points, droites, coniques, etc.) à partir de correspondances de telles primitives entre différentes images d'une même scène rigide. Une des plus importantes difficultés lors de la conception d'un tel algorithme est la paramétrisation du problème. Nous nous sommes intéressés à trouver une paramétrisation permettant d'utiliser un nombre minimal de paramètres, ce qui réduit le coût de l'algorithme (en terme de temps machine) et fixe les ambiguïtés présentes lors de l'utilisation d'une paramétrisation redondante [39][38]. Par ailleurs, nous avons proposé un schéma d'optimisation quasi-linéaire, permettant une mise en place simple de l'ajustement de faisceaux.

6.1.7. Reconstruction de surfaces réfléchissantes

Une méthode de reconstruction 3-D de surfaces réfléchissantes a été développée pendant le stage de DEA de Thomas Bonfort. La reconstruction s'effectue à partir de plusieurs images, acquises par une caméra, de la réflection, dans la surface, d'un objet connu (une mire plane en pratique). La surface réfléchissante est reconstruite dans un espace discrétisé (voxels). Par rapport à la plupart des méthodes existantes, la nôtre ne procède pas par grossissement itératif de la surface estimée, et n'a pas tendance à diverger, grâce à l'absence d'hypothèses sur la continuité de la surface.

6.1.8. 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. En outre, une seule image peut alors parfois 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 (travaux effectués en 2001).

Récemment, ces travaux ont été étendus au cas multi-images. Notamment, nous avons montré comment des contraintes géométriques peuvent être combinées avec des contraintes d'auto-calibrage, afin d'effectuer un calibrage multi-images flexible [55][56]. De plus, nous avons proposé une approche linéaire pour la construction de modèles à l'aide de contraintes géométriques. Cette approche utilise des contraintes géométriques introduites interactivement par un utilisateur pour construire, de manière itérative, un modèle d'une scène respectant ces contraintes. L'intérêt est d'offrir une approche robuste pour produire des modèles à partir de quelques images uniquement.

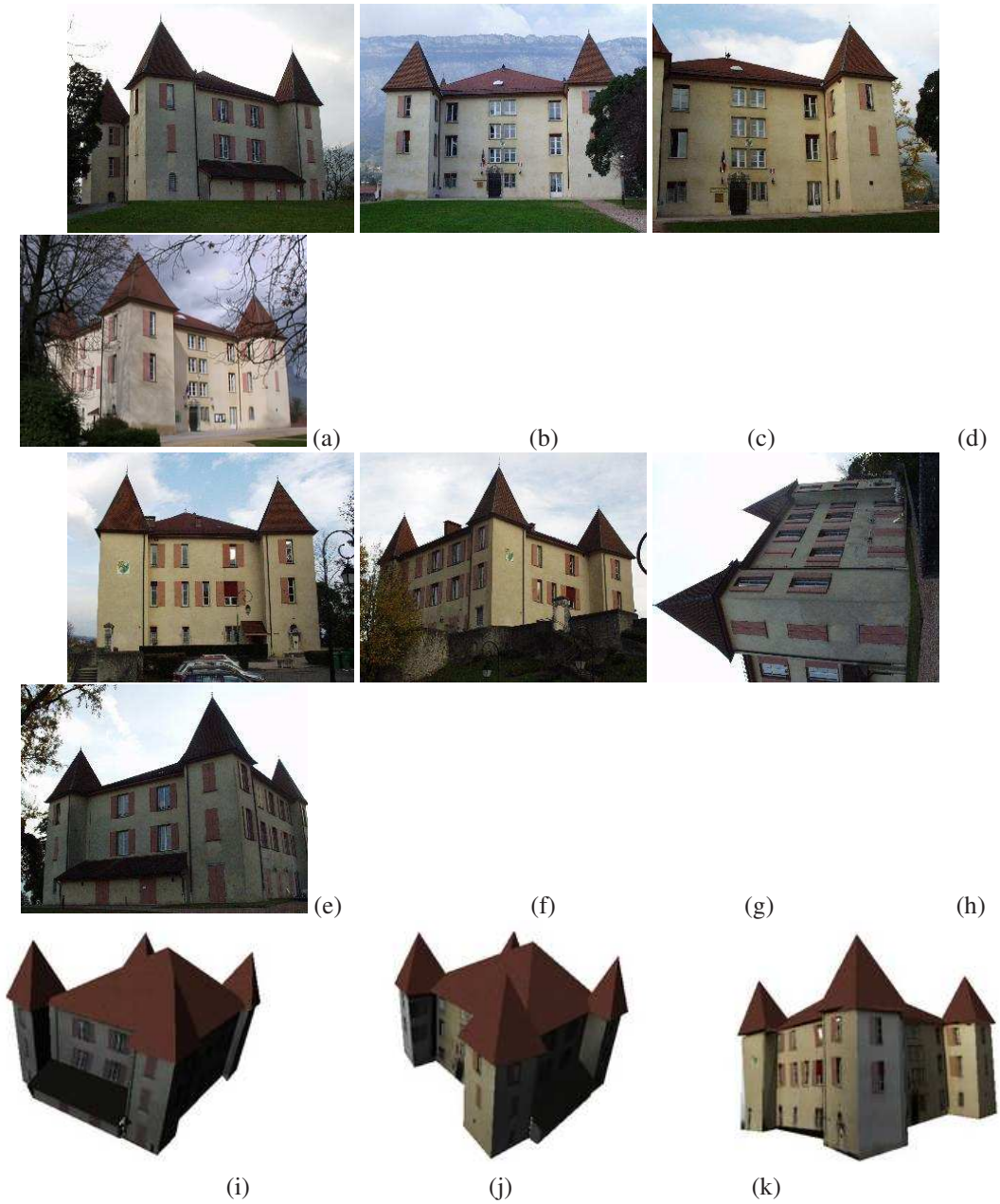


Figure 3. Exemple de modèle obtenu à partir de quelques images et de contraintes géométriques : (a)–(h) Les photos utilisées pour la modélisation ; (i)-(k) Les images du modèle texturé.

6.1.9. 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 [42]. 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.

6.1.10. Acquisition de modèles en temps réel

Le projet MOVI fournit actuellement un effort important sur l'acquisition de modèles à partir de plusieurs caméras, en particulier dans un contexte temps réel. De récents résultats ont été obtenus en utilisant l'information fournit par les silhouettes dans les images. Les silhouettes délimitent la partie visible, dans une image, de l'objet ou des objets en considération. Cette information s'avère robuste et permet de construire une approximation de la surface des objets observés : *l'enveloppe visuelle*. Pour obtenir les silhouettes, des logiciels qui identifie les objets du fonds ont été développés, notamment dans le cadre du stage de DEA de Jean-Sébastien Franco. Plusieurs méthodes de calculs de l'enveloppe visuelle ont ensuite été implémentées, et des contributions importantes ont été apportées sur ce thème en terme de précision et de robustesse. Par ailleurs, plusieurs stages actuellement en cours concernent ce thème et portent notamment sur l'acquisition d'images en temps réel ou la parallélisation des algorithmes d'estimation de modèles.

6.2. Couplage vision/robotique

Participants : Radu Horaud, Frédérick Martin, Joao Barreto.

Mots clés : *asservissement visuel*, *modélisation de robots*, *suivi multi-caméras*.

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'élargisssement 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'euclidien et le passage du projectif à l'euclidien n'est, ni plus ni moins, une forme d'auto-étalonnage de l'ensemble robot-caméras. Un des éléments clé du couplage entre vision et robotique est le tracking d'objets et de robots avec une ou plusieurs caméras.

6.2.1. Suivi d'objets en mouvement avec plusieurs caméras

Toujours dans le cadre du projet européen VIGOR et de la thèse de doctorat de Frédérick Martin 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 [31]. Les résultats d'un tracking d'une pièce de bateau avec trois caméras sont illustrés sur la figure 4.

6.2.2. Asservissement visuel avec plusieurs caméras

Malgré les nombreux progrès dans le domaine de l'utilisation des caméras multiples, la grande majorité des systèmes de contôle visuel de robots utilisent une seule caméra. Nous avons introduit la géométrie épipolaire

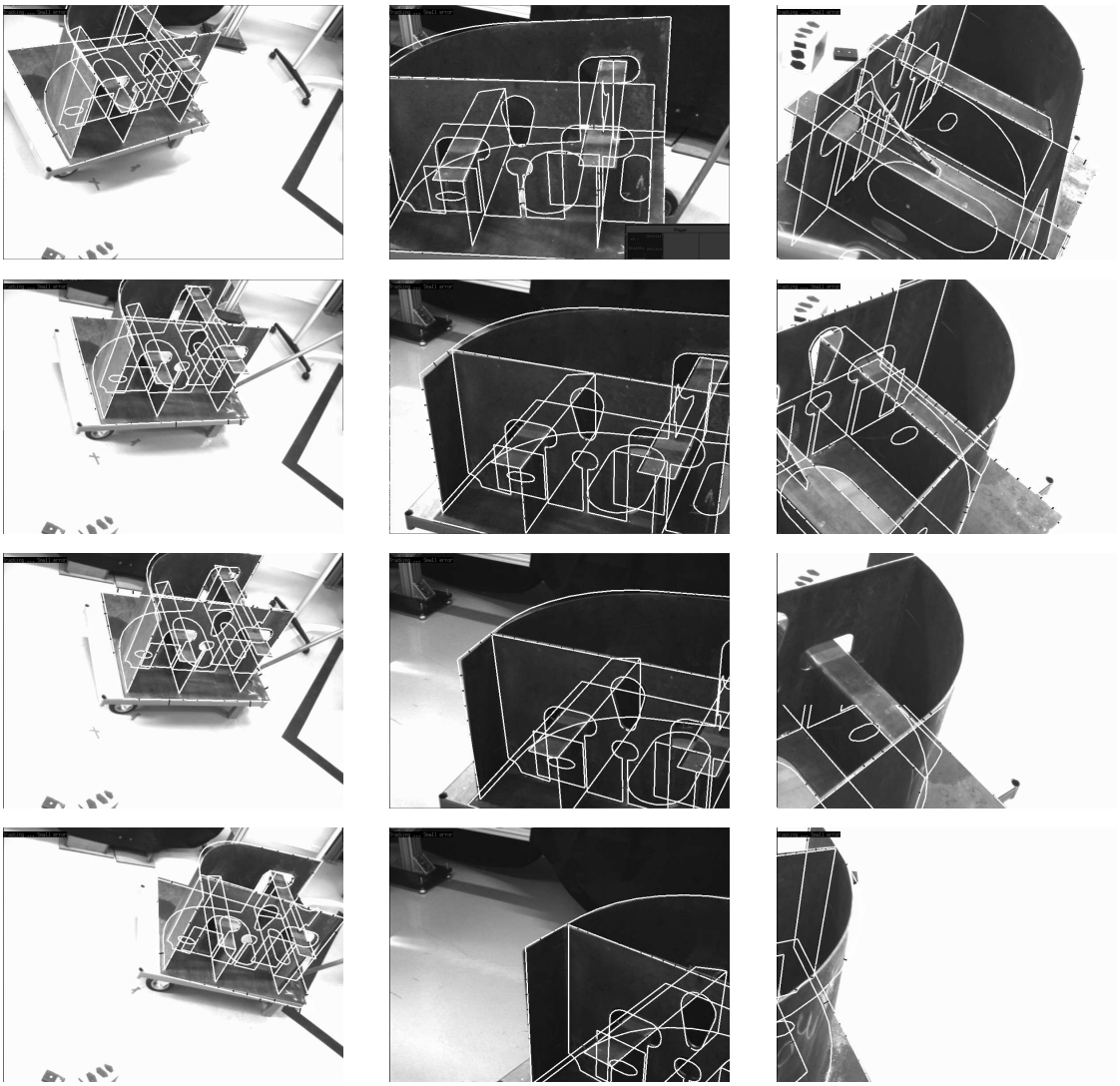


Figure 4. Une séquence de tracking avec trois caméras.La première caméra (colonne gauche) a un large champ de vue ce qui permet de maintenir une information continue concernant la position de l'objet. Les deux autres caméras peuvent ainsi tolérer des occlusions partielles, voire totales.

dans la boucle d'asservissement de manière formelle et nous avons montré qu'il n'y a pas, dans ce cas, des singularités de commande (comme avec une seule caméra). Nous avons également étudié la situation pratique lorsqu'une caméra perd le signal : nous parlons alors d'asservissement stéréoscopique virtuel.

6.2.3. Asservissement visuel avec une caméra catadioptrique

Les caméras catadioptriques sont des caméras munies d'une optique ET d'un mirroir. Dans notre cas il s'agit d'un miroir parabolique ce qui confère à la caméra une image panoramique. On parle également de caméra omni-directionnelle.

Nous avons étudié le tracking d'objets avec une telle caméra. En particulier nous avons montré que l'utilisation d'une telle caméra revient à généraliser le modèle sténopé classique. Le résultat théorique obtenu est le suivant [37] : le tracking catadioptrique ne contient aucune singularité supplémentaire par rapport au tracking avec des caméras traditionnelles.

6.3. Indexation d'images et reconnaissance d'objets

Participants : 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*.

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.

6.3.1. 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 calculés les descripteurs. L'utilisation de contraintes géométriques entre images et l'estimation robuste de ces contraintes permet un rejet important des faux appariements [57]. 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 affines a également été traité. Nous avons développé des points d'intérêt invariants aux transformations affines [46][27]. La détection de tels points est basée sur trois idées clé, : 1) La matrice des moments d'ordre deux en un point permet de normaliser une région pour devenir invariant aux transformations affines. 2) L´échelle d'une structure locale est déterminée par les extrema locaux de dérivées normalisées (Laplacien). 3) Une version adaptée affine du détecteur de Harris détermine la localisation des points d'intérêt. Un algorithme itératif modifie localisation, échelle et forme du voisinage pour chaque point et converge vers des points invariants affine.

Pour l'appariement et la reconnaissance, chaque image est caractérisée par un ensemble de points invariants affine ; la transformation affine associée avec chaque point permet de calculer un descripteur invariant affine. Des résultats expérimentaux montre une bonne performance en présence de transformations perspective importantes, figure 5.

6.3.2. Détection des humains

Nous avons commencé cette année à travailler sur la détection des humains dans les images fixes et la video. Ce travail poursuit l'effort engagé sur la détection des visages, mais introduit des difficultés nouvelles, car les autres parties du corps présentent une plus grande variabilité d'apparences. Nous avons choisi une approche par machine à support vecteurs (SVM) pour chaque partie du corps, et un modèle structural estimé au maximum de vraisemblance, qui combine les scores de détection des parties avec un modèle géométrique du corps humain. Nous avons vérifié le bon comportement de cette généralisation de l'algorithme de Viterbi au cas d'une structure articulaire du corps humain sur la base de données des piétons du MIT [47].

Figure 5. Exemple d'une recherche dans une base avec 2000 images. L'image correspondante a été retrouvé correctement. On montre les régions correspondante après vérification avec la matrice fondamentale estimée de facon robuste à partir des appariements initiaux.

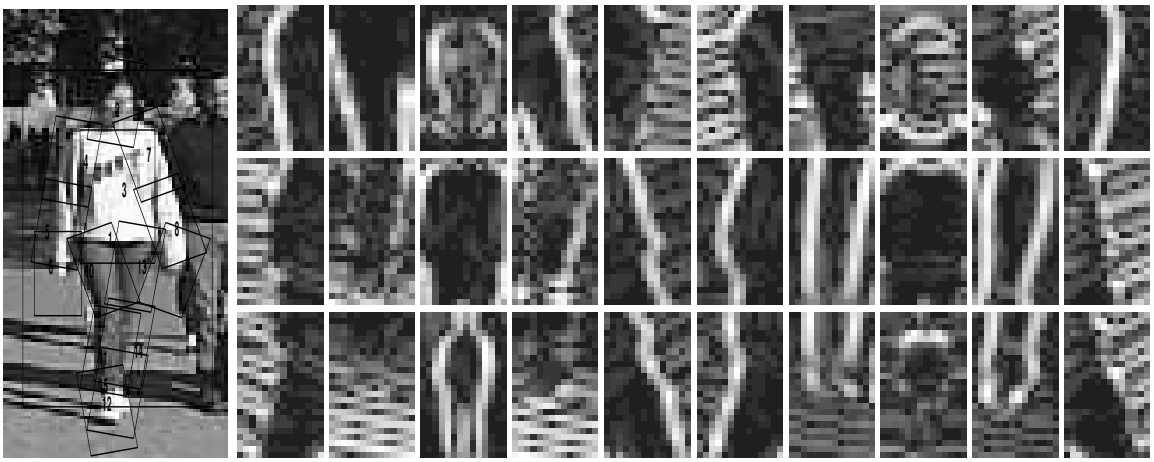


Figure 6. Apprentissage supervisé des modèles du corps humain : annotation manuelle et extraction des signatures d'images de chaque partie du corps (gradient horizontal et vertical). De haut en bas et de gauche à droite : bras, avant-bras et main gauche ; cuisse, jambe et pied gauche ; tête, torse, corps entier ; cuisse, jambe et pied droit ; bras, avant-bras et main droite.

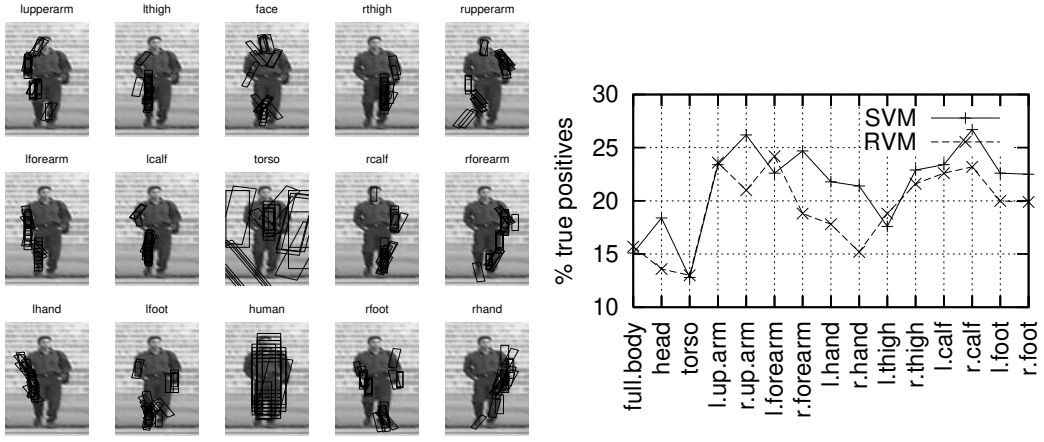


Figure 7. Détection individuelle des parties du corps. (a) Même avec des taux de détections correctes de l'ordre de 5 % il est possible d'obtenir un bon résultat parmi les 20 meilleures réponses. (b) Graphe des taux de détections correctes des classifieurs individuels sur la base de donnés des piétons du MIT.

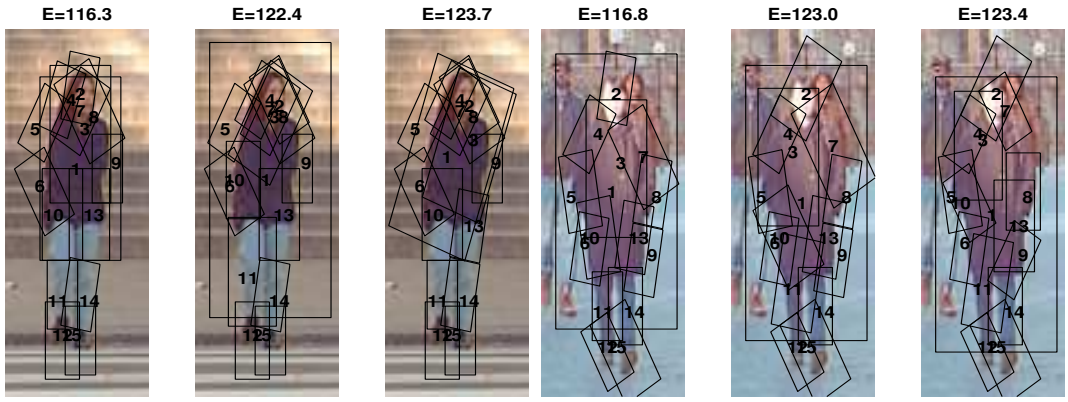


Figure 8. Exemples de détection des personnes à partir des vingt meilleures détections SVM de chaque partie du corps - les trois configuration de plus basse énergie sont présentées pour chaque exemple.

Cette combinaison de méthodes d'apprentissage statistiques et structurelles est très prometteuse et nous travaillons à une version multi-résolution afin de l'appliquer au film de cinéma, qui introduit des variations de point de vue et d'échelle spectaculaires.

6.3.3. 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 qui permet de construire des modèles à partir d'un ensemble d'images positives et négatives, [48]. 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 : Krystian Mikolajczyk, Roger Mohr, Cordelia Schmid, Rémi Ronfard, Navneet Dalal, Radu Horaud.

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.

6.4.1. Mouvements de caméra

Le stage de DEA de Caroline Le Corvec poursuit cette année les travaux engagés par Navneet Dalal et Adrien Bartoli concernant la modélisation et la compensation des mouvements de caméra. Nous nous interéresserons au cours de ce stage aux travellings, par analyse des parallaxes de mouvement.

6.4.2. Reconnaissance des visages

Nous avons poursuivi le travail réalisé l'année dernière sur la détection des visages, par des expérimentations sur deux films (plus de 300.000 images), qui ont montré l'importance de la détection des profils (plus de 80 % des visages détectés) et la bonne qualité des résultats obtenus par les méthodes développées par Krystian Mikolajczyk.

Le stage de DEA de Benjamin Chastaigner sera consacré au traitement des séquences de visages obtenusm, en vue de l'identification des acteurs. En particulier, nous aborderons le problème de séparation de la géométrie du visage de l'acteur et de ses déformations (poses et expressions) à l'aide de méthodes de factorisation multi-modes.

6.4.3. Mouvements déformables

Nous souhaitons également aborder cette année le problème de la capture des mouvements déformables pour l'animation d'acteurs virtuels. Nous envisageons d'utiliser pour ces études un modèle de surfaces de subdivisions déformables, qui a été introduit avec succés pour la sculpture en temps réel [58]. Ces surfaces minimisent leurs déformations, tout en respectant aux moindres carrés les contraintes imposées par correspondances ponctuelles depuis une ou plusieurs caméras. Le modèle est intrinsèquement multi-échelle, ce qui permet d'envisager des solutions approchées en temps réél.

7. Contrats industriels

7.1. Surveillance vidéo

Participants : Radu Horaud, Peter Sturm, Markus Michaelis.

Depuis le 1 novembre 2000 le projet accueille Markus Michaelis dans le cadre d'une convention entre l'INRIA et la société allemande Plettac Electronics. Cette société commercialise déjà des systèmes automatiques de détection de personnes à partir de l'analyse d'une vidéo issue d'une seule caméra. L'objectif principal de cette collaboration est d'étudier et concevoir des systèmes de surveillance multi-caméras. Le savoir-faire du projet en termes de calibration mono et multi caméras ainsi qu'une étude spécifique de la problématique de la surveillance automatique ont contribué à l'obtention de résultats nouveaux. Il s'agit de combiner une caméra grand champ avec une caméra équipé d'un objectif à focal varialbe et pouvant tourner sur elle-même (deux rotations). Le concept a conduit à une réalisation algorithmique et expérimentale.

Le prototype disponible à l'INRIA a été jugé suffisamment prommetteur par la société Plettac pour que la réalisation d'un produit commercial se mette en place.

Un brevet est également en cours de dépôt.

8. Actions régionales, nationales et internationales

8.1. Actions nationales

- Dans le cadre du GdR ISIS, le projet MOVI a un projet jeunes chercheurs avec l'IRIT (Institut de Recherche en Informatique, Toulouse). La collaboration porte sur l'auto-calibrage de caméras à partir d'images d'une scène plane.
- Dans le cadre du programme ROBEA, MOVI est partenaire dans un projet avec les projets SHARP, PRIMA, VISTA (Rennes) et RIA (LAAS, Toulouse). La collaboration porte sur l'interprétation de scènes dynamiques complexes et la planification réactive de mouvements dans de tels scènes.
- Rémi Ronfard participe à la nouvelle action sécifique sur les méthodes à vecteurs support mise en place cette année par le department STIC du CNRS.
- 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.
- 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 interviens sur ces aspects. Un ingénieur expert travaille actuellement sur ce projet et plusieurs stagiaires et thésards sont impliqués partiellement.

8.2. Actions financées par la Commission Européenne

8.2.1. 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 publique. 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).

8.2.2. 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 scenarios 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.

8.2.3. 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 (Israel). 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.

8.2.4. Lava.

Le projet EU 5e cadre IST-2001-34405 « LAVA – Learning for Adaptable Visual Assistants » a débuté en mai 2002 pour 3 ans. Le but est de développer des techniques d'apprentissage statistique adaptées à la reconnaissance des formes en vison, en particulière dans le contexte des assistants électroniques intelligents dotés d'une caméra. Le projet est interdisciplinaire, impliquant des équipes de l'apprentissage, de la vision, et de la modélisation cognitive. Les partenaires sont Xerox Research Centre Europe (coordinateur, Grenoble, France) ; MOVI (GRAVIR-INRIA-CNRS-INPG, Grenoble, France) et VISTA (IRISA-INRIA, Rennes, France) ; Royal Holloway College, Université de Londres (RHUL, Egham, Angleterre) ; l'Université de Lund (Lund, Suède) ; l'Université Technique de Graz (Graz, Autriche) ; l'Institut Dalle Molle d'Intelligence Artificielle Perceptive (IDIAP, Martigny, Suisse) ; et l'Université National d'Australie (ANU, Canberra, Australie). LAVA représentera 51 personne-ans d'effort, pour un budget total de 4,3 MEu, avec un support Européen de 2,4 MEu. MOVI travaillera principalement sur les aspects développement des descripteurs images statiques, l'interface vision-apprentissage, et l'apprentissage semi-supervisée.

8.3. Relations bilatérales internationales

8.3.1. Europe

8.3.1.1. 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 un séjour d'une semaine à Kingston, et P. Remagnino a visité MOVI pendant une semaine également.

8.3.2. Amérique

8.3.2.1. CNRS/UIUC.

Dans le cadre du programme CNRS/UIUC (The Univ. of Illinois, U.S.A.), nous collaborons avec l'équipe de J. Ponce sur plusieurs thèmes dont les enveloppes visuelles ainsi que la reconnaissance d'objets.

8.3.2.2. NSF/INRIA.

Par ailleurs nous avons soumis une proposition de collaboration avec l'équipe VIP de MIT AI Lab (Trevor Darell) dans le cadre du programme NSF-INRIA, et sur le thème de la modélisation et du suivi à l'aide de caméras numériques.

8.3.3. Asie

8.3.3.1. 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 a séjourné à Montbonnot pour une durée d'un mois et P. Sturm a visité le partenaire Chinois pendant 10 jours.

8.3.4. Océanie

W. Triggs à visité Prof R. Hartley et Dr A. Smola de l'Université National d'Australie (ANU, Canberra) et NICTA (le nouveau « National centre of excellence on Information and Communications Technology, Australia » - en quelque sorte l'équivalent australien de l'INRIA) pendant une semaine, et il est « collaborateur externe » sur leur projet « Pattern Recognition and Scene Analysis via Machine Learning », la partie australienne de notre projet européen LAVA.

9. Diffusion des résultats

9.1. Animation de la communauté scientifique

9.1.1. 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)

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

- ECCV'02 (B. Triggs et C. Schmid, area chairs et P. Sturm, programme committee)
- ICPR'02 (P. Sturm)
- ICIP'02 (P. Sturm)
- VI'02 (P. Sturm)
- ICRA'02 (R. Horaud)
- FG'02 (R. Horaud)
- ICVGIP'02 (C. Schmid et B. Triggs)

9.1.3. Autres :

P. Sturm fait partie du comité du Prix de Thèse SPECIF 2002.

C. Schmid fait partie de la commission d'évaluation et la commission des emplois scientifiques.

9.2. Enseignement universitaire

- 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.
- Base d'images, ENSIMAG, INPG, 10h, C. Schmid.
- Vision stéréoscopique, Mastère photogrammétrie Numérique, ENSG, 14h, F. Devernay.
- Synthèse d'images, MAGISTÈRE INFORMATIQUE, UNIV. JOSEPH FOURIER, RICM, ISTG, 100h, E. Boyer
- Analyse d'images, DESS INFORMATIOQUE, UNIV. JOSEPH FOURIER, 30H, E. Boyer.
- Géométrie projective, DEA IVR, INPG, 6h, E. Boyer.

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 prononcé des séminaires à Hong Kong (CUHK et HKUST, octobre 2002) et à Xi'an (octobre 2002).
- P. Sturm a fait une présentation invitée au « Pattern Recognition and Computer Vision Colloquium » à Prague (novembre 2002).
- C. Schmid a fait une présentation invitée au séminaire Dagstuhl, Content-Based Image and VIdeo Retrieval, janvier 2002.
- C. Schmid a fait une présentation a la journée INTECH, recherche par le contenu de documents multi-medias, mars 2002.
- C. Schmid a donne un tutorial dans le cadre de l'école d'été en image et robotique, juillet 2002.

10. Bibliographie

Bibliographie de référence

[1] S. CHRISTY, R. HORAUD. *Euclidean Shape and Motion from Multiple Perspective Views by Affine Iterations.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 11, volume 18, November, 1996, pages 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.* in « Computer Vision and Image Understanding », numéro 2, volume 69, 1998, pages 135-155.

[3] R. HARTLEY, P. STURM. *Triangulation.* in « Computer Vision and Image Understanding », numéro 2, volume 68, 1997, pages 146-157.

[4] R. HORAUD, G. CSURKA, D. DEMIRDJIAN. *Stereo Calibration from Rigid Motions.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 12, volume 22, December, 2000, pages 1446-1452, ftp://ftp.inrialpes.fr/pub/movi/publications/HoraudCsurkaDemirdjian-pami2000.ps.gz.

[5] R. HORAUD, F. DORNAIKA, B. ESPIAU. *Visually Guided Object Grasping.* in « IEEE Transactions on Robotics and Automation », numéro 4, volume 14, August, 1998, pages 525-532.

[6] R. HORAUD, F. DORNAIKA. *Hand-Eye Calibration.* in « International Journal of Robotics Research », numéro 3, volume 14, June, 1995, pages 195-210.

[7] R. HORAUD, F. DORNAIKA, B. LAMIROY, S. CHRISTY. *Object Pose : The Link between Weak Perspective, Paraperspective, and Full Perspective.* in « International Journal of Computer Vision », numéro 2, volume 22, March, 1997, pages 173-189.

[8] R. HORAUD, O. MONGA. *Vision par ordinateur : outils fondamentaux.* Editions Hermès, Paris, 1995, *Deuxième édition revue et augmentée*.

[9] L. HÉRAULT, R. HORAUD. *Figure-ground discrimination : a combinatorial optimization approach.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 9, volume 15, September, 1993, pages 899-914.

[10] 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, Dec, 2001, http://www.inrialpes.fr/movi/publi/Publications/2001/LBP01.

[11] K. MIKOLAJCZYK, C. SCHMID. *Indexing based on scale invariant interest points.* in « Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada », pages 525-531, 2001, http://www.inrialpes.fr/movi/publi/Publications/2001/MS01.

[12] R. MOHR, B. BOUFAMA, P. BRAND. *Understanding Positioning from Multiple Images.* in « Artificial Intelligence », volume 78, 1995, pages 213-238.

[13] S. PETITJEAN, E. BOYER. *Regular and Non-Regular Point Sets : Properties and Reconstruction.* in « Computational Geometry - Theory and Application », numéro 2-3, volume 19, 2001, pages 101-126, http://www.inrialpes.fr/movi/publi/Publications/2001/PB01.

[14] L. QUAN, T. KANADE. *Affine Structure from Line Correspondences with Uncalibrated Affine Cameras.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 8, volume 19, août, 1997, pages 834-845.

[15] L. QUAN. *Invariants of Six Points and Projective Reconstruction from Three Uncalibrated Images.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 1, volume 17, janvier, 1995, pages 34-46.

[16] L. QUAN. *Conic Reconstruction and Correspondence from Two Views.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 2, volume 18, février, 1996, pages 151-160.

[17] L. QUAN. *Self-Calibration of An Affine Camera from Multiple Views.* in « International Journal of Computer Vision », numéro 1, volume 19, mai, 1996, pages 93-105.

[18] A. RUF, R. HORAUD. *Visual Servoing of Robot Manipulators, Part I : Projective Kinematics.* in « International Journal of Robotics Research », numéro 11, volume 18, November, 1999, pages 1101-1118, http://www.inria.fr/rrrt/rr-3670.html.

[19] C. SCHMID, R. MOHR, C. BAUCKHAGE. *Evaluation of Interest Point Detectors.* in « International Journal of Computer Vision », numéro 2, volume 37, 2000, pages 151-172, http://www.inrialpes.fr/movi/publi/Publications/2000/SMB00.

[20] C. SCHMID, R. MOHR. *Object Recognition Using Local Characterization and Semi-Local Constraints.* in « IEEE Transactions on Pattern Analysis and Machine Intelligence », numéro 5, volume 19, mai, 1997, pages 530-534.

[21] 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 », pages 432-437, June, 1999.

[22] 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 », pages 1100-1105, Juin, 1997.

[23] 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 », pages 709-720, Avril, 1996.

[24] B. TRIGGS. *Matching Constraints and the Joint Image.* in « IEEE Int. Conf. Computer Vision », éditeurs E. GRIMSON., pages 338-43, Cambridge, MA, juin, 1995.

[25] B. TRIGGS. *Autocalibration and the Absolute Quadric.* in « IEEE Conf. Computer Vision & Pattern Recognition », Puerto Rico, 1997.

Thèses et habilitations à diriger des recherche

[26] M.-A. AMELLER. *Applications de la géométrie algébrique effective à la vision.* thèse de doctorat, Institut National Polytechnique de Grenoble, jul, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Ame02, (Applications of effective algebraic geometry in vision).

[27] K. MIKOLAJCZYK. *Detection of local features invariant to affines transformations.* thèse de doctorat, INPG, Grenoble, juillet, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Mik02.

[28] C. SMINCHISESCU. *ESTIMATION ALGORITHMS FOR AMBIGUOUS VISUAL MODELS - Three Dimensional Human Modeling and Motion Reconstruction in Monocular Video Sequences.* thèse de doctorat, INPG, Juillet, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Smi02.

Articles et chapitres de livre

[29] N. ANDREFF, B. ESPIAU, R. HORAUD. *Visual Servoing from Lines.* in « International Journal of Robotics Research », numéro 8, volume 21, August, 2002, pages 769-700, http://www.inrialpes.fr/movi/publi/Publications/2002/AEH02.

[30] E. BOYER, P. STURM. *Traité IGAT : Synthèse d'images géographiques.* éditions Hermes, 2002, chapitre Modélisation à partir d'images, pages 57-89.

[31] F. MARTIN, R. HORAUD. *Multiple Camera Tracking of Rigid Objects.* in « International Journal of Robotics Research », numéro 2, volume 21, February, 2002, pages 97-113, http://www.inrialpes.fr/movi/publi/Publications/2002/MH02.

[32] G. OLAGUE, R. MOHR. *Optiaml camera placement for accurate reconstruction.* in « Pattern Recognition », volume 35, 2002, pages 927-944, http://www.inrialpes.fr/movi/publi/Publications/2002/OM02.

[33] P. STURM. *Critical Motion Sequences for the Self-Calibration of Cameras and Stereo Systems with Variable Focal Length.* in « Image and Vision Computing », numéro 5-6, volume 20, Mar, 2002, pages 415-426, http://www.inrialpes.fr/movi/publi/Publications/2002/Stu02c.

Communications à des congrès, colloques, etc.

[34] M.-A. AMELLER, A. BARTOLI, L. QUAN. *Minimal Metric Structure and Motion from Three Affine Images.* in « In Proceedings of the Fifth Asian Conference on Computer Vision », volume I, pages 356-361, Jan, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/ABQ02b.

[35] M.-A. AMELLER, A. BARTOLI, L. QUAN. *Reconstruction metrique minimale a partir de trois cameras affines.* in « In Actes du Congres Francophone de Reconnaissance des Formes et Intelligence Artificielle », volume II, pages 471-477, Jan, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/ABQ02.

[36] M.-A. AMELLER, L. QUAN, B. TRIGGS. *Le calcul de pose : de nouvelles méthodes matricielles.* in « Reconnaissance des Formes et Intelligence Artificielle », jan, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/AQT02.

[37] J. BARRETO, F. MARTIN, R. HORAUD. *Visual Servoing/Tracking Using Central Catadioptric Cameras.* in « International Symposium on Experimental Robotics », série Advanced Robotics Series, Springer-Verlag, éditeurs B. SICILIANO, P. DARIO., July, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/BMH02.

[38] A. BARTOLI. *A Unified Framework for Quasi-Linear Bundle Adjustment.* in « In Proceedings of the Sixteenth IAPR International Conference on Pattern Recognition », volume II, pages 560-563, Aug, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Bar02b, Quebec City, Canada..

[39] A. BARTOLI. *On the Non-Linear Optimization of Projective Motion Using Minimal Parameters.* in « In Proceedings of the Seventh European Conference on Computer Vision », volume II, pages 340-354, May, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Bar02.

[40] A. BARTOLI. *The Geometry of Dynamic Scenes - On Coplanar and Convergent Linear Motions Embedded in 3D Static Scenes.* in « In Proceedings of the Thirteenth British Machine Vision Conference », volume I, pages 394-403, Sep, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Bar02c, Cardiff, UK..

[41] A. BARTOLI, N. DALAL, B. BOSE, R. HORAUD. *From Video Sequences to Motion Panoramas.* in « Proceedings of the IEEE Workshop on Motion and Video Computing », IEEE Computer Society Press, pages 201-207, 5-6 December, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/BDBH02, Orlando, Florida.

[42] A. BARTOLI, P. STURM. *La matrice de mouvement pour droites 3D, application à l'alignement de recon-*

structions de droites.. in « Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle », volume I, pages 29-37, january, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/BS02, Angers, France.

[43] N. DALAL, R. HORAUD. *Indexing Key Positions between Multiple Videos.* in « Proceedings of IEEE Workshop on Motion and Video Computing », IEEE Computer Society Press, pages 65-71, 5-6 December, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/DH02, Orlando, Florida.

[44] G. DEWAELE, M.-P. CANI, R. HORAUD. *Argile virtuelle temps-reel : le cas bidimensionnel.* in « Neuvieme reunion du groupe de travail Animation et simulation », juin, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/DCH02.

[45] S. LAZEBNIK, A. SETHI, C. SCHMID, D. KRIEGMAN, J. PONCE, M. HEBERT. *On Pencils of Tangent Planes and the Recognition of Smooth 3D Shapes from Silhouettes.* in « eccv02 », volume III, pages 651-665, 2002.

[46] K. MIKOLAJCZYK, C. SCHMID. *An affine invariant interest point detector.* in « European Conference on Computer Vision », Springer, pages 128-142, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/MS02, Copenhagen.

[47] R. RONFARD, C. SCHMID, B. TRIGGS. *Learning to parse pictures of people.* in « European Conference on Computer Vision », Jun, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/RST02, Copenhagen.

[48] C. SCHMID. *Apprentissage de modèles pour la recherche d'images.* in « RFIA », volume III, pages 781-789, 2002.

[49] C. SMINCHISESCU. *Consistency and Coupling in Human Model Likelihoods.* in « IEEE International Conference on Automatic Face and Gesture Recognition », pages 27-32, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Smi02b.

[50] C. SMINCHISESCU, A. TELEA. *Human Pose Estimation from Silhouettes. A Consistent Approach Using Distance Level Sets.* in « WSCG International Conference on Computer Graphics, Visualization and Computer Vision », 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/ST02c.

[51] C. SMINCHISESCU, B. TRIGGS. *Building Roadmaps of Local Minima of Visual Models.* in « European Conference on Computer Vision », volume 1, pages 566-582, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/ST02.

[52] C. SMINCHISESCU, B. TRIGGS. *Hyperdynamics Importance Sampling.* in « European Conference on Computer Vision », volume 1, pages 769-783, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/ST02b.

[53] P. STURM. *Mixing Catadioptric and Perspective Cameras.* in « Workshop on Omnidirectional Vision, Copenhagen, Denmark », pages 37-44, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Stu02.

[54] P. STURM. *Structure and Motion for Dynamic Scenes - The Case of Points Moving in Planes.* in « European Conference on Computer Vision, Copenhagen, Denmark », volume 2, pages 867-882, May, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/Stu02b.

[55] M. WILCZKOWIAK, E. BOYER, P. STURM. *3D Modeling Using Geometric Constraints : A Parallelepiped Based Approach.*. in « Proceedings of the Seventh European Conference on Computer Vision, ECCV'02, Copenhagen, Denmark », volume IV, pages 221-237, May, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/WBS02.

[56] M. WILCZKOWIAK, P. STURM, E. BOYER. *Calibrage et Reconstruction à l'aide de Parallélépipèdes et de Parallélogrammes.* in « Actes du 13ème Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle, RFIA'02, Angers, France », pages 849-857, January, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/WSB02.

Rapports de recherche et publications internes

[57] Y. DUFOURNAUD, C. SCHMID, R. HORAUD. *Image Matching with Scale Adjustment.* rapport technique, numéro RR 4458, INRIA Rhône-Alpes, Montbonnot Saint-Martin, May, 2002, http://www.inria.fr/rrrt/rr-4458.html.

[58] I. MARTIN, R. RONFARD, F. BERNARDINI. *Detail-preserving variational surface design with multi-resolution constraints.* rapport technique, numéro RC22499, IBM T.J. Watson Research Center, June, 2002, http://www.inrialpes.fr/movi/publi/Publications/2002/MRB02.

[59] M. PERSONNAZ, R. HORAUD. *Camera calibration : estimation, validation and software.* rapport technique, numéro RT-0258, INRIA Rhone Alpes, Grenoble, March, 2002, http://www.inria.fr/rrrt/rt-0258.html.

[60] M. PERSONNAZ, P. STURM. *Calibration of a stereo-vision system by the non-linear optimization of the motion of a calibration object.* rapport technique, numéro RT-0269, INRIA, September, 2002, http://www.inria.fr/rrrt/rt-0269.html.

[61] M. PERSONNAZ, P. STURM. *Specifications of the software videomatch 1.1.* rapport technique, numéro RT-0260, INRIA, INRIA Rhone Alpes, March, 2002, http://www.inria.fr/rrrt/rt-0260.html.