Synthèse des travaux et activités scientifiques

présentée par

Peter Franz STURM

pour obtenir le grade de
Habilitation à diriger des Recherches
de l'**INSTITUT NATIONAL POLYTECHNIQUE DE GRENOBLE**
Spécialité : **Informatique**

Recherches en vision par ordinateur

Présentée publiquement le 16 mai 2006

devant le jury composé de

Président :	M. Roger Mohr	INPG
Rapporteurs :	M. Kostas Daniilidis	University of Pennsylvania
	M. Michel Dhome	LASMEA
	M. Jean Ponce	Beckman Institute et University of Illinois
Examinateurs :	M. Luc Van Gool	Katholieke Universiteit Leuven et ETH Zürich
	M. Richard Hartley	The Australian National University
	M. Radu Horaud	INRIA
	M. Long Quan	Hong Kong University of Science and Technology

Avant-propos

Dans ce document, je décris mes activités professionnelles, pour la période allant de 1998 à 2005. La plus grande partie de ce document (parties II à VI) est rédigée en anglais. Elle contient la description des résultats scientifiques que j'ai obtenus durant cette période. Un résumé en français en est donné dans la première partie du manuscrit. Il est précédé d'une description de mes activités d'animation de la recherche (organisations de colloques, participations à des comités de programme, responsabilités scientifiques, participations à des projets, communications invitées, etc.) et des tâches administratives et d'enseignement dont j'ai eu la charge. Ce document est accompagné par une sélection des articles les plus représentatifs de mes activités de recherche pour la période concernée.

Foreword

In this document, I describe my professional activities, concerning the years from 1998 to 2005. The description of my scientific results is given in English (parts II to VI). The first part of this document is in French; it contains a description of my research related activities and of teaching and administrational duties, as well as a summary of my scientific results. This document is accompanied by a collection of my most representative papers, published in the relevant period of time.

Table des matières

Première partie

Synthèse des activités scientifiques et administratives et des travaux scientifiques

Chapitre 1

Activités scientifiques et administratives

1.1 Encadrement

J'ai encadré et co-encadré une vingtaine de personnes jusqu'à ce jour, dont six doctorants. Trois thèses ont déjà été soutenues, en 2003, 2004 et 2006 respectivement. J'ai également encadré une dizaine de doctorants inscrits à l'étranger, pendant les séjours qu'ils ont effectués à l'INRIA. Le plus souvent, il s'agissait de séjours de trois mois ou plus, organisés de manière assez informelle. Les encadrements sont décrits plus en détails dans la suite. La plupart de mes travaux scientifiques depuis six ans ont été menés *via* ces encadrements.

1.1.1 Thèses

Pour chaque étudiant, j'indique si j'ai été officiellement directeur de thèse (par dérogation de l'École Doctorale) ou co-encadrant. Ma participation active vis-à-vis de ces encadrements est quantifiée en termes de pourcentages. Tous mes doctorants ont été inscrits à l'INPG. Le doctorat de Srikumar Ramalingam s'effectue par ailleurs selon une convention de co-tutelle (voir ci-dessous).

Adrien Bartoli (2000-03), co-encadrement (90 %), directeur de thèse : Radu Horaud. Titre de la thèse : *Reconstruction et alignement en vision 3D : points, droites, plans et caméras*. Adrien a entre autres travaillé sur des paramétrages et algorithmes pour une reconstruction 3D optimale en utilisant des contraintes géométriques de coplanarité et colinéarité. Ces travaux sont décrits dans les paragraphes 2.1.2 et 2.3.1 et les chapitres 5 et 10.
Adrien a reçu le **Prix de Thèse INPG 2004**.

Marta Wilczkowiak (2000-04), co-encadrement (50 %) avec Edmond Boyer (50 %), directeur de thèse : Radu Horaud. Titre de la thèse : *3D Modelling From Images Using Geometric Constraints*. Comme Adrien Bartoli, Marta a travaillé sur l'utilisation de contraintes géométriques pour la reconstruction 3D et le calibrage de caméras. Elle a étudié des approches basées sur l'utilisation d'un nombre minimal de contraintes, avec la possibilité d'en combiner de divers types (colinéarité ou coplanarité de points, connaissance de distances entre points, orthogonalité ou parallélisme entre droites et plans, etc.). Ces travaux sont décrits dans le paragraphe 2.3.1 et le chapitre 10. Une partie de ces travaux a été effectuée en collaboration avec Gilles Trombettoni de l'INRIA Sophia-Antipolis, collaboration initiée et animée par Marta.

Thomas Bonfort (2002-06), co-encadrement (100 %), directeur de thèse : Radu Horaud. Titre de la thèse : *Reconstruction de surfaces réfléchissantes à partir d'images*. Thomas a travaillé sur la reconstruction

3D de surfaces spéculaires. Les approches originales proposées, ainsi que les méthodes développées, sont décrites dans le paragraphe 2.3.4 et le chapitre 13.

Pau Gargallo (2003-06), directeur de thèse (100 %). Pau travaille sur la reconstruction 3D dense à partir d'images. Son travail sur l'utilisation de cartes de profondeur pour la modélisation de la scène est décrit dans le paragraphe 2.3.3 et le chapitre 12.

Srikumar Ramalingam (2004-06), directeur de thèse (90 %). Cette thèse s'effectue selon une convention de co-tutelle entre l'INPG et l'Université de Californie à Santa Cruz (encadrant californien : Suresh Lodha). Srikumar travaille sur l'utilisation de modèles très généraux de caméras, associant à l'ensemble des pixels un ensemble de « rayons visuels », non nécessairement contraints géométriquement. Ses travaux concernent le calibrage et l'autocalibrage de telles caméras et leurs utilisations pour la reconstruction 3D et l'estimation du mouvement. Ils sont décrits dans la section 2.2 et dans la partie III.

Aude Jacquot (2003-06), co-encadrement (90 %), directeur de thèse : Radu Horaud. Aude travaille dans le cadre d'un contrat CIFRE avec Thales Optronique (encadrant chez Thales : Olivier Ruch). Son sujet de thèse concerne le suivi d'objets dans des séquences d'images. Elle a développé des approches basées sur les filtres à particules, voir le paragraphe 2.4 et le chapitre 15.

1.1.2 Post-docs, stages de DEA ou de fin d'études, ingénieurs

Dana Cobzaş, post-doctorante (2004-05), bourse du NSER Canada. Dana travaille sur la reconstruction simultanée de la géométrie d'un objet et de ses propriétés de réflectance. Ces travaux ont commencé lors du séjour post-doctoral de Dana et continuent depuis, en collaboration avec Martin Jägersand et Neil Birkbeck, de l'Université de l'Alberta. Ils sont décrits dans le paragraphe 2.2 et le chapitre 14.

Matthieu Personnaz, ingénieur expert (2000-03). Matthieu a travaillé, entre autres, sur le projet Européen VISIRE (voir la section 1.6). Il a développé des outils (bibliothèques et interfaces graphiques) d'extraction et de suivi de points dans des séquences vidéo, en particulier pour le calibrage de caméras. Il a également mis en œuvre une bibliothèque de méthodes d'appariement dense et a traduit des algorithmes, développés par moi-même, de `C` en `Visual C++` (algorithmes de plaquage de textures, de maillage, d'autocalibrage, de génération de fichiers VRML, etc.).

Pau Gargallo, stage de DEA (2002-03). Lors de son stage de DEA, Pau a développé des méthodes de reconstruction de surfaces à partir d'images. Ces méthodes s'appuyaient sur la « tétrahédrisation » de Delaunay de l'espace 3-D, définie sur des nuages de points 3-D reconstruits et utilisaient des critères géométriques, de visibilité et de photo-consistance. Les techniques qu'il a utilisées s'apparentent à celles dites du *voxel carving*.

Salvatore Notarangelo, stage de Magistère/Maîtrise (2002-03), co-encadrement avec Edmond Boyer. Salvatore a mis en œuvre une approche pour le calibrage d'un système multi-caméras, en utilisant un bâton comportant des marqueurs colorés, comme objet de calibrage.

Lucile Prin-Zanet (École Supérieure des Procédés Elecroniques et Optiques, Orléans), stage de fin d'études (2002-03). Lucile a mis en œuvre un système pour le calibrage de caméras de surveillance. Ce système, comme celui décrit au paragraphe précédent, utilise un bâton comportant des marqueurs. La spécificité de son travail réside dans le fait que le calibrage d'un système multi-caméras est possible à partir de mouvements généraux du bâton alors que, dans son cas, le calibrage d'une seule caméra exige un mouvement contraint de cet objet.

Marta Wilczkowiak, projet de fin d'études (1999-2000), co-encadrement avec Edmond Boyer. Lors de son stage, Marta a développé des méthodes de reconstruction 3D à partir d'une seule image, en utilisant des contraintes géométriques décrites par l'utilisateur. Ces travaux ont été poursuivis lors de sa thèse.

Thomas Bonfort, stage de DEA (2001-02). Lors de ce stage, Thomas a travaillé sur une première approche pour la reconstruction 3D de surfaces spéculaires à partir d'images. Ces travaux ont été poursuivis lors de sa thèse.

Anthony Garcia, stage de Magistère/Maîtrise (2001-02). Anthony a mis en œuvre une méthode pour le calcul de la trajectoire d'une caméra dans un studio virtuel. Pour ce faire, une caméra auxiliaire a été attachée à la caméra principale (celle qui filme une action) de manière à ce qu'elle voit le plafond, sur lequel des cibles ont été déployées. Après un calibrage de ce système de caméras, le calcul de pose de la caméra auxiliaire, à savoir sa position et son orientation, est effectué en observant les cibles sur le plafond. De ce calcul, on peut alors déduire la pose, et donc la trajectoire, de la caméra principale. Cette connaissance est utile dans un contexte de réalité augmentée, lorsque l'on souhaite insérer de façon réaliste des objets virtuels dans un film ou une séquence vidéo.

Magdalena Urbanek, stage de DEA (2000-01) et projet de fin d'études (1999-2000), co-encadrement avec Radu Horaud. Magda a étudié, entre autres, les effets de la mise au point et du zoom d'une caméra sur les paramètres intrinsèques du modèle sténopé associé. Ces travaux sont décrits dans le paragraphe 2.1.1 et le chapitre 4.

Laurent Verschueren, (KU Louvain, Belgique), projet de fin d'études (2000). Laurent a développé une méthode de reconnaissance de bâtiments, fondée sur des descripteurs locaux d'images issus des travaux de Cordelia Schmid.

Frank Althoff, (Université de Bielefeld, Allemagne), projet de maîtrise (1996). Frank a mis en œuvre une bibliothèque et une interface graphique pour la génération et la visualisation de données simulées pour la vision 3D (simulation de scènes et de caméras).

1.1.3 Visiteurs

Kiyoung Kim, doctorant, Gwangju Institute of Science and Technology, Corée du Sud (six mois, 2005/06). Kiyoung développe une méthode d'autocalibrage d'un système multi-caméras portable.

Neil Birkbeck, étudiant en MSc, Université de l'Alberta, Canada (six semaines, 2005). Neil a travaillé sur le problème de la reconstruction conjointe de la géométrie et de propriétés de réflectance. Le travail décrit dans [17], le paragraphe 2.3.5 et le chapitre 14 a été accompli ensuite au Canada, lors de son projet de MSc, sous la direction de Dana Cobzaş et Martin Jägersand.

Diego Aguilera, doctorant, Universidad de Salamanca, Espagne (trois mois, 2005). Diego a travaillé sur l'utilisation conjointe de scanners laser et de caméras, pour la modélisation 3D.

Jean-Philippe Tardif, doctorant, Université de Montréal, Canada (trois mois en 2005 et en 2006). Jean-Philippe a travaillé sur le calibrage et l'autocalibrage d'une caméra, en prenant en compte un modèle général de distorsion radiale. Cette collaboration a continué en dehors de ses séjours à l'INRIA. Ces travaux sont décrits dans le paragraphe 2.2 et les chapitres 6 et 7.

Srikumar Ramalingam, étudiant Master, University of California at Santa Cruz, USA (trois mois en 2003 et six mois en 2004). La collaboration avec Srikumar, démarrée lors de ses deux premiers séjours à l'INRIA, a été poursuivie dans le cadre d'une convention de co-tutelle pour sa thèse, voir le paragraphe 1.1.1.

Tomislav Pribanić, doctorant, Université de Zagreb, Croatie (trois mois, 2004). Tomislav a mis en œuvre et comparé différentes méthodes pour le calibrage d'un système multi-caméras, à l'aide d'un ou plusieurs bâtons comportant des marqueurs. Ces travaux sont décrits dans le paragraphe 2.1.1 et le chapitre 4.

Pär Hammarstedt, doctorant, Université de Malmö, Suède (deux mois, 2004). Pär a étudié le problème du calibrage mono-caméra à partir d'une mire 1-D, poursuivant les travaux de Zhengyou Zhang [131].

Il a proposé des solutions directes et a identifié les mouvements critiques vis-à-vis de ce problème, voir le paragraphe 2.1.1 et le chapitre 4.

Ferran Espuny, doctorant, UPC Barcelone, Espagne (deux mois, 2004). Ferran a étudié le problème de l'autocalibrage dans le cas de mouvements plans.

Nadia Lasri, stage d'été, (trois mois, 2000). Nadia a travaillé sur la reconnaissance de bâtiments à partir d'images.

Tomáš Svoboda, doctorant, Université Technique de Prague, République Tchèque (six mois, 1996). Tomáš a étudié l'influence des erreurs de calibrage sur l'estimation du mouvement d'une caméra. Ces travaux ont été publiés dans [107, 108].

1.2 Enseignement

Univ. de Zaragoza, Espagne	Cours doctoral, vision par ordinateur 3D	2×20h	2005 et 2006
INPG	Cours, DEA/M2R, vision par ordinateur 3D	6×12h	depuis 2000
INPG	Cours, DEA/M2R, optimisation numérique	6×6h	depuis 2000
Univ. Joseph Fourier, Grenoble	Cours, Magistère, vision par ordinateur	18h	1999
Reading Univ., Royaume-Uni	Cours, MSc, traitement d'images	4h	1998
TU Karlsruhe, Allemagne	Travaux Pratiques : C, Pascal, Assembleur	3×20h	1990-93

1.3 Co-organisation de colloques

Atelier BenCOS 2005 (Workshop Towards Benchmarking Automated Calibration, Orientation and Surface Reconstruction from Images), organisé lors du congrès ICCV à Pékin, en octobre 2005. Co-organisation avec Camillo Ressl (TU Vienne), Daniel Scharstein (Middlebury), Olaf Hellwich (TU Berlin) et Ilkka Niini (Mapvision Ltd., Finlande). Cet atelier a été organisé dans le cadre des activités du groupe de travail III/1 de l'ISPRS (cf. la section 1.5).
http://www.ipf.tuwien.ac.at/isprs/wgiii1/ws2005_wg12.html

Atelier OMNIVIS 2004 (Fifth Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras), organisé lors du congrès ECCV à Prague, en mai 2004. Co-organisation avec Tomáš Svoboda (Prague) et Seth Teller (MIT).
http://cmp.felk.cvut.cz/~svoboda/Omnivis2004/

Workshop on Mathematical Methods in Computer Vision, Banff International Research Station (BIRS) Alberta, Canada, en septembre/octobre 2006. L'organisateur principal est Martin Jägersand (Université de l'Alberta) et le comité d'organisation est composé de Bill Triggs (CNRS/GRAVIR), Anders Heyden (Université de Malmö, Suède), Steve Zucker (Yale), Jim Little (Université de British Columbia), Dana Cobzaş (Université de l'Alberta) et moi-même. Cet atelier est parrainé par le BIRS et fonctionnera selon un schéma d'invitation (il n'y aura pas d'appel à communications).
http://www.cs.ualberta.ca/~vis/vision06/

1.4 Participation à des comités de programme et activités de relecture

Courant 2006, je deviendrai membre de l'***Editorial Board*** de la revue *Image and Vision Computing*.

Je suis ***Area Chair*** pour ECCV 2006 (European Conference on Computer Vision, Graz, Autriche). J'ai participé à des comités de programme de plus de quarante congrès, symposiums ou ateliers :

ICCV	IEEE International Conference on Computer Vision	2005 (Pékin) 2003 (Nice)
ECCV	European Conference on Computer Vision	2004 (Prague) 2002 (Copenhague)
CVPR	IEEE Conference on Computer Vision and Pattern Recognition	2006 (New York) 2005 (San Diego) 2003 (Madison) 2001 (Kauai)
RFIA	Congrès de Reconnaissance des Formes et Intelligence Artificielle	2006 (Tours) 2004 (Toulouse)
ACCV	Asian Conference on Computer Vision	2006 (Hyderabad) 2004 (Jeju-Do)
ICPR	International Conference on Pattern Recognition	2006 (Hongkong) 2002 (Québec) 2000 (Barcelone)
ICIP	IEEE International Conference on Image Processing	2006 (Atlanta) 2005 (Gênes) 2004 (Singapour) 2003 (Barcelone) 2002 (Rochester) 2001 (Thessalonique)
BMVC	British Machine Vision Conference	2006 (Édimbourg) 2005 (Oxford) 2004 (Kingston)
ICASSP	Int. Conference on Acoustics, Speech and Signal Processing	2004 (Montréal) 2003 (Hongkong) 2002 (Orlando)
OMNIVIS	Workshop on Omnidirectional Vision, Camera Networks and Non-classical Cameras	2005 (Pékin) 2003 (Madison)
DV	Workshop on Dynamical Models for Computer Vision	2006 (Graz) 2005 (Pékin)
3DPVT	Symp. on 3D Data Processing, Visualization and Transmission	2006 (Chapel Hill)
VISAPP	Int. Conference on Computer Vision Theory and Applications	2006 (Setúbal)
ICVGIP	Indian Conference Computer Vision, Graphics and Image Proc.	2006 (Madurai)
ISVC	International Symposium on Visual Computing	2006 (États-Unis)
PCV	Photogrammetric Computer Vision (ISPRS Symposium)	2006 (Bonn)
CIPA	Vision Techniques Applied to the Rehabilitation of City Centres	2004 (Lisbonne)
ORASIS	Journées Jeunes Chercheurs	2005 (Fournol)
SACV	IEEE Workshop on Statistical Analysis in Computer Vision	2003 (Madison)
AmbInt	Workshop on Ambient Intelligence	2003 (Pise)
VI	Vision Interface Conference	2002 (Calgary)

J'ai effectué des relectures d'articles pour les revues suivantes :

IJCV	International Journal on Computer Vision
NATURE	
PAMI	IEEE Transactions on Pattern Analysis and Machine Intelligence
RA	IEEE Transactions on Robotics and Automation
RO	IEEE Transactions on Robotics
SMC	IEEE Transactions on Systems, Men and Cybernetics
MM	IEEE Transactions on Multimedia
IP	IEEE Transactions on Image Processing
JOE	IEEE Journal of Oceanic Engineering
VISP	IEE Proceedings Vision, Image & Signal Processing
CVIU	Computer Vision and Image Understanding
IJPRAI	International Journal of Pattern Recognition and Artificial Intelligence
IVC	Image and Vision Computing Journal
MVA	Machine Vision and Applications
JOSA	Journal of the Optical Society of America A
JCST	Journal of Computer Science and Technology (Chine)
JMIV	Journal of Mathematical Imaging and Vision
OE	Optical Engineering
PERS	Photogrammetric Engineering and Remote Sensing
PRL	Pattern Recognition Letters
JEI	Journal of Electronic Imaging
EL	Electronics Letters
RTI	Journal of Real-Time Imaging
TdS	Traitement du Signal
IJST	Iranian Journal of Science & Technology

J'ai été relecteur auxiliaire pour les congrès suivants :

SIGGRAPH	Conference on Computer Graphics and Interactive Techniques	2005
CVPR	IEEE Conference on Computer Vision and Pattern Recognition	1997, 2000
ICRA	International Conference on Robotics and Automation	2003
FSR	International Conference on Field and Service Robotics	2005
ISMAR	International Symposium on Mixed and Augmented Reality	2003
EG	Eurographics Conference	2002
ICCV	IEEE International Conference on Computer Vision	1998, 1999, 2001
ECCV	European Conference on Computer Vision	1996, 2000
CAIP	Int. Conference on Computer Analysis of Images and Patterns	1997, 1999
RFIA	Congrès de Reconnaissance des Formes et IA	2000

1.5 Responsabilités scientifiques

Animateur de l'action « Géométrie et Image » du GdR ISIS, 2006-09. Parmi les thèmes principaux de cette action se trouvent la reconstruction 3D à partir d'images, pour des scènes rigides ou déformables, et plus généralement tout ce qui peut toucher à une modélisation géométrique des capteurs, des images, de scènes 3D. L'action tente aussi d'établir des liens avec la communauté « Informatique Graphique », pour ce qui est de la visualisation de modèles, de la modélisation de scènes déformables, de la modélisation de propriétés de réflectance, etc. La tâche principale liée à l'animation de

cette action consistera à proposer et organiser des journées thématiques.
http://gdr-isis.org/

Animateur du Groupe de Travail III/1 « Automatic Calibration and Orientation of Optical Cameras » de l'ISPRS (International Society for Photogrammetry and Remote Sensing), 2004-08, avec Camillo Ressl (TU Vienne) et Ilkka Niini (Mapvision Ltd., Finlande). L'un des buts principaux de ce groupe de travail est de développer des *benchmarks* pour des méthodes automatiques d'estimation du mouvement ou d'autocalibrage à partir de séquences d'images. Ceci reposera sur la définition d'un protocole d'évaluation, en termes de critères et procédés, et sur la préparation et publication de données de référence (données « test »). Les animateurs de ce groupe de travail ainsi que ceux du groupe de travail III/2 (Surface Reconstruction) ont co-organisé l'atelier BenCOS 2005 (voir le paragraphe 1.3).
http://www.commission3.isprs.org/wg1/

Membre du « Task Group Single Images in Conservation » du CIPA (Comité International de la Photogrammétrie Architecturale). Le but de ce groupe de travail était de regrouper les chercheurs travaillant sur la reconstruction 3D à partir d'une seule image.

Membre du comité du Prix de Thèse SPECIF en 2000, 2001 et 2002 (prix attribué annuellement à une thèse en informatique, soutenue en France).

Rapporteur de thèses :
- John Mallon, Dublin City University, Irlande, 2005.
- Diego Ortín Trasobares, Universidad de Zaragoza, Espagne, 2005.
- Branislav Mičušik, Czech Technical University, Prague, 2004.
- Sébastien Cornou, Université Blaise Pascal, Clermont-Ferrand, 2004.
- Carlos Hernández Esteban, École Nationale Supérieure des Télécommunications, 2004.

Examinateur de thèses :
- Sio-Hoï Ieng, Université Pierre et Marie Curie (Paris 6), 2005.
- Srikumar Ramalingam, University of California, Santa Cruz, 2005 (qualifying examination).
- Neil Birkbeck, University of Alberta, Canada, 2005 (MSc examination).
- Omar Tahri, Université de Rennes 1, 2004.
- Jonathan Fabrizio, Université Pierre et Marie Curie (Paris 6), 2004.
- Vincent Frémont, École Centrale de Nantes, 2003.
- Nathalie Pessel, Université Montpellier II, 2003

1.6 Participation à des projets

CAVIAR (vision catadioptrique pour robots aériens), projet ANR – Agence Nationale de la Recherche, 2006-09. Partenaires : CREA (Amiens, coordinateur), LAAS (Toulouse), LE2I (Le Creusot), ICARE (INRIA Sophia-Antipolis) et MOVI (INRIA Rhône-Alpes). Le but de ce projet est de permettre l'utilisation de la vision omnidirectionnelle pour le contrôle de robots aériens et la reconstruction de terrains à partir d'acquisitions d'images aériennes.

VISIRE, projet Européen, 2000-03. Partenaires : Eptron SA (Espagne, coordinateur), Université de Lund (Suède), Université Polytechnique de Madrid (Espagne), Giunti Multimedia SA (Italie) et MOVI (INRIA Rhône-Alpes). Dans ce projet, un système de reconstruction 3D à partir de séquences vidéo a été développé. Nos contributions ont été l'implémentation et le transfert d'algorithmes d'autocalibrage, d'ajustement de faisceaux, de maillage de surfaces, de plaquage de textures et de génération de fichiers VRML. Elle sont issues de recherches sur la reconstruction 3D interactive (thèses d'Adrien Bartoli et Marta Wilczkowiak).

ParkNav, projet ROBEA – Programme Robotique et Entités Artificielles, 2002-2005. Partenaires : e-Motion (INRIA Rhône-Alpes, coordinateur), LAAS (Toulouse), LAGADIC (INRIA Rennes), PRIMA et MOVI (INRIA Rhône-Alpes). Le but du projet est de permettre l'utilisation de la vision pour la planification de trajectoires de véhicules dans des environnements dynamiques. Nos contributions concernent le calibrage de caméras dites « de surveillance », le suivi d'objets et le calcul des zones occupées par des obstacles mobiles.

CALIPSOO, projet Jeunes Chercheurs du GdR ISIS, 2002-2003. Projet bilatéral entre l'IRIT (Institut de Recherche en Informatique de Toulouse) et MOVI (INRIA Rhône-Alpes). Le sujet principal du projet était l'autocalibrage de caméras à partir de scènes planes (voir le paragraphe 2.1.2 et le chapitre 5).

Projets PRA – Programme Franco-Chinois de Recherches Avancées, 2001-02 et 1997-98. Projet bilatéral entre l'Université Xidian de Xi'an (Chine) et MOVI (INRIA Rhône-Alpes). Il s'agissait principalement de programmes d'échange, durant lesquels j'ai été accueilli par le laboratoire chinois, afin d'y donner des séminaires.

Projet PAI – Programme d'Actions Intégrées, 2001-02. Projet bilatéral entre l'Université de Kingston (Royaume-Uni) et MOVI (INRIA Rhône-Alpes). Nous avons développé une approche de calibrage de caméras dites « de surveillance ». Le dépôt d'un brevet sur notre approche a été préparé, mais n'a finalement pas abouti pour des raisons administratives, notamment parce que les modalités du co-financement n'ont pu être réglées.

VECTOR, projet EPSRC – Engineering and Physical Sciences Research Council, 1997-99. Mon post-doc à l'Université de Reading (Royaume-Uni) s'est déroulé dans le cadre de ce projet. Nous avons développé des méthodes de calibrage de caméras (basées sur l'utilisation de mires planes), de reconstruction 3D et de sélection de modèles pour la vision stéréoscopique (voir le paragraphe 2.4 et le chapitre 16).

CUMULI, projet Européen, 1996-2000. Partenaires : MOVI (INRIA Rhône-Alpes, coordinateur), ROBOTVIS (INRIA Sophia-Antipolis), Fraunhofer-IGD (Allemagne), Université de Lund (Suède), Imetric SA (Suisse) et Image Systems AB (Suède). Ma contribution a résidé dans l'étude des mouvements critiques pour l'autocalibrage, effectuée dans le cadre de ma thèse.

1.7 Communications invitées et tutoriaux

Avec mon doctorant Srikumar Ramalingam et Rahul Swaminathan (Deutsche Telekom Laboratories at the Technical University of Berlin), je donnerai un **tutorial dans le cadre de ECCV 2006** (European Conference on Computer Vision, Graz, Autriche) intitulé *General Imaging Design, Calibration and Applications*.
`http://perception.inrialpes.fr/people/Sturm/tutorial.html`

J'ai effectué les communications suivantes, en tant que conférencier invité, lors de congrès ou colloques :

- *Generic Models and Algorithms for Omnidirectional Vision*. Sanken International Symposium 2006 on Advanced Science and Technology for Materials, Biology, and Information, Osaka, Japon, février 2006.
- *Structure-from-Motion with Generic Camera Models*. Computer Vision Colloquium (à l'occasion de la réunion du comité scientifique – *Area Chair meeting* – de ECCV 2006), Graz, Autriche, janvier 2006.
- *A Generic Framework for Structure & Motion and Camera Calibration*. 12th Seminar on Theoretical Foundations of Computer Vision – Imaging Beyond the Pin-hole Camera, Schloß Dagstuhl, Allemagne, juin 2004.
- *Autocalibrage en vision par ordinateur : de la théorie à la pratique*. Journées Jeunes Chercheurs ORASIS, Gérardmer, France, mai 2003.

- *Calibration and Orientation for Omnidirectional Vision*. Workshop on New Developments in Close Range Photogrammetry, Bonn, Allemagne, mars 2003.
- *Mixing Catadioptric and Perspective Cameras*. Pattern Recognition and Computer Vision Colloquium, Prague, République Tchèque, novembre 2002.
- *Autocalibrage de caméras et reconstruction 3D à partir d'images*. Journées Nationales de la Recherche en Robotique, Hyères, France, octobre 2001.

J'ai donné des séminaires invités aux endroits suivants :
- Sanken, Osaka University, Japon, février 2006.
- CREA Lab, Amiens, France, novembre 2005.
- National Lab for Pattern Recognition, Pékin, Chine, octobre 2005.
- National Technical University of Athens, Grèce, avril 2005.
- Czech Technical University, Prague, République Tchèque, novembre 2004.
- KAIST (Korea Advanced Institute of Science and Technology), Daejeon, Corée du Sud, septembre 2004.
- GIST (Gwangju Institute of Science and Technology), Gwangju, Corée du Sud, septembre 2004.
- Seoul National University, Corée du Sud, septembre 2004.
- Sogang University, Department of Mathematics, Séoul, Corée du Sud, août 2004.
- Sogang University, Department of Media Technology, Séoul, Corée du Sud, août 2004.
- FGAN-FOM, Ettlingen, Allemagne, mai 2003.
- Malmö Högskola, Suède, avril 2003.
- Xidian University, Xi'an, Chine, octobre 2002.
- Chinese University of Hong Kong, octobre 2002.
- Hong Kong University of Science and Technology, octobre 2002.
- ETH Zürich, Suisse, avril 2001.
- LASMEA, Clermont-Ferrand, France, mars 2001.
- National Lab for Pattern Recognition, Pékin, Chine, mai 2000.
- Kingston University, Royaume-Uni, juin 1999.
- Siemens Corporate Research, Princeton, États-Unis, mai 1999.
- University of Zaragoza, Espagne, septembre 1999.
- Xidian University, Xi'an, Chine, avril 1997.

1.8 Responsabilités administratives et collectives

Je suis membre du **Groupe de Travail Actions Incitatives (GTAI)** de l'INRIA, qui est associé au **COST (Conseil d'Orientation Scientifique et Technologique)**. Ces organes ont été créés en 2004 ; la mission du COST est de superviser et de coordonner la structure scientifique d'ensemble de l'institut et de contribuer à la dynamiser et la développer. Pour ce faire, quatre groupes de travail y sont associés, dont le GTAI. Ses tâches principales sont d'organiser la sélection, le suivi et l'évaluation scientifique des actions incitatives de l'INRIA, telles que les ARC (Actions de Recherche Coopérative) et les ODL (Opérations de Développement Logiciel).

http://www.inria.fr/inria/cost/index.fr.html

Depuis 2000, je suis membre du **CLHS (Comité local hygiène et sécurité)** de l'INRIA Rhône-Alpes, qui est, entre autres, chargé de veiller sur les conditions de travail, d'hygiène et de sécurité au sein de l'institut, de proposer des solutions pour leur amélioration et de mener des enquêtes en cas d'accidents de travail.

J'ai été membre du **CUMI (Comité des utilisateurs des moyens informatiques)** de l'INRIA Rhône-Alpes, de 2001 à 2004.

Chapitre 2

Synthèse des travaux scientifiques

Dans ce document, je décris très succinctement mes activités de recherche depuis 1998. La grande majorité des travaux décrits ici ont été effectués depuis mon recrutement à l'INRIA, en tant que Chargé de Recherche (novembre 1999). Quelques travaux réalisés pendant mon séjour post-doctoral (1997-1999, à l'Université de Reading, Royaume-Uni) sont rappelés. Pour chaque travail, je citerai les noms des collaborateurs y ayant contribué, ceux-ci étant pour la plupart des doctorants (voir le tableau 3.1 de la page 26).

Mon domaine de recherche est la vision par ordinateur et la plupart de mes travaux ont trait aux problèmes de modélisation géométrique, qu'il soient inhérents à ce domaine ou plus « en périphérie ». J'ai travaillé sur un éventail de problèmes assez représentatif de cette thématique. Un des fils conducteurs de mes recherches a toujours été, non seulement de proposer des algorithmes permettant de résoudre de nouveaux problèmes ou de mieux résoudre des problèmes connus, mais aussi de bien comprendre la nature théorique des problèmes abordés.

Mes travaux peuvent être regroupés en quatre parties principales :
- calibrage et autocalibrage de caméras perspectives ;
- modèles génériques de caméras et unification de la vision 3D ;
- reconstruction 3D ;
- quelques autres travaux.

Calibrage et autocalibrage de caméras perspectives. Mes « travaux géométriques » d'avant 2004 ont été surtout placés sous l'hypothèse du modèle sténopé (« caméra perspective »), même si j'ai fait quelques excursions dans le monde de la vision dite omni-directionnelle, issue des nouvelles technologies de caméras à très grands champs de vue.

Les travaux de cette partie concernent le calibrage et l'autocalibrage de caméras. Les motivations principales étaient, d'une part, de proposer des méthodes pratiques pour différents scénarios d'applications et, d'autre part, d'étudier, à la fois, la faisabilité et les cas singuliers de ces méthodes, afin d'en guider l'emploi. Concernant le calibrage, nous avons proposé des méthodes utilisant des mires planes ou linéaires, d'un emploi très flexible, et avons étudié les solutions singulières relatives à ces méthodes. Concernant l'autocalibrage, nous avons considéré plusieurs cas de figure : le cas « minimal », très fréquent en pratique, où seule la distance focale est à calculer ; le cas de scènes planes, en particulier dans le cas où elles contiennent des cercles ; le cas d'une caméra effectuant des mouvements plans. L'étude des solutions singulières associées aux « mouvements critiques » a été menée pour deux de ces scénarios, complétant ainsi une étude déjà menée sur l'approche classique fondée sur les équations de Kruppa.

Modèles génériques de caméra et unification de la vision 3D. En 2003, je me suis posé la question de savoir s'il était possible de proposer une approche générique pour le calibrage de caméras. Cette question était suscitée par l'existence dans la littérature d'une pléthore de modèles et d'algorithmes pour le calibrage de différents types de caméra : caméra perspective « classique » ; caméra catadioptrique (système caméra+miroir) ; caméra de type *push-broom* ; caméra équipée d'un objectif *fish-eye*, etc. La réponse initiale n'était pas très compliquée : il suffit d'adopter un modèle générique de caméra, en l'occurrence un modèle qui associe un rayon visuel à chaque pixel. Le calibrage consiste à calculer les paramètres de ces rayons. Des méthodes pour cela existaient déjà, mais nous avons proposé la première qui permette de calibrer de façon « classique », c'est-à-dire en utilisant plusieurs images d'une mire de calibrage, acquises selon différents points de vue inconnus. Ce premier travail a été prolongé et son aboutissement a été de proposer quatre modèles génériques de caméras : caméra non centrale ; caméra axiale ; caméra centrale ; caméra radialement symétrique. Nous avons également posé les briques de base permettant de résoudre le problème du *structure from motion* pour ces modèles de caméras : estimation du mouvement ou de la pose ; reconstruction 3D ; étude de la géométrie d'images multiples.

Reconstruction 3D. Nos travaux sur la reconstruction 3D ont évolué ces dernières années : de méthodes produisant des modèles grossiers de la scène et nécessitant une interaction humaine, nous sommes arrivés à des méthodes automatiques et produisant des modèles denses de points. Les premières approches avaient pour but d'exploiter au mieux des contraintes géométriques sur la scène, décrites par un utilisateur ou connues *a priori*. Les contraintes considérées concernaient la coplanarité ou la colinéarité de points, l'orthogonalité ou le parallélisme de droites ou de plans, etc. Parmi les différentes méthodes que nous avons développées, certaines sont optimales, vis-à-vis du problème sous-jacent d'estimation de paramètres, tandis que d'autres donnent des solutions dans des scénarios où l'information est « minimale », par exemple pour la reconstruction à partir d'une seule image. Les modèles 3D obtenus sont plans par morceaux ; leur niveau de détail reste donc limité. Depuis, nous avons développé des méthodes pour une modélisation plus dense d'une scène. Le travail le plus récent concerne la modélisation « quasi-dense » de la géométrie d'une scène et de ses propriétés de réflectance. Préalablement à ceci, nous avons également considéré le cas de surfaces réflechissantes, ou bien spéculaires, et proposé des méthodes de reconstruction 3D de celles-ci.

Autres travaux. Deux autres travaux sont décrits dans ce mémoire : le suivi d'objets dans une séquence vidéo (travaux effectués dans le cadre d'une convention CIFRE) et la sélection de modèle pour la stéréovision.

Organisation. Dans ce qui suit, nous donnons un résumé en français très sommaire de tous ces travaux. Dans le chapitre 3 une introduction à ces travaux est fournie en anglais, suivie dans les parties II à V d'une description un peu plus détaillée de ceux-ci, toujours en anglais. Comme il est dit dans l'avant-propos, ce document est accompagné par un recueil de mes articles les plus représentatifs en rapport avec les travaux décrits. Ce recueil a la même structure que les parties II à V de ce document.

2.1 Calibrage et autocalibrage de caméras perspectives

2.1.1 Calibrage de caméras

Le calibrage de caméras est en principe un problème résolu depuis longtemps. Pourtant, la plupart des méthodes existantes étaient conçues pour une utilisation industrielle ou en laboratoire et nécessitaient un équipement particulier. Avec l'avènement des appareils photographiques et caméras numériques « grand public », le besoin de nouvelles méthodes de calibrage est devenu évident : nécessité de flexibilité de mise en œuvre, afin de pouvoir être accessibles au tout un chacun, mais aussi nécessité de modèles plus généraux, permettant de s'adapter à différent types d'appareils, comme ceux équipés d'objectifs spécifiques, tels le zoom.

Calibrage utilisant des mires planes. Les méthodes classiques pour le calibrage requièrent une mire de calibrage tridimensionnelle. Ceci n'est pas pratique : fabriquer des mires 3D est compliqué et coûteux ; les utiliser est contraignant. Dans [84, 97], nous avons montré comment calibrer des caméras en utilisant des mires planes. Ce type de méthode, proposée quasiment en même temps aussi par Z. Zhang [130], chercheur à Microsoft Corp. aux États-Unis, fait aujourd'hui office de standard dans la recherche pour le calibrage de caméras en vision par ordinateur.

Calibrage utilisant des mires linéaires. Ce sont typiquement des bâtons comportant des marqueurs. Cette approche est le standard industriel actuel pour le calibrage de systèmes multi-caméras dédiés à la capture du mouvement. Le point faible des systèmes existants est souvent le besoin d'interaction humaine pour initialiser ou contraindre le problème. Nous avons proposé et évalué différentes méthodes d'initialisation, inspirées des recherches de la dernière décennie en vision par ordinateur [69, 70].

En réalité, les méthodes décrites ci-dessus ne fonctionnent que pour des systèmes multi-caméras ; elles ne sont pas adaptées au calibrage d'un système mono-caméra. Z. Zhang a proposé, en 2002, une méthode qui résoud ce problème. Dans [42], nous avons proposé des solutions directes pour cette approche et avons mis en évidence les configurations dégénérées.

Calibrage de caméras équipées d'un zoom. L'utilisation du zoom, ainsi que la mise au point d'une caméra, modifient les paramètres dits intrinsèques (distance focale etc.) de celle-ci. Dans [115, 116], nous avons évalué ces effets. Nous en avons déduit dans quels cas les utilisations combinées du zoom et de la mise au point nécessitent ou non un calibrage spécifique.

2.1.2 Autocalibrage de caméras

L'autocalibrage de caméras ne requiert pas de mire en utilisant uniquement des images d'une scène quelconque. En photogrammétrie, le principe de l'autocalibrage est connu depuis des décennies. Pourtant, il a plutôt été traité dans le cadre de l'ajustement de faisceaux – c'est-à-dire de l'optimisation simultanée des paramètres des caméras et de la scène – en faisant l'hypothèse qu'une relativement bonne initialisation de ces paramètres est disponible. L'autocalibrage a été introduit en vision par ordinateur par Maybank et Faugeras en 1992 [55] ; son objectif principal était de pouvoir estimer tous ces paramètres sans avoir recours à des valeurs initiales.

Dans ce qui suit, je décris mes contributions théoriques et pratiques au problème de l'autocalibrage (depuis 1998).

Mouvements critiques dans l'approche basée sur les équations de Kruppa. Dans ma thèse [93], j'ai introduit la notion de mouvements critiques pour l'autocalibrage, c'est-à-dire des mouvements de caméra pour lesquelles l'autocalibrage est impossible, quelle que soit la méthode. En dehors des singularités inhérentes à ces mouvements critiques, certaines méthodes particulières peuvent être sujettes à des singularités supplémentaires. C'est le cas de l'approche basée sur les équations de Kruppa [129]. Dans [86], nous expliquons que cette approche possède plus de mouvements critiques que l'approche basée sur la quadrique duale absolue [112] qui, elle, en possède déjà plus que le problème de l'autocalibrage en lui-même.

Autocalibrage de la distance focale. La plupart des méthodes d'autocalibrage sont assez génériques, dans le sens où elles permettent de calibrer tous les paramètres intrinsèques. Néanmoins, en pratique, on connaît souvent de bonnes approximations de certains de ces paramètres et, typiquement, seule la distance focale est à calculer. Nous avons alors proposé des méthodes spécifiques pour ce problème (plus simple) qui sont d'une grande utilité en pratique. Dans [90, 94], nous avons abordé le cas d'une distance focale qui varie durant l'acquisition d'une séquence d'images et dans [87, 104] celui d'une distance focale constante. Pour les deux cas, nous avons décrit, de façon exhaustive, les mouvements critiques de caméra, vis-à-vis desquels l'autocalibrage échoue.

Autocalibrage dans le cas de mouvements plans. Toute connaissance sur la structure de la scène ou le mouvement de la caméra permet de contraindre le problème de l'autocalibrage. Nous avons étudié plusieurs cas de figure : prise en compte de la planéité de la scène, incluant le cas particulier où celle-ci contient des cercles (voir les deux paragraphes suivants) ; prise en compte du fait que la caméra effectue des mouvements plans. Par mouvements plans, nous faisons référence aux mouvements se composant de translations dans un plan et de rotations autour de la normale à ce plan. Dans [32], nous montrons que le problème de l'autocalibrage peut alors essentiellement être ramené à l'autocalibrage d'une caméra linéaire (caméra avec un segment de droite comme rétine, évoluant dans un monde 2D). Ce dernier problème est assez facile à résoudre, grâce à l'introduction du tenseur trilinéaire entre trois vues de la caméra linéaire qui, lui, est beaucoup plus simple à caractériser et à estimer que celui entre trois vues planes.

Autocalibrage utilisant des scènes planes. Comme pour le calibrage (paragraphe 2.1.1), la prise en compte de la planéité de la scène est également utile pour l'autocalibrage. Le concept de l'autocalibrage à partir de scènes planes a été introduit par Triggs [113] et nous avons proposé différentes extensions théoriques et algorithmiques, en vue d'une application pratique plus aisée [40, 52, 60, 96].

Autocalibrage utilisant des images de cercles. Dans [100], nous avons introduit l'idée de calibrer une caméra en utilisant des images de cercles coplanaires ou avec des plans de support parallèles. Une des applications les plus intéressantes est le calibrage d'une caméra acquérant une séquence d'images d'un objet posé sur une table tournante : chaque point sur la surface de l'objet suit une trajectoire circulaire en 3D et toutes ces trajectoires ont des plans de support parallèles entre eux.

Le problème dual de ce type de calibrage est la détermination de la structure euclidienne des plans supportant les cercles considérés. À partir des projections de deux cercles, on obtient en général deux solutions pour cette structure ; selon la configuration (cercles ayant ou pas des points d'intersection réels, cercle incluant l'autre, etc.), la solution incorrecte peut être éliminée. Dans [41], nous donnons un traitement complet pour ce problème, pour toutes les configurations de deux cercles. Nous proposons également un algorithme permettant de résoudre ce problème à partir des projections de plusieurs cercles prallèles dans une image, *via* la résolution d'un système surdéterminé d'équations linéaires.

Calcul optimal de la matrice fondamentale. Dans [7] nous proposons un algorithme pour l'optimisation non linéaire efficace de la matrice fondamentale entre deux images. Cet algorithme utilise un paramétrage de la matrice fondamentale, par les facteurs de la décomposition en valeurs singulières de celle-ci (deux matrices orthogonales et une matrice diagonale) dont la mise à jour lors de l'optimisation non linéaire est aisée.

2.2 Modèles de caméra génériques et unification de la vision 3D

Depuis 3 ans, nous travaillons sur des extensions de la vision 3D par ordinateur, permettant de passer des modèles de caméras perspectives à des modèles plus généraux de caméras. Nous avions fait le constat que la littérature du domaine propose une multitude de méthodes pour le calibrage et la résolution du problème du *structure from motion* et que ces méthodes sont pour la plupart conçues pour un seul type de caméra à la fois. Ainsi, des méthodes différentes existent pour les caméras perspectives, catadioptriques, *fish-eye*, *push-broom*, etc. Il nous a semblé désirable de disposer d'une théorie et d'algorithmes unifiés, qui seraient applicables à n'importe quelle caméra. Dans ce qui suit, nous décrivons une série de travaux effectués dans cette lignée. Certains de ces travaux, dont ceux sur le calibrage, l'autocalibrage et la géométrie d'images multiples, sont les premiers dans la littérature qui atteignent un tel niveau de généralité.

Calibrage. Nous avons adopté un modèle générique de caméra qui consiste à associer, à chaque pixel de l'image, un rayon visuel dans l'espace 3D. Ce modèle regroupe quasiment toutes les technologies de caméras utilisées en vision par ordinateur. Dans le cas le plus général, les rayons visuels peuvent être quelconques, nous parlons alors d'une *caméra non centrale*. Nous considérons également les deux cas particuliers suivants : pour une *caméra centrale* tous les rayons visuel se coupent en un seul point tandis que pour une *caméra axiale* ils coupent tous une seule droite 3D. Nous avons proposé des algorithmes de calibrage pour ces trois modèles génériques [72, 73, 75, 76, 101, 102]. Par rapport aux quelques approches existantes, ces algorithmes présentent l'avantage d'être plus flexibles (acquisition selon des points de vue quelconques).

Les travaux précédents concernaient des modèles très génériques de caméras. Nous avons également exploré un modèle légèrement plus spécifique, mais qui s'applique néanmoins à la plupart des caméras courantes : il décrit les distorsions radialement symétriques auxquelles sont sujettes beaucoup de caméras. L'utilisation de cette contrainte rend le calibrage possible à partir d'une seule image et il devient en général plus aisé et stable [109].

Autocalibrage. Dans [74], nous proposons une première approche d'autocalibrage pour le modèle de caméra centrale. Elle ne requiert que l'extraction du flot optique ou bien de correspondances denses, entre des images d'une scène inconnue.

Nous avons également exploré l'autocalibrage pour le modèle de caméra à distorsions radialement symétriques, mentionné ci-dessus. Une méthode de type *plumb-line* est d'abord introduite, c'est-à-dire une méthode qui utilise les projetés de segments droits. Cette méthode est ensuite utilisée pour l'autocalibrage à partir d'une scène plane, de structure inconnue [110]. Deux vues suffisent en général pour réaliser l'autocalibrage de la distorsion. Pour réaliser un autocalibrage complet, on se ramène au cas d'une caméra perspective « classique », ayant la distance focale comme seule inconnue ; une solution existe à partir d'au moins trois images (cf. [113]).

Vision 3D. Dans [71, 103, 105], nous décrivons des approches génériques pour les tâches de base du *structure from motion* : estimation de la pose d'un objet, estimation du mouvement d'un objet et reconstruction d'un objet. Ces approches sont simples et applicables pour tout modèle de caméra.

Géométrie d'images multiples. Des correspondances de points ou d'autres primitives entre des images donnent des informations sur : la structure de la scène, le calibrage de la caméra et ses mouvements. L'une des briques de base de la « géométrie d'images multiples » est constituée des « tenseurs d'appariement » : tenseurs qui encapsulent le calibrage de la caméra et ses mouvements. La connaissance de ces tenseurs pour un ensemble d'images permet de contraindre la mise en correspondance de primitives. Réciproquement, des correspondances permettent le calcul des tenseurs, desquels peuvent ensuite être extraits les paramètres de calibrage et du mouvement de la caméra. Dans [91, 103] nous développons cet aspect de la géométrie d'images multiples, pour les modèles génériques de caméras introduits ci-dessus. Concrètement, nous avons montré comment établir les tenseurs d'appariement pour ces caméras ; ces tenseurs sont définis par rapport aux coordonnées de Plücker des droites représentant les rayons visuels, tandis que les tenseurs d'appariement « classiques » le sont par rapport aux coordonnées homogènes de points image.

L'ensemble de ces contributions théoriques et des algorithmes des paragraphes précédents constitue donc un traitement unifié de la géométrie d'images multiples et du *structure from motion* ; il généralise ainsi une certaine partie de la littérature existante sur ces sujets.

Avant de considérer les modèles génériques de caméras, nous avons effectué un premier travail visant à donner une vision unifiée de certains concepts relatifs à la géométrie d'images multiples, qui reste valable pour différents modèles de caméras. Notamment, nous avons montré que, pour des systèmes constitués de n'importe quel assemblage de caméras perspectives, affines et para-catadioptriques (capteur omnidirectionnel constitué d'une caméra affine et d'un miroir parabolique), des tenseurs d'appariement existent [89]. De plus, nous avons donné des exemples de géométrie épipolaire et introduit le concept d'autocalibrage pour de tels systèmes de caméras hétérogènes.

2.3 Reconstruction 3D

Beaucoup de travaux en reconstruction 3D visent à formuler le problème de la façon la plus générale qui soit : traitements complètement automatiques et abandon d'hypothèses sur le contenu de la scène. Au paragraphe 2.3.3 nous décrivons les travaux que nous avons effectués dans cette optique. Pourtant, la reconstruction 3D est encore aujourd'hui un problème très difficile et la résolution du problème – dans le sens d'une totale généralité – est loin d'être atteinte. En parallèle, nous avons donc aussi exploré une voie alternative, renonçant à l'une ou l'autre des exigences mentionnées ci-dessus. Les scènes urbaines et, plus généralement, les objets produits par l'humain n'ont le plus souvent pas des formes quelconques. Par la suite, nous décrivons nos différents travaux qui exploitent des hypothèses plus ou moins fortes sur la forme des scènes dont on veut obtenir un modèle 3D à partir d'images. Ces travaux correspondent à peu près aux travaux de thèse de Marta Wilczkowiak (soutenue en 2004) et d'Adrien Bartoli (soutenue en 2003).

Certains des travaux décrits dans le paragraphe précédent ont abouti à des modèles 3D de scènes, modèles plutôt grossiers (typiquement constitués de plans dominants). Nous avons également travaillé sur comment obtenir des modèles 3D plus denses à partir d'images. Il s'agit d'un problème très difficile, puisque les images dépendent de la géométrie de la scène et des caméras, mais également de l'illumination et des propriétés de réflectance de la scène. Nous avons donc, tout d'abord (comme il est de coutume dans le domaine) adopté l'hypothèse de réflectance lambertienne, mais nous avons également travaillé sur le cas diamétralement opposé des surfaces spéculaires (travaux de thèse de Thomas Bonfort). Depuis peu, nous avons commencé à nous pencher sur l'estimation conjointe de la géométrie et de la réflectance de la scène.

2.3.1 Utilisation de contraintes géométriques en vision 3D

Scènes planes par morceaux. Dans [2, 5, 10–13, 27, 28] nous proposons des représentations et méthodes pour reconstruire des scènes qui contiennent des structures planes dominantes (par exemple, des murs). Nous avons montré que la reconstruction 3D est plus stable dans ce cas, même si ces structures ne sont pas parfaitement planes dans la réalité.

Nous avons montré que le fait que la surface d'un objet comporte des parties planes est également très utile pour le suivi de cet objet dans une vidéo [29].

Vision 3D pour des segments de droite. La plupart des travaux en reconstruction 3D et estimation du mouvement s'appuient sur des correspondances de *points* entre images. Néanmoins, les surfaces de beaucoup d'objets artificiels contiennent des arêtes et il semble donc utile de profiter de cette connaissance. Un certain nombre de travaux ont été faits sur ce sujet et nous les avons étendus selon plusieurs directions. Dans [3, 4, 8], nous proposons un éventail de méthodes pour l'estimation du mouvement d'une caméra, basées sur des correspondances de segments de droite entre les images. Ceci pour les cas de caméras calibrées et non calibrées. Dans [6, 9], nous proposons des méthodes pour la reconstruction 3D.

Contraintes géométriques plus générales. La plupart des résultats décrits dans les deux paragraphes précédents, donnaient lieu à des traitements automatiques. Nous résumons maintenant nos travaux qui supposent une certaine interaction de la part d'un utilisateur. Concrètement, nous supposons qu'un utilisateur apporte des annotations sur des images, correspondant à des contraintes géométriques telles que : deux segments de droites sont parallèles ou orthogonaux, tels points sont coplanaires, un segment de droite est parallèle ou orthogonal au plan engendré par ces points, etc. La prise en compte de telles contraintes permet d'obtenir des reconstructions 3D plus stables et surtout de les obtenir avec beaucoup moins d'images que dans le cas général, parfois même avec une seule image. La théorie complète de notre approche est décrite dans [118–123, 125, 126]. Dans [85, 98, 117], le cas particulier de la reconstruction 3D à partir d'une seule image est traité.

Nos travaux ont vite quitté le domaine purement géométrique et nous avons commencé une collaboration avec des spécialistes des techniques dites *constraint decomposition techniques* (collaboration qui a été initiée et animée complètement par M. Wilczkowiak). Des techniques pour trouver automatiquement de bons paramétrages pour un problème à partir d'un ensemble non organisé de contraintes ont été adoptées, permettant des reconstructions 3D, à partir de plusieurs images [124].

2.3.2 Reconstruction 3D de scènes dynamiques

Points se déplaçant dans des plans. La plupart des « travaux géométriques » en vision par ordinateur considèrent des scènes statiques. Nous avons étudié un scénario dynamique : la scène est constituée de points qui bougent indépendamment les uns des autres, la seule contrainte étant que chaque point bouge dans un plan et que ces plans forment un faisceau. Dans [88], nous avons montré que même dans ce scénario dynamique, l'estimation du mouvement et la reconstruction 3D sont possibles.

Utiliser la contrainte de gravité. Un travail un peu anecdotique était de répondre à la question suivante : peut-on calibrer un système multi-caméras en observant des objets mobiles en vol libre ? À titre d'exemple, notre idée serait de pouvoir calibrer en utilisant les images des trajectoires des objets manipulés par un jongleur ou d'une balle de tennis. Nous avons proposé une formulation théorique de ce problème et des algorithmes le résolvant [99].

2.3.3 Reconstruction 3D dense à partir d'images multiples

Le projet Européen VISIRE avait pour but la reconstruction de modèles 3D à partir de vidéos, acquises à l'aide d'un caméscope porté à l'épaule. Au sein de ce projet, une chaîne complète de tâches de vision par ordinateur a été mise en œuvre par les partenaires, allant de l'extraction et de l'appariement de primitives, *via* l'autocalibrage et la reconstruction 3D de points, jusqu'à un maillage de surface en 3D. Quant au maillage, nous avons dans un premier temps développé des méthodes prenant en compte des contraintes de visibilité et de photo-consistance [78, 79]. Ce maillage était défini sur les points 3D reconstruits et était donc encore relativement grossier.

Dans [36, 37], nous proposons une approche de reconstruction *dense* de surfaces ; la représentation de surface que nous utilisons est l'union des cartes de profondeur, associées aux images d'entrée. Une formulation bayésienne pour l'estimation de ces cartes de profondeur et des couleurs associées a été mise en œuvre et des résultats très prometteurs sur des scènes difficiles ont été obtenus.

2.3.4 Reconstruction de surfaces spéculaires

Nous avons proposé deux approches pour la reconstruction de surfaces spéculaires à partir d'images. Elles nécessitent l'utilisation d'un objet de référence. Dans la pratique, nous utilisons une mire de calibrage plane. Des images de sa réflexion sur la surface donnent des informations sur la forme de cette dernière.

Nos deux méthodes supposent des conditions d'acquisition complètement maîtrisées : caméra calibrée et poses de la caméra et de la mire connues dans un repère global (pour des déplacements de la caméra et/ou de la mire). Nous avons proposé plusieurs approches pour le calcul de pose (voir plus bas). L'une des méthodes de reconstruction procède comme suit. La caméra est stationnaire et acquiert deux images ou plus de la réflexion de la mire, mise dans autant de positions différentes. Nous supposons qu'un appariement dense entre les pixels de l'image et des points de la mire a été établi pour chaque image[1]. Pour chaque pixel, nous construisons deux droites 3D : celle du rayon visuel associé au pixel (définie par le calibrage de la caméra) et la droite reliant les points appariés, pour les différentes positions de la mire. L'intersection de ces deux droites est alors un point sur la surface du miroir [21]. Ainsi, la surface est reconstruite comme un nuage de points dense. De bonnes précisions (erreur de reconstruction en-dessous du millième de la taille de l'objet reconstruit) ont été obtenues sur des images réelles de scènes simples (ensembles de miroirs plans). Notons que Kutulakos et Steger ont récemment proposé une approche plus générale, s'appliquant également à des réfractions et des doubles réflexions [53].

L'autre approche que nous avons développée s'appuie sur une discrétisation du volume de travail en *voxels* [19, 20] et représente une extension du *voxel coloring* [80] aux surfaces spéculaires.

En dehors de ces méthodes de reconstruction, nous avons également proposé plusieurs méthodes pour le calcul de pose dans notre scénario. Soulignons que les meilleures reconstructions sont obtenues si la surface spéculaire couvre au mieux le champ de vue de la caméra ; ceci exclut que la caméra ait une vue directe de la mire en plus de celle de sa réflexion. Il s'agit alors de calculer la pose de la mire par rapport à la caméra, sans vue directe. La méthode la plus générale calcule la pose à partir des réflexions dans une surface miroir inconnue, pour deux positions de la mire [18]. Il s'agit d'une méthode très élégante et qui donnerait lieu à une grande flexibilité de reconstruction d'une surface spéculaire : il faut simplement placer une mire à deux ou plus de positions inconnues et d'en observer les réflexions dans la surface spéculaire. Malheureusement, les calculs se sont avérés trop peu stables en pratique. Nous avons développé deux autres méthodes plus contraignantes mais qui donnent de bons résultats. L'une d'elle s'apparente au calibrage dit *hand-eye* [18]. La deuxième s'appuie sur l'observation de la mire dans un miroir plan, selon trois ou plus positions différentes [95].

2.3.5 Reconstruction conjointe de la géométrie et de propriétés de réflectance

Nous sommes convaincus que la prise en compte des propriétés de réflectance des objets à reconstruire et de l'illumination environnante est importante, à plusieurs titres. Tout d'abord, l'estimation des propriétés de réflectance d'un objet mènera à des modèles plus riches permettant des rendus de meilleure qualité. Plus généralement, les images sont fonction de tous ces facteurs : géométrie de la scène et de la caméra, propriétés de réflectance de la scène et radiométriques de la caméra, illumination. En principe, il faudrait donc les modéliser tous même si l'on en veut reconstruire qu'un sous-ensemble. Pourtant, ceci n'est pas possible sans informations supplémentaires en dehors des images ou sans faire des hypothèses simplificatrices. Il existe beaucoup de possibilités pour ce faire : adoption d'un modèle de réflectance particulier (lambertien, de Phong, ...), modélisation hors ligne de l'illumination environnante (*environment maps*), hypothèses sur la géométrie de la scène (absence d'inter-réflexions, conformité à une famille de modèles appris au préalable, par exemple l'hypothèse qu'il s'agit d'une scène urbaine), etc. Dans mes recherches à venir, je veux explorer ces possibilités.

Mon premier travail dans cette direction a été de proposer une approche pour la reconstruction conjointe de la géométrie et de propriétés de réflectance. Les données en entrée sont des séquences d'images, représentant un objet posé sur une table tournante, donc effectuant un mouvement circulaire, acquises par une

[1] Dans la pratique, nous utilisons un écran plat comme mire et y affichons une série de motifs binaires (codes *Gray*). Cela permet une mise en correspondance très dense et assez facile à mettre en œuvre.

caméra fixe. L'objet est éclairé par une source de lumière « ponctuelle » placée à une position différente pour chaque séquence. La position de la source de lumière est pré-calibrée et le mouvement de l'objet est estimé. Chaque séquence est acquise à l'aide de marqueurs sur la table tournante. Un modèle 3D initial de l'objet est obtenu à partir de son enveloppe visuelle. Puis, un maillage triangulaire en est extrait et on le fait évoluer. L'évolution est contrôlée par une mesure de photo-consistance et un *a priori* habituel sur la géométrie de l'objet. L'évolution du modèle géométrique est alternée avec l'estimation d'un modèle de réflectance spatialement variable (modèle de Phong). Des résultats prometteurs ont été obtenus [17].

2.4 Autres travaux

Suivi d'objets. Dans le cadre d'un contrat CIFRE, nous travaillons sur le suivi d'objets dans des séquences d'images. Les contraintes pratiques sont que l'on ne dispose d'aucun modèle précis des objets à suivre et que le suivi doit se faire en partant d'une initialisation manuelle extrêmement simple (l'utilisateur clique au milieu de l'objet dans une image ou dessine une ellipse englobante). La principale approche que nous avons mise en œuvre s'appuie sur un filtrage à particules ; ainsi, une estimation de la densité de probabilité de l'état de l'objet (position, taille, forme) est propagée et mise à jour d'une image à l'autre. L'un des principaux problèmes sur lesquels nous nous sommes penchés est la représentation de l'objet. Nous avons utilisé plusieurs représentations ; représentation de l'apparence (histogrammes pondérés de couleurs) ou de la forme (formes de base planes ou volumiques). Deux aspects importants de nos travaux sont : *(i)* de permettre aux algorithmes de changer de représentation de l'objet au cours du temps ; ainsi, si l'on s'approche de plus en plus d'un objet, une représentation initiale par une forme plane peut évoluer en une représentation volumique, plus précise [49] ; *(ii)* de choisir automatiquement une dimension appropriée (nombre de cases) des histogrammes de couleurs, en effect, celle-ci est souvent donnée explicitement dans les approches existantes ; pour ce faire, nous appliquons un critère de sélection de modèles [47, 48], issu de la littérature statistique.

Sélection de modèle. Beaucoup de problèmes en vision par ordinateur ont comme clef de voûte la question de sélection de modèle. Le problème est de choisir un modèle à partir d'observations, selon les deux critères antagonistes : simplicité du modèle et ajustement fin de celui-ci aux données. Le problème que nous avons considéré est celui de l'estimation du mouvement et de la reconstruction 3D à partir de deux images. Les deux modèles concurrents sont celui d'une scène plane et celui d'une scène volumique, modèles qui sont validés par l'existence d'une homographie ou d'une matrice fondamentale qui soient en accord avec des correspondances de points entre les deux images. Pour le choix du meilleur modèle, nous avons utilisé le concept de *minimum description length* (MDL) [56, 57].

Chapter 3

Introduction in English

This document gives a summary of my research and related activities, over approximately the last 7 years. Research related activities, including the supervision of students, involvement in projects, reviewing, organization of events, as well as teaching and administrational duties, are given in French in chapter 1, followed by a concise description of my research. I now switch to English, for an extended summary of my research. Most of "my" research is actually the outcome of collaborations with my students and colleagues in other institutes, see further below.

My research area has been Computer Vision and most of my research concerns its geometrical branch. Its general aim is to obtain descriptions of the "world", which for us consists of cameras and objects (the "scene"), from images acquired by these cameras. The way objects look like in images depends on many factors: their shape, their reflectance properties, the illumination, the position of the camera and its intrinsic properties (geometrical projection law and radiometric properties). Many types of image-based object descriptions have been studied in computer vision; they all have some degree of invariance to (combinations of) the above factors. Different branches of computer vision study the extraction and use of different such object descriptions. Most recent approaches to object recognition for example, use descriptions of object parts that are intended to be to some degree invariant to viewpoint, illumination, camera intrinsic parameters, object shape (e.g. are applicable to deformable objects), or to be invariant within a class of similar objects. In addition, many approaches incorporate loose geometric descriptions, characterizing the relative location of object parts (usually in the image plane); these descriptions are designed to bring about a better discriminative power to object recognition, while being somewhat invariant within object classes, as well as to object shape, viewpoint, etc.

Structure from motion or other modalities (shading, focus, defocus, ...) usually aims at obtaining explicit geometric object descriptions: shape of objects (invariant to euclidean or other transformations), motion thereof, camera positions, etc. These descriptions typically should exactly characterize specific objects, i.e. they are not intended to be invariant to object shape or even within object classes. However, the notion of object class may still play an important role; indeed, characterizations of an object class (e.g. buildings of a certain architectural style, faces, cars) may be used as priors (or, regularizers) to guide the recovery of a specific object's geometrical model. Such approaches are successfully used in city modelling from aerial images or 3D modelling of faces. However, these typically use hand-made models of object classes; few approaches exist that would rely on automatically learnt object class descriptions. One of the most noticeable examples is the family of approaches for marker-less human body motion estimation from images; many promising approaches rely on learnt descriptions of typical body motions.

Most works in image-based 3D modelling, like a large part of ours, have considered rigid objects. In addition, many approaches have addressed the modelling of primitives, most often points and sometimes line segments or other primitives, leading to sparse 3D models. Often, only geometrical laws have been used to obtain results, in the form of feasibility theorems and algorithms. It is a common opinion within the computer vision community that the potential of geometry-only approaches has mainly been exhausted today. This is mostly true, although for example the case of deformable or piecewise rigid objects continues to be actively researched.

As for image-based 3D modelling as a general topic, it is by far not accomplished; to my knowledge, no reasonably working system exists today that would produce complete surfacic or volumetric object descriptions under a variety of conditions with respect to illumination, complexity of object shape, reflectance properties of objects, or non-rigid object motion. There may be various avenues to reach that goal. One consists in adopting richer models (non-Lambertian reflectance models, non-sparse surface models, models for non-rigid motion, etc.) and estimating them. Another consists in providing prior knowledge of various kinds, e.g. a pre-calibrated model of the illumination, a catalogue of typical materials with their reflectance properties, models of typical object shapes or motions. Such prior knowledge can consist in a collection of examples, or can be models learnt from observations. Much remains to be done...

After this general but brief introduction, let us come back to the main focus of this document, my works over the last few years. The three main threads of work correspond to the titles of the following three parts of this document:

- Calibration and Self-Calibration of Perspective Cameras
- Generic Camera Models and Unified Treatment of Structure from Motion
- 3D Reconstruction

Part II on (self-) calibration of perspective cameras, as well as the first two chapters in part IV on 3D reconstruction, describe works in the classical multi-view geometry style: estimating models of camera and scene geometry from primitives extracted in images and correspondences thereof. We consider four main scenarios for this question. *(i) Calibration* using images of objects of known structure (here, planar or linear objects). *(ii) Self-calibration* using images of arbitrary, rigid scenes. *(iii) Using geometric constraints* on relative positions of primitives (collinearity or coplanarity of points, known distances between points, orthogonality or parallelism of lines and planes, etc.), for 3D reconstruction, but also for estimating camera calibration and pose. *(iv) 3D reconstruction of dynamic scenes.* This scenario is not dealt with as generally as for the above three; we only consider two special and idealized cases of dynamic scenes.

The last three chapters of part IV describe works on image-based 3D modelling outside the geometry-only multi-view approach. We have worked on three different aspects. *(i) Dense 3D modelling* from multiple images, i.e. reconstruction of surface models. Approaches using triangular mesh-based models and depth maps are described. *(ii) 3D reconstruction of specular surfaces.* We describe our geometry-based approaches for the opposite of the Lambertian reflectance model – specular reflection. *(iii) Modelling of 3D geometry and reflectance properties.* Here, we describe our first work on estimating dense surface and reflectance models from images, based on pre-calibrating the illumination.

Part III is dedicated to our works on using highly general camera models in structure from motion related problems: calibration, self-calibration, 3D modelling, pose and motion estimation, multi-view geometry. We use different levels of general camera models: in the most general one, the set of pixels is associated with a set of projection rays, that may not have a common optical center (non-central camera). This is then specialized to general models for central and so-called axial cameras (all projection rays touch one point or one line), as well as radially symmetric models (the association between pixels and projection rays is radially symmetric around a distortion center in the image and an optical axis in 3D).

In part V, we describe a few other works, on object tracking and model selection.

Part VI contains concluding remarks and describes some ideas for our future work.

The scope of this document is to give a concise overview of my research (and related activities, see chapter 1); for complete and detailed descriptions, I refer to the accompanying collection of the most representative publications. It has the same structure as the following parts of this document.

The majority of the works described in this document are the result of collaborations with students and colleagues at INRIA or other labs. A synopsis of these collaborations is given in table 3.1. For many works, the main credit is due to my collaborators.

Calibration and self-calibration of perspective cameras			
Calibration			
• Planar calibration grids	[84, 97]	S. Maybank	Reading (UK)
• Linear calibration objects	[69, 70]	T. Pribanić	Zagreb (Croatia)
	[42]	P. Hammarstedt	Malmö (Sweden)
• Calibration of zoom lenses	[115, 116]	M. Urbanek, R. Horaud	MOVI
Self-calibration			
• Critical motions for Kruppa	[86]		
• Planar motions	[32]	O. Faugeras, L. Quan	ROBOTVIS, MOVI
• Variable focal length	[90, 94]		
• Constant focal length	[87, 104]	Z. Cheng	Singapore
• Planar scenes	[40, 96]	P. Gurdjos	IRIT (Toulouse)
	[52, 60]	F. Jurie, E. Noirfalise	LASMEA (Cl.-Fd.)
• Using images of circles	[41]	P. Gurdjos, Y. Wu	IRIT, NLPR (China)
• Optimal computation of F	[7]	A. Bartoli	MOVI
Generic camera models and unified treatment of Structure from Motion			
Calibration			
• General framework	[72, 73, 101, 102]	S. Ramalingam, S. Lodha	MOVI, UCSC (USA)
• Axial cameras	[75, 76]	S. Ramalingam, S. Lodha	MOVI, UCSC (USA)
• Radially symmetric cameras	[109]	J.-P. Tardif	Montréal (Canada)
Self-calibration			
• Central cameras	[74]	S. Ramalingam, S. Lodha	MOVI, UCSC (USA)
• Radially symmetric cameras	[110]	J.-P. Tardif, S. Roy	Montréal (Canada)
Structure from motion			
• Algorithms	[71, 103, 105]	S. Ramalingam, S. Lodha	MOVI, UCSC (USA)
• Multi-view geometry	[89, 91, 103]	S. Ramalingam	MOVI, UCSC (USA)
3D Reconstruction			
Using geometric constraints for 3D vision			
• Piecewise planar scenes	[2, 5, 10–13]	A. Bartoli, R. Horaud	MOVI
	[27, 28]	O. Chum, T. Pajdla	Prague
	[29]	D. Cobzaş	MOVI
• Lines	[3, 4, 6, 8, 9]	A. Bartoli	MOVI
• More general constraints	[85, 98, 117–123, 125, 126]	M. Wilczkowiak, E. Boyer	MOVI
	[124]	··· and G. Trombettoni	COPRIN (Sophia-A.)
3D Reconstruction of dynamic scenes			
• Points moving in planes	[88]		
• Exploiting gravity	[99]	L. Quan	MOVI
Multi-view dense 3D reconstruction			
• Mesh-based models	[78, 79]		
• Dense models	[36, 37]	P. Gargallo	MOVI
3D reconstruction of specular surfaces			
• *Voxel carving*	[19, 20]	T. Bonfort	MOVI
• Triangulation	[21]	T. Bonfort, P. Gargallo	MOVI
• Pose computation	[95]	T. Bonfort	MOVI
Modelling of 3D geometry	[17]	D. Cobzaş,	MOVI
and reflectance properties		N. Birkbeck, M. Jägersand	Alberta, Canada
Other works			
• Object tracking	[47–49]	A. Jacquot, O. Ruch	MOVI, Thales
• Model selection	[56, 57]	S. Maybank	Reading (UK)

Table 3.1: Synopsis of collaborations and publications for the individual works.

Part II

Calibration and Self-Calibration of Perspective Cameras

Introduction

Camera calibration is in principle solved since dozens of years, see e.g. [24, 82]. Perhaps the most widely used approach is based on acquiring images of a calibration object that usually carries a set of markers with accurately known 3D coordinates. Best results are achieved when these markers are well distributed over the camera's field of view, i.e. over some depth range and such that their projections cover well the image area. All this implies that appropriate calibration objects have a three-dimensional structure and are carefully manufactured or measured (e.g. using theodolites). These requirements are acceptable for industrial applications and more generally whenever highly accurate calibration matters or whenever the user does not trust other approaches.

In other circumstances it may be preferable to alleviate the above requirements. Various approaches have been proposed to this end, using calibration objects with less accurate knowledge of 3D coordinates, calibration objects of some constrained structure, or no calibration object at all.

The first approach, using calibration objects with less accurate knowledge of 3D coordinates, is well described in [54]. The main idea is to use the standard calibration approach just for getting initial values of parameters of interest: intrinsic parameters and camera poses. These parameters are then optimized, together with the 3D coordinates of the calibration markers, which are now considered as unknowns. This is feasible if sufficiently many images are acquired from sufficiently general and appropriate viewpoints. In photogrammetry, this optimization process is called *self-calibrating bundle adjustment*; as for *self-calibration*, as typically understood in computer vision, it refers to calibration from the same amount of information as above: observations of features with unknown 3D coordinates, in different images. In addition, it is implied that no initial values are available, e.g. no approximate 3D coordinates of markers; see below for more. Note that in self-calibrating bundle adjustment, initial 3D coordinates are typically not completely forgotten about, but are used to define soft constraints on the estimated ones.

Let us come back to computer vision-type self-calibration. As mentioned above, it refers to calibration without any initial knowledge of 3D coordinates. Moreover, no markers or other artificial landmarks are used; the input are simply a set of images of some scene. Self-calibration theory and algorithms are typically based on optical flow or matches between interest points. Let us briefly explain why self-calibration, i.e. computation of camera intrinsics and poses and 3D scene structure, is possible from such poor input. This is mainly due to the well-known coplanarity constraint: consider a match between two points in different images. This gives a constraint on the intrinsic and extrinsic parameters of these two images: hypothetical values for these parameters are only potentially correct if the 3D projection rays of our two points, computed using these values, intersect. This is a strong constraint since randomly drawn intrinsic and extrinsic parameters have zero probability of conforming even to a single match. We now have to examine if matches between images allow to compute a *single* solution for the intrinsic and extrinsic parameters, which would necessarily be the correct one. This turns out not to be the case: Faugeras and Hartley have proven independently from one another that projection matrices, and thus 3D scene structure, can only be estimated up to a global projective transformation [31, 45][1]. The essence of self-calibration

[1]This has been known by photogrammetrists at the end of the 19th century [34].

algorithms is to eliminate this ambiguity using additional information. This can be any information about the 3D scene structure (distances between points, geometric constraints such as parallelism or orthogonality of line segments, etc.), camera extrinsics (e.g. the knowledge that camera displacement is a planar motion) or intrinsics. Self-calibration usually refers to the latter. It has been shown that the information that a single intrinsic parameter remains constant throughout image acquisition, is sufficient for self-calibration [46, 68].

We now mention a third approach for alleviating requirements on the calibration process, besides self-calibration and using calibration objects with inaccurately known structure: using calibration objects of some *constrained* structure. One may distinguish "marker-less" calibration objects, such as linear objects (plumbline calibration of non-perspective distortions), spheres [62] or circles (see chapter 5), from planar or linear objects with attached or imprinted markers. The above "marker-less" calibration objects are easy to come by or observe. For example, when acquiring turntable sequences, individual points on an object's surface follow circular trajectories, making calibration from images of circles an interesting approach. As for planar or linear calibration objects with markers, they are useful in several respects. They are easier to manufacture, handle and store than traditional calibration objects and can nevertheless give accurate results if correctly used (imaging them in sufficiently many and sufficiently general positions). Linear objects are in addition very useful for calibrating multi-camera systems, e.g. camera-based motion capture systems, since their markers are visible from nearly everywhere, as opposed to opaque planar calibration grids for example.

This section is of course not meant to give an exhaustive treatment of camera calibration and self-calibration, but to give a short introduction to our works in this area. More detailed texts are found e.g. in [1, 33, 44, 82].

In the following two chapters, we describe our works on camera calibration. In chapter 4, we describe our approaches using calibration objects. The first set of approaches use objects of constrained structure – planar or linear calibration objects. We also present an experimental study on the calibration of cameras with zoom lenses. In chapter 5, we present our works on self-calibration: algorithms for various scenarios (planar scenes, planar motions, using images of circles, etc.) and theoretical analyses of critical motions. We have also proposed methods that exploit geometric constraints such as orthogonality or parallelism of line segments, for 3D reconstruction and camera (self-)calibration. These are described in chapter 10 in part IV.

Finally, let us note that all the works described above and in the following two chapters, address the perspective or pinhole camera model or classical extensions thereof (e.g. radial or tangential distortion models). In part III, we consider more general camera models and propose "generic" calibration and structure from motion approaches, i.e. which are applicable to a wide range of camera types.

Chapter 4

Camera Calibration

Plane-based calibration.

Our most significant work on camera calibration concerns the use of planar calibration objects (also called "grids" in the following). The fact that calibration using planar grids is possible, is known at least since Tsai's work [114] and may have been known to photogrammetrists before that (although I'm not aware of any reference). In [114], the full potential of using planar grids was not developed yet: the only case dealt with was computation of a single intrinsic parameter (focal length) from one view of a planar calibration grid[1].

The first works, to the best of my knowledge, that fully exploit multiple views of planar grids taken from unknown viewpoints, are ours [97] and Zhang's [130], which were done independently and simultaneously. In [97, 130], it is shown how to obtain linear constraints on the intrinsic parameters[2] from homographies between scene and image planes. In [84], we further develop methods for computing pose parameters, for the general case of multiple cameras looking at multiple planar grids (or, equivalently, moving cameras and/or grids). Mainly due to an executable distributed by Zhang and an Open Source implementation within the OpenCV library[3], the plane-based approach has become a standard tool for calibrating cameras.

In [97] we describe, besides the actual calibration algorithm, singularities, i.e. acquisition conditions for which calibration fails or is unstable. To obtain an accurate calibration, it is essential to avoid these conditions, together with respecting other guidelines, especially covering the entire image area with the grid. The singularity or not of acquisition conditions depends on the amount of prior knowledge on the intrinsic parameters (e.g. known values of certain intrinsics or knowledge that certain intrinsics are constant). In [98], we give the singularities for all practically relevant cases of calibration from one or two images of a planar grid. A general rule of thumb is that the more "structured" or symmetric the acquisition conditions (relative poses of camera and grid) are, the more likely they are to be singular.

▷ Papers 1 and 2 in the accompanying collection [97] [84]

Using linear calibration objects.

So-called calibration "wands", i.e. sticks with attached markers, are routinely used for the calibration of commercial multi-camera-based motion capture systems. Commercial systems typically require two steps: first, acquisition of the wand(s) in special positions, to initialize the estimates of camera poses. Then, a so-called "wand dance", the image sequences of which are used to optimize pose and intrinsic

[1] Multiple views of a planar grid were also considered, but only for the case of a known displacement of the grid (achieved using a translation stage); the setup was thus used as a 3D calibration object.

[2] Actually, on the IAC – the image of the absolute conic – which is equivalent to the set of intrinsic parameters.

[3] `http://www.intel.com/technology/computing/opencv/index.htm`

parameters. It is interesting to shorten the procedure and relax the requirements on the user's cooperation, by doing everything from the wand dance only. During and after his internship with me, Tomislav Pribanić developed, implemented and tested a variety of methods for initializing the calibration process using images of one, two or three rigidly linked wands. The different methods are based on different ways of exploiting information on distances between markers and right angles between wands, for the computation of intrinsic parameters [69, 70]. Obtained results are more accurate than those of a popular commercial system[4], while using a simpler procedure.

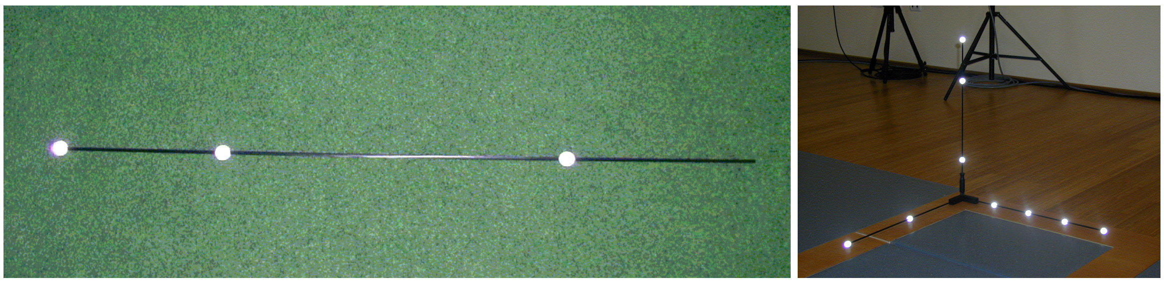

Figure 4.1: A calibration wand and a set of rigidly attached wands.

The methods described above are applicable to *multi-camera* systems, but not for calibrating a *single* camera: the wand's pose can be computed from a single calibrated image, but there is no information left for constraining the camera's intrinsic parameters. This can be explained as follows. The wand's 3D pose has 5 degrees of freedom: 3 for the position of one of the markers plus 2 for the orientation of the wand. The wand's image accounts for 5 independent pieces of information: the image positions of all its markers are encoded by the position of one of them (2 parameters), the orientation of the imaged wand (1 parameter) and the positions of two additional markers along the imaged wand (2 parameters). The image position of any further marker is already computable from those of the first three markers, using the knowledge of the distances between the original markers and the cross-ratio. Hence, the image of a wand provides exactly the information required to compute pose but not calibration.

In [131], Zhang noted that wands can nevertheless be used for calibrating single cameras, if their motions are restricted. Namely, if one of the markers is kept stationary, then images from multiple orientations of the wand do provide information that is useful for calibration. In [42], we describe closed-form solutions for this calibration approach, for the practically relevant cases of prior knowledge on intrinsic parameters (known/unknown aspect ratio and/or principal point). We also reveal the singularities that cause the approach to fail; this is the case if the wand traces out a cone in 3D. This calibration approach is probably mainly of conceptual interest.

▷ PAPER 3 IN THE ACCOMPANYING COLLECTION [42]

Calibration of zoom lenses.

Zooming and focusing modify to some extent the intrinsic parameters of the pinhole model (focal length etc.). We performed an experimental study in order to determine ranges of zoom and focus settings within which intrinsic parameters are (not) constant [115, 116]. The motivation was to derive a guideline for when a re-calibration is necessary due to zooming/focusing. For the camera model tested we found that with minimal zoom, the effective focal length was rather insensitive to focusing. For a maximal zoom, this was found to be true above a certain focus setting (scene farther away than some threshold).

We also developed an associated self-calibration algorithm, for the scenario of an acquisition session with a camera that is initially calibrated but whose zoom/focus settings are modified during the session. The

[4] eMotion/Smart, http://www.emotion3d.com, http://www.bts.it/eng/proser/elisma.htm

proposed algorithm computes the focal length of a view, given matches with another, fully calibrated view. This is a special case of self-calibration (cf. next chapter), with the benefit of suffering from just one simple type of critical motion: when the optical center of the fully calibrated view lies on the optical axis of the other view.

▷ PAPER 4 IN THE ACCOMPANYING COLLECTION [115]

Chapter 5

Camera Self-Calibration

As mentioned in the introduction to this part, self-calibration is considered here as calibration without a calibration grid with known marker coordinates. In computer vision, the concept has been introduced by Maybank and Faugeras in [55] where it was shown that the so-called Kruppa equations give constraints on intrinsic camera parameters, using only image matches at the outset. This first approach was limited to fixed intrinsic parameters. Since then, several other approaches have been developed, some of which coping with varying intrinsics (or, images of the same scene acquired by different cameras). The most fundamental findings on self-calibration have probably been obtained before 1999: the initial approach [55], the absolute quadric formulation [112], self-calibration for varying intrinsics [46, 68], study of critical motions [92]. A complete state-of-the-art is beyond the scope of this document; please refer to [33, 44, 67, 93].

In the following, our works on self-calibration since 1999 are summarized. They represent a mixture of algorithms for various special cases (constrained camera motions, scene geometry or sets of unknown intrinsics) and theoretical studies of associated singularities (critical motions).

Critical motions for the Kruppa equation based approach.

In my PhD thesis [93], I have introduced the notion of critical motions, i.e. motions of the camera that prevent its self-calibration, whatever method is used. Individual self-calibration algorithms may be subject to additional singularities. This is for example the case for absolute quadric based approaches that neglect the rank-3 constraint on the absolute dual quadric. In [86], we show that the classical approach based on Kruppa equations [55, 129] suffers from these singularities, plus additional ones. This is illustrated in figure 5.1.

▷ PAPER 5 IN THE ACCOMPANYING COLLECTION [86]

Focal length self-calibration.

Most self-calibration approaches are rather generic in that they allow to estimate all intrinsic parameters of the pinhole model. In practice however, approximate values of certain intrinsics are often available. It is often enough to only estimate the focal length in an initial self-calibration step, and then to optimize other parameters during bundle adjustment if required. We thus have studied this special case of self-calibration, for the cases of constant or varying focal length.

In [90, 94], we show the critical motions in the case of a varying focal length and in [87, 104] for a constant one. In case of a varying focal length, the most notable critical motion consists of a camera moving along a conic with the optical axis being always tangent to that conic. As for a constant focal length, the most relevant singularities concern two views whose optical axes are parallel or intersect in a point that is equidistant from the optical centers.

In [87, 104], we also give equations for computing the constant focal length in terms of the fundamental

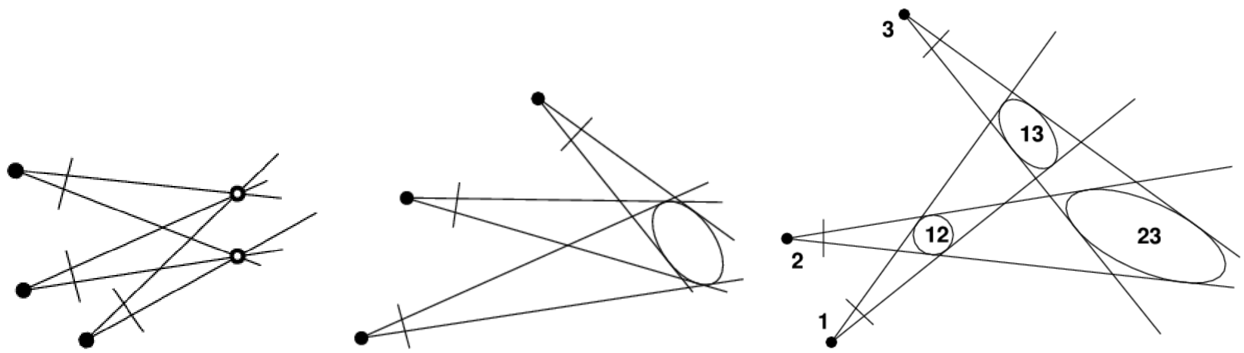

Figure 5.1: 2D sketches of different levels of singularity, affecting different types of self-calibration methods. Each singularity subsumes the one(s) shown to the left of it. Left: generic singularity, i.e. existence of a *conic* in 3D (represented by two points here) that is different from the absolute conic, but whose projections are identical in all images. Middle: singularity affecting all methods that neglect the rank-3 constraint on the absolute dual quadric; occurs whenever there exists a non-degenerate quadric (represented by an ellipsis here) with identical projections in all views. Right: singularity affecting approaches based on Kruppa equations; occurs when for each pair (i,j) of views there exists a quadric such that for every view i, all associated quadrics have the same image.

matrix' SVD (singular value decomposition), and show their associated singularities.

▷ PAPERS 6 AND 7 IN THE ACCOMPANYING COLLECTION [90] [104]

Self-calibration for planar motions.

As mentioned above, any information on the scene or the camera's motion can be used for self-calibration, in addition to information on intrinsic parameters. In this and the following two paragraphs, we consider three such scenarios: planar scenes, scenes containing circles, and cameras carrying out planar motions. The latter correspond to motions composed of translations in a plane and rotations about that plane's normal. This type of motion is quite common, e.g. many photographs are taken this way.

In [32], we show that the self-calibration problem can be reduced to that of self-calibration of a linear camera, living in a plane. Consider the intersection line of the image plane with the plane of motion; together with the optical center, this forms a linear camera, or, 1D camera. From correspondences between the original 2D images, it is simple to hypothesize correspondences between the associated linear images. Then, trifocal tensors between triplets of linear images can be estimated. In contrast to the trifocal tensors between 2D images, those of linear images do not obey any internal consistency constraints, i.e. estimation using linear equations is appropriate. Finally, self-calibration of the linear camera from trifocal tensors can be done in closed form. This results in the estimation of the images of the motion plane's circular points, which can then be used as constraints on the original camera's image of the absolute conic, allowing to achieve its self-calibration. Overall, this approach is elegant and "light-weight", avoiding the computation of the classical trifocal tensors between 2D images. Critical motions for the linear camera self-calibration problem were already given in my PhD thesis [93].

▷ PAPER 8 IN THE ACCOMPANYING COLLECTION [32]

Self-calibration for planar scenes.

Like for calibration, planar scenes or objects are also useful for self-calibration. The main benefit is somewhat different here. For self-calibration, i.e. in the absence of specific markers, it is interesting to use planar objects since image matching becomes much easier, due to the absence of self-occlusions. Using planar objects for self-calibration has been introduced by Triggs [113]. We propose a theoretical study of the problem [40, 96] and several practical algorithms: non-linear estimation of intrinsic parameters using a minimum of parameters [40, 96] and iterative estimation using video sequences [52, 60]. We also give closed-form solutions for the special case where one of the images is taken in fronto-parallel position [40, 96]; this was found to give sufficiently accurate initial values in practice to make a subsequent bundle adjustment converge to an accurate calibration.

An example is shown in figure 5.2. The non-linear estimation, or bundle adjustment, involves the computation of the camera's pose with respect to the scene plane. This allows to rectify its images as if taken by a fronto-parallel camera.

▷ PAPER 9 IN THE ACCOMPANYING COLLECTION [40]

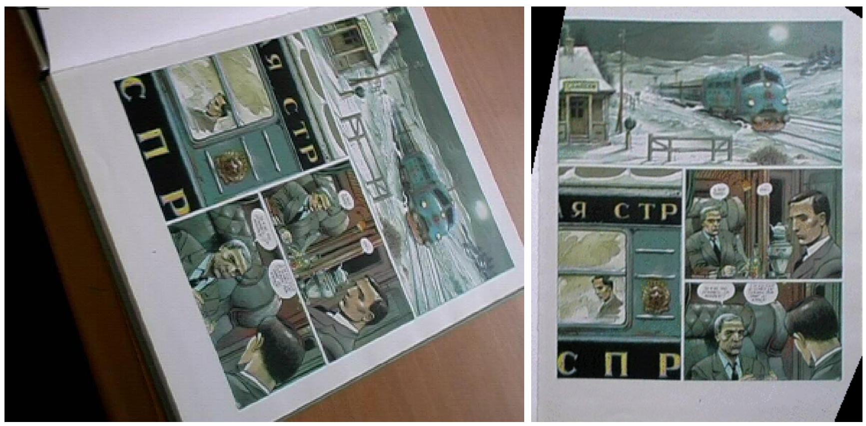

Figure 5.2: Left: an original input image of the planar object used for self-calibration. Right: image after rectification using the self-calibration result. Imperfections are (at least partly) due to the page being slightly bent close to the crease.

Calibration using images of circles.

In [100], we have introduced the idea of calibrating a camera using images of circles that are coplanar or whose support planes are parallel. An application for this is the calibration of turntable sequences: each point on the object surface follows a circular trajectory and the set of trajectories take place in parallel planes. Hence, image point tracks allow to estimate the images of circles with parallel support planes[1].

A somewhat dual problem to self-calibration is the computation of the Euclidean structure of the circles' support planes. In general, the images of two circles give two solutions. Depending on the circles' relative position, the wrong solution can be easily eliminated. In [41], we give a complete treatment of all special cases for this problem, and propose an algorithm using multiple circles and that only requires to solve a linear equation system. Examples are shown in figure 5.3.

Optimal fundamental matrix computation.

In [7] we propose a parameterization and an algorithm for the efficient non-linear optimization of the fundamental matrix between two images. It is parameterized via the factors of its SVD (singular value decomposition), i.e. two orthogonal and one diagonal matrix, whose update during the non-linear optimization is rather simple.

▷ PAPER 10 IN THE ACCOMPANYING COLLECTION [7]

[1] Note that for turntable sequences, an additional constraint is that all circular trajectories are centered around an axis orthogonal to the plane's of motion. This constraint is used e.g. in [50].

Figure 5.3: Top: photograph of the cover of some comic book and rectified images corresponding to the plane's Euclidean structure, obtained using 2 respectively 9 small black circles put on the object. Bottom: photograph of a kitchen scene and rectified image obtained using 6 circles (shown in blue on the original image).

Part III

Generic Camera Models and Unified Treatment of Structure from Motion

Introduction

Many approaches for camera calibration and structure from motion exist in the literature. Most of them are associated with the pinhole camera model, but many were also developed for other models, such as classical distortion models (e.g. radial and tangential distortions), models for large distortions (fish-eye lenses), for catadioptric cameras (camera + mirror) and other particular camera/image types, e.g. push-broom cameras, non-central mosaics and stereo systems. Three years ago I wondered about the possibility of developing a single calibration approach that would be applicable to all such camera models. The basic idea is extremely simple; it uses a generic camera model that consists of a lookup table attributing to each pixel the coordinates of the associated line of sight. This model subsumes all the above ones, for the prize of having 4 parameters per pixel (3D line coordinates) instead of a few parameters for the whole image (e.g. 5 for the pinhole model). The model is based on idealistic assumptions, e.g. that a pixel's response depends on light travelling along a single line of sight, but such assumptions are common and typical remedies are applicable (e.g. use of a point-spread function).

Having defined the camera model, the next question is how to calibrate it, i.e. how to compute the line of sight for each pixel. We wanted to achieve this as usual, i.e. by taking images of a calibration object, from unknown viewpoints[2]. This is described in chapter 6. Our initially developed generic calibration approach has eventually been specialized to a hierarchy of four generic camera models: the above general model where lines of sight may be distributed arbitrarily, the central model where there exists an effective optical center through which all lines of sight pass, an intermediate so-called axial model, and a radially symmetric model. All this is described in more detail in chapter 6.

It was natural to develop structure from motion methods and theory for our generic camera model. Again, the literature contains a large number of methods e.g. for pose estimation. Often, relationships between different methods are not revealed, especially for methods developed for different camera types. For example, pose estimation methods developed for various catadioptric camera models, are usually straightforward adaptations of existing methods for pinhole cameras.

We thus develop approaches for what we consider the basic tasks of structure from motion: pose estimation, motion estimation, 3D reconstruction, and multi-view geometry (matching tensors). These are described in chapters 8 and 9. Most of them are simple, have been known previously and are obvious to many people. Nevertheless, we found it useful to give a compact treatment of these issues in a single place.

Finally, we have also developed self-calibration methods for the central and the radial symmetric camera models. These are described in chapter 7.

Before proceeding, let us make an observation on our generic camera model. It associates to each *pixel* the coordinates of its line of sight. This can be interpreted as a discrete sampling of a continuous function, mapping any image *point* to its line of sight. One could rewrite the following chapters using the continuous model, but this is not within our plans for future work. Further, instead of sampling this "back-projection

[2]We found out later that at least three labs have independently from one another proposed to use the same model for calibration, but using images taken from known viewpoints [25, 38, 39].

function” per pixel, one could also choose a sub-pixel or super-pixel sampling, depending on the precision and amount of matches between the calibration object and the image. The discrete model needs to be completed by an interpolation scheme if the line of sight of some arbitrary image point needs to be computed. Any classical interpolation method such as nearest neighbor or bilinear interpolation is applicable.

Planned future work along the lines of the following chapters, is described in part VI, with the general conclusions of this document.

Chapter 6

Calibration

Introduction.

We consider the generic camera model introduced above: to each pixel is associated a 3D line representing its line of sight. A simple approach for calibrating this model from two or more images of a calibration object has been developed independently by at least three groups [25, 38, 39]. There, between image takings the calibration object is moved by a known displacement. This is for example achieved by mounting a planar calibration object on a translation stage and carrying out translations by known distances. Each image is matched to the calibration object. Let us suppose that for each pixel, we can compute a corresponding "calibration point" i.e. a point on the calibration object (matching is discussed in more detail below). Since we know the motion between different positions of the calibration object, we can express all calibration points in the same coordinate frame. We can now compute the line of sight for each pixel, by fitting a 3D line to its matching calibration points.

Variants of this approach exist. For example, CNES[1] uses a directional light source mounted on a robot arm. The robot arm is moved such that the light source sweeps a desired portion of the space of directions. For each image, the pixel corresponding to the observed light source is easily extracted and the direction of its line of sight computed using the robot arm's forward kinematics. Note that here, a central camera model is assumed, i.e. all lines of sight are supposed to go through a common optical center. Only their directions need to be computed, whereas the above approach allows for non-central cameras. Let us also note that Grossberg and Nayar further investigate radiometric calibration of the generic camera model [39].

Let us briefly discuss benefits and disadvantages of such calibration approaches. The main advantage is that no specific camera model is required; radial or other distortions are handled automatically, as well as local optical aberrations for example. Due to the latter, the calibration can be more accurate than what is achievable with any standard parametric model. However, the local accuracy of the calibration depends directly on the matching's accuracy. As for pinhole cameras, a high accuracy can be obtained using e.g. circular markers [23]. Here, matching is based on extracting the markers' elliptical images. In some sense, all pixels within or touching such an image ellipsis, typically a few dozens, "collaborate" to achieve highly accurate matching. Whereas for our generic camera model, matching has to be achieved more locally, thus less accurately.

Matching.

In practice, we use planar calibration objects. We have used two approaches for matching pixels to points on the calibration plane. The first one uses circular or square markers and extracts their image positions as if the camera were a pinhole, e.g. by extracting an elliptical image of circular markers. Then, for each pixel,

[1] Centre National d'Études Spatiales, http://www.cnes.fr

we consider the 4 closest image points corresponding to an extracted marker, and interpolate the position of the corresponding point on the calibration plane using the planar homography defined by these 4 image points and the corresponding marker positions. This approach is not very accurate.

The second approach is of structured light type [21, 109]. A flat screen is used as calibration object; for each viewpoint, a set of Gray codes are displayed on the screen (vertical and horizontal black-and-white striped patterns, cf. figure 6.1). Each screen pixel is characterized by a unique sequence of blacks and whites. For each image pixel, one can thus determine a matching screen pixel. This matching is not perfect, since precision is limited to one screen pixel; sub-pixel precision can be obtained using some regularization [21]. Another possibility to achieve sub-pixel precision would be to display “continuous” grey-level patterns on the screen. However, a significant number of images may be required and in order to distinguish grey-levels, some radiometric calibration of the camera and e.g. its vignetting, may have to be done in addition.

Nevertheless, we believe that this can be done, and that ultimately, matches can be obtained with a precision and accuracy that are sufficient for a highly accurate calibration of our generic camera model. We may investigate this in future work.

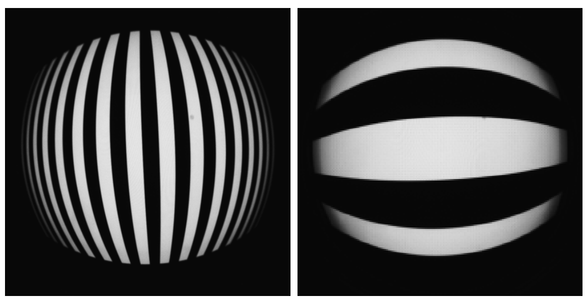

Figure 6.1: Two images of a flat screen with structured light type patterns on display, acquired using a fish-eye lens.

Proposed calibration approach.

Let us come back to the calibration problem. As mentioned above, our approach does not require knowledge of displacements of the camera or the calibration object between image acquisitions. Rather, it first computes them using the provided input (matches between image pixels and calibration points) and then applies the method explained in the first paragraph of this chapter, to compute the pixels’ lines of sight. The computation of displacements goes as follows.

Let $\mathbf{Q}$, $\mathbf{Q}'$ and $\mathbf{Q}''$ be the calibration points matched to the same pixel in three images, acquired for different positions of the calibration object. These are 4-vectors of homogeneous coordinates, expressed in the local coordinate frame of the calibration object. Without loss of generality, we choose as global coordinate frame the one associated with the object’s first position. The unknown displacements between the second and third frames and the first one, are given by 3×3 rotation matrices R' and R'' and translation vectors $\mathbf{t}'$ and $\mathbf{t}''$. Mapping the calibration points to the common frame gives points

$$\mathbf{Q} \qquad \begin{pmatrix} \mathsf{R}' & \mathbf{t}' \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{Q}' \qquad \begin{pmatrix} \mathsf{R}'' & \mathbf{t}'' \\ \mathbf{0}^\mathsf{T} & 1 \end{pmatrix} \mathbf{Q}''$$

With the correct displacements, these three points are collinear, since they all lie on the same line of sight. Let us consider the 4×3 matrix whose columns are the above coordinate vectors. Collinearity of the points is then equivalent to that matrix having column-rank less than 3. This in turn is equivalent to the vanishing of the determinants of all four of its 3×3 submatrices. Each such determinant can be written as a trilinear equation in the calibration points:

$$\sum_{ijk} Q_i Q'_j Q''_k \mathcal{T}_{ijk} = 0 \tag{6.1}$$

whose coefficients make up a trilinear tensor $\mathcal{T}$ that depends exactly on the unknown displacements. Using sufficiently many triplets of matches, these trifocal tensors can thus be estimated by solving linear equations (6.1). In [101, 102] it is shown how to extract the displacements from these tensors and to finalize the calibration.

This approach allows *in principle* to calibrate any camera that can be cast in the generic model used here; that this is not entirely the case, is explained in the next paragraph. Note that even a stereo system may be calibrated using this approach: it suffices to consider it as a single camera, whose set of pixels is the union of the individual cameras' pixels…

This basic approach works with exactly three images and only calibrates the pixels which are matched to calibration points for all three positions of the calibration object. Especially for cameras with a wide field of view, it is difficult to cover the entire image area using a typical calibration object. In [72, 73], we describe approaches for achieving a calibration for the entire image area, using multiple images.

Variants of the calibration approach.

When using linear equations to compute the tensors introduced above, unique solutions are only obtained if the calibration object is of three-dimensional structure. This is easy to prove and can be intuitively explained using the analogy with fitting a fundamental matrix to a set of matching points between two perspective images: if the matches stem from a planar object in the scene, then the fundamental matrix can not be estimated uniquely (there is a 3-degree-of-freedom family of solutions). We have thus developed a variant of our calibration approach, dedicated to planar calibration objects [101, 102]. It is slightly more complicated than the general approach, but is much more practical (cf. chapter 4 on calibration of perspective cameras).

There are other circumstances that cause ambiguities in the estimation of our tensors, related to the type of the camera to be calibrated. For example, if the camera is central (all lines of sight go through a single point), then the above linear equation systems give again ambiguous solutions for our tensors. This observation extends to a hierarchy of camera types, as follows. As mentioned, a **central camera**'s lines of sight all go through a single point, the optical center. As for an **axial camera**, they all cut a single 3D line, the camera axis. This is the case for example for stereo systems composed of cameras with collinear optical centers, and non-central catadioptric devices constructed using a surface of revolution type mirror and a camera placed on the axis of revolution. A special case of axial cameras is the **x-slit camera**, where there exist two camera axes that cut all lines of sight [61, 132] (also called two-slit or crossed-slits cameras). Linear push-broom cameras [43] fall into that class. Finally, cameras not fitting into these categories, are termed **fully non-central**.

If the actual camera is not fully non-central, then the above tensors can not be estimated uniquely using linear equations only. We thus have developed variants of the calibration approach for the axial camera model [75, 76] and the central one [101, 102]. The x-slit camera has not yet been fully dealt with, since it is less important than the others.

Overall, we thus end up with three generic calibration methods, instead of a single one as desired at the outset. Nevertheless, these should cover all relevant camera types used routinely.

Let us make a note on the calibration's stability. As explained, when using the general (non-central) method to calibrate a central camera for example, then the tensors involved can not be estimated uniquely and the method fails. If the camera is only "slightly" non-central, i.e. if all lines of sight pass trough some small volume, then the calibration is likely to be unstable. Stability or not depends on several factors:

how non-central a camera is but also the accuracy of the matches and the number of input images. We empirically examined if a non-central camera model and calibration is justified, by calibrating fish-eyes and catadioptric cameras. They were calibrated using the central and the non-central approach; the quality of each calibration was then assessed by evaluating pose estimation results obtained using the respective calibration. In one case, the non-central model gave significantly better results [73].

To summarize, we have proposed a set of generic calibration approaches. As for now, we consider this mainly as a conceptual contribution. Especially, we do not claim that these approaches should replace existing ones in practice. Nevertheless, we are confident that especially the central model may be useful in practice; its calibration can be made stable and accurate with an appropriate calibration protocol. We are working on a calibration toolbox for this model that hopefully will become a useful tool for others.

The first experimental results we have obtained are shown in figures 6.2 and 6.3.

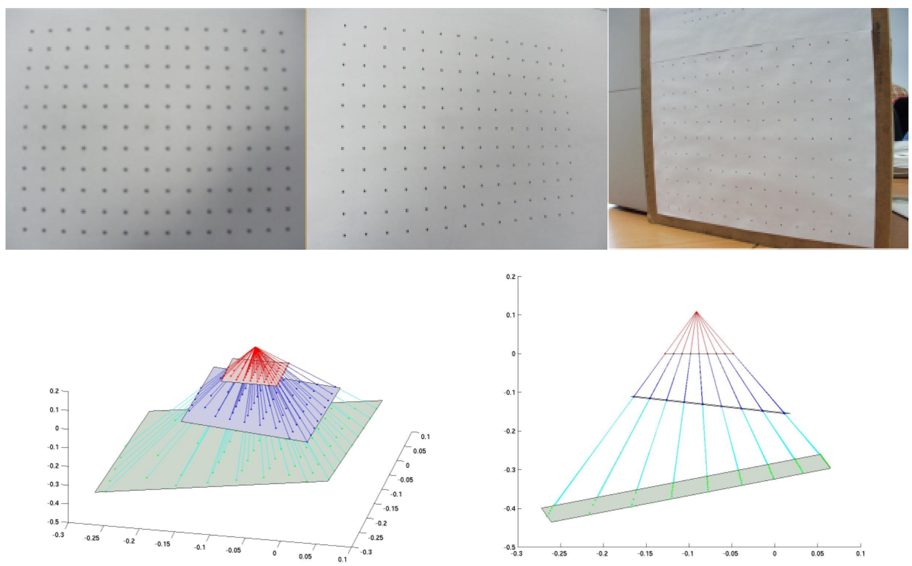

Figure 6.2: Calibration of a camera with slight radial distortion, using the central camera model. Top: images of 3 calibration grids of different sizes, captured by the camera. Bottom: two views of the calibrated lines of sight and estimated pose of the calibration grids. The colored line segments join matching points between different grids; they being collinear is a qualitative indicator that the calibration is correct.

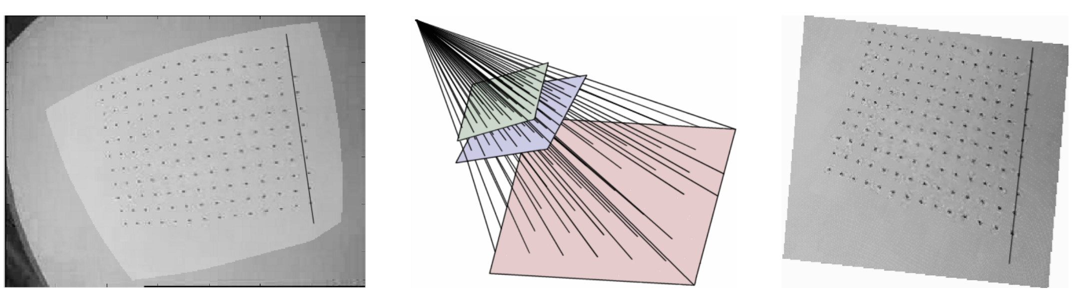

Figure 6.3: Calibration of a fish-eye's central image region using the central camera model. Left: one of three input images (in white the region that was calibrated). Center: calibrated lines of sight and estimated pose of calibration grids. Right: input image after distortion correction based on the calibration result. The distortion corrected for is not spectacularly large, but note that no parametric distortion model has been used here.

Radially symmetric cameras.

Even the most constrained one among the above generic camera models – the central model – is extremely general in that the mapping between pixels and the directions of their lines of sight can in theory be arbitrary. This is more generality than often needed. A good compromise between such a general model and the classical low-parametric ones is to additionally consider radial symmetry of the mapping from pixels to lines of sight. By this we mean the existence of a *distortion center* in the image and an *optical axis* in 3D such that the lines of sight associated with pixels on a *distortion circle* centered in the distortion center, span a right *viewing cone* centered around the optical axis (cf. figure 6.4). Further, for pixels on a radial line, i.e. a line passing through the distortion center, the lines of sight lie on a plane containing the optical axis. Angles between radial lines are supposed to be identical to angles between the associated planes[2]. One may interpret this as a generic model for radial distortion.

We consider both, central and non-central versions of that model: in the central model, all viewing cones have the same vertex (the optical center) whereas in the non-central version, the vertices may "slide along" the optical axis. This is thus an *axial* camera model, according to the definition given in the previous paragraph. It may be used for fish-eyes or non-central catadioptric cameras.

Calibration of our radially symmetric camera models comes down to estimating the distortion center, the *distortion function* (mapping from radius of distortion circles to opening angle of associated viewing cones) and the positions of the viewing cones' vertices (for the non-central version). In [109] we give two calibration approaches. One of them allows to calibrate already from a single image of a calibration plane. Very promising results have been obtained, cf. figure 6.5.

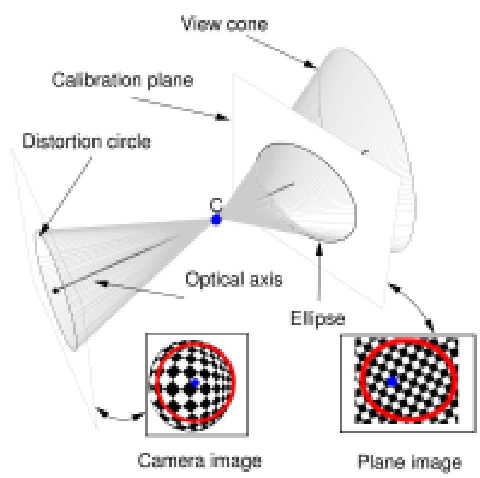

Figure 6.4: Radially symmetric camera model. Also shown is a calibration plane.

▷ PAPERS 11 TO 14 IN THE ACCOMPANYING COLLECTION [102] [73] [76] [109]

[2] This comes down to assuming an aspect ratio of 1. The model may be generalized to non-unit aspect ratios.

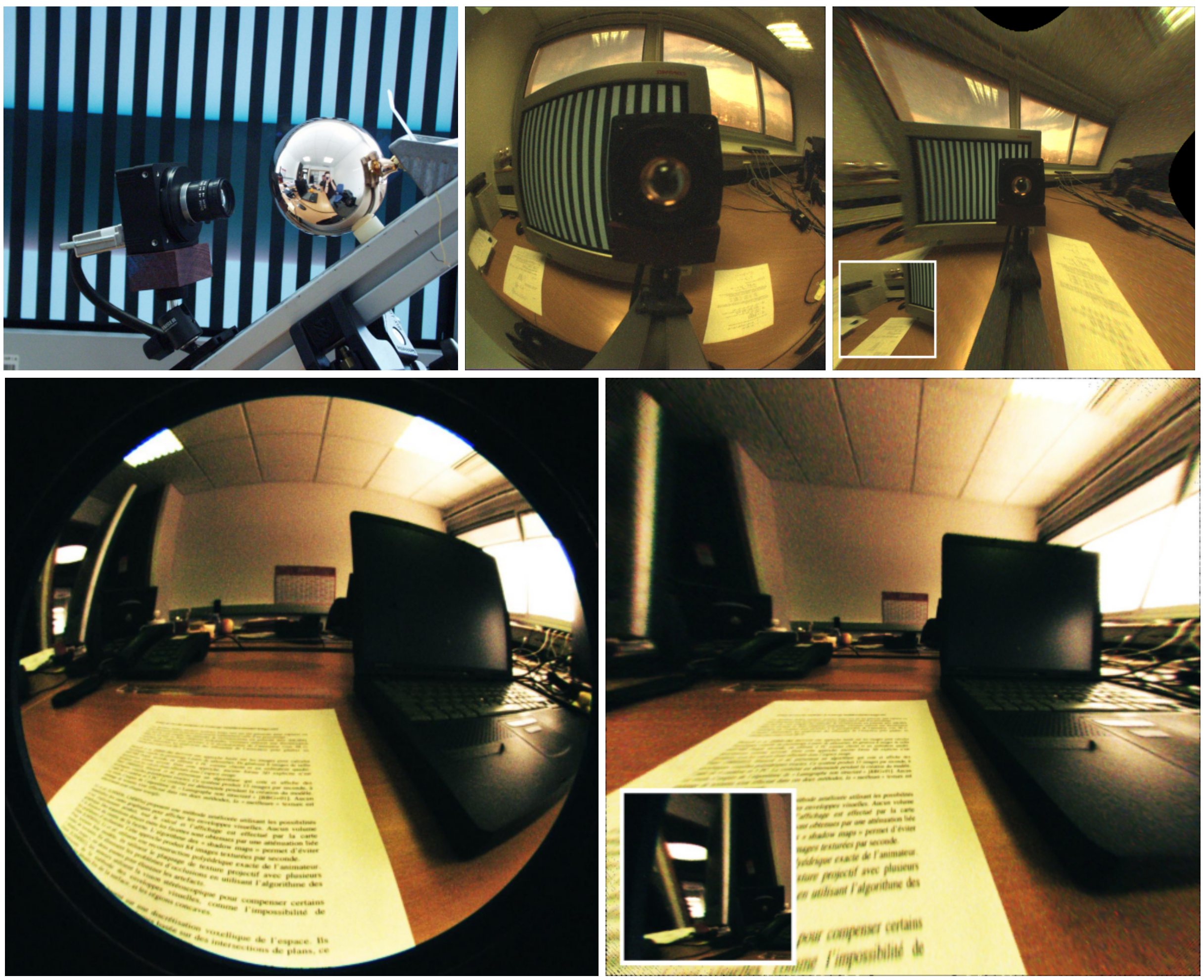

Figure 6.5: Calibration results for the radially symmetric camera model. Top: "home-made" catadioptric camera, using a Christmas ornament; an image acquired using that camera and its rectification based on the calibration result. Note the small inset image showing the rectification of a region close to the image border. Bottom: original and rectified images for a fish-eye lens.

Chapter 7

Self-Calibration

Central camera model.

Self-calibration of the generic non-central camera model introduced in the previous chapter, seems to be possible but difficult. So far, the only self-calibration approaches for generic camera models are recent ones by Nistér et al. [59] and ourselves [74]. They both are restricted to the central camera model and furthermore require particular camera motions, namely rotations about the optical center or translations. Nistér et al. mainly deal with the case of infinitesimal motions, whereas we consider finite motions.

In [74], we explain the constraints on self-calibration arising from pure translations and rotations. Based on these, we give a simple algorithm for self-calibration from point tracks along images sequences corresponding to two rotations and one translation.

It is the self-calibration problem where we encountered the theoretical limitations of our discrete camera model, due to associating lines of sights only to (discrete) pixels. The problem is that this prevents proper reasoning on matches between images, as explained in the following. As for *calibration*, the supposed input were matches between camera pixels and points on calibration objects. However, for *self-calibration* using our model, matches between camera pixels are required. Suppose we want to simulate such matches. Let us first simulate the camera's calibration, i.e. define, e.g. randomly, the lines of sight for all pixels. Now, let us simulate a displacement of our camera, i.e. of its set of lines of sight. Suppose we take images, before and after the displacement. Independently of the simulated scene, two pixels, one in the first and one in the second image, are possible matches exactly if their lines of sight intersect. However, the probability for two lines of sight intersecting one another, is zero in general (since we consider *lines* of sight, not *volumes* of sight, so to speak). This implies that it is in general impossible to even simulate matches between pixels for the discrete generic camera model! Let alone to provide a theoretical proof for the feasibility of self-calibration...

To prove the feasibility of self-calibration, one thus needs to consider a continuous camera model (future work). Note that these statements also hold for the pinhole model: exact matches between *pixels* have zero probability to occur in general. However, theoretical reasoning about self-calibration or other problems is no problem since the camera model is continuous (for example, the calibration matrix allows to compute a line of sight for each sub-pixel image point). In practice, self-calibration of the pinhole model works even if matches are only determined for pixels or another discrete sampling of the image plane. The reason is that the non-exactness of such matches can be treated as noise... For the same reason, our self-calibration algorithm for the generic camera model [74] does work in practice. First experimental results are shown in figure 7.1.

▷ PAPER 15 IN THE ACCOMPANYING COLLECTION [74]

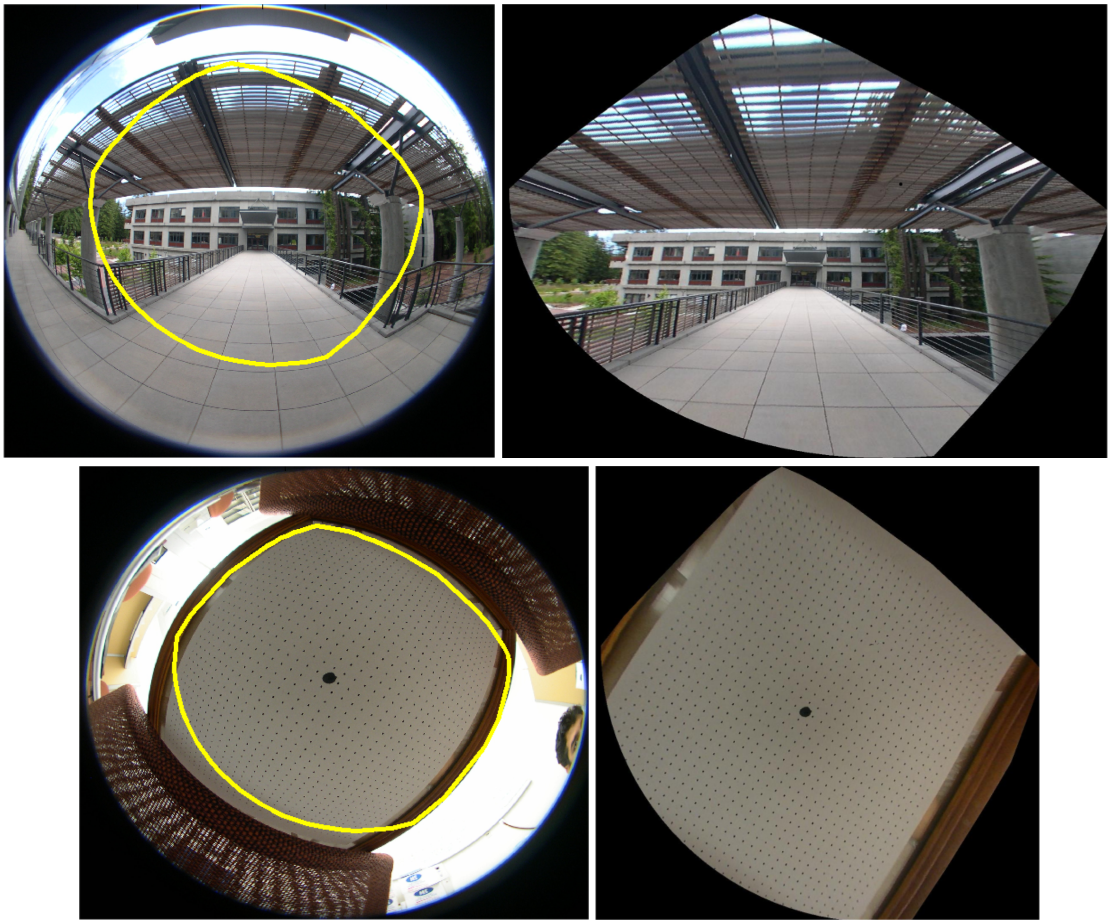

Figure 7.1: Distortion correction results based on self-calibration of the generic central camera model. Left: original images; the calibrated image region, i.e. the set of pixels for which the directions of associated lines of sight were computed, is outlined in yellow. Right: distortion correction of the calibrated image region. Results are not perfect (some linear features remain slightly bent), but note that this experiment probably concerns the most general self-calibration scenario so far in the literature.

Radially symmetric camera model.

We have also developed a self-calibration approach for the radially symmetric camera model [110], introduced in chapter 6. It is based on a plumbline type method: from images of line segments, one can calibrate the distortion function by solving a single linear equation system.

This can be applied for self-calibration using a planar object, as follows. Consider two images of the planar object and suppose that dense matches have been obtained (e.g. using optical flow along an image sequence). Let us for the moment assume to know the distortion center. Consider one line through the distortion center in the first image, and gather all pixels lying "on" that line. By definition of the camera model, the scene points projecting to these pixels are collinear (on the scene plane), i.e. they lie on a line segment. Since we know the matches of all these pixels in the second image, we can thus identify the projection of that line segment. By repeating this procedure for many lines through the distortion center, and going in both directions (from first to second image and vice-versa), we can thus gather a significant number of line images, and apply the plumbline method.

Note that it is nowhere required that the scene contain actual straight line segments and that dense matches are actually not required (this was just stated for ease of explanation). Further, in [110], we show how to estimate the position of the distortion center, which for simplicity was supposed to be known in the above explanation.

This approach is considerably simpler than the approach by Thirthala and Pollefeys [111], imposes radial symmetry directly, and works with fewer images. First experimental are shown in figure 7.2.

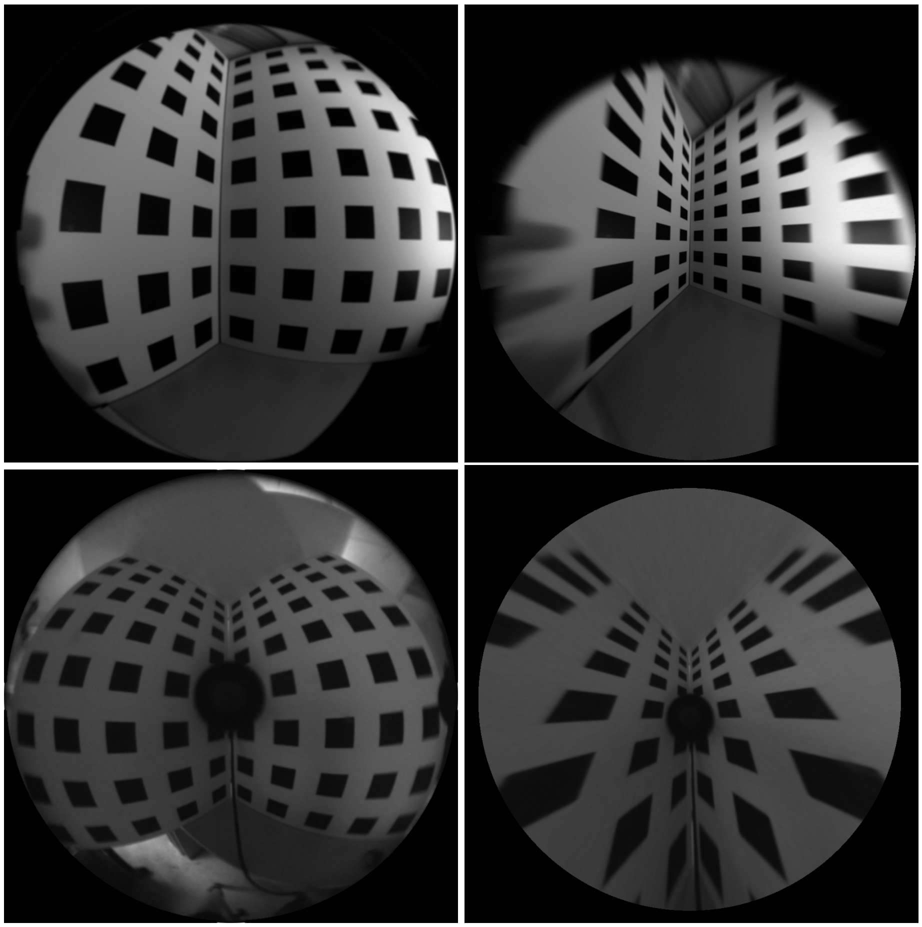

Figure 7.2: Rectification examples based on self-calibration of the radially symmetric camera model. Top: original and distortion corrected image acquired using a 3.5mm fish-eye lens. Bottom: same for a catadioptric camera, with a field of view above 180°. The principal distortion circle (corresponding to the opening angle of 180°) was estimated, and distortion correction is necessarily only shown for an image region inside this circle.

▷ Paper 16 in the accompanying collection [110]

Chapter 8

Structure from Motion

It is relatively straightforward to formulate basic structure from motion tasks for the general camera model defined in chapter 6. We consider the problems of pose and motion estimation and 3D point reconstruction (triangulation).

Pose estimation is usually defined as the computation of the position and orientation of an object with known structure, relative to a calibrated camera, given one image. As for the pinhole model, it is well known that from 3 point correspondences between the object and the image, a finite number of solutions for the pose can be computed. In general, a unique solution is found with 4 correspondences or more. The same statements apply for the generic camera model (and others, as a matter of fact), even in the non-central case [26, 58, 71, 103, 105]. The main difference compared to the pinhole case is that 3-point-pose comes down to solving a degree 8 polynomial, compared to degree 4 for pinhole cameras. Further, in the pinhole case, 4 solutions exist, together with mirrored solutions, whereas in the non-central case there is no such symmetry in general.

As for **motion estimation** with the generic camera model, the counterpart to the perspective essential matrix was introduced by Pless [66]. Lines of sight are represented using Plücker coordinates. Two pixels match one another if the associated lines of sight intersect; this can be expressed nicely in terms of a generalized epipolar constraint with a 6×6 essential matrix. Like in the perspective case, the essential matrix can be computed from image matches and camera motion extracted from it. In [103] we show the essential matrices for the different camera models introduced in chapter 6: non-central, axial, x-slit, and central models.

The last problem, **3D point triangulation**, is the simplest one to solve; it comes down to intersecting lines of sight, or, in the presence of noise, to estimating the closest point to a set of lines.

Note that here, we do not discuss optimal solutions (in the usual sense of minimizing reprojection errors), rather methods that may be used in a RANSAC scheme to find initial solutions. Minimizing reprojection errors, i.e. doing bundle adjustment, is not straightforward for the generic, discrete camera model due to the absence of a differentiable projection function. In [71] we propose a heuristic solution for this problem.

▷ PAPERS 17 AND 18 IN THE ACCOMPANYING COLLECTION [105] [106]

Chapter 9

Multi-View Geometry

Matches of points or other features across images provide information on the structure of the scene, the cameras' calibration and pose. One of the building blocks of multi-view geometry is the set of matching tensors that encapsulate camera calibration and pose. These tensors give constraints that are useful for feature matching. Reciprocally, given matches one may compute the tensors and extract information on camera calibration and pose from them. This has been well studied for the perspective camera model [33, 44]. Bifocal, trifocal, and quadrifocal tensors have been shown to exist, i.e. tensors and associated matching constraints that link two, three, or four cameras. In their usual definition, they act on homogeneous coordinates of image features, typically 3-vectors of homogeneous point or line coordinates.

In [89], we first introduced matching tensors for a set of cameras of different types. We showed that for any mixture of perspective, affine and para-catadioptric[1] cameras, bifocal, trifocal and quadrifocal tensors do exist. The bifocal case, i.e. the epipolar geometry of such heterogeneous systems, was analyzed and approaches for self-calibration were proposed.

In [91, 103], we then examined the existence of matching tensors for the generic camera models considered throughout the previous chapters. Only the calibrated case can be handled, as explained below. In that case, matches between pixels translate to matches between lines of sight. When representing these using Plücker coordinates, one can again derive bifocal to quadrifocal matching tensors acting on them. The bifocal case corresponds to the essential matrix already discussed in the previous chapter. Like for perspective cameras, no matching tensors exist that link more than four cameras at a time.

As mentioned above, only the calibrated case be handled for our discrete camera model. This is because matching constraints essentially express that lines of sight associated with matching image points/pixels, intersect in 3D. With the discrete camera model, no analytical relationship between pixels and lines of sight exists; hence matching tensors acting on pixels can not be established.

So far, we have only established matching tensors related to scene *points*, i.e. that constrain the matching of image pixels (via their associated lines of sight). Similarly to the case of perspective cameras, matching tensors related to scene *points and lines*, can probably be established. We have intuitions on how to do this but did not yet have the time to develop this further.

▷ PAPERS 19 TO 21 IN THE ACCOMPANYING COLLECTION [91] [103] [89]

[1] Orthographic or affine camera looking at a mirror whose shape is a parabola-based surface of revolution.

Part IV

3D Reconstruction

Introduction

We have worked on several aspects and levels of generality of 3D modelling from images. Our first works are purely geometric and lead to coarse 3D models; they are based on sparse interest points or line segments extracted in images which are matched in other images or to which geometric constraints are added interactively. These approaches are described in chapter 10. In chapter 11, we consider similar approaches for the modelling of two special types of dynamic scenes.

In chapter 12, we describe approaches for obtaining denser models. A first approach automatically constructs a triangular mesh on the 3D points reconstructed from interest points that were extracted in the input images and used for structure from motion computations. A second approach gives very dense models, parameterized via depth maps.

Besides modelling scene geometry, we are increasingly interested in modelling reflectance characteristics of objects going beyond the frequently used Lambertian/diffuse model. In chapter 13, we first consider the "opposite" of the diffuse model, namely the reconstruction of perfectly or dominantly specular objects. Recently, we started to consider the joint estimation of geometry and "hybrid" reflectance properties, i.e. with diffuse and specular components. This is described in chapter 14.

Chapter 10

Using Geometric Constraints for 3D Vision

Geometric constraints on the scene can be used in different ways. They can be actively annotated on images by a user or may represent some prior on the scene's structure, e.g. that the scene is piecewise planar. A well-known example for the first case is the Façade system of Debevec et al. [30]. The user constructs a 3D model consisting of a set of basis shapes, by selecting them from a catalogue of CAD models and adjusting their parameters using the available images and calibration information. Such a basis shape can be seen as a set of joint geometric constraints on the 3D location of a set of points and/or line segments for example. In our work, we use more local constraints, for example coplanarity or collinearity of points, orthogonality or parallelism of line segments or planes, equal distances between two pairs of points, etc. Some of these constraints are compactly encoded using a simple 3D basis shape, namely the parallelepiped. This line of work is discussed in the third paragraph below.

In the following two paragraphs, we describe two other sets of works, concerning coplanarity and collinearity constraints.

Piecewise planar scenes.

In [2, 5, 10–13, 27, 28] we propose representations and algorithms for the reconstruction of the piecewise planar parts of scenes. The information on which points are coplanar is supposed to be given (can be provided interactively, but also determined automatically). Among others, we show how to minimally parameterize sets of coplanar points, for subsequent bundle adjustment type optimization of structure and motion. The parameterization was chosen such that individual points may belong to more than one such plane (e.g. corner point of a house). Using experiments, we evaluated the quality of structure and motion estimation, especially when the ground truth 3D points are not exactly coplanar, as it is the case in real scenes (windows, uneven walls, ...). We evaluated up to which amount of non-planarity the enforcement of coplanarity constraints allows a better structure and motion estimation than without coplanarity constraints.

In another work, we exploited an object's piecewise planarity for tracking it in video sequences [29].

▷ PAPERS 22 TO 24 IN THE ACCOMPANYING COLLECTION [5] [28] [29]

Structure from motion for lines.

In [3, 4, 6, 8, 9] we propose a series of algorithms for motion estimation and 3D reconstruction using lines or line segments. This work does not fit exactly in this chapter, since we mostly work directly with line features extracted in images, without using geometric constraints explicitly; nevertheless, it is naturally embedded here, representing collinearity among coplanarity and other geometric constraints dealt with in this chapter. We formulated motion estimation using correspondences between line features, for projective, affine and euclidean motion, i.e. for uncalibrated and calibrated cases. Various algorithms have been developed, based on different optimization criteria (3D or 2D residuals). We also devised methods for optimal triangulation of line features and full bundle adjustment of structure and motion.

▷ PAPERS 25 AND 26 IN THE ACCOMPANYING COLLECTION [8] [9]

More general geometric constraints.

In the following, we describe our approaches to using geometric constraints that are provided interactively by a user, e.g. by annotating them on input images. Constraints we use are for example the coplanarity or collinearity of points, orthogonality or parallelism of line segments or planes, the fact that distances between two pairs of points are equal, etc. The motivations for using such constraints are mainly twofold. First, incorporating constraints in structure and motion estimation may lead to more accurate results, as already mentioned in the first paragraph of this chapter. Second, one can obtain 3D models from a single image already or from multiple images taken from very different viewpoints. In the latter case, automatic image matching may be difficult, but a few manually given matches plus constraints can allow to estimate structure, pose, and camera calibration if required. The complete theory and associated algorithms we developed are described in [118–123, 125, 126]. In some of our work, we use parallelepipedic basis shapes to compactly encode geometric constraints such as orthogonality or coplanarity. Our approaches for 3D reconstruction from a single image are given in [85, 98, 117]; a few results are shown in figures 10.1 and 10.2.

In most of the works cited in this paragraph, we used geometric constraints to compute initial values of structure and motion parameters. As for their subsequent optimization (bundle adjustment), we examined ways of obtaining minimal structure parameterizations satisfying the given constraints. This was done via a collaboration with Gilles Trombettoni from INRIA Sophia Antipolis who works on constraint decomposition techniques[1]. A technique for automatically extracting such parameterizations from an unstructured and possibly redundant set of constraints was adopted and used to generate 3D models from larger image sets [124]. Results are shown in figures 10.3 and 10.4.

▷ PAPERS 27 TO 29 IN THE ACCOMPANYING COLLECTION [98] [85] [126]

[1]This collaboration was initiated and managed by our PhD student Marta Wilczkowiak.

Figure 10.1: Results of 3D reconstruction from single images. Three original images are shown, together with renderings of the reconstructed 3D models.

Figure 10.2: Results of 3D reconstruction from a single panoramic image, taken with a catadioptric camera. Top: input image and 3D model. Middle and bottom: renderings of the texture-mapped 3D model.

Figure 10.3: Results of 3D reconstruction of “Place Notre Dame” (Grenoble) from 15 images. Top: 4 of the input images. Shown are parallelepipeds used for computing initial estimates of structure and motion. Middle-left: estimated parallelepipeds and camera positions using a proposed factorization method. Middle-right: final result of non-linear optimization. Bottom: texture-mapped model rendered from different viewpoints.

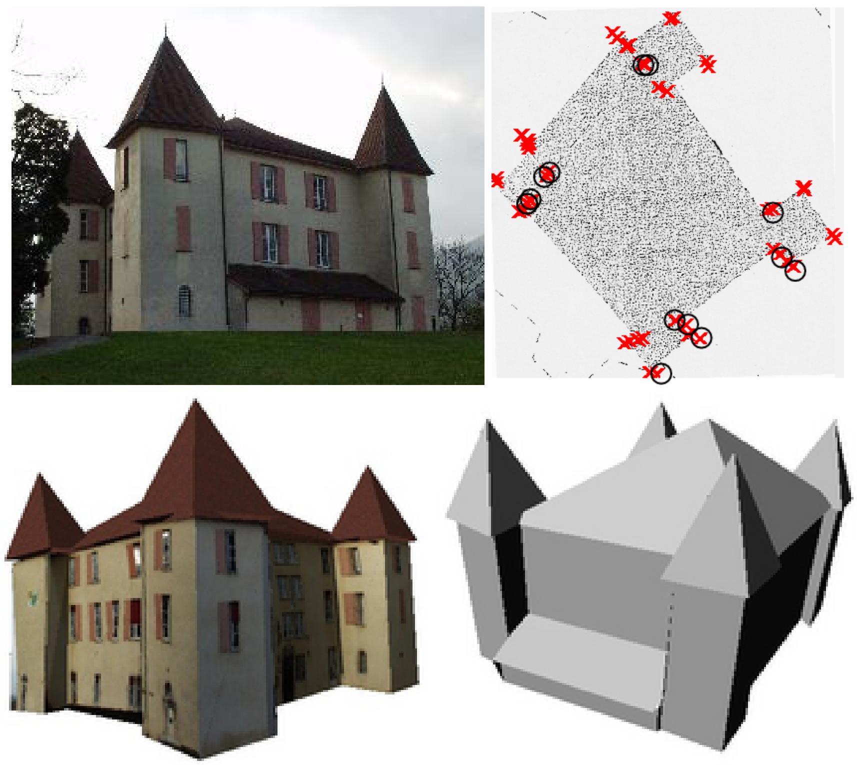

Figure 10.4: Results of 3D reconstruction of “Château Miribel” (Montbonnot St Martin) from 7 images. Top: an input image and a coarse map of the castle, used as an additional input image (modeled as orthographic image). Bottom: final 3D model.

Chapter 11

3D Reconstruction of Dynamic Scenes

We have considered two rather special cases of dynamic scenes.

Points moving in planes.

This work was inspired by works of Shashua and Wolf [81, 127], where the two following scenarios were considered. In [81], the scene consists of a single plane in which points move independently from one another, but on linear trajectories. This scene is observed by a moving camera and it is shown how the scene, i.e. all points and their motions, can be reconstructed from three images (projective reconstruction). In [127], the scene consists of points moving independently in 3D, again on linear trajectories. Here, 3D images are considered, e.g. acquired using a stereo system. Again, the motions of all points' can be reconstructed in some common coordinate system.

In [88], we considered a similar scenario. The scene is again supposed to consist of independently moving points. Each point moves in some plane and the set of planes are supposed to form a pencil. One motivation for studying this scenario was as follows. In urban scenes for example, objects move on a ground plane and individual points on objects remain often at the same height, i.e. each point moves in a plane and the set of planes form a pencil (being parallel to one another). This assumption holds for rigid objects performing planar motions but also to some extent for walking humans. Still, the main motivation was to study an original scientific problem.

We have shown that when points do not have linear trajectories, the scene can be reconstructed up to a few degrees of freedom. With additional information, on a few points that are static or that follow linear trajectories, the scene can be completely reconstructed.

▷ PAPER 30 IN THE ACCOMPANYING COLLECTION [88]

Exploiting gravity for calibration and pose estimation.

In [99], we studied the following question: what information on camera calibration and pose can be extracted from image sequences of moving objects, if their trajectories are dictated by gravity? By this, we mean either free falling objects or free flying ones, i.e. objects that are launched and follow a parabolic trajectory. We have shown what can be obtained on the calibration and relative pose of either a single or multiple cameras observing such objects. It was also shown that video sequences acquired by different cameras can be synchronized using this information.

▷ PAPER 31 IN THE ACCOMPANYING COLLECTION [99]

Chapter 12

Multi-View Dense 3D Reconstruction

Mesh-based models.

Our first works on automatic 3D surface modelling were within the European project VISIRE. One of its goals was to reconstruct interior scenes, e.g. a hall of a museum. Figure 12.1 shows our most complete result. The scene was reconstructed from hand-held video sequences. An important practical problem is that it is rather difficult to capture the entire scene in a single video sequence, such that it can be reconstructed: each part of the scene has to be filmed from at least two different viewpoints. We thus acquired about 35 video sequences. From each sequence, a metric reconstruction of a part of the hall was obtained automatically, using the "standard" processing chain of feature extraction and tracking, projective reconstruction, self-calibration and bundle adjustment. These models were then registered semi-automatically, starting with a few manually defined point matches between pairs of models, followed by a refinement using ICP (Iterative Closest Point) [15].

Finally, a triangular mesh was computed on the complete set of 3D points. We have developed several methods, consisting in using different criteria for accepting/rejecting hypothetical triangles: photoconsistency, constraints on triangle shape, visibility constraints (e.g. no triangles should lie between a 3D point and the optical centers of cameras where the 3D point was extracted) [78, 79].

▷ PAPER 33 IN THE ACCOMPANYING COLLECTION [79]

Figure 12.1: 3D model of the upper part of a hall in Palazzo Pitti, Florence (dimensions: about $25m \times 10m \times 10m$). The gap visible in the leftmost image is due to insufficient overlap in the video sequences. Especially the rightmost image highlights inaccuracies, possibly resulting from too small baselines in many places.

Models represented by depth maps.

In [36, 37], we propose an approach for dense 3D modelling, that was mainly inspired by that of Strecha et al. [83]. 3D structure is represented via the union of depth maps for all images, with associated colors: per pixel per input image, a 3D point is estimated along the associated line of sight, together with its color. Like Strecha et al. [83], we handle occlusions and unmodeled effects (e.g. specularities) using hidden visibility variables: for each 3D point and each input image, we estimate the probability of that point being visible in that image.

A Bayesian formulation is established for this estimation problem. Further ingredients are a likelihood term and priors on the visibility and depth variables. The likelihood is based on photoconsistency; it is important to note that ideally, one would like to compute the likelihood of the variable estimates as the product of the likelihoods computed for all pixels. Many previous approaches though, compute products of likelihoods per 3D point; this is not desirable, as the observations (pixels) are not used equally: depending on the current estimate of 3D structure, some pixels may contribute to many more such individual likelihoods than others. We propose an approximation to the ideal case of per-pixel likelihoods. Further details on this and other aspects of our approach are given in [36].

Figures 12.2 to 12.4 show 3D models obtained with our approach, from a simple to a rather difficult scene.

Although we have obtained a set of very satisfying results, we also have experienced problems with our approach. On the one hand, the depth map representation is very attractive since it allows for highly dense 3D models and in addition is naturally well adapted to the resolution of the input. On the other hand, it does not represent a surface. It is desirable to formulate 3D modelling as a surface estimation problem; whatever the parameterization of the surface used, the desired evolution of the latter should be transposable to the evolution of the underlying representation. This is not entirely possible with our parameterization. For example, consider the case of an occluding contour in an image and suppose that its position is wrongly estimated, at the initialization or another stage of the estimation (cf. the upper "rim" of the statue in figure 12.4). The desired surface evolution would consist in displacing it orthogonally to the lines of sight corresponding to the occluding contour. Due to the depth map parameterization, individual points can only move along their line of sight, not laterally. To obtain the desired surface evolution, individual depth values would have to make huge jumps, and these are usually not found by a local optimization.

Having said that, we are nevertheless satisfied with the results obtained so far, in our first work on dense 3D modelling. We have several ideas for future work, concerning other parameterizations and stereoscopic segmentation [51] for example.

▷ PAPER 32 IN THE ACCOMPANYING COLLECTION [36]

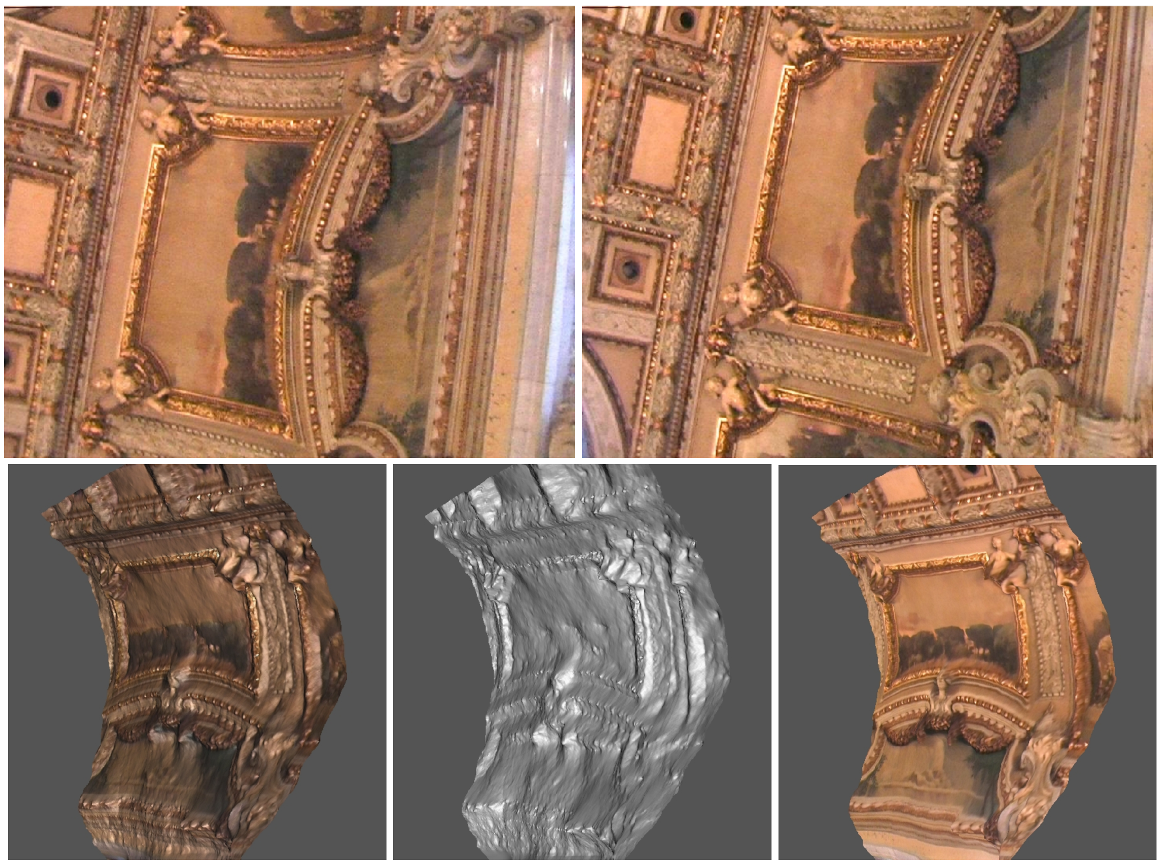

Figure 12.2: Top: two of a total of five input images. Bottom: reconstructed 3D model. From left to right: texture-mapped with additional shading, only shading, only texture-mapped.

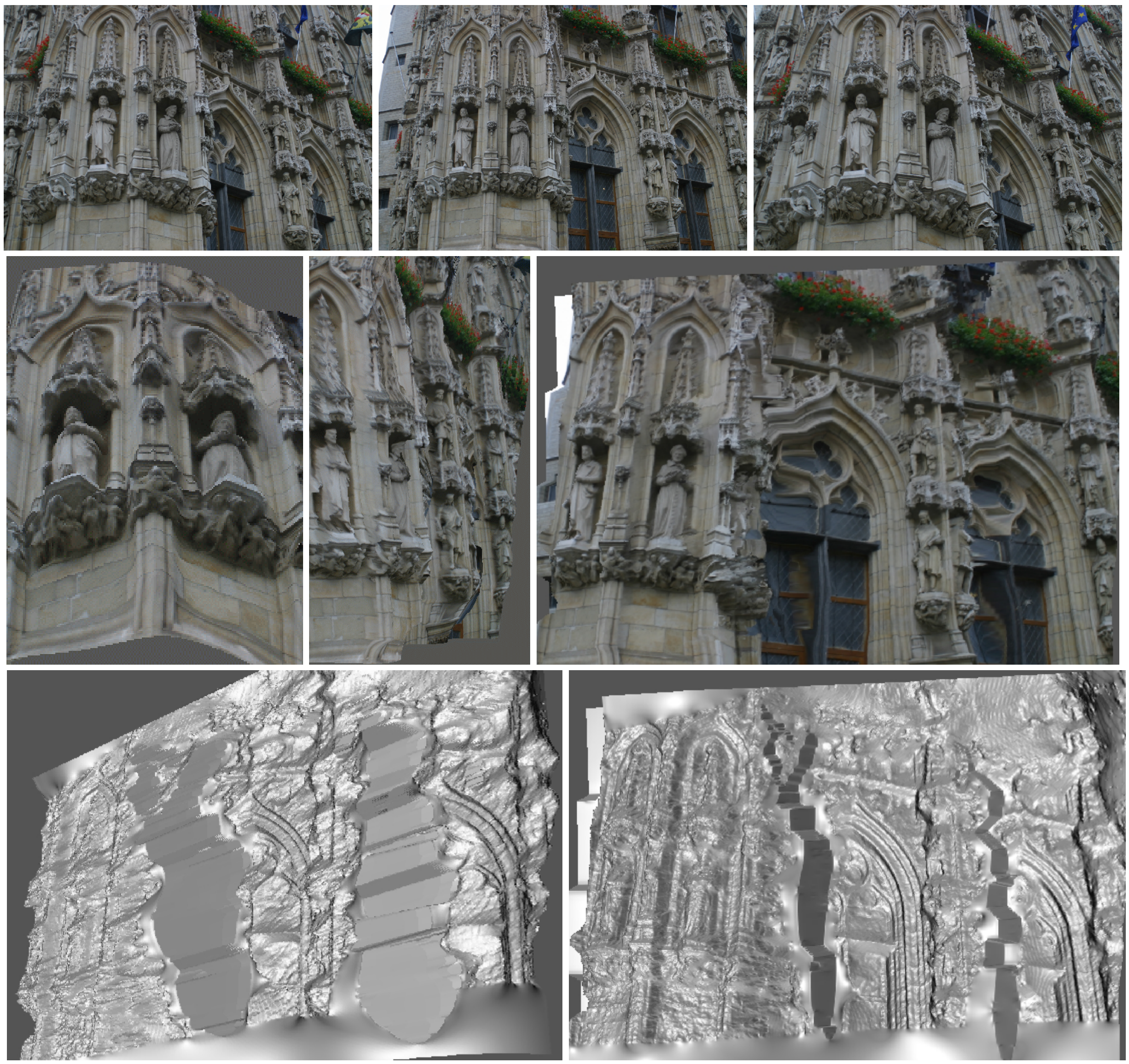

Figure 12.3: Top: three input images (from http://www.esat.kuleuven.ac.be/∼cstrecha/testimages/). Middle: texture-mapped 3D model. Bottom: shaded 3D model, highlighting estimated depth discontinuities. Also note the smooth parts of the model, corresponding to parts visible in a single image only. The depth prior we use seemingly allows all, strong discontinuities, sharp details and smooth surface parts.

Figure 12.4: Top: five input images. Second row: initial model (obtained by "smoothing" the cloud of reconstructed interest points) and the estimated visibility map for the first input image relative to the third one (white means that the associated pixel of the first image is deemed likely to be visible in the third image). Bottom rows: texture-mapped 3D model. Parts of the background are only visible in one or two images and are necessarily badly reconstructed. The rest of the scene is quite well reconstructed though; especially, the large discontinuity between the statue and the background has been found completely.

Chapter 13

3D Reconstruction of Specular Surfaces

We have developed two approaches for the 3D reconstruction of specular surfaces. They are based on images of a reference object's reflections in the specular surface. In practice, we use a planar calibration grid, but other objects may be used too. Both methods require a completely calibrated setup: calibrated camera and known poses of the camera and reference object with respect to some global reference frame. Methods for obtaining these poses are discussed further below. Further, both methods require the matching between images and the reference object (we use the approach described in chapter 6).

Voxel carving.

The first approach we developed is an extension of *voxel coloring* [80], from Lambertian to specular surfaces. Photoconsistency measures can not be directly used for specular surfaces: when looking at the same specular surface point from different viewpoints, one in general sees each time the reflection of a different point of the surrounding scene (reference object), i.e. a different color. We thus use other consistency measures to decide upon the voxels' occupancy, as follows. Our voxel carving approach proceeds in two steps, a first step where data is accumulated on the voxel grid, using all input images and a second step where voxels are declared as belonging to the specular surface or not. In the first step, the following operation is carried out for each point in each image for which a match to a point on the reference object has been obtained. The image point's line of sight is traced in 3D; for each voxel that is traversed, we compute a hypothetical surface normal vector: if the voxel were on the true specular surface, then the surface normal at that place would have to be the bisector of the line of sight and the line joining the voxel and the matched point on the reference object. With sufficiently many images and matches, voxels will be traversed several times and several hypothetical normal vectors will be associated to each of them. Only for voxels lying on the actual specular surface, the associated normal vectors will be identical (modulo noise). In the second step of the algorithm, we thus declare voxels as being on the surface if their normal vectors are sufficiently consistent with one another. Several heuristic consistency criteria have been used, see [19, 20] for more details.

This approach relaxes assumptions on the continuity of the specular surface that were typically made by other approaches, which integrate the surface whereas we reconstruct it point by point. Nevertheless, the approach suffers from a rather small spatial resolution. This is remedied by the following approach.

Triangulation approach.

In the above approach, no constraints were put on the camera's and the reference object's displacements between image takings, besides that the resulting poses can be somehow determined (see below). For the following approach, we require the camera to remain stationary (with respect to the specular surface). Then, from two or more images of the reference object's reflections, when putting it in different positions, we may reconstruct points on the specular surface as follows. Let us consider an image point and the matching points on the reference object, for two or more of its positions, i.e. the points whose reflections are seen in that image point. Since we assume to know the poses of the reference object, we can compute the line in 3D joining these matching points. Its intersection with the image point's line of sight gives a point on the specular surface [21].

This approach is extremely simple, but can achieve a rather high resolution and accuracy, depending of course on the density of the matches, the camera's resolution, etc. As for matching, we use the structured light approach described in chapter 6. Note that Kutulakos and Steger have formulated the same approach for more general scenarios, handling refractive surfaces in addition to specular ones and handling double reflections [53].

Figures 13.2 and 13.3 on page 78 show experimental results, for the reconstruction of four toy mirror surfaces and a car windscreen.

Pose computation.

The above methods require the pose of the camera and the reference object for all images, in some global reference frame. We have developed several methods to compute that. The simplest one is applicable when the reference object is positioned such that the camera has a direct view of it. In that case classical pose estimation can be performed. This scenario is not very practical though: part of the camera's resolution is "spent" on the direct view of the object and thus lost for the specular surface. Further, positioning of the objects in order to obtain both, a good direct and reflected view, may be cumbersome. We thus have developed other methods.

For experiments with our voxel carving approach, we attached the camera rigidly to the reference object. The camera did not have a direct view of the object, so we computed their relative pose off-line, using a hand-eye calibration type method [19, 20]. Then, during image acquisition, the reference object's displacements (and hence, those of the rigidly attached camera), were computed using a fixed stereo system, cf. figure 13.1.

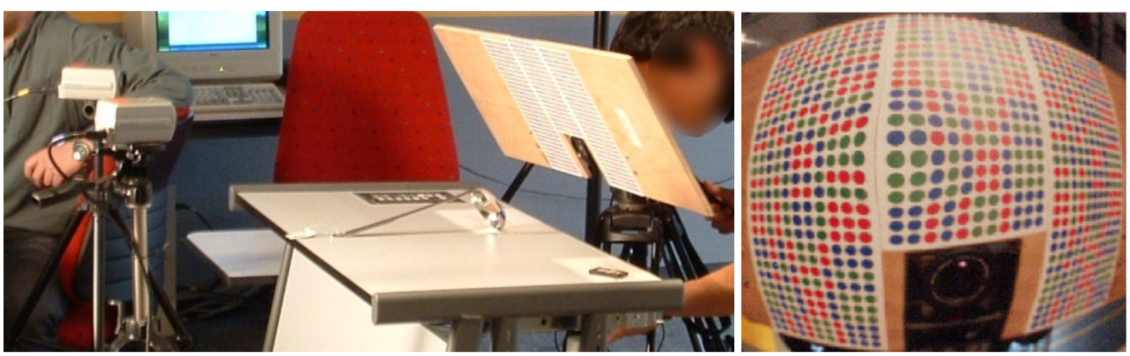

Figure 13.1: Left: experimental setup used in [19, 20]. The reference object is rigidly attached to the camera. The stereo system on the left is used to compute the displacement of the reference object, and thus of the camera. Right: a typical image of the reference object's reflection in the specular spoon.

This approach was still relatively cumbersome, due to the required stereo system and hand-eye calibration. Ideally, one would like to compute the reference object's pose directly from the images of its reflections. If the specular surface and its pose are already known, this becomes a classical pose estimation problem which can be solved for each individual image. In our case this is not possible. Nevertheless, we have shown that pose estimation is indeed possible, from two or more images of an object's reflections in an unknown specular surface [18]. This involves a formulation akin to trilinear tensors between three perspective images, as summarized in the following.

Consider matches between one image point and the points on the reference object, for two of its positions. The goal is to estimate the two poses of the reference object, in the camera's coordinate frame. Without loss of generality, we thus suppose that the camera is in canonical position. Let the line of sight associated with the image point be represented by the optical center (the origin) and a direction $\mathbf{D}$. Let $\mathbf{Q}_i = (X_i, Y_i, 0, W_i)^\mathsf{T}, i = 1, 2$ be the homogeneous coordinates of the reference points, in the reference object's local coordinate system. The poses of the reference object are represented by rotation matrices R_i and translation vectors $\mathbf{t}_i$. The line of sight and the two reference points must be coplanar, which can be written as:

$$\det \begin{pmatrix} \mathbf{0} & \mathbf{D} & W_1\mathbf{t}_1 + \mathsf{R}_1 \begin{pmatrix} X_1 \\ Y_1 \\ 0 \end{pmatrix} & W_2\mathbf{t}_2 + \mathsf{R}_2 \begin{pmatrix} X_2 \\ Y_2 \\ 0 \end{pmatrix} \\ 1 & 0 & W_1 & W_2 \end{pmatrix} = 0$$

This simplifies to:

$$\det \begin{pmatrix} \mathbf{D} & W_1\mathbf{t}_1 + \mathsf{R}_1 \begin{pmatrix} X_1 \\ Y_1 \\ 0 \end{pmatrix} & W_2\mathbf{t}_2 + \mathsf{R}_2 \begin{pmatrix} X_2 \\ Y_2 \\ 0 \end{pmatrix} \end{pmatrix} = 0$$

This equation is trilinear in the coordinates of $\mathbf{D}$ and the $\mathbf{Q}_i$ and can be written as

$$\sum_{ijk} D_i Q_{1j} Q_{2k} \mathcal{T}_{ijk} = 0$$

with $\mathcal{T}$ being a "trifocal" tensor depending exactly on the two poses to be determined. After estimating $\mathcal{T}$ from matches, the poses can be extracted relatively easily from it.

This approach, together with the triangulation method described above, allows to reconstruct a specular surface from only two images of a reference object's reflections, without requiring a priori pose information. Unhappily, the estimation of the tensor turned out to be too unstable in the presence of noise. In some sense, this minimal scenario seems to represent a practical limit, although solutions for even more general circumstances should exist, e.g. without using a known reference object. Let us also note that the shape of specular objects can be perceived from reflections of unknown scenes, using e.g. statistics on natural images [35].

A last approach for pose estimation represents some compromise between flexibility and accuracy. In [95] we show that when putting a planar mirror in three or more positions such that the reference object's reflection in it is seen by the camera, we can compute its pose.

▷ PAPERS 34 TO 36 IN THE ACCOMPANYING COLLECTION [19] [21] [95]

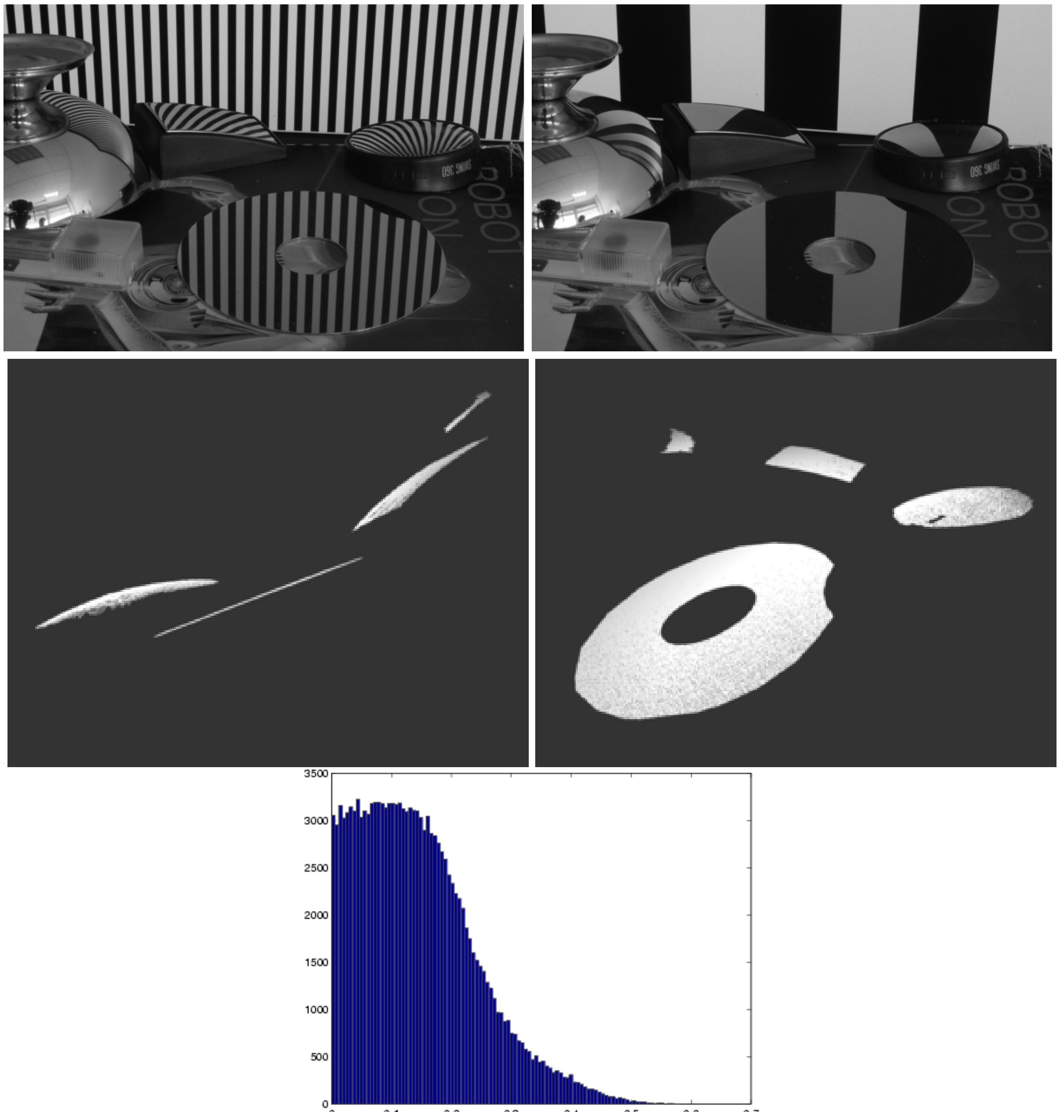

Figure 13.2: Experimental results on 3D reconstruction of specular surfaces. Top: original images, showing two of the patterns displayed on a flat screen, used in the structured light type matching approach. Middle: reconstructed point clouds. Bottom: quantitative evaluation; histogram of residual distances of points to the best-fit plane, for the ring-shaped planar object (a hard drive platter). The unit on the x-axis is mm. The total number of reconstructed points for that object was over 500,000.

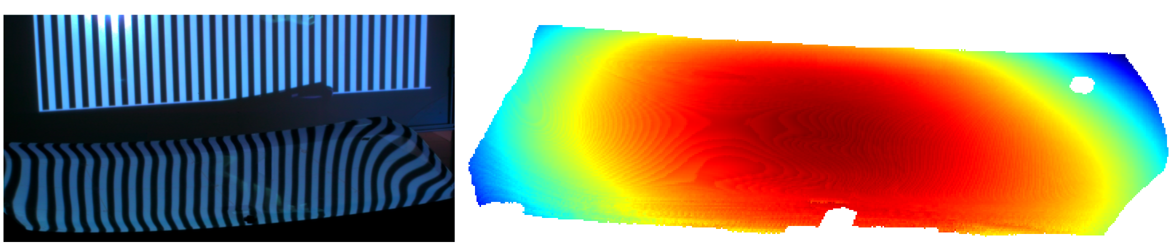

Figure 13.3: Reconstruction of a car's windscreen. Left: one input image. Right: height map of the 3D model. A quantitative evaluation was not possible (a reference CAD model corresponded to another position and thus to another bending of the object).

Chapter 14

Modelling of 3D Geometry and Reflectance Properties

We are convinced that it is important for 3D modelling to take into account reflectance properties of objects to be reconstructed and the illumination. On the one hand, estimating reflectance properties leads to richer models, allowing for renderings of better quality and under simulated lighting conditions. On the other hand, images are a function of several factors: scene and camera geometry, the scene's reflectance and the camera's radiometric properties, and the illumination. Hence, neglecting one of these factors will in general bias the estimation of the others. However, estimating all those factors is not possible without additional information besides images and/or without making simplifying assumptions. There are many possibilities for doing so: adopting a particular reflectance model (Lambertian, Phong, . . .), off-line modelling of the illumination, making hypotheses on the scene's geometry (absence of inter-reflections, objects being of one among a set of learnt models), putting objects in front of some artificial background, etc.

In the work described in this chapter, we use some of these possibilities: an artificial light source is used and its position is calibrated off-line, we adopt the Phong reflectance model, and use an artificial background [17]. From image sequences acquired under these conditions, we then estimate both, the scene's geometry and reflectance properties. The geometry is represented by a triangular mesh and reflectance properties are estimated for a densely sampled set of points on the mesh. For each individual point, we estimate reflectance coefficients; it is well known though that the specular coefficients can only be estimated if specular highlights are among the observations. With typical numbers of light sources and images, this is not the case for all surface points. Hence, for points with insufficient observations, only the albedo is computed; as for the specular coefficients we heuristically attribute those of another surface point for which they could be estimated and which has a similar albedo.

Scene geometry is estimated using a variational approach; this is alternated with the estimation of reflectance coefficients. The geometry is initialized by the visual hull, computed using silhouettes extracted due to the artificial background. For image acquisition, objects to be reconstructed were put on a simple turntable, whose motions were computed using markers tracked in the images. Two sequences were acquired, for different positions of the light source, in order to provide richer observations for the estimation of reflectance. Promising results on a variety of objects have been obtained, cf. figures 14.1 and 14.2.

▷ PAPER 37 IN THE ACCOMPANYING COLLECTION [17]

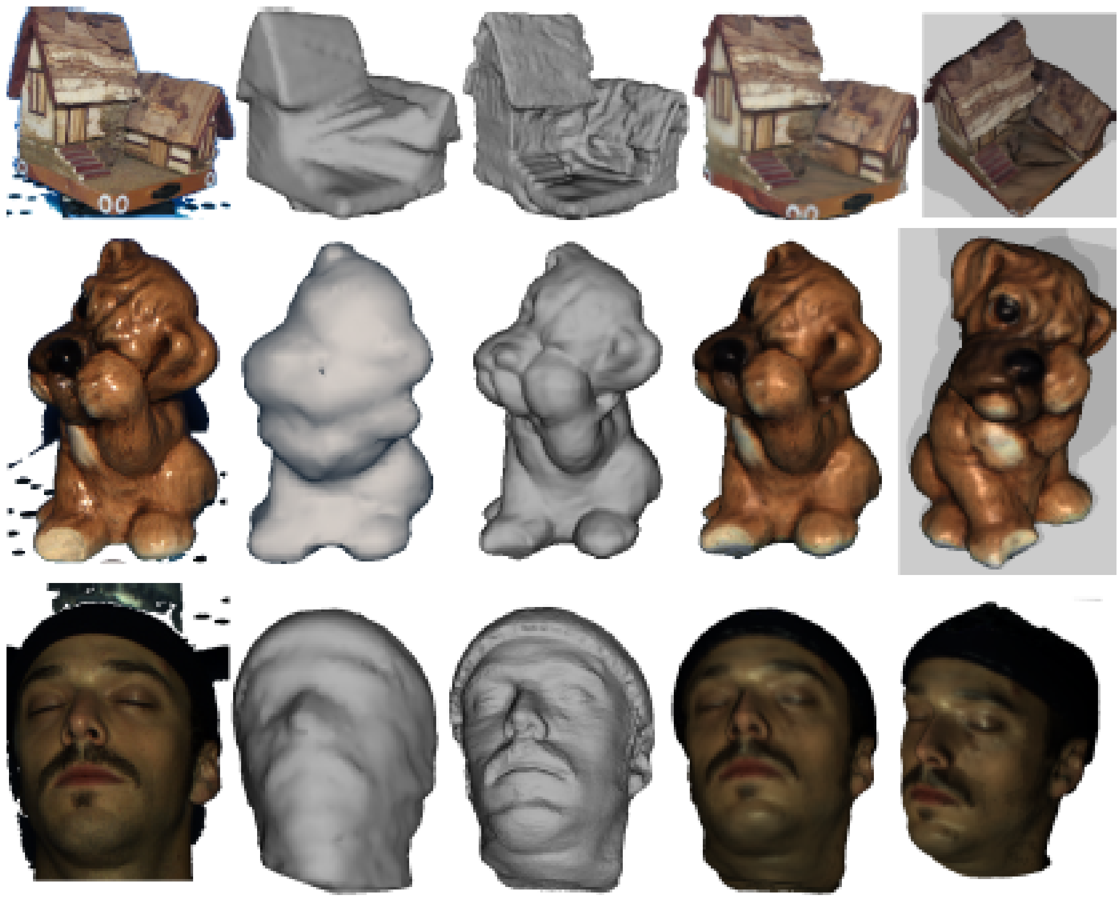

Figure 14.1: Modelling results for three objects. From left to right: input image, visual hull, final shape, texture-mapped model, relighted model.

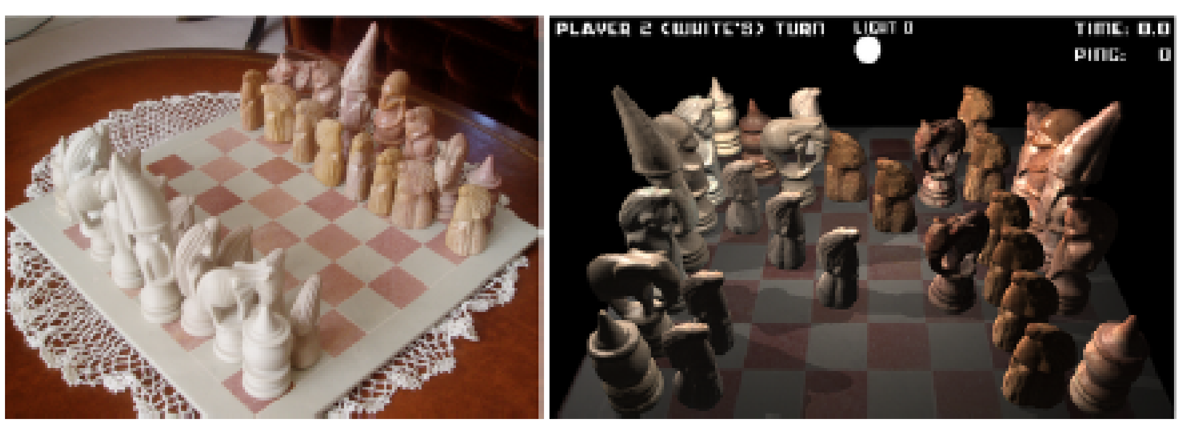

Figure 14.2: Right: image of a chess game. Right: Relighted model. All figures were modelled separately.

Part V

Other Works

Chapter 15

Object Tracking

Within a CIFRE contract (partial sponsorship of Aude Jacquot's PhD thesis by Thales Optronique), we work on object tracking in video sequences. Application constraints impose that no accurate and detailed model for objects to be tracked is available and that tracking has to start with an extremely simple manual initialization (a user clicks in the middle of the image region corresponding to the object and/or draws a bounding ellipsis around the object). The general approach we use is based on particle filtering, which allows to estimate a probability density function on the object's state (position, size, shape) and propagate it from one image to the next. One of the main problems is the choice of representation for the object. We have been using several of them: representation of "appearance" (grey-level or color histograms) or structure (one of a set of planar or three-dimensional basis shapes: rectangle, ellipsis, parallelepiped, etc.). Two aspects of our works are: *(i)* the object representation may change during tracking. For example, when approaching an object from a far position, an initial planar model may evolve into a more precise three-dimensional one, once enough information is contained in the images [49]. *(ii)* many approaches decide in advance on the number of bins in color histograms; we use a model selection criterion [16] for setting it automatically [47, 48].

▷ PAPER 38 IN THE ACCOMPANYING COLLECTION [47]

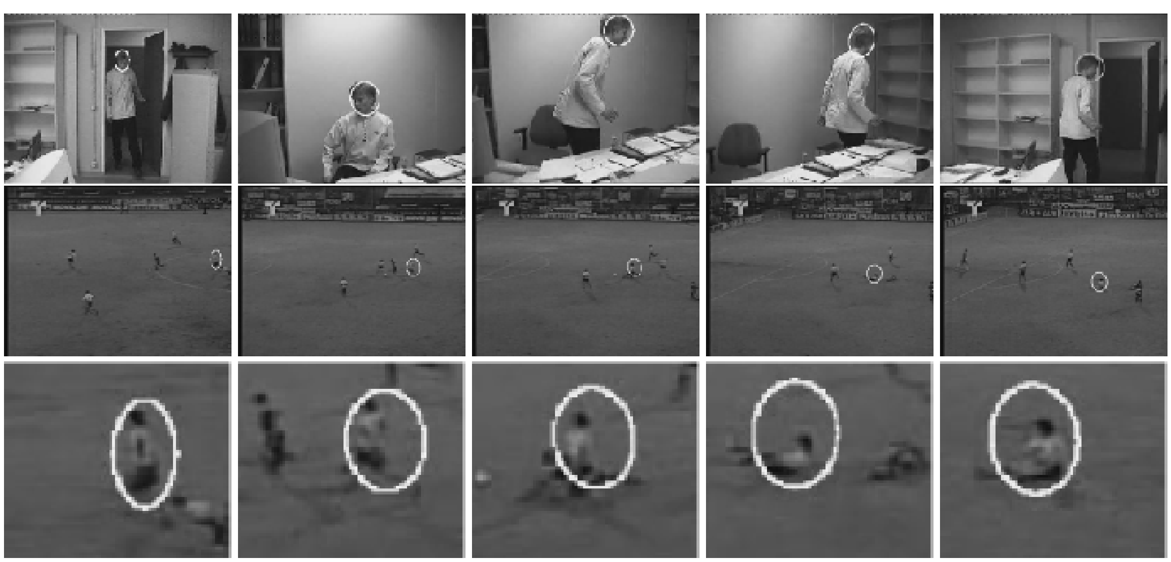

Figure 15.1: Tracking results using grey-level histograms as object representation [47].

Chapter 16

Model Selection for Two-View Geometry

Many computer vision problems require to solve a model selection problem, i.e. to choose between different models to explain observations, or better, to attribute probabilities to such models. The aim is to obtain a good compromise between the quality of fit to observations (likelihood) and the complexity of the model ($\sim$ prior on models). Several generic approaches have been proposed in the literature, e.g. BIC (Bayes Information Criterion), AIC (An Information Criterion, often called Akaike Information Criterion), and MDL (Minimum Description Length). We used the latter to choose between models for two-view geometry, i.e. models that explain point matches between two images. We did not use the typical approximate MDL criterion found in the literature that attributes a score to each model according to a function depending only on a likelihood term and the number of dimensions of the model. Rather, we tried to use MDL literally, by encoding the observed data using each model in turn, together with the model parameters. The shortest such code would correspond to the best model.

The models used were the homography (corresponding to a planar scene), fundamental matrix, affine fundamental matrix, and "background" (no global geometric relationship between the matches). Encoding matches using the homography model can be done by encoding the coordinates of the points in the first image, the coefficients of the homography, and the residuals between points in the second image and those of the first image mapped by the homography. As for the fundamental matrix model, points in the second image are encoded via a coordinate corresponding to their position along their epipolar line, plus the residual (distance to epipolar line).

Ideally, if the scene is planar then the homography model should win and if sufficiently non-planar, the fundamental matrix should lead to a shorter code. However, our method nearly always preferred the homography model [56, 57]. This can be explained as follows. If the scene is highly non-planar, then encoding matches using the homography model requires encoding large residuals in the second image, i.e. two large numbers per match (x and y residuals). However, even for the fundamental matrix model, encoding position along the epipolar line requires encoding one large number per match, usually larger than the two numbers for the homography model. Hence, the two models do not give rise to codes of significantly different lengths.

In [56], we gave arguments that for more than two views, a model corresponding to non-planar scenes has a better chance to become the preferred model with the proposed approach. Nevertheless, it was disappointing that for matches between two views, our method would nearly always decide that the scene is planar even if a human observer, when looking at the images, would opt without hesitation for the non-planar model. Above, we gave a technical explanation, reasoning about code lengths, but in hindsight, another explanation may be as follows. When looking at images and interpreting them, humans use of course a lot of prior information. A typical prior may be that a typical scene is composed of a set of smooth surfaces. Using such a prior in the encoding procedure for the fundamental matrix model would be possible: positions on epipolar lines should not be encoded independently for each point, but e.g. by encoding differences in these positions for points whose matches in the first image are neighbors. With such a scheme, the result may match more closely the conclusions made by a human observer.

▷ PAPER 39 IN THE ACCOMPANYING COLLECTION [56]

Part VI

Conclusions

In this document, I have provided an overview of my research in the period of roughly 1998 to 2005 as well as research related and administrational activities. The description of my scientific results is completed by an accompanying collection of representative papers. Conclusions and perspectives are outlined in the following.

My research has mostly been targeted towards the representation, understanding and estimation of geometrical descriptions from images. This concerns the internal geometry of cameras (calibration), displacements (motion estimation) and descriptions of scene geometry (3D reconstruction). My research has been following three main threads, corresponding to parts II to IV of this document:

- Calibration and self-calibration of perspective cameras.
- Generic camera models and unified treatment of structure from motion.
- 3D reconstruction.

Calibration and self-calibration of perspective cameras.

Most of my activities in this area were carried out up to 2003. The results obtained, together with those of my PhD thesis, constitute a recognized set of contributions to the area. Our most significant work in this area, in the period covered by this document, is probably that on plane-based camera calibration. As for self-calibration, the set of our results contribute to complete the mosaic of different facets of the problem, via the study of critical motions and the development of algorithms for special cases, such as planar scenes, planar motions, focal length self-calibration, etc. Like many others in the computer vision community, I think that the main theoretical results one can expect in this area have already been found.

As for self-calibration in practice, meaning self-calibration of internal parameters but also, and maybe more importantly, motion estimation, it would be good to establish benchmark data and evaluation procedures, similar to what exists for binocular stereo matching[1]. This is one of the goals of the ISPRS Working Group "Automatic Calibration and Orientation of Optical Cameras" which I co-chair.

One work I would have liked to do since several years, but never had the time or manpower to do so, consists in developing an interactive calibration system that would work as follows. A flat screen would play the role of the calibration grid. Based on the already acquired images, the system would compute the next best viewpoint, i.e. the viewpoint such that the acquired image would most reduce the calibration's uncertainty. The user can be guided in the appropriate positioning of the camera for example by displaying the desired image, overlaid on the current one, on the flat screen. Finally, the calibration pattern, displayed on the screen, may be dynamically adapted in order to achieve the highest calibration accuracy. Such a system would allow for foolproof, accurate, and quick camera calibration.

Generic camera models and unified treatment of structure from motion.

We have studied generic camera models, consisting of lookup tables that associate coordinates of lines of sight to individual pixels. Our contributions cover the topics of calibration and self-calibration, structure from motion and multi-view geometry. For several of these topics, one may consider our works as the most general ones that exist to date. We consider these for now mainly as conceptual contributions: the possibility of using the same algorithm for calibration, motion estimation, etc. for any type of camera, is appealing and clarifies similarities between many existing works in the literature.

Let us discuss their potential relevance in practical applications. The algorithms for pose and motion estimation with minimum amount of data, i.e. point correspondences, are useful general tools. Often, researchers working on omnidirectional vision devise new such algorithms for each new camera technology,

[1] Middlebury Stereo Vision Page `http://cat.middlebury.edu/stereo/`.

overlooking that all that is usually needed (for the geometry part of the problem) is such a minimal method, embedded in a RANSAC-type scheme, and a bundle adjustment. As for calibration using generic camera models, the potential usefulness depends on the following factors. For calibration with low accuracy requirements and/or of cameras for which a classical model is trusted, associated classical methods are certainly appropriate. Generic models may be interesting for calibrating the rather complex thus vulnerable catadioptric cameras or other cameras with a very large field of view. Also, they may contribute to a highly accurate calibration, including the estimation of local optical aberrations, similarly in spirit to [23, 77].

We continue to work on generic camera models for one or two more years. Our goals are to finalize several issues, among which theoretical proofs for self-calibration and the formulation of matching tensors for lines and points. We are also developing a generic calibration toolbox, currently under Matlab. Our aim is to distribute it as Open Source software and to make it a useful tool for others.

3D reconstruction.

Prior to about 2003, our approaches on 3D reconstruction were mostly of geometry-only type, generating coarse 3D models from sparse feature matches, usually together with camera motion estimation. To this end, we have exploited geometric constraints on the scene structure, to get accurate results and/or be able to work with few images. We then started to work towards obtaining denser models, both in terms of geometry and "appearance" (here, reflectance properties), with promising results.

We think that modelling of 3D geometry and appearance from images is still an unsolved problem in general. A complete system should be able to work in the presence of semitransparent objects (windows, trees, curtains), shiny or uniformly colored surfaces, shadows, etc. We would like to work towards that goal. This will require progress on several levels:

- developing hybrid representations for models. For example, for some parts of a scene, the best "appearance" model may be a stochastic texture, whereas for other parts, one of a set of candidate reflectance models is more appropriate (e.g. estimating a texture map or the parameters of a low-parametric reflectance model with a specular component, or selecting a complex reflectance model in a catalogue of expected building materials). This will involve segmentation and model selection. On a more abstract level, one may think in the following terms: the goal of the modelling is to "explain" the input, i.e. a set of images or videos, given available prior knowledge. The explanation, or model, could be a mixture of e.g. object recognition results and explicit 3D models: parts of an image may be explained by recognizing them as showing objects contained in some database whereas to explain others, models like those mentioned above may have to be generated.
- working with more realistic models, e.g. taking into account ambient light when considering surface shading, using reflectance models, considering interreflections, etc.
- incorporating prior knowledge, for example: statistics on local 3D structures (similar to local statistics on natural images), joint statistics of image appearance and local 3D structure (learnt e.g. by analyzing "stereo" images acquired by a laser range scanner and a color camera), using a catalogue of typical building materials and their reflectance properties, etc.

These will be among the main topics for our future work.

Bibliography

[1] K.B. Atkinson, editor. *Close Range Photogrammetry and Machine Vision*. Whittles Publishing, 1996.

[2] A. Bartoli and P. Sturm. Constrained structure and motion from N views of a piecewise planar scene. In *Proceedings of the First International Symposium on Virtual and Augmented Architecture, VAA'01, Dublin, Ireland*, pages 195–206, June 2001.

[3] A. Bartoli and P. Sturm. The 3D line motion matrix and alignment of line reconstructions. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, volume I, pages 287–292. IEEE Computer Society Press, December 2001.

[4] A. Bartoli and P. Sturm. La matrice de mouvement pour droites 3d, application à l'alignement de reconstructions de droites. In *13ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Angers, France*, pages 29–37, January 2002.

[5] A. Bartoli and P. Sturm. Constrained structure and motion from multiple uncalibrated views of a piecewise planar scene. *International Journal of Computer Vision*, 52(1):45–64, 2003.

[6] A. Bartoli and P. Sturm. Multiple-view structure and motion from line correspondences. In *Proceedings of the 9th International Conference on Computer Vision, Nice, France*, volume I, pages 207–212, October 2003.

[7] A. Bartoli and P. Sturm. Non-linear estimation of the fundamental matrix with minimal parameters. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(4):426–432, 2004.

[8] A. Bartoli and P. Sturm. The 3D line motion matrix and alignment of line reconstructions. *International Journal of Computer Vision*, 57(3):159–178, 2004.

[9] A. Bartoli and P. Sturm. Structure from motion using lines: Representation, triangulation and bundle adjustment. *Computer Vision and Image Understanding*, 100(3):416–441, December 2005.

[10] A. Bartoli, P. Sturm, and R. Horaud. A projective framework for structure and motion recovery from two views of a piecewise planar scene. Research Report 4070, INRIA, Grenoble, France, October 2000.

[11] A. Bartoli, P. Sturm, and R. Horaud. Structure and motion from two uncalibrated views using points on planes. In *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling, 3DIM'01, Québec City, Canada*, pages 83–90, May 2001.

[12] A. Bartoli, P. Sturm, and R. Horaud. Projective structure and motion from two views of a piecewise planar scene. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, volume 1, pages 593–598, July 2001.

[13] A. Bartoli, P. Sturm, and R. Horaud. Triangulation projective contrainte par multi-coplanarité. In *Actes des Journées Jeunes Chercheurs ORASIS, Cahors, France*, pages 47–56, June 2001.

[14] A. Bartoli, M. Coquerelle, and P. Sturm. A framework for pencil-of-points structure-from-motion. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, volume 3022 of *Lecture Notes in Computer Science*, pages 28–40, May 2004.

[15] P.J. Besl and N.D. McKay. A method for registration of 3D shapes. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992.

[16] L. Birgé and Y. Rozenholc. How many bins should be put in a regular histogram. Technical Report PMA-721, Laboratoire Probabilités et Modèles Aléatoires, Université Pierre et Marie Curie, Paris, France, April 2002.

[17] N. Birkbeck, D. Cobzaş, P. Sturm, and M. Jägersand. Variational shape and reflectance estimation under changing light and viewpoints. In H. Bischof and A. Leonardis, editors, *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria*, Lecture Notes in Computer Science, May 2006.

[18] T. Bonfort. *Reconstruction de surfaces spéculaires à partir d'images*. PhD thesis, Institut National Polytechnique de Grenoble, February 2006.

[19] T. Bonfort and P. Sturm. Voxel carving for specular surfaces. In *Proceedings of the 9th International Conference on Computer Vision, Nice, France*, volume 1, pages 591–596. IEEE Computer Society Press, October 2003.

[20] T. Bonfort and P. Sturm. Reconstruction de surfaces réflechissantes à partir d'images. In *Actes des Journées Jeunes Chercheurs ORASIS, Gérardmer, France*, pages 59–65, 2003.

[21] T. Bonfort, P. Sturm, and P. Gargallo. General specular surface triangulation. In *Proceedings of the Asian Conference on Computer Vision, Hyderabad, India*, volume II, pages 872–881, January 2006.

[22] E. Boyer and P. Sturm. Modélisation à partir d'images. In F. Sillion, editor, *Synthèse d'Images Géographiques*, Géomatique, pages 57–89. Hermès Science Publications, 2002.

[23] P. Brand. *Reconstruction tridimensionnelle d'une scène à partir d'une caméra en mouvement : de l'influence de la précision*. Thèse de doctorat, Université Claude Bernard, Lyon I, October 1995. URL `ftp://ftp.imag.fr/pub/Mediatheque.IMAG/theses/95-Brand.Pascal/notice-francais.html`.

[24] D.C. Brown. Close-range camera calibration. *Photogrammetric Engineering*, 37(8):855–866, 1971.

[25] G. Champleboux, S. Lavallée, P. Sautot, and P. Cinquin. Accurate calibration of cameras and range imaging sensors: the NPBS method. In *Proceedings of* IEEE *International Conference on Robotics and Automation, Nice, France*, pages 1552–1558, May 1992.

[26] C.-S. Chen and W.-Y. Chang. On pose recovery for generalized visual sensors. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(7):848–861, July 2004.

[27] O. Chum, T. Pajdla, and P. Sturm. On geometric error for homographies. Technical Report CTU-CMP-2003-13, Center for Machine Perception, Czech Technical University, Prague, 2003.

[28] O. Chum, T. Pajdla, and P. Sturm. The geometric error for homographies. *Computer Vision and Image Understanding*, 97(1):86–102, January 2005.

[29] D. Cobzas and P. Sturm. 3D SSD tracking with estimated 3D planes. In *Proceedings of the Second Canadian Conference on Computer and Robot Vision, Victoria, Canada*, May 2005.

[30] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry-and image-based approach. In SIGGRAPH '96*, New Orleans*, August 1996.

[31] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[32] O. Faugeras, L. Quan, and P. Sturm. Self-calibration of a 1d projective camera and its application to the self-calibration of a 2d projective camera. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1179–1185, October 2000.

[33] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, March 2001.

[34] S. Finsterwalder. Die geometrischen Grundlagen der Photogrammetrie. *Jahresbericht Deutscher Mathematik*, 6:1–44, 1899.

[35] R.W. Fleming, A. Torralba, and E.H. Adelson. Specular reflections and the perception of shape. *Journal of Vision*, 4:798–820, 2004.

[36] P. Gargallo and P. Sturm. Bayesian 3D modeling from images using multiple depth maps. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, USA*, volume 2, pages 885–891, June 2005.

[37] P. Gargallo and P. Sturm. Modélisation 3D à partir d'images en utilisant plusieurs cartes de profondeur. In *Actes des Journées Jeunes Chercheurs ORASIS, Fournol, France*, 2005.

[38] K.D. Gremban, C.E. Thorpe, and T. Kanade. Geometric camera calibration using systems of linear equations. In *Proceedings of* IEEE *International Conference on Robotics and Automation, Philadelphia, Pennsylvania, USA*, pages 562–567. IEEE Computer Society Press, 1988.

[39] M.D. Grossberg and S.K. Nayar. A general imaging model and a method for finding its parameters. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, volume 2, pages 108–115. IEEE Computer Society Press, July 2001.

[40] P. Gurdjos and P. Sturm. Methods and geometry for plane-based self-calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA*, volume 1, pages 491–496, June 2003.

[41] P. Gurdjos, P. Sturm, and Y. Wu. Euclidean structure from n>=2 parallel circles: Theory and algorithms. In H. Bischof and A. Leonardis, editors, *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria*, Lecture Notes in Computer Science, May 2006.

[42] P. Hammarstedt, P. Sturm, and A. Heyden. Closed-form solutions and degenerate cases for camera calibration with one-dimensional objects. In *Proceedings of the 10th International Conference on Computer Vision, Beijing, China*, October 2005.

[43] R.I. Hartley and R. Gupta. Linear pushbroom cameras. In J.-O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, volume 800 of *Lecture Notes in Computer Science*, pages 555–566. Springer-Verlag, May 1994.

[44] R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.

[45] R.I. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[46] A. Heyden and K. Åström. Minimal conditions on intrinsic parameters for euclidean reconstruction. In *Proceedings of the Third Asian Conference on Computer Vision, Hong Kong*, volume II, pages 169–176, January 1998.

[47] A. Jacquot, P. Sturm, and O. Ruch. Adaptive tracking of non-rigid objects based on color histograms and automatic parameter selection. In *Proceedings of the* IEEE *Workshop on Motion and Video Computing, Breckenridge, Colorado, USA*, pages 103–109, January 2005.

[48] A. Jacquot, P. Sturm, and O. Ruch. Suivi adaptatif d'objets non rigides basé sur des histogrammes de couleurs et une sélection automatique de paramètres. In *Actes des Journées Jeunes Chercheurs ORASIS, Fournol, France*, 2005.

[49] A. Jacquot, P. Sturm, and O. Ruch. 3D object tracking based on multi-model particle filtering. In *Proceedings of the Sixth IEEE International Workshop on Visual Surveillance, Graz, Austria*, May 2006.

[50] G. Jiang, L. Quan, and H.-T. Tsui. Circular motion geometry using minimal data. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(6):721–731, 2004.

[51] H. Jin, D. Cremers, A. Yezzi, and S. Soatto. Shedding light on stereoscopic segmentation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Washington, USA*, volume 1, pages 36–42, 2004.

[52] F. Jurie, E. Noirfalise, and P. Sturm. Calibrage itératif de caméras à partir de scènes planes. In *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Toulouse, France*, volume 1, pages 261–269, January 2004.

[53] K.N. Kutulakos and E. Steger. A theory of refractive and specular 3d shape by light-path triangulation. In *Proceedings of the 10th International Conference on Computer Vision, Beijing, China*, pages 1448–1455, October 2005.

[54] J.-M. Lavest, M. Viala, and M. Dhome. Do we really need an accurate calibration pattern to achieve a reliable camera calibration? In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, volume 1, pages 158–174, May 1998.

[55] S.J. Maybank and O.D. Faugeras. A theory of self calibration of a moving camera. *International Journal of Computer Vision*, 8(2):123–151, 1992.

[56] S.J. Maybank and P. Sturm. MDL, collineations and the fundamental matrix. In T. Pridmore and D. Elliman, editors, *Proceedings of the 10th British Machine Vision Conference, Nottingham, England*, pages 53–62. British Machine Vision Association, September 1999.

[57] S.J. Maybank and P.F. Sturm. Minimum description length and the inference of scene structure from images. In *Proceedings of the IEE Colloquium on Applications of Statistics to Pattern Recognition, Birmingham, UK*, April 1999.

[58] D. Nistér. A minimal solution to the generalized 3-point pose problem. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Washington, USA*, volume 1, pages 560–567, 2004.

[59] D. Nistér, H. Stewénius, and E. Grossmann. Non-parametric self-calibration. In *Proceedings of the 10th International Conference on Computer Vision, Beijing, China*, pages 120–127, October 2005.

[60] E. Noirfalise, F. Jurie, and P. Sturm. Autocalibration itérative de caméras à partir de structures planes. In *Actes des Journées Jeunes Chercheurs ORASIS, Gérardmer, France*, pages 157–164, 2003.

[61] T. Pajdla. Geometry of two-slit camera. Technical Report CTU-CMP-2002-02, Center for Machine Perception, Czech Technical University, Prague, March 2002.

[62] M.A. Penna. Camera calibration : a quick and easy way to determine the scale factor. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 13(12):1240–1245, December 1991.

[63] M. Personnaz and P. Sturm. Calibration of a stereo-vision system by the non-linear optimization of the motion of a calibration object. Technical Report 0269, INRIA, Grenoble, France, September 2002.

[64] M. Personnaz and P. Sturm. Specifications of the software videomatch 1.1. Technical Report 0260, INRIA, Grenoble, France, March 2002.

[65] M. Personnaz, P. Sturm, and F. Devernay. Design of a library for dense matching. Technical Report 0278, INRIA, Grenoble, France, April 2003.

[66] R. Pless. Using many cameras as one. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Madison, Wisconsin, USA*, volume II, pages 587–593, June 2003.

[67] M. Pollefeys. *Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences*. PhD thesis, Katholieke Universiteit Leuven, Belgium, Faculteit Toegepaste Wetenschappen, Arenbergkasteel, B-3001 Heverlee, Belgium, May 1999.

[68] M. Pollefeys, R. Koch, and L. Van Gool. Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters. In *Proceedings of the 6th International Conference on Computer Vision, Bombay, India*, pages 90–95, January 1998.

[69] T. Pribanić, P. Sturm, and M. Cifrek. Camera parameter initialization for 3d kinematic systems. In *Proceedings of the International Symposium on Image and Signal Processing and Analysis, Zagreb, Croatia*, September 2005.

[70] T. Pribanić, P. Sturm, and M. Cifrek. Practical way to initialize camera parameters using absolute conic. In *Proceedings of the Fifth IASTED International Conference on Visualization, Imaging, and Image Processing, Benidorm, Spain*, September 2005.

[71] S. Ramalingam, S. Lodha, and P. Sturm. A generic structure-from-motion algorithm for cross-camera scenarios. In *Proceedings of the 5th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Prague, Czech Republic*, pages 175–186, May 2004.

[72] S. Ramalingam, P. Sturm, and S.K. Lodha. Theory and experiments towards complete generic calibration. Research Report 5562, INRIA, Grenoble, France, April 2005.

[73] S. Ramalingam, P. Sturm, and S.K. Lodha. Towards complete generic camera calibration. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, USA*, volume 1, pages 1093–1098, June 2005.

[74] S. Ramalingam, P. Sturm, and S.K. Lodha. Towards generic self-calibration of central cameras. In *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, pages 20–27, October 2005.

[75] S. Ramalingam, P. Sturm, and S.K. Lodha. Generic calibration of axial cameras. Research Report 5827, INRIA, Grenoble, France, December 2005.

[76] S. Ramalingam, P. Sturm, and S. Lodha. Theory and calibration algorithms for axial cameras. In *Proceedings of the Asian Conference on Computer Vision, Hyderabad, India*, volume I, pages 704–713, January 2006.

[77] S. Rémy, M. Dhome, N. Daucher, and J.T. Lapresté. Estimation de la distorsion radiale d'un système optique. In *Actes du 9ème Congrès AFCET de Reconnaissance des Formes et Intelligence Artificielle, Paris, France*, pages 99–108, January 1994.

[78] T. Rodriguez, P. Sturm, M. Wilczkowiak, A. Bartoli, M. Personnaz, N. Guilbert, F. Kahl, M. Johansson, A. Heyden, J.M. Menendez, J.I. Ronda, and F. Jaureguizar. VISIRE – photorealistic 3d reconstruction from video sequences. In *Proceedings of the* IEEE *International Conference on Image Processing, Barcelona, Spain*, volume III, pages 705–708, September 2003.

[79] T. Rodriguez, P. Sturm, P. Gargallo, N. Guilbert, A. Heyden, J.M. Menendez, and J.I. Ronda. Photorealistic 3d reconstruction from handheld cameras. *Machine Vision and Applications*, 16(4):246–257, 2005.

[80] S.M. Seitz and C.R. Dyer. Photorealistic scene reconstruction by voxel coloring. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1067–1073. IEEE Computer Society Press, June 1997.

[81] A. Shashua and L. Wolf. Homography tensors: On algebraic entities that represent three views of static or moving planar points. In D. Vernon, editor, *Proceedings of the 6th European Conference on Computer Vision, Dublin, Ireland*, volume 1842 of *Lecture Notes in Computer Science*, pages 507–521. Springer-Verlag, June 2000.

[82] C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.

[83] C. Strecha, R. Fransens, and L. Van Gool. Wide-baseline stereo from multiple views: a probabilistic account. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Washington, USA*, volume 1, pages 552–559, June 2004.

[84] P. Sturm. Algorithms for plane-based pose estimation. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina, USA*, pages 1010–1017, June 2000.

[85] P. Sturm. A method for 3D reconstruction of piecewise planar objects from single panoramic images. In *Proceedings of the* IEEE *Workshop on Omnidirectional Vision, Hilton Head Island, South Carolina*, pages 119–126. IEEE, June 2000.

[86] P. Sturm. A case against Kruppa's equations for camera self-calibration. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1199–1204, October 2000.

[87] P. Sturm. On focal length calibration from two views. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii, USA*, volume II, pages 145–150. IEEE Computer Society Press, December 2001.

[88] P. Sturm. Structure and motion for dynamic scenes – the case of points moving in planes. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 2351 of *Lecture Notes in Computer Science*, pages 867–882. Springer-Verlag, May 2002.

[89] P. Sturm. Mixing catadioptric and perspective cameras. In *Proceedings of the Workshop on Omnidirectional Vision, Copenhagen, Denmark*, pages 37–44, June 2002.

[90] P. Sturm. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. *Image and Vision Computing*, 20(5-6):415–426, 2002.

[91] P. Sturm. Multi-view geometry for general camera models. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, San Diego, USA*, volume 1, pages 206–212, June 2005.

[92] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 1100–1105, June 1997.

[93] P. Sturm. *Vision 3D non calibrée : contributions à la reconstruction projective et étude des mouvements critiques pour l'auto-calibrage*. Thèse de doctorat, Institut National Polytechnique de Grenoble, December 1997. URL ftp://ftp.imag.fr/pub/Mediatheque.IMAG/theses/97-Sturm.Peter/notice-francais.html.

[94] P. Sturm. Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length. In T. Pridmore and D. Elliman, editors, *Proceedings of the 10th British Machine Vision Conference, Nottingham, England*, pages 63–72. British Machine Vision Association, September 1999.

[95] P. Sturm and T. Bonfort. How to compute the pose of an object without a direct view? In *Proceedings of the Asian Conference on Computer Vision, Hyderabad, India*, volume II, pages 21–31, January 2006.

[96] P. Sturm and P. Gurdjos. Méthodes et fondements géométriques pour l'autocalibrage plan. In *14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Toulouse, France*, volume 1, pages 351–360, January 2004.

[97] P. Sturm and 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.

[98] P. Sturm and S.J. Maybank. A method for interactive 3D reconstruction of piecewise planar objects from single images. In T. Pridmore and D. Elliman, editors, *Proceedings of the 10th British Machine Vision Conference, Nottingham, England*, pages 265–274. British Machine Vision Association, September 1999.

[99] P. Sturm and L. Quan. Camera calibration and relative pose estimation from gravity. In A. Sanfeliu, J.J. Villanueva, M. Vanrell, R. Alquézar, J.-O. Eklundh, and Y. Aloimonos, editors, *Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain*, volume 1, pages 72–75, September 2000.

[100] P. Sturm and L. Quan. Affine stereo calibration. Technical Report LIFIA-29, LIFIA–IMAG, Grenoble, France, June 1995.

[101] P. Sturm and S. Ramalingam. A generic calibration concept: Theory and algorithms. Research Report 5058, INRIA, Grenoble, France, December 2003.

[102] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In T. Pajdla and J. Matas, editors, *Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, volume 3022 of *Lecture Notes in Computer Science*, pages 1–13. Springer-Verlag, May 2004.

[103] P. Sturm and S. Ramalingam. Géométrie d'images multiples pour des modèles de caméra généraux. *Traitement du Signal*, 22(5), October 2005.

[104] P. Sturm, Z.L. Cheng, P.C.Y. Chen, and A.N. Poo. Focal length calibration from two views: Method and analysis of singular cases. *Computer Vision and Image Understanding*, 99(1):58–95, July 2005.

[105] P. Sturm, S. Ramalingam, and S.K. Lodha. On calibration, structure-from-motion and multi-view geometry for general camera models. In R. Reulke and U. Knauer, editors, *Proceedings of the 2nd ISPRS Panoramic Photogrammetry Workshop, Berlin, Germany*. International Society for Photogrammetry and Remote Sensing, February 2005. Published in the Int. Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Vol. XXXVI-5/W8.

[106] P. Sturm, S. Ramalingam, and S.K. Lodha. On calibration, structure from motion and multi-view geometry for generic camera models. In K. Daniilidis, R. Klette, and A. Leonardis, editors, *Imaging Beyound the Pinhole Camera*. Kluwer Academic Publishers, 2006.

[107] T. Svoboda and P. Sturm. What can be done with a badly calibrated camera in ego-motion estimation ? Technical Report CTU-CMP-1996-01, Center for Machine Perception, Czech Technical University, Prague, November 1996.

[108] T. Svoboda and P. Sturm. A badly calibrated camera in ego-motion estimation, propagation of uncertainty. In G. Sommer, K. Daniilidis, and J. Pauli, editors, *Proceedings of the 7th International Conference on Computer Analysis of Images and Patterns, Kiel, Germany*, volume 1296 of *Lecture Notes in Computer Science*, pages 183–190. Springer-Verlag, September 1997.

[109] J.-P. Tardif and P. Sturm. Calibration of cameras with radially symmetric distortion. In *Proceedings of the 6th Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras, Beijing, China*, pages 44–51, October 2005.

[110] J.-P. Tardif, P. Sturm, and S. Roy. Self-calibration of a general radially symmetric distortion model. In H. Bischof and A. Leonardis, editors, *Proceedings of the 9th European Conference on Computer Vision, Graz, Austria*, Lecture Notes in Computer Science, May 2006.

[111] S. Thirthala and M. Pollefeys. Multi-view geometry of 1d radial cameras and its application to omnidirectional camera calibration. In *Proceedings of the 10th International Conference on Computer Vision, Beijing, China*, pages 1539–1546, October 2005.

[112] B. Triggs. Autocalibration and the absolute quadric. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Puerto Rico, USA*, pages 609–614. IEEE Computer Society Press, June 1997.

[113] B. Triggs. Autocalibration from planar scenes. In *Proceedings of the 5th European Conference on Computer Vision, Freiburg, Germany*, 1998.

[114] R.Y. Tsai. An efficient and accurate camera calibration technique for 3D machine vision. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Miami Beach, Florida, USA*, pages 364–374, 1986.

[115] M. Urbanek, R. Horaud, and P. Sturm. Combining off- and on-line calibration of a digital camera. In *Proceedings of the Third International Conference on 3D Digital Imaging and Modeling, Québec City, Canada*, pages 99–106, May 2001.

[116] M. Urbanek, R. Horaud, and P. Sturm. Calibration of digital amateur cameras. Technical Report 4214, INRIA, July 2001.

[117] M. Wilczkowiak, E. Boyer, and P. Sturm. Camera calibration and 3D reconstruction from single images using parallelepipeds. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, volume 1, pages 142–148. IEEE Computer Society Press, July 2001.

[118] M. Wilczkowiak, E. Boyer, and P. Sturm. Calibrage de caméra et reconstruction 3D à l'aide de parallélépipèdes. In *Actes des Journées Jeunes Chercheurs ORASIS, Cahors, France*, pages 117–126, June 2001.

[119] M. Wilczkowiak, E. Boyer, and P. Sturm. 3D modelling using geometric constraints: A parallelepiped based approach. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Proceedings of the 7th European Conference on Computer Vision, Copenhagen, Denmark*, volume 2353 of *Lecture Notes in Computer Science*, pages 221–236. Springer-Verlag, May 2002.

[120] M. Wilczkowiak, P. Sturm, and E. Boyer. Calibrage et reconstruction à l'aide de parallélépipèdes et de parallélogrammes. In *13ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, Angers, France*, pages 849–857, January 2002.

[121] M. Wilczkowiak, P. Sturm, and E. Boyer. The analysis of ambiguous solutions in linear systems and its application to computer vision. In *Proceedings of the 14th British Machine Vision Conference, Norwich, England*, volume II, pages 53–62, September 2003.

[122] M. Wilczkowiak, P. Sturm, and E. Boyer. Une approche pour l'analyse et exploitation des contraintes géométriques en modélisation 3D. In *Actes des Journées Jeunes Chercheurs ORASIS, Gérardmer, France*, pages 41–50, 2003.

[123] M. Wilczkowiak, P. Sturm, and E. Boyer. Using geometric constraints through parallelepipeds for calibration. Research Report 5055, INRIA, Grenoble, France, December 2003.

[124] M. Wilczkowiak, G. Trombettoni, C. Jermann, P. Sturm, and E. Boyer. Scene modeling based on constraint system decomposition techniques. In *Proceedings of the 9th International Conference on Computer Vision, Nice, France*, volume 2, pages 1004–1010. IEEE Computer Society Press, October 2003.

[125] M. Wilczkowiak, P. Sturm, and E. Boyer. Using geometric constraints for camera calibration and positioning and 3d scene modelling. In *Proceedings of the CIPA International Workshop on Vision Techniques Applied to the Rehabilitation of City Centres, Lisbon, Portugal*, October 2004.

[126] M. Wilczkowiak, P. Sturm, and E. Boyer. Using geometric constraints through parallelepipeds for calibration and 3D modelling. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 27 (2):194–207, February 2005.

[127] L. Wolf, A. Shashua, and Y. Wexler. Join tensors: On 3D-to-3D alignment of dynamic sets. In A. Sanfeliu, J.J. Villanueva, M. Vanrell, R. Alquézar, J.-O. Eklundh, and Y. Aloimonos, editors, *Proceedings of the 15th International Conference on Pattern Recognition, Barcelona, Spain*, volume 1, pages 388–391, September 2000.

[128] C. Wu, Z. Chen, and P. Sturm. 3D reconstruction and metrology from uncalibrated image sequences. In *Proceedings of the PRA Workshop on Information Science and Technology, Beijing, China*, 2002.

[129] C. Zeller and O. Faugeras. Camera self-calibration from video sequences: the kruppa equations revisited. Rapport de recherche 2793, INRIA, February 1996.

[130] Z. Zhang. A flexible new technique for camera calibration. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334, 2000.

[131] Z. Zhang. Camera calibration with one-dimensional objects. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 26(7):892–899, 2004.

[132] A. Zomet, D. Feldman, S. Peleg, and D. Weinshall. Mosaicing new views: The crossed-slit projection. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 25(6):741–754, June 2003.