Esprit Long Term Research Project 21914

CUMULI

Computational Understanding of Multiple Images

Final Report

Status: Public

21 September 2000

The CUMULI Consortium

MOVI, INRIA Rhône-Alpes, Grenoble, France
ROBOTVIS, INRIA, Sophia-Antipolis, France
Mathematical Imaging Group, Lund University, Sweden
Imetric S.A., Courgenay, Switzerland
Image Systems A.B., Linköping, Sweden
Fraunhofer IGD, Darmstadt, Germany

Executive Summary

The main aim of Esprit Reactive Long Term Research project 21914 CUMULI— *Computational Understanding of Multiple Images* — was to develop advanced methods for accurate 3D industrial measurement from images, with particular emphasis on: (*i*) understanding and exploiting the underlying multi-image geometry of the problem, for both discrete images and image streams; (*ii*) increasing system flexibility and automation by developing alternative initialization techniques; (*iii*) exploiting higher-level geometric reasoning to improve the 3D reconstruction process. CUMULI built on the very significant advances in the geometry of multi-image vision made in the computer vision community over the past decade, and applied them to real world visual measurement problems. It was also an opportunity for some very fruitful collaboration and cross-fertilization between the computer vision, industrial photogrammetry, geometric reasoning, and 3D modelling communities.

Academically, CUMULI's main achievements are the following:

- A very substantial increase in our understanding of the process known as *auto*- or *self-calibration*, which recovers (*i*) full camera calibrations and poses and (*ii*) Euclidean 3D scene structure (up to a similarity transformation), from initial uncalibrated projective images, using only some limited 'qualitative' information about the camera or scene, such as the time constancy of some camera parameters, or knowledge of some 3D parallelism or orthogonality relationships. Within CUMULI, very much stabler basic algorithms have been developed and applied to a wide range of different autocalibration constraint sets. The singularities of these methods have also been thoroughly studied and categorized, both in the case of discrete images and for video sequences.
- A consolidation of our understanding of projective methods, especially as these relate to image streams (*e.g.* the two approaches to differential matching constraints, and the work on incremental projective reconstruction), and to calibration information (*e.g.* the work on autocalibration and special motions).
- A wide range of routines have been developed for camera initialization (pose and often partial calibration) from various combinations of features (points, lines, cylinders, conics...). These routines have been integrated into the systems of all three industrial partners, in many cases greatly increasing the overall system flexibility and automation.
- A fruitful confrontation took place between formal algebraic methods for geometric reasoning on the one hand, and the practical realities of computer vision on the other (most notably, the need to deal with many, and uncertain, features). Although only limited progress was made on solving these difficult problems during CUMULI, the exchange was fruitful as it has given each field a new perspective on its own problems and on the needs of the other, which is already leading to further developments.

On the industrial side, CUMULI produced the following main achievements:

- All three of the industrial partners used CUMULI camera initialization technology to improve the critical initialization stage — and hence the overall reliability and automation — of their products.

- In the case of IMETRIC, these methods were a small but essential component in making the fully automated TI^2 vision controlled machining system feasible. This technology has won Swiss awards, and is the only automated vision controlled 3D machining process ever certified by Boeing.
- Image Systems have included both these initialization methods, and improved tracking methods developed under CUMULI, in their TrackEye tracking software, and they are moving towards a more complete 3D tracking product for use in aerospace and car crash testing applications.
- The capabilities of Fraunhofer IGD's Augmented Reality system, which is based in part on CUMULI methods, contributed to the formation of ARVIKA, the world's largest industrial Augmented Reality consortium. ARVIKA includes Fraunhofer IGD and over 20 leading German manufacturers. The Fraunhofer IGD AR system was also used to add AR capabilities to the commercial Fraunhofer IGD / VRcom "Virtual Design 2" Virtual Reality system, at the request of the Volkswagen company which uses this system for product design.
- Methods for 3D curve reconstruction developed in CUMULI were modified to serve as a basis for a system for 2D handwriting recognition. The technology is under consideration for Swedish and international patents, and is being commercialized by the start-up Decuma AB at Ideon, the Lund Science and Technology Park.

The CUMULI Consortium

CUMULI ran from September 1996 to February 2000 and represents about 17 person-years of work. The consortium contained six partners, three academic and three industrial:

- MOVI, INRIA Grenoble, France
- ROBOTVIS, INRIA Sophia-Antipolis, France
- Mathematical Imaging Group, Dept. of Mathematics, Lund University, Lund, Sweden
- IMETRIC SA, Porrentruy, Switzerland
- Image Systems AB, Linköping, Sweden
- Fraunhofer IGD, Darmstadt, Germany.

Grenoble, Sophia and Lund are all academic research groups with expertise in computer vision and geometric reasoning. IMETRIC and Image Systems are applications-oriented Small and Medium Enterprises with expertise in close range industrial photogrammetry (IMETRIC) and motion measurement from high-speed image sequences (Image Systems). Fraunhofer IGD is a developer and integrator of 3D modelling systems for augmented and virtual reality, and also performs some basic and applied research.

Contents

Chapter 1

Introduction

This document is the final report of the Esprit Reactive Long Term Research project 21914 CUMULI— *Computational Understanding of Multiple Images*. CUMULI's main aim was to develop advanced methods for accurate 3D industrial measurement from multiple images, with particular emphasis on: (*i*) understanding and exploiting the underlying multi-image geometry of the problem, for both discrete images and image streams; (*ii*) increasing system flexibility and automation by developing alternative initialization techniques; (*iii*) exploiting higher-level geometric reasoning to improve the 3D reconstruction process. CUMULI built on the very significant advances in the geometry of multi-image vision made in the computer vision community over the past decade, and applied them to real world visual measurement problems. It was also an opportunity for some very fruitful collaboration and cross-fertilization between the computer vision, industrial photogrammetry, geometric reasoning, and 3D modelling communities.

CUMULI was built around an inter-disciplinary consortium uniting six partners: three are academic research institutes with expertise in computer vision and geometric reasoning (INRIA Grenoble, INRIA Sophia-Antipolis, Lund); two are applications-oriented Small and Medium Enterprises with expertise in close range industrial photogrammetry (IMETRIC) and motion measurement from high-speed image sequences (Image Systems); and one is a developer and integrator of 3D modelling systems for augmented and virtual reality (Fraunhofer IGD).

Box 1 gives some key statistics relating to CUMULI, and box 2 lists some of the successes and awards that CUMULI technology has won. Box 3 gives some details and contact points for the CUMULI consortium.

1.1 Project Structure

CUMULI was divided into three main workpackages (WP's), which ran in parallel throughout the project:

- WP 1. Multi-camera geometry, discrete images;
- WP 2. Image streams and 3D motion;
- WP 3. Algebraic symbolic reasoning.

The themes of these workpackages were partly inspired by the applications of CUMULI's three industrial partners Imetric, Image Systems, and Fraunhofer IGD [1]. Each workpackage was led by one of CUMULI's three academic partners Lund, Grenoble, and Sophia-Antipolis, and

---

[1] Fraunhofer IGD perform both longer term research and technology development. Within CUMULI, they acted mainly as a technology integrator and were therefore classed as an industrial partner.

Box 1: CUMULI vital statistics

- **Consortium:** 3 industrial/development partners, 3 academic partners.
- **Expertise:** industrial photogrammetry, computer vision, geometric reasoning, augmented and virtual reality.
- **Period:** 1 September 1996 — 29 February 2000
  - 42 months, including a 6 month extension.
- **Total effort:** approximately 17 person-years
  - 11 at academic sites, 6 at industrial ones.
- **Total budget:** 2.1 MEu, including:
  - 1.1 MEu of European Union support;
  - 0.1 MEu of Swiss government support;
  - 0.9 MEu funded by the partners themselves.
- **Scientific production:** approximately 95 publications, mostly in refereed international conferences and journals.

finished with a technology demonstrator integrated by the work package's industrial partner. The partners contributed to the workpackages as follows:

Workpackage Name	Scientific Leader	Industrial Co-leader	Person-months					
			Gre	Soph	Lund	I.S.	Imet.	IGD
WP 1. Discrete Images	Lund	Imetric	17	12	17	2	17	
WP 2. Image streams	Grenoble	Im. Syst.	15	11	14	20		
WP 3. Reasoning	Sophia	IGD	3	24	7			22

Workpackages 1 and 2 were closely related, and both concentrated on core topics in geometric vision and photogrammetry: projective and uncalibrated structure; constraints, calibration and Euclidean structure; and measurement of non-point-like geometric features such as lines, conics and other curves. WP 1 focused on discrete images and the algebraic aspects of these problems. It was directly linked to the IMETRIC demonstrator whose main goal was more flexible methods of recovering multi-camera orientation. However the camera initialization methods developed in WP 1 proved very useful for all three industrial demonstrators and have been used by all three industrial partners. WP 2 concentrated on specializing the general results of WP 1 to continuous image sequences, tracking, and motion estimation. It was linked to the Image Systems car crash-testing demonstrator.

WP 3 was more speculative. It started from the viewpoint that geometric models should be more than just coordinates — they contain rich networks of incidence relations, constraints, *etc.*, which need to be exploited in advanced applications. To measure or reconstruct such a model from uncertain data, we should therefore develop methods of recognizing and mobilizing these constraints to ensure consistency, reduce uncertainty, improve registration, generate new matching hypotheses . . . . Hence, we feel that geometric reasoning under uncertainty will play a central role in future large-scale vision systems. WP 3 was an initial attempt to adapt existing geometric theorem proving technology to this sort of application. It was linked to the Fraunhofer IGD demonstrator, which focused on model-image calibration, registration and occlusion problems for augmented reality applications.

Box 2: CUMULI successes and awards

- **Industrial awards and honours:**
  - IMETRIC's TI^2 product was the first 3D Image Metrology system ever certified by Boeing for CNC machine guidance, and also won a Swiss *Innovation Award for Technology Location Switzerland* for this achievement.
- **New products using CUMULI results:**
  - CUMULI extensions to IMETRIC's core IMSlib software were used in its IMS Image Metrology software, its TI^2 vision based machine control system, its new Icam range of metrology cameras, and a surface scanning system jointly developed by Daimler-Chrysler and IMETRIC.
  - The next release of Image Systems' TrackEye™ visual tracking software will contain CUMULI camera initialization and 3D reconstruction routines.
  - The next release of Fraunhofer IGD's Virtual Reality system "Virtual Design 2 (VD2)" (commercialized by its spin-off company VRCom — `www.vrcom.de`) will contain a new module for camera calibration and image augmentation based on CUMULI results, as requested by the Volkswagen company.
  - Methods for 3D curve reconstruction developed in CUMULI were modified to serve as a basis for a system for 2D handwriting recognition. The technology is under consideration for Swedish and international patents, and is being commercialized by the start-up Decuma AB at Ideon, the Lund Science and Technology Park.
- **Academic awards and honours:**
  - The CUMULI-sponsored paper [66] jointly won the best paper prize at the 1998 European Conference on Computer Vision.
- **New projects and consortia using CUMULI results:**
  - ARVIKA — the world's largest industrial Augmented Reality consortium.
  - VISIRE, VIBES, EVENTS, CarSense — new E.U. research projects on vision related themes.

Box 3: The CUMULI consortium at a glance

- Web page: *http://www.inrialpes.fr/CUMULI*

1. Project leader: MOVI, INRIA Grenoble, France

- MOVI (MOdelling for VIsion) is led by Dr Radu HORAUD (formerly by Prof. Roger MOHR).
- Expertise: computer vision, especially vision geometry.
- Other major projects: FIRST, SECOND, VIVA, VIGOR, VISIRE, EVENTS . . .
- Main contributors: Roger MOHR (project leader), Bill TRIGGS (deputy leader), Long QUAN.
- Info: *http://www.inrialpes.fr/movi*, $\{$*Bill.Triggs,Long.Quan*$\}$*@inrialpes.fr*
- Prof. Dongming WANG (then of the LEIBNIZ laboratory in Grenoble, now at LIP6, Université Paris VI) also collaborated on CUMULI, on geometric reasoning topics.

Continued on next page . . .

Box 3 continued . . .

2. ROBOTVIS, INRIA Sophia-Antipolis, France

- The ROBOTVIS group is led by Prof. Olivier FAUGERAS.
- Expertise: computer vision, especially vision geometry; geometric reasoning.
- Other major projects: VIVA, Realise, IMPROOFS, EPSIS, CarSense . . .
- Main contributors: Theo PAPADOPOULO, Thierry VIÉVILLE, Didier BONDYFALAT, Bernard MOURRAIN, Olivier FAUGERAS
- Bernard MOURRAIN and Didier BONDYFALAT are members of the SAGA (Systèmes Algébriques, Géométrie et Applications) team at INRIA Sophia-Antipolis, who specialize in applications of algebra and geometry.
- Info: *www-sop.inria.fr/robotvis*, {*Theodore.Papadopoulo,Thierry.Vieville*}*@sophia.inria.fr* *www-sop.inria.fr/saga*, *Bernard.Mourrain@sophia.inria.fr*

3. Mathematical Imaging Group, Lund University of Technology, Sweden

- The Mathematical Imaging Group is led by Prof. Gunnar SPARR.
- Expertise: computer vision, especially vision geometry; medical imaging.
- Other major projects: VIVA, VISIRE, VISIT . . .
- Main contributors: Gunnar SPARR, Anders HEYDEN, Kalle ÅSTRÖM, Fredrik KAHL.
- Info: *http://www.maths.lth.se/matematiklth/vision*, {*gunnar,andersp,kalle*}*@maths.lth.se*

4. Image Systems AB, Linköping, Sweden

- Image Systems AB (formerly Innovativ Vision Image Systems AB) specializes in systems for motion analysis and high resolution film digitization.
- Main products: TrackEye and GoldenEye digital film scanners and image processing software.
- Main clients: automotive and aircraft/military testing, film/media industry.
- Main contributors: Magnus OLSSON, Anders KÄLLDAHL.
- Info: *http://www.trackeye.com*, *info@imagesystems.se*

5. Imetric SA, Courgenay, Switzerland

- IMETRIC produces systems for high precision 3D metrology based on digital photogrammetry.
- Main products: ICam 6 and 28 metrology cameras, TI^2 integrated manufacturing system.
- Main clients: aerospace, automotive, shipbuilding.
- Main contributors: Horst BEYER, Graeme VAN DER VLUGT.
- Info: *http://www.imetric.com*, *info@imetric.com*.

6. Fraunhofer IGD A4, Darmstadt, Germany

- The Fraunhofer IGD A4 group is led by Dr Stefan MÜLLER.
- Expertise: augmented and virtual reality, scientific visualization.
- Other major projects: Realise, ARVIKA . . .
- Main contributors: Didier STRICKER.
- Info: *http://www.igd.fhg.de/igd-a4*, {*Didier.Stricker,Stefan.Mueller*}*@igd.fhg.de*

Chapter 2

WP 1: Multi-Camera Geometry, Discrete Images

Workpackage 1 concerns the theoretical and practical aspects of deriving very accurate estimates of 3D geometrical structure from a discrete set of images of a static scene. The practical work focuses on the needs of IMETRIC, who also supplied the test data. The theoretical and algorithmic results are used not only in this work package, but also form a basis for workpackages 2 and 3. Hence, Image Systems and Fraunhofer IGD have also followed the work in this package and absorbed some of the results.

2.1 State of the art before CUMULI

Before CUMULI, the geometry and algebra of multiple views were to a large extent understood, but the research community had been mainly focusing on the case of point features. Some well known algorithms for point features developed before CUMULI are, *e.g.*, the 8 point fundamental matrix algorithm, and the affine factorization method of Tomasi & Kanade. In CUMULI, we have extended this knowledge to other geometric features such as lines, curves and surfaces.

At the start of CUMULI, self-calibration was a known but not very well-investigated research topic. The existing algorithms often worked poorly, and their failure modes were to a large extent unknown. The exact conditions for which self-calibration is possible were not known, either on a theoretical or a practical level.

2.2 State of the art after CUMULI

The state of the art within multiple view vision has advanced considerably within the last few years, in particular within the following areas, where CUMULI has played a significant role.

Within CUMULI there are several examples where solutions have been given to previously unsolved structure and motion problems. For example structure and motion problems with incomplete data, affine views of lines, combinations of points, lines and conics, curves and surfaces.

In general there is now a better understanding of multiple view geometry through new theoretical results on multiple view geometry and matching tensors.

Several novel factorization algorithms have been developed for cases of uncalibrated camera models, for curves and for combinations of points, lines and conics.

Within CUMULI both theoretical results and practical algorithms have been developed for determining intrinsic camera calibration parameters (*e.g.* the focal length or camera constant) using very weak assumptions. For example, it was shown that the knowledge that even a single intrinsic parameter is constant during the motion suffices for recovering all of the intrinsic parameters, and thence the entire 3D scene structure up to a 3D similarity. Another success is a method for finding these parameters even in the case where the scene is planar.

However, all of these methods fail under certain common conditions, in particular if the camera motion is restricted to so called critical motions. During the project these critical motions have been classified for several situations.

2.3 Open problems

Although the state of the art has advanced during the project there still are several unsolved specific structure and motion problems, *e.g.* the multiplicity of solutions and algorithms for finding the solutions with 9 lines in 3 uncalibrated views.

Robust methods of feature tracking and structure and motion estimation for image pairs have been studied extensively, and to some degree the same problem for continuous image streams. However wide-baseline feature matching is still an open problem.

2.4 Task 1.1: Projective Multi-camera Geometry

This worktask focused on the projective part of the general vision geometry and 3D reconstruction problems. This includes all results that are valid for general uncalibrated perspective cameras.

In this task, Lund contributed a number of theoretical results on multi-camera geometry. One of the key results is an analysis of the differences in using only the bilinear constraints or the trilinear constraints [36]. Another key result is [29], in which all algebraic relations between different tensor components are revealed in a common framework. The connections between algebraic, subspace and shape methods are better understood, and the different approaches have been unified in [34, 10]. This has also resulted in efficient and robust structure and motion algorithms [57, 28, 58]. These shape based factorization methods give estimates of structure and motion that are independent of coordinate systems, ordering of points, and images. In the joint paper [20] (deliverable 1.1/24), different tools for multi-view geometry, like affine shape, multilinear constraints and P-matrices have been brought together.

Grenoble and Lund collaborated on a study of minimal cases for projective reconstruction from incomplete point correspondences [54]. When all points are visible in all images the minimal cases are well-known, but in practice correspondences in some of the images are often missing. The study introduces a framework for the missing data problem, derives all the minimal cases for 3 and 4 images, and gives practical algorithms for them. The methods are a useful foundation for RANSAC-style approaches to camera initialization.

Grenoble studied optimization technique for accurate matching tensor estimation [67]. Matching tensors are an implicit, close-to-the-image representation of 2–4 camera geometry, extremely useful for feature correspondence and the intermediate stages of 3D reconstruction. However estimating them accurately is quite complicated as there are complex nonlinear constraints, and many possible constraint and feature types, error models and parametrizations, special or degenerate cases, and numerical initialization and optimization methods. Our main idea is to decouple the problem, so that different *(i)* feature types and error models; *(ii)* tensor types and parametrizations; *(iii)* initialization and optimization methods, can be compared quantitatively, and combined to handle new problems.

Grenoble also specialized its multi-image matching tensor formalism to the case of plane + parallax (images homographically warped to align a common reference plane). This greatly simplifies the geometry, and allows an efficient and stable multi-image projective reconstruction method based on a rank-1 factorization of a residual parallax matrix into point depths and camera center vectors (rather than the conventional rank 4 factorization into 3D points and projection matrices) [70].

Sophia-Antipolis developed a formalism to allow the detection of special feature configurations under either projective, affine or Euclidean geometry, with a simple uniform representation of points and lines. This front-end processing ensures that a robust set of data is input to subsequent motion algorithms, and hence improves their performance.

Sophia-Antipolis also studied the internal consistency constraints on the coefficients of the trifocal tensor, which encodes the projective geometry of three cameras. This has led to a set of algebraic equations that have been used to give a new one-to-one minimal parameterization of the tensor. The parameterization was used in a new estimation method. A geometric description of these constraints was given in [25, 24, 51].

2.5 Task 1.2: Euclidean and Affine Multi-camera Geometry

This task was dedicated to applying various types of additional, calibration related, constraints to the projective results discussed under WP 1.1 above, in order to recover Euclidean scene structure (*i.e.* up to a 3D similarity transformation) and full camera calibrations. In terms of abstract projective geometry, Euclidean scene reconstruction can be regarded as a special case of projective reconstruction in which a special 3D conic called the absolute conic is singled out. This object essentially encodes 3D angle information. Autocalibration methods recover the conic or equivalent information — and hence camera calibrations and Euclidean scene structure — by applying qualitative scene or camera constraints, such as the fact that certain camera parameters are the same (but perhaps unknown) across several images.

Within CUMULI, such autocalibration methods were developed greatly. An object called the absolute dual quadric — the algebraic dual of the absolute conic — was introduced by Grenoble [65] and related to a parametrization introduced independently by Lund [35]. These new parametrizations allowed a much simpler formulation of the autocalibration problem. The initial work was for the case where all of the internal parameters of the camera remain the same (but initially unknown) during an image sequence, but the theory was subsequently developed to allow autocalibration under very weak assumptions on the camera. For example, with sufficiently many general images, the knowledge that the camera skew is zero (which is always true to a good approximation for real cameras) suffices [37, 31].

However, autocalibration turns out to be infeasible for certain special types of camera motion called critical motions. These motions arise quite often in practice, so it is important to study them and understand the nature of the difficulty. During CUMULI, Lund and Grenoble collaborated on the task of categorizing the critical motions for various popular sets of autocalibration constraints [63, 38, 44, 39, 45]. Computer algebra and geometric arguments were used to derive the critical motions (motions for which the autocalibration constraints are intrinsically too weak to allow all calibration parameters to be recovered) for several calibration constraints including vanishing skew, known aspect ratio and full internal calibration modulo unknown focal lengths.

Grenoble also reformulated the absolute quadric method in terms of ‘direction frames’ — triplets of 3D projective points representing a basis for the 3D Euclidean structure — and used this to develop an autocalibration algorithm for a moving projective camera viewing a *planar* scene [66]. Planes are common primitives in built environments, for which feature extraction and matching are relatively easy. However, they are singular cases for projective reconstruction, so *(i)* there is more need for calibration, and *(ii)* conventional autocalibration can not be initialized as the projective 3D structure in not available. The new method makes a nonlinear least squares search over all unknown calibration parameters, and also over the unknown Euclidean structure of the 3D plane parametrized by its two circular points or direction frame vectors. Several initialization techniques have been studied. The main disadvantage is that relatively many images are needed (at least 5 if none of the five camera intrinsic parameters are known).

Further results on autocalibration include Lund’s result that projective reconstruction is possible with only 5 points and one conic in two images. In this minimal case there are 10 solutions [43], similar to the case of 5 points in two calibrated images [50].

Sophia-Antipolis and Grenoble studied the autocalibration of a 1D projective camera in a 2D world, and described a method for uniquely determining its two internal parameters using the trifocal tensor of three 1D images [23]. The 1D trifocal tensor can be estimated linearly from point correspondences with no approximations, unlike the usual 2D one. Given the tensor, calibration reduces to finding the roots of a cubic equation in one variable. The main interest here is that certain common configurations of standard 2D cameras can be reduced to the 1D model. In particular, we deduced a new method for self-calibrating an ordinary 2D camera undergoing planar motion.

Sophia-Antipolis considered the problem of autocalibration from uncalibrated image sequences where the motion or camera parameters are known to take certain particular forms. They developed minimal parameterizations for these cases, *e.g.* when the camera displacement is a fixed axis rotation, a pure translation, ... They are currently using this formalism to study the problem of choosing a displacement that allows a robot mounted camera to achieve a given perceptual task.

Sophia-Antipolis also developed a related formalism for detecting special types of camera displacement and scene structure in projective, affine and Euclidean geometry. They derived the combined model and camera singularities of uncalibrated monocular image sequences, for which the fundamental matrix correspondence model degenerates to a homographic one [47, 48, 49].

For cameras viewing a relatively compact scene, and in other cases where perspective effects are small, the perspective camera projection law is often well approximated by the simpler **affine camera** law. In CUMULI, we studied several methods based on this practically useful approximation. Lund developed a factorization method for structure and motion from point, line and conic features [42, 41]. Grenoble developed a new method for Euclidean structure from motion from three affine views [55]. The method is based on intrinsic three-view properties, in particular an equivalence between 2-D affine cameras and 1-D projective ones operating on the plane at infinity [53]. We show that the relative camera orientations are entirely encoded by the 1-D trifocal tensor of the plane at infinity, and from this derive two new algorithms for scene reconstruction from three views. One uses just the minimal 1-D trifocal tensor, the other the full affine three-view constraints. In contrast, previous three view affine scene reconstruction algorithms use only the two-view constraints. The algorithms have been demonstrated on real image sequences.

Grenoble made a study of photogrammetric of bundle adjustment, reported in the survey [71]. The aim of this study was partly educational. Owing mainly to the unfamiliarity of the photogrammetry literature and terminology, the vision community is still unaware of some of the basic photogrammetric developments, which is causing a good deal of duplication of effort. Grenoble have also begun an investigation of sparse and iterative numerical algorithms for the linear update prediction step in bundle adjustment, aimed at improving the performance of bundle methods on large and difficult problems.

Finally, Grenoble developed several new quasi-linear methods for camera initialization from a single image of a few known 3D points [68]. These belong to the family of n-point quasi-linear 'pose + X' methods, where 'X' is some combination of calibration parameters. Existing methods give pose of a calibrated camera ($n = 4$, X empty), and pose + 5 parameter internal calibration (Direct Linear Transformation, $n = 6$). The new methods give pose + focal length ($n = 4$) and pose + focal length + principal point ($n = 5$). One of the main motivations was to provide useful camera initialization methods for the WP2 and WP3 demonstrators.

2.6 Task 1.3: Geometric Features and Uncertainty

The aim of this worktask was to develop calibration and reconstruction methods based on non-point features such as curves and surfaces, including 3D lines, conics and quadrics.

Lund continued their previous theoretical work [18, 59] on the recovery camera motion using only the deformation of apparent contours in images [40]. They implemented algorithms which calculate both projective and Euclidean motion from multiple images containing silhouettes of unknown general surfaces [40].

Lund also extended their shape based factorization methods for points to the case of curves. This theory enables efficient structure and motion estimates using only the images of general curves [11, 8]. As in the point case, the shape based algorithms are independent of coordinate systems and the ordering of the images. In fact, the algorithm also solves the point correspondence problem for the curves.

The shape based factorization methods for points can be refined using nonlinear 'bundle adjustment' techniques [56]. Lund have generalized these from points to curves. Thus it is possible to obtain statistically optimal estimates of structure and motion for curves if the characteristics of the noise are known [12]. Bundle adjustment has also been generalized to combinations of points, lines, conics, curves and patches [1, 5]. This requires the development of techniques for calculating and handling noise characteristics for extracted points, edge curves, correlation patches and other features. This has been possible due to new developments on stochastic models of image acquisition and low level image processing [60, 61, 12] and the work on finding affine correlations [6].

Lund also continued its work on understanding the statistics of shape, which might potentially be used for tracking and recognition, [9, 7]. Line based scene reconstruction was considered in [4], where the minimal cases of 5 lines in 4 affine views and 6 lines in 3 affine views were solved. Work also continued on flexible calibration using little information [3].

Grenoble developed a number of methods for camera pose and relative orientation using unconventional features, including lines, circles, and the sides of circular cylinders.

2.7 Task 1.4: Automatic Estimation of Camera Pose

This task was dedicated to WP 1 work related directly to industrial transfer, in particular the IMETRIC demonstrator. A particular concern here was to allow IMETRIC to initialize the camera poses in their metrology bundle adjustment routines, without having to first install a special "orientation cross" in the workspace. Placing such a cross is time-consuming and hence expensive, it often creates undesirable safety hazards, and in some automated applications it is simply infeasible.

Lund put a considerable effort into the implementation and transfer of existing techniques as well as the more recent results from the group. A MATLAB-toolbox for structure and motion, resection and intersection of points, lines, conics and curves was developed [62]. They also continued to research and implement routines for Euclidean structure and motion in minimal or close to minimal situations and under degenerate configurations. In particular the methods of Philip [52] (for 6 or more non-coplanar points) and of Wunderlich [75] (for 4 or more coplanar points) have been implemented, tested and transferred to the industrial partner. The MATLAB toolbox was also tested on data provided by IMETRIC, to solve the relative orientation problem. This is expected to have a substantial impact on the industrial application.

Grenoble developed a complementary MATLAB library, which includes various methods for structure from motion and for estimating camera pose and relative orientation. The pose and orientation methods include both a range of conventional point-based algorithms for general and coplanar points, and routines that use less-conventional features such as lines and conics.

Grenoble has now developed and tested quite a large number of alternative methods for conventional calibrated camera pose from known 3D points. Many methods exist for this well-studied problem, but all are rather delicate and improved methods for this would have been useful in all three CUMULI demonstrators. So far, the "standard" 3 point method (of which there are a great many variants) still dominates the field, but Grenoble will continue working on this problem and its relatives, as they are an excellent testbed for new ideas about polynomial solving which should apply to many other small geometric problems in vision.

IMETRIC implemented direct solutions for estimating camera pose, and evaluated their impact on work procedures in both standard metrology systems and automated systems for machine control and in-process inspection. The direct pose solution is in practical use and has proven to be very reliable. IMETRIC has found that the typical success rate is approximately 99.5%. In the remaining cases the solution fails due to erroneous identifications of targets by the IMETRIC software, which are not recovered in later steps. IMETRIC will improve the algorithms in order to make these solutions even more robust. These automated systems have been well received in industry as evidenced by customer feedback.

IMETRIC also evaluated the impact of direct solutions for relative orientation on work procedures in the different systems it manufactures. Despite the fact that IMETRIC has alternative procedures that circumvent the need for relative orientation in a number of automated systems, some applications show major benefits, in particular manual measurement of objects, and machine control. In both cases there is no repeat situation, and the placement of an orientation cross is an additional effort in the first case and a major security risk in the second. IMETRIC recently incorporated Grenoble's 5 point relative orientation method into its software and is currently testing this with a view to inclusion in its product line.

IMETRIC's WP 1.4 demonstrator is described more fully in appendix A.

Chapter 3

WP 2. Image Streams and 3D Motion

This workpackage closely parallels WP 1, but emphasizes image streams of moving scenes rather than discrete images of static ones. Its final demonstrator focuses on 3D motion tracking for Image Systems's car crash-testing application, but the results are also of interest for scene and object tracking in Fraunhofer IGD's augmented reality applications.

The underlying theory of image formation and 3D reconstruction is of course the same for continuous images as for discrete ones, so it must be emphasized that almost all of the scientific work listed under WP 1 is also very relevant to WP 2. In particular, WP 1 techniques for initializing and calibrating cameras, and for estimating their motion relative to static or rigidly moving objects in the scene, have proven very useful in the WP 2 (and also WP 3) demonstrator. However, to avoid duplication we will only discuss work specific to WP 2 here.

Although the differences between discrete image problems and image stream ones are mainly methodological, they do often lead to quite different implementation strategies. When there are only a few discrete images, the emphasis is on refined algebraic and statistical techniques that make the most of the limited amount of image data. For image streams, the sheer volume of data tends to make optimal approaches infeasible, and the emphasis is on fast but relatively coarse incremental numerical algorithms. Moreover, even with many images, the accuracy of 3D depth recovery depends strongly on the width of the overall stereo baseline. If there are only a few images, the inter-image motion must be large, which makes it difficult to obtain reliable feature correspondences without a time-consuming combinatorial search. For image streams, the inter-image motion is much smaller and correspondence between adjacent images is relatively easy using local feature tracking methods. However, large-baseline correspondence is still not easy, as error-free tracks must now be processed and maintained through many images, which requires low mistracking and drop-out rates.

3.1 Task 2.1: Incremental Projective Camera Geometry

This task aimed to clarify the underlying geometric structure of reconstruction from multiple closely-spaced images, in the case of completely uncalibrated perspective cameras. This case is somewhat idealized, but advanced tools from projective geometry make it significantly more tractable than the calibrated case and allow relatively simple reconstruction methods. In fact, although projective reconstruction methods only recover the alignment and incidence structure of the 3D geometry (up to a 3D projective deformation), this is already a large proportion of the total recoverable scene information, which may by itself suffice for some applications.

At the heart of multi-image projective vision lies a family of mathematical objects known as **matching tensors**. These are the "multi-image signature" of the projective 3D camera geometry. They carry an implicit representation of this geometry, which can be used both for inter-image feature correspondence and as a stepping-stone to projective 3D scene reconstruction. When CUMULI started, the global structure of the multi-image geometry and its matching tensors had recently been understood for the case of discrete images of point features, thanks largely to the combined efforts of CUMULI's three academic partners [22, 21, 30, 64]. Within CUMULI, we extended this work to deal with the closely-spaced images common in image streams. This required more than just a straightforward application of the existing theory because the geometry becomes increasingly singular as the camera centres coalesce — a fact which simplifies some formulae and derivations, but greatly complicates others.

In the end, two distinct approaches were developed. The Lund group (building on work from Sophia [73]) developed a formalism based on Taylor series expansion of the camera motion [2]. A key result is that third order constraints are needed to fully reconstruct the scene and the camera motion: second order ones (*i.e.* optical flow) are not sufficient.

This approach is well adapted to problems in which the motion model is defined by a differential equation. However, it rapidly becomes intractable when applied to problems with *discretely sampled* motions, where it leads to an infinite series of matching tensors and constraints of ever-increasing order and complexity. For such problems, Grenoble developed an alternative approach based on finite difference expansions [69], and showed how to apply it to 'tensor tracking' — the updating of a matching tensor along a sequence of images. This approach is much closer in spirit to the original discrete image one, whose relative simplicity it maintains.

Several projective 3D scene reconstruction methods designed for use with image streams were also developed under this work task. These are able to deal efficiently with large numbers of images by working recursively, extracting the implicit structural information from each image in turn, and integrating it into a running 3D reconstruction. Both Lund [32] and Grenoble (see deliverable MATLAB libraries) developed recursive projective scene reconstruction methods based on the factorization paradigm. The Lund method works iteratively for each image, whereas the Grenoble method is direct (non-iterative) but requires initial estimates of either discrete or continuous inter-image matching tensors (which can be extracted from the given image correspondences).

Sophia-Antipolis developed a front-end module for low-level motion analysis of long video sequences (MPEG animations, surveillance recordings, television programs, . . .). The system is based on their first order motion formalism for continuous uncalibrated monocular image sequences. The first step is an image stabilization process which iteratively cancels the dominant rotational and calibration-change disparities. This works even for arbitrarily moving cameras with unknown and varying intrinsic parameters viewing moving scenes. Regions whose stabilized disparity is non-negligible represent either moving objects, or nearby ones with significant stereo parallax. These regions are segmented and labelled with a projective indicator of their relative location and size.

3.2 Task 2.2: Continuous Constraints and Euclidean Structure

This goal of this task was to take the projective methods developed in the previous task WP 2.1, and to add further scene or camera constraints sufficient to give a calibrated 3D reconstruction (*i.e.* up to a Euclidean similarity transformation, or change of 3D coordinates and scale). The relationship between WP 2.2 and WP 2.1 thus mirrors that between WP 1.2 and WP 1.1. The constraint enforcement work for discrete images presented under WP 1.2 is therefore highly relevant to the current task. Indeed, one of our main conclusions here is that when it comes to enforcing additional constraints *a posteriori* on an initial projective reconstruction, it makes little algorithmic difference whether the reconstruction was obtained from discrete images or image streams. The place where the continuous/discrete distinction does make a difference to constraint enforcement is in on-line methods, where feature correspondence and tracking can be considerably stabilized by incremental constraint enforcement.

In this task, Lund extended their theory of continuous matching constraints for image streams (see WP 2.1) to the case of Euclidean structure from calibrated cameras [2]. They also developed a method that obtains a Euclidean reconstruction from an approximate affine one. The affine camera matrices are calculated using "closure constraints" based on affine matching tensors estimated from the image data. This procedure makes it easy to cope with missing data and puts equal weight on each image and feature. Finally, the affine reconstruction is upgraded to a Euclidean reconstruction assuming zero skew and unit aspect ratio [33, 41].

Sophia-Antipolis developed their simplified parameterization of the motion analysis problem also in the case of active vision. Here, a robotic system controls the camera translation while at the same time finding the rotation that optimally stabilizes the image by minimizing the overall retinal displacement. Even when the alignment is not exact, they showed that the simplified motion parametrization applies within the foveal part of the visual field. One result that follows from this is that camera self-calibration is very easy in this case: it is easy to recover the subset of the camera intrinsic parameters that is required for 3D-reconstruction.

3.3 Task 2.3: On-line Calibration and 3D Motion from Image Streams

This task was devoted to work directly associated with the Image Systems demonstrator: tracking and 3D motion estimation in high-speed film and video cameras, applied to estimate the rigid and non-rigid motions of marked points during car crash safety testing. New standards and consumer pressure are forcing new cars to pass an increasingly wide range of safety tests, designed to ensure safety in a variety of common crash scenarios. The classic test of head-on collision into a wall is still used, but is now supplemented with angled collisions, multi-car collisions, tests for individual components such as children's chairs and wheelchair restraints, *etc*. This — and the high cost of testing which means that as much information as possible must be extracted from each test — is forcing the industry to transition from the traditional largely 2D tracking and measurement mode to a fully 3D one.

This transition is not without its difficulties, as it requires substantial changes of procedure in the crash-testing labs, as well as a considerable investment in new equipment and software. As in other industries, it is much easier to adopt new methods when they are tailored to fit into the existing workflow. One significant outcome of the CUMULI collaboration was an extended dialogue between the academic partners, Image Systems, and their clients, which greatly clarified both the industry constraints and the technical possibilities and opportunities for 3D reconstruction in crash testing. The benefits of this are still somewhat intangible, but we think that it will ultimately help to speed the adoption and flexibility of 3D methods in crash testing labs.

Image Systems collaborated with the Swedish National Institute for Road and Traffic Research (VTI) to produce two large sets of test data for this application[1]. In the original "cars"

---

[1] We are willing to make these data sets available for comparative testing purposes, but they are much too large

Figure 3.1: Images from four cameras in the Image Systems “cars” crash test sequence.

data set (fig. 3.1), five cameras running at about 500 Hz view a moving car crashing into the side of a stationary one. In the more recent “wheelchair” data set (fig. 3.2), three cameras running at 500 Hz view the side of a wheelchair mounted on a car chassis, which simulates the effect of a head-on collision at about 40 km/h. This test was designed to evaluate the methodology for testing compliance with new EC regulations for fastening wheelchairs in buses. In this case, the fastening survived but the dummy was nearly decapitated by a poorly adjusted seat belt . . .

3.3.1 Tracking Difficulties

One of the disappointments of CUMULI was the difficulty of extracting and tracking the circular targets in the original “cars” data set. Although it may look relatively easy from the images in fig. 3.1, this task proved to be beyond the capabilities of the simple feature extraction and tracking routines that we developed. CUMULI’s principal focus was vision geometry and 3D reconstruction, and we decided right from the start that low-level image processing and tracking were not to be among its research topics. More effective routines could no doubt have been developed, but this would have required the diversion of resources from core CUMULI topics, and we decided not to do this. The “wheelchair” data set was taken partly in response to this difficulty.

The problems were caused by a combination of several factors. Correlation tracking works up to a point given a suitable initialization, but suffers from both drop-outs and track jumping. These problems are exacerbated by the similarity of the targets, the random film motion

---

to put on the WWW.

Figure 3.2: Images 20,40,60,80 from a digital video camera in Image Systems “Wheelchair” test sequence.

caused by the high-speed prism cameras, the large lighting and shadowing variations, the rapid brightness variation of the retro-reflective central disk of the targets under small changes in lighting direction, and the many flying fragments of glass. Feature extraction is made more difficult by these factors, and the fact that the black band around the central target is too narrow, which makes segmentation difficult by allowing bridges between the target centre and the background. Targets for this application should ideally be a dense brilliant matt white on matt black, with the diameter of the black region at least 3–4 times that of the white one.

3.3.2 The Image Systems Demonstrator

This demonstrator is described more fully in appendix B. It is based on Image Systems’s existing TrackEye software and uses the “wheelchair” data set. The workflow in TrackEye centres around 2D analysis of a relatively modest number of manually initialized tracks. Image Systems were keen to preserve this interactive style in the demonstrator as it fits well with both their clients needs and the current structure of TrackEye, so we did not attempt to produce a fully automated method.

Features (primarily marked points) are identified manually and tracked in each camera view using continuity constraints. High-speed moving prism cameras are subject to random image motions caused by film vibrations, so in this case the images also need to be stabilized by tracking fixed points in the lab frame. The stabilized tracks from all cameras are then resampled at common times to allow 3D analysis. (The image streams from different cameras are time-stamped but not synchronized).

Initial estimates of the 3D camera positions (and optionally, of some of their internal parameters) are then extracted from the resampled tracks. A number of CUMULI pose and relative orientation algorithms are available for this, depending on what is known about the scene. Typically, a number of marked points will have been measured before the test, either fixed reference points in the scene, or points on the moving wheelchair or car. The camera and scene geometry of a test setup varies significantly from test to test, so it is useful to have maximum flexibility for this stage. Once the camera geometry has been initialized, initial 3D tracks can be recovered by standard resection at each resampled time step. Finally, the 3D tracks and the camera parameters are refined by a bundle adjustment over all parameters. (The bundle adjustment step has not yet been integrated into the TrackEye system. This will be done by Q1 of 2001).

Chapter 4

WP 3: Algebraic Symbolic Reasoning

This workpackage focused on applying modern geometric reasoning techniques to computer vision scene reconstruction problems, and especially to the Augmented Reality applications of Fraunhofer IGD. Here we discuss mainly the theoretical progress. For more information about Fraunhofer IGD's WP 3.2 demonstrator, see appendix C.

4.1 The "vision" behind WP3

The main goal of WP3 was to investigate the possibility of using automatic geometric reasoning in computer vision. The underlying philosophical idea is to add a new level of flexibility in computer vision by defining algorithms that "mimic" better some properties of the human visual system. Obviously, the human visual system is too complex and too poorly understood to be used as a computational model in computer vision. However, it does not seem to require a lot of experience to convince oneself that we *see* so well (compared to what can be achieved with computer vision) because we *know* what we are looking at. In other words, we do not look at the world and recover its structure solely from the images we get. We already have a generalized model of the world that we are observing, and our everyday task is to make our perceptions fit with this model by adapting it to our perceptions.

Indeed, it is well known that the more informative the model you use, the better the results you can get. However, usually models that carry a lot of information tend to be very specific and so can not be applied to a variety of situations. This is exactly the dilemma that is currently facing computer vision. Thus, the philosophy underlying WP3 is to provide a flexible way to handle complex models. To do so, we work with very simple objects but allow these to be combined to create complex models. Flexibility is achieved through tools that allow the manipulation and use of the various relations that link the basic objects together.

The "grand view" here would be to develop a computer vision system that is capable of making hypotheses from images of the viewed scene (*e.g.* that some lines correspond to verticals, that some features are coplanar, that some planes are orthogonal . . .) on the basis of some heuristics, and then using some very general tools to process these, validate or reject them, and go on to infer further geometric information and ultimately build a plausible 3D reconstruction of the scene. Obviously, this is a very challenging goal that is far beyond the current state of the art. But still WP3 was a small step towards it.

4.2 A Small but Important Step

Many of the tools that would be required to implement the scheme sketched above were not considered at all during WP3. For example, the problem of generating hypotheses was not studied and little was done towards the validation or rejection of hypotheses. Still, WP3 laid down some important foundations for the envisioned goal. The main advances were three-fold:

- Confrontation of geometric reasoning methods with computer vision requirements has significantly and fruitfully motivated the current investigations on high-level algebraic methods for geometric reasoning. The standard geometric reasoning tools were originally developed for rigorous mathematical theorem proving. They were designed to to deal with relatively small numbers of primitives (*e.g.* points or lines) arranged in specific, somewhat "contrived" fashions and often introduced in a given order (the so-called constructive order). In some applications of computer vision, this is not possible: primitives and relations are numerous and given in an un-orderly fashion. Extensions of the classical tools had to be studied to cope with such situations.
- On a more practical side (for computer vision), CUMULI brought various new tools that allow the use of some given geometric knowledge to constrain the 3D reconstruction process.
- A specific application considered the case of 3D reconstruction from a camera and a map for urban scenes. The map is used as a rich source of Euclidean constraints. This case is very interesting because the viewing constraints of urban environments often make large scale 3D reconstruction difficult, and such modelling is very important for augmented reality applications (see a sample of the London data set provided by IGD in figure 4.1). The use of geometric knowledge has proven to be feasible even in the early stages of the reconstruction process, and gives very promising results.

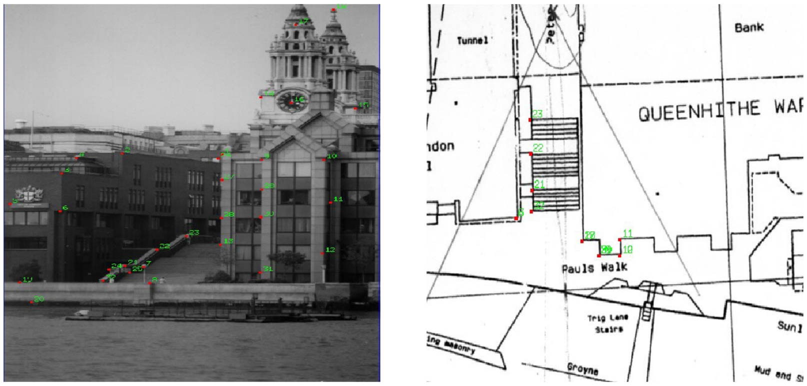


Figure 4.1: The map of an urban environment is a rich source of information that can help the process of 3D reconstruction.

More details are given below.

4.3 Achievements of CUMULI WP3

The achievements of WP3 are summarized in figure 4.2. The main results are depicted briefly hereafter.

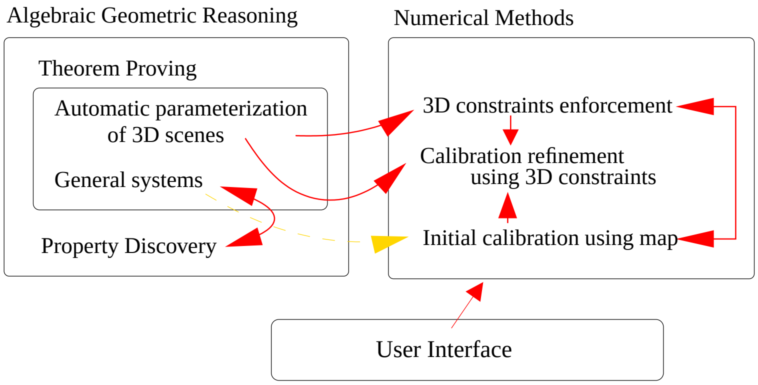


Figure 4.2: An overview of CUMULI WP3 work. Arrows represent the relations between the various topics that have been studied. The dotted arrow represents a “weak relation”.

4.3.1 Computer Vision

As we mentioned above, the dilemma between strong models and flexibility is one of the very difficult problems that faces computer vision. This is particularly true of the sub-field that deals with 3D scene geometry. Basically, two trends are present:

- In model based computer vision (very characteristic of the work done in computer vision in the 70’s and early 80’s), specific models are used to achieve perceptual tasks. For example, people have used broad wire frame models of a car to detect, track and obtain orientation information about cars in urban traffic scenes. Because these models contain quite a lot of information about what a car is, this leads to algorithms that are robust enough to be used in industrial applications. Of course, the model will not in general work for vans or trucks, but the variety of shapes of road vehicles is small enough to be described reasonably well by a small number of models. The limitations of this approach clearly appear when more complex environments need to be handled, *e.g.* houses or more general urban scenes mixing various architectural elements such as buildings, roads, stairs, bridges, ... In such cases, the number of models required needed for genericity is simply too high to be practical.

- A good example of the other trend is the work done within WP1 and WP2. Geometrical models are still present, but they are generally very simple, *e.g.* a few points and/or lines, or just a Euclidean assumption. This has the advantage of being very general, so it can be applied to many different situations. But the cost is a lower overall quality of results, as less information is given to the algorithm.

Between these two extremes, little work has been done. One noticeable exception is the Facade system developed at Berkeley [19], which describes objects using various building blocks (such as cubes, cylinders, . . .) and relations between these (one block is on top of another or aligned with it or . . .). This allows for both complicated and flexible models and gives very good results. The key feature of the method is not so much the building blocks as its ability to handle the complex relations between these, as this is what brings adaptability into the process.

The importance of adding some kind of geometric reasoning to computer vision was recognized quite some time ago, for example in the ACRONYM project at MIT (Brooks and Binford) in the early 80’s, and later in the workshop [46]. However, the algebraic tools and the computer power available at that time limited the effort to very simple toy problems.

From the computer vision point of view, the CUMULI WP3 work brought a number of advances in the context of highly constrained urban scenes:

Constraining 3D reconstructions: Several methods were designed to numerically enforce constraints during 3D reconstruction. The basic idea is always to embed a few iterations of an optimization method that enforces the 3D constraints into the loop that does 3D reconstruction. This is completely new in the computer vision field to our knowledge and allows the removal of a lot of small defects that appear with traditional reconstruction methods.

Lund has extended its affine shape method to enforce affine and Euclidean constraints [58]. The method is based on standard linear algebra, can deal with any number of point in any number of images, treated uniformly. It is also independent of the point coordinates representation and knows out to deal with occlusions. An result example is shown in figure 4.3.

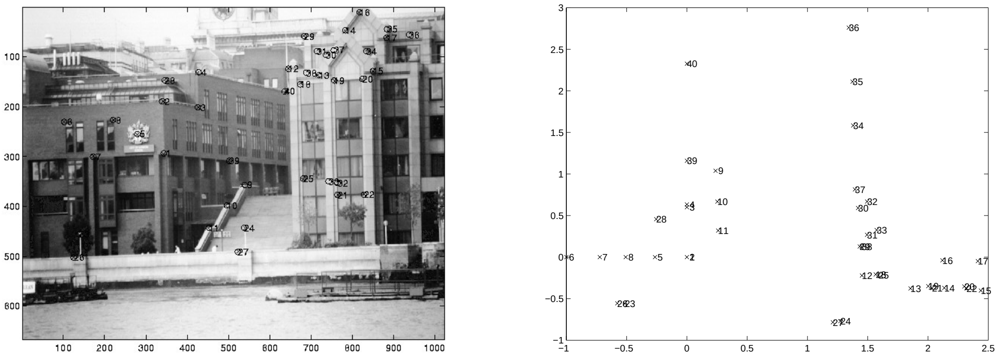


Figure 4.3: On the left: a London image with the primitives used by the extended affine shape method. On the right: a top view of the reconstructed 3D points. With a little imagination, one can see that the walls (virtual lines passing through aligned points), are orthogonal or parallel to each other depending on the constraints that were defined.

Sophia has studied two similar methods: one, analogous to Lund's but not based on the affine shape formalism, was tailored to the specific case of constraining a camera using a map, and the other [16] was based on automatic parameterization of the primitives of the scene given the constraints (see below). Figure 4.4 gives a small example of the results that can be obtained with this method.

Calibration of a camera using a single view and a map: Calibration is a procedure that computes the characteristics of a camera (position, orientation, zoom, . . .). It is often a prerequisite for 3D reconstruction. Sophia has studied how some constraints (coming from the knowledge of the map or given by a human) can be used to provide an improved calibration [15]. The London sequence provided by IGD proved to be very difficult to handle using standard calibration tools, as it is hard to get a sufficiently accurate initialization. Figure 4.5 shows how simple considerations related to the specific case of an image and a map (basically, knowledge of the image lines that correspond to known map verticals) can be used to stabilize this step. Actually, even more constraints can be used to obtain the calibration of the camera up to 4, 2 or zero parameters (depending on the nature of the constraints that are taken into account).

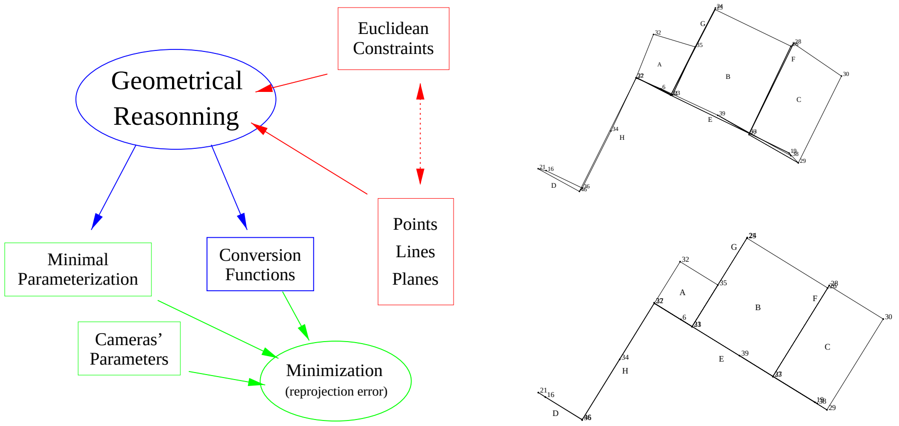


Figure 4.4: *Left:* The basis of the method. The tool shown in figure 4.7 is used to obtain a minimal parameterization of the scene and the "conversion functions" needed to deduce the remaining parameters. These are fed to an optimization procedure which minimizes the reprojection error in the images. *Right:* The reconstruction results before (top) and after (bottom) applying the constraints. Both of these results are the top view of a house. One can easily see that facades that were neither coplanar nor orthogonal in the unconstrained case are corrected by imposing the constraints.

Figure 4.5: Calibration of a camera given a map: The map in the middle shows the result obtained with the standard method whereas the one on the right is obtained using the method specifically designed for that case. The dotted lines overlayed on the maps represent the light rays captured by the camera as computed by the algorithm. The point of intersection of these lines is the position of the camera in the map. The standard method locates the camera in the Thames! The new method, more correctly, locates it on the bank on the opposite side of the scene.

4.3.2 Geometric Reasoning

Geometric reasoning has made tremendous progress over the last few years. Still, at the beginning of CUMULI, there were some major drawbacks to the existing methods that prevented their use in computer vision:

- Most of the theorem proving systems were available only for 2D problems. For some of them, this was an intrinsic limitation of the underlying mathematical method. For others based on explicit coordinates, the method was in principle easy to apply to the 3D case, but limited to a very small number of 3D objects.
- Standard theorem proving systems were built to check whether a given property was true given certain hypotheses. They had no way to deduce general properties of a given situation that might be of interest for computer vision applications.
- Another limitation of the existing methods was their poor ability to handle Euclidean properties such as orthogonality or distances in certain situations. This is all the more problematic because these properties are very common in human made environments.
- Finally, no existing method was able to handle the very large number of unsorted primitives that arise in computer vision problems.

Some of these problems are really very difficult ones. In the next few paragraphs, we show that CUMULI's research has brought at least partial solutions to some of them.

Better algebraic reasoning tools: Having accomplished good theoretical results in particular in creating a geometric reasoning system that can work both in 2D and 3D, without using coordinates and that can deal with Euclidean constraints [27, 26, 74], the Grenoble group experienced the explicit challenge of how to apply these results successfully and effectively to practical problems from computer vision. This challenge has led to more and unexpected effort and exploitation on the interaction of theoretical studies with real-world problems in geometric reasoning and computer vision and resulted in several interesting applications, as can be seen from recent CUMULI-related publications. In fact, attacking computer vision problems has been one of the remarkable applications presented at the International Workshops on Automated Deduction in Geometry (Toulouse 1996, Beijing 1998 [17] and Zurich 2000).

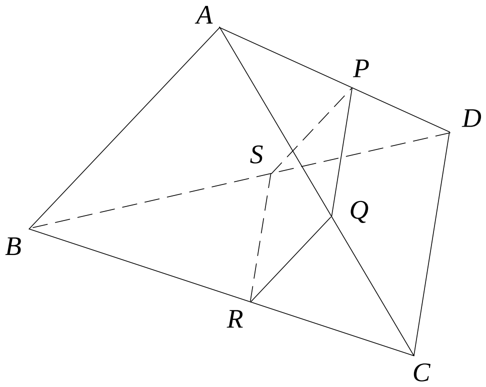


Figure 4.6: If a plane intersects the edges AB, AC, CD and DB of a tetrahedron $ABCD$ at four points M, N, E and F, and if MNEF is a parallelogram, then the center of this parallelogram is on the line connecting the midpoints of AD and BC.

In addition, the Grenoble system depicted in the previous paragraph can be used to discover properties in 2D or 3D. Another such method has also been developed at Sophia but only for the 2D case [13]: this system was able to find all the properties of a given type for a small number of types of properties. However, this system was not computationally efficient. Being able to discover all properties efficiently as well as extending the types of properties that can be discovered is an open problem.

Constructive order: Sophia has developed a method to do automatic parameterization of 3D constrained models [14]. This method is illustrated in figure 4.7. In contrast to standard geometric reasoning methods, this algorithm is able to handle large numbers of primitive objects (more than 50).

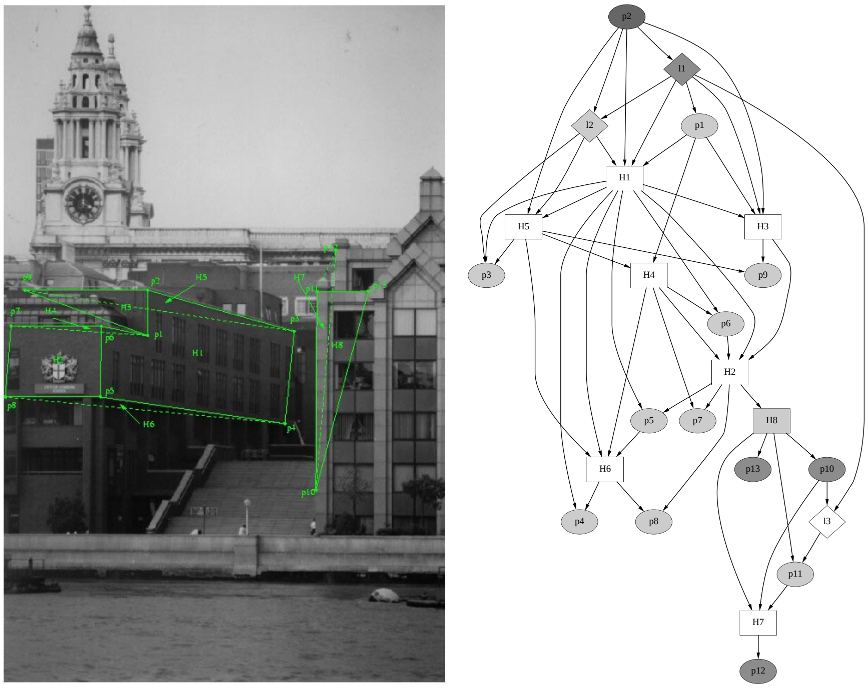


Figure 4.7: *Left:* A view of the London scene with some marked primitives. *Right:* An oriented graph depicting a minimal parameterization of the scene. The box type refers to the type of the primitive (rectangles for planes, diamonds for lines, ovals for points). The edges entering a node represent constraints. Colours represent the degree of freedom of the primitives, dark meaning completely free and white completely defined.

The method relies on some heuristics, but seems to give good results with all the cases that have been tested. Obtaining a provable, efficient method for finding a minimal parameterization is an open problem.

4.4 Conclusions

WP3 was a very speculative workpackage. It required a good deal of effort to understand what could be done to combine geometric reasoning and computer vision. Even though we cannot be sure yet that the “grand view” depicted above is exactly what needs to be done or whether it will be achievable some day, this study has been very fruitful for both disciplines:

- For computer vision, it opened a new way of addressing the problem of 3D reconstruction and, we hope, of adding more “intelligence” in that process. Also, thinking in terms of constraints has lead to some work that is immediately applicable, *e.g.* in the field of augmented reality.

- For geometric reasoning, it gave the opportunity of considering other applications than theorem proving, by studying the types of problems that arise in the situations encountered in computer vision.

There are many open problems that still need to be solved in order to be able to fully exploit geometric reasoning in computer vision. Some of them have already been mentioned above. Some of them are just a question of computer engineering, as most methods for geometric reasoning require some symbolic manipulation capabilities, which are often difficult to mix with the more numerical nature of standard computer vision procedures. One problem in the original workplan that still remains open is the possibility of doing geometric reasoning in the presence of uncertainty. The main cause of difficulty here is simply to define what geometric reasoning with uncertain hypotheses should actually mean.

Nonetheless, WP3 opened a fresh road for 3D reconstruction in computer vision, and a set of very “real” problems for the field of automated geometric reasoning. This has already led to some practical results, but further work is required to get the full benefits of combining these two techniques.

Chapter 5

Industry workshops & wider dissemination in CUMULI

During CUMULI, we made several efforts towards wider dissemination of results. Perhaps the greatest success is Fraunhofer IGD's involvement in the formation of the world's largest Augmented Reality consortium ARVIKA (www.arvika.de), which involves about 20 major German manufacturers, including Volkswagen, Daimler-Chrysler, Airbus, DASA, Ford Germany and Siemens. It was formed in part as a result of Augmented Reality work at IGD, including results from CUMULI. The main application sectors are the development, production and servicing of complex machinery such as cars and aircraft. See the Fraunhofer IGD section of the CUMULI Technology Implementation Plan for more details on ARVIKA. Fraunhofer IGD is also pursuing further industrial AR work outside of ARVIKA.

Other contributions to dissemination during the project include:

- Lund and Image Systems presented CUMULI at an industrial session of the annual Swedish workshop SSAB, arranged by the Swedish Association for Automated Image Analysis, in Gothenburg, March 1999. Presentations were made by Image Systems (Anders Källdahl) and Lund (Gunnar Sparr), and also included a video demonstration from IMETRIC. The audience was about 100 people, about 30 of whom were industrial.

- IMETRIC won the Swiss *Innovation Award for Technology Location Switzerland* for its TI^2 technology which includes some critical methods from CUMULI. TI^2 was certified by Boeing for CNC machining of aerospace components. It is the first and only vision system they have ever certified for this application.

- IGD was twice co-organizer of the International Workshop on Augmented Reality (IWAR'98 and IWAR'99). These events were in the USA, but this year the workshop will become the International Symposium on Augmented Reality ISAR 2000, to be held October 5-6 2000 in Munich Germany.

- Grenoble was the principal organizer of a very successful academic workshop *Vision Algorithms: Theory and Practice* [72] at the International Conference on Computer Vision, 21-22 September 1999, in Corfu, Greece. Although not explicitly linked to CUMULI, this workshop embodied many CUMULI themes.

We had originally planned to host an "Industrial Workshop on 3D Measurement from Multiple Images" in Paris at the end of CUMULI. In part, this was intended to be a follow-up to a very successful workshop with a similar theme which was held in Paris in 1994, in which

Grenoble and IMETRIC participated. Unfortunately, this workshop had to be cancelled at the last-minute, owing to organizational problems beyond the consortium's control.

Chapter 6

General Comments and Perspectives

We finish with some general comments about various aspects of CUMULI, and some perspectives that were opened up by the project.

6.1 Research Directions

Scientifically, the period of CUMULI was one of rapid progress, to which CUMULI itself contributed heavily. But it was also in a sense the end of an era. Over the last decade our understanding of vision geometry has been revolutionized by the systematic adoption of tools from synthetic projective geometry. Much of this advance originated from Europe, aided by a string of major European research projects such as VIVA, Realise, Vanguard and CUMULI. There is a clear feeling in the community that although some consolidation still remains to be done, the period of rapid progress in geometry has come to an end. The geometric results are still being applied in more development-oriented projects, but the focus of attention in the vision community has shifted towards: (*i*) applications (especially media-based ones such as vision based interfaces, augmented reality and virtual studios); and (*ii*) less geometric (but equally mathematical) topics, notably a revival of “image understanding” based on modern statistical analysis and learning techniques, often applied to modelling human motion or appearance.

6.2 Pose and Polynomials

One of the surprises of CUMULI was the extent to which small algebraic routines for camera initialization from a few known features were still welcome. The academic partners originally considered this to be a rather dry and largely solved topic, at least for point features. However, all three industrial partners found our absolute and relative orientation routines very useful for their demonstrators, and are actively developing products that incorporate them. The routines that we developed for initialization from non-point feature combinations have not yet seen so much use, but they also may come into their own as the industrial partners meet applications where they are needed.

This success is gratifying, but slightly embarrassing because even the best of these initialization routines are not nearly so reliable as we would like them to be. They mostly involve the reduction of a minimal or near-minimal set of observations to a system of polynomials, which is then solved by a numerical method. Owing to (near-)minimality, singular or near-singular configurations of the data for which the problem can not be uniquely solved are common (in practice often annoyingly so). Many of these failures are intrinsic and can not be avoided except by using more or better data. However, too many others are created artificially by the particular algebraic formulation or polynomial solver used. It must be admitted that despite much research, even our best current tools for manipulating and solving polynomial systems are often very clumsy, particularly when the system is overspecified (has more equations than variables, and hence generically no roots at all) and has coefficients that are themselves uncertain owing to measurement or rounding error. More research is needed to find efficient numerical methods for solving polynomial systems with uncertain coefficients.

Note that exactly the same conclusion arose in WP 3, where it proved very difficult to combine geometric reasoning with uncertainty, owing at least partly to the lack of effective methods for handling uncertain polynomials. This is a difficult field, but we think that there is much progress to be made here in the future.

6.3 Industrial Transfer

The transfer of technology and know-how between academia and industry remains difficult. Within CUMULI, the goal of our transfer effort was largely educational. We aimed for each partner (both academic and industrial) to build up an active, applicable know-how of the basic methods, concerns and application constraints of the other partners. We did of course develop and exchange software as well, which was indeed used in the technology demonstrators. But we always viewed the software more as a concrete aid to knowledge transfer than as an end in its own right.

Perhaps CUMULI was unusual in this, but none of the industrial partners were interested in conventional “commercialization” of the software we developed. Their goal was to *understand* the methods developed, and then reimplement them within their own systems. No code that wasn’t fully understood and hence fully maintainable was acceptable, and every routine had to be adapted to the geometric conventions and coding and documentation standards of the relevant system. This may seem like a great deal of duplication of effort, but we believe that technology transfer that is based mainly on bodily transfer of code is often somewhat illusory: without a thorough understanding of its methods and limitations, the code is essentially unmaintainable by the industrial partner, so that it can not be safely used in long-lived product lines.

Indeed, CUMULI confirmed our experience that adapting working academic-level algorithms to make them robust enough to use in industrial systems can often take as much effort as developing the original prototypes, if not more. The method must be made to work in the context of an existing system, with all of the conventions and compromises that that implies. Thus, in many cases it was necessary to embed the original algorithms in a large amount of additional code to provide a sufficient robustness for routine industrial use.

With the educational aim in view, we originally decided to transfer code in MATLAB, on the grounds that this allowed significantly more compact and legible implementations than C/C++. MATLAB certainly allows algorithms to be rapidly prototyped and tested, it does produce somewhat more legible routines than C, and it is relatively independent of issues such as the library used for linear algebra. But on the other hand, it is too far from C to allow easy, bug-free reimplementation by the industrial partners. On balance, we do not think that there is a clear case on either side.

Appendices

These appendices contain descriptions of the three main CUMULI demonstrators, and a list of scientific papers published under CUMULI. The demonstrators are:

- WP 1. “Automatic estimation of camera pose”: methods for camera pose and relative orientation from various different feature types, integrated by Imetric SA for flexible system initialization;
- WP 2. “On-line calibration and 3D motion from image streams”: 3D motion tracking for crash testing applications, integrated by Image Systems AB;
- WP 3. “Image and Video Augmentation”: methods for camera pose and calibration for use in Augmented Reality applications, integrated by Fraunhofer IGD.

Appendix A

Demonstrator 1: Automatic Estimation of Camera Pose

A.1 Demonstrator: Automatic Estimation of Camera Pose

This demonstrator was integrated by Imetric SA into the IMSlib 3D Image Metrology Library. This library is the basis for the ImetricS software, the ICam metrology cameras, and the TI^2 machine control systems. The demonstrator shows the impact of the pose estimation and relative orientation techniques on the performance of these systems.

The IMS lines of systems are general 3D Image Metrology Systems used in industries ranging from Aerospace to Engineering. The TI^2 systems have been developed in cooperation with aerospace partners with Boeing on one hand and a grouping of BAe Systems, Rolls Royce, and Bombardier on the other hand. The system was certified by Boeing to control an NC machine in 1999 and will go into production during Q3 2000. It will be the first system in the world where a highly accurate NC machine is controlled in 6 DOF for machining operations by a 3D Image Metrology System. There are a number of programs under way where this technology will be applied in aerospace projects (Eurofighter, 3XX, A400M, JSF) as well as in special applications in the shipbuilding industry.

The IMS technology is also used in a surface scanning system jointly developed by Daimler-Chrysler and Imetric.

A.2 Pose Estimation

The improvements in pose estimation have had a major impact on the performance of the IMETRIC systems. Originally, users had to manually select 4 points in each image and to provide approximate XYZ-coordinates of the camera in order to enable the software to compute the pose for one image. The need for approximate camera positions made automation impossible for systems using hand-held cameras.

Three items were used to alleviate this situation. Firstly a so called “orientation cross” was developed. This consists of 2 bars with at least 4 targets which are mounted together. Secondly Imetric developed so called “coded targets”. These were placed on the 4 locations on the “orientation cross”. Thirdly the pose from 4 points algorithm developed in CUMULI was used to automatically compute the pose. The first implementation showed that under certain conditions, if the coded targets are misidentified, the algorithms could not work. Thus the number of targets on the orientation cross were increased to 6 targets and the performance was

Figure A.1: A typical image of a ship bulkhead instrumented with retroreflective photogrammetric targets. The Imetric orientation cross is visible in the centre. Image courtesy of Bath Iron Works.

improved via the use of "robust" solutions to provide a reliability of better than 99.9%. Further enhancements of the implementation showed that during testing of the system by an aerospace partner during several months as well as measurements performed during 1.5 years of operation of an automated system, no cases at all were detected where a wrong orientation was computed, and the cases where no orientation was computed even though a sufficient number of targets with known coordinates was available was below 1 in 1000. Performance characteristics like these can be considered a "minimum" requirement in applications like TI^2, where this kind of technology is used to control NC machines. The aerospace applications of TI^2 mostly involve machining of parts which cost already over 100,000 Eu, and very often a scrap part would lead to unacceptable delays in a manufacturing program.

A.3 Imetric 3D Image Metrology Systems

Systems of this type have found acceptance in a large area of applications within the aerospace, automotive, shipbuilding, machinery industry, and the engineering field. The practical use of the systems was initially hindered by the difficulty to use these systems. One of these elements was the requirement of the user to provide initial values for the position from where the user was taking images. Figure A.1 shows a typical unit of a ship. In the middle of the unit one can see a so called "orientation cross". One can furthermore discern some coded targets and a number of standard targets. The task to be performed here is to determine the excess amount of steel to be trimmed so that the unit fits precisely onto the ship. Typically several hundred points are measured for this purpose. Originally users were required to provide approximations for the positions from which the images were acquired. Given that the images are taken with a cherry picker, is very difficult to take pictures and at the same time note the XYZ-coordinates of one's position in the ship coordinate system with any level of reliability. Using the algorithms developed in CUMULI and implemented by Imetric the user only has to press one button and

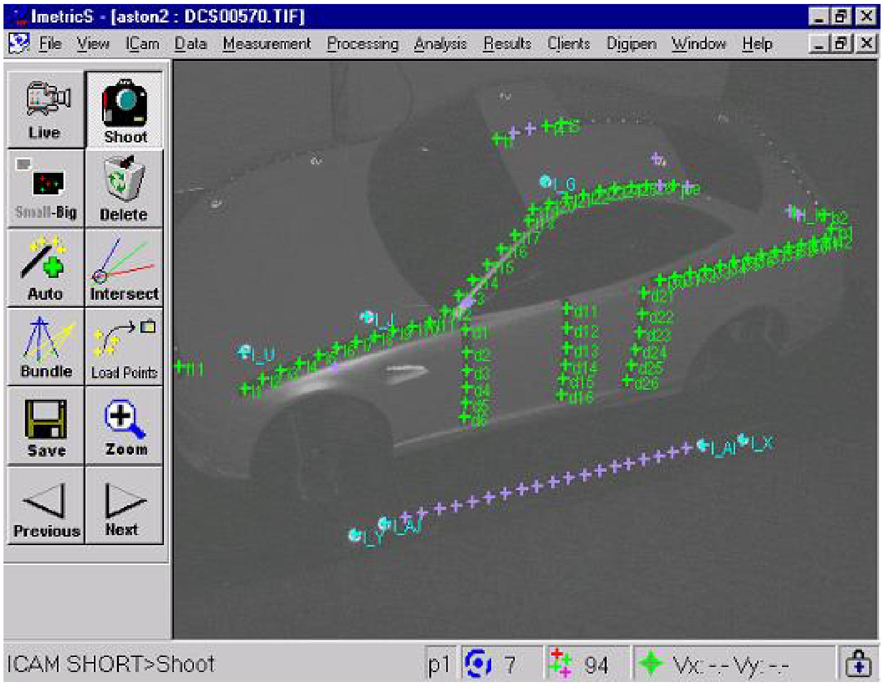


Figure A.2: The screen of Imetric's new ICam28 metrology camera during the measurement of a car model.

all images are automatically measured and oriented. For this purpose the coded targets on the orientation cross are detected by the software. The images are initially oriented with respect to the coordinate system defined by the orientation cross. The measurement speed on a 300 MHz Pentium based computer is in the order of a few seconds per image. Traditionally the measurement of one image required 3 to 4 minutes. The high level of automation and the robustness of the software allow the use of relatively unskilled personnel to perform the measurements, which is another benefit besides the pure improvement in measurement speed.

This application also shows very effectively the need for the relative orientation algorithm developed within CUMULI. As one can see from figure A.1, the orientation and scale cross must be placed on the unit by the operator. The cross sometimes falls, leading to significant delays. It would thus be much easier for the user to simply fix the coded targets (as is already being done) and not to have to do anything else. The relative orientation software is in the process of being tested at this shipyard as part of the delivery of a new ImetricS software release in Q3 2000.

A.4 ICam Metrology Camera Systems

Figure A.2 shows the screen of an ICam28 camera during the measurement of a car model. The complete camera is controlled via a specially designed user interface and two buttons. The figure shows the screen of the ICam after the measurement of one image. One can see that the software has detected 7 coded targets and 97 other targets. The image has been oriented using values for the coded targets from an earlier measurement. This is indicated by the green cross on the lower right message area. The user can thus judge immediately after acquiring an image (within 2 sec for the ICam6 and 8 sec for the ICam28) whether or not the image measurement

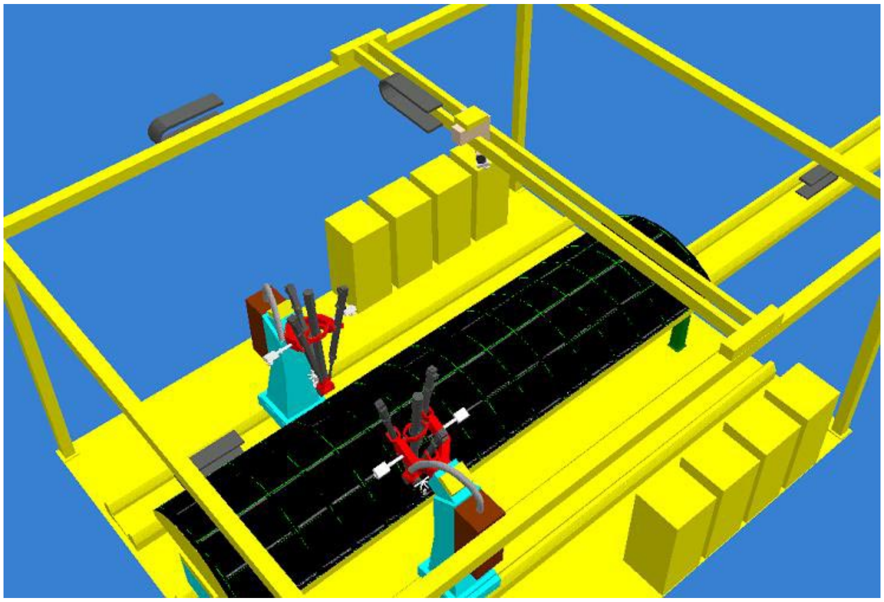


Figure A.3: A CAD simulation of the TI^2 vision controlled machining system for drilling holes in aircraft skin panels.

worked and the image was oriented. This is a major advantage for the user and is only possible among other things due the algorithms developed within CUMULI.

A.5 TI^2 System

The basic idea of the TI^2 System existed in 1995. At that time it was clear that the basic technology was available, but that the reliability and automation of some essential components would have to be significantly improved. Among them were cameras sufficiently reliable to perform for several years (or even several months) without failure, and a large number of algorithms that were needed to make the system totally automatic. One of these was the determination of camera pose and/or the determination of the relative orientation of images. Figure A.3 shows a system design for a machine, which is to drill over 1000 holes into a skin panel. The operation of the system consists of the following steps. First targets (coded and standard) are placed on the skin in arbitrary positions. Scale bars and tooling hole adapters are integrated into the fixture but not visible in the figure. A camera mounted on a gantry takes several images from several positions, measures them and computes 3D coordinates for all targets on the skin. Using the scale bars and tooling hole adapters the 3D coordinates are transformed into the part coordinate system. Thereafter, the two NC machines are accurately positioned using the two cameras mounted on each machine and targets placed on the end-effector. In this application all operations must occur completely automatically. Thus for the measurement of all targets the relative orientation is essential as one does not want to rely on reference targets on the part/tool or on precise positions of the cameras. The two camera systems on the NC machines only require the pose from point algorithms as they “see” a large number of coded and standard targets.

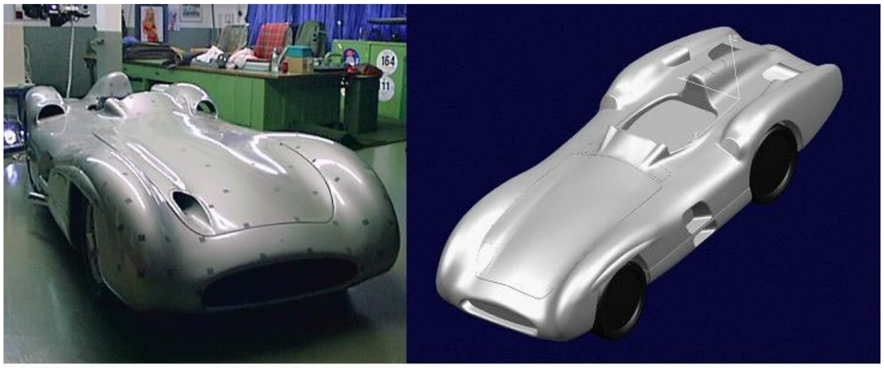


Figure A.4: A historic car prepared for scanning with the Daimler-Chrysler / Imetric surface scanning system, and the resulting surface model.

A.6 Surface Scanning System

Daimler-Chrysler have been cooperating since 1993 on the development of a surface scanning system. The systems consists of a camera and a calibrated pattern projection device. Figure A.4 shows on the left a picture of a (museum) car during scanning. One can see the targets on the car which are used to align the different patches. On the right hand one can see a surface representation of the car generated from some 30 million points which were measured on the car within $\frac{1}{2}$ day. Part of Imetric's contribution to the system are the calibration and positioning algorithms, which include algorithms developed within CUMULI.

A.7 Relative Orientation using Non-Point Features

Although Imetric decided not to include these capabilities within the current demonstrator, it continues to work with clients who are interested in using such algorithms in their systems. Given that initial routines developed under CUMULI are now available for several problems of this type, the introduction of this technology is currently more a question of work flow and certification issues than one of availability of the basic technology. Imetric currently envisions augmenting the "orientation capabilities" of its IMSlib software during Q3 and Q4 of 2001.

A.8 Reliability and Practical Use

CUMULI has confirmed Imetric's experience that adapting "working" algorithms to make them robust enough to use in industrial metrology systems can often take as much effort as developing the original prototypes, if not more. It is not enough for the method to work in isolation. It must be made to work in the context of an existing system, with all of the conventions and compromises that that implies. Thus, it was necessary to embed many of the algorithms in a large amount of additional code to provide a level of robustness sufficiently high for routine industrial use.

Appendix B

Demonstrator 2: On-line Calibration and 3D Motion from Image Streams

This demonstrator was integrated by Image Systems AB within their TrackEye^{TM} software. It shows how camera calibration and 3D reconstruction techniques from CUMULI allow more flexible system initialization and improved 3D motion estimation in a car crash testing application.

B.1 Background

For many years, the analysis of high-speed film and video images from crash tests has been an important tool in the car industry. Points on cars or on dummies are tracked through a sequence of images (typically about 100 images from each camera with a frame frequency of up to 1000 Hz) and parameters such as displacements and accelerations are calculated.

In most cases this analysis is 2-dimensional: cameras are mounted so that they view the scene perpendicular to the direction of motion, and only motion in the image plane is studied. The image scale is set using a known distance in the scene. This works because in traditional crash testing, where a car is crashed straight into an obstruction, there is little or no sideways motion due to the symmetry of the setup.

However, with the advent of new car safety regulations, there is a requirement for tests, such as side impacts, where it is important to measure motion in three dimensions. This means that there will be an increased demand for software to compute 3D trajectories of tracked points.

Traditional methods for 3D analysis require calibrated cameras. The calibration procedure is performed using a special calibration setup, and calibrates the internal camera parameters focal length, principal point (the intersection between the camera's optical axis and the image plane), aspect ratio and skew, as well as lens distortion parameters.

Compared to 2D testing, this adds a number of complications to the testing procedures. Not only is the calibration itself time-consuming, but it also becomes necessary to keep track throughout the entire analysis procedure of which camera individual was used to produce each image sequence. Cameras must also be re-calibrated every time the lens is changed, and a calibration is valid only for one particular focus setting.

For these reasons it is highly desirable to be able to work with uncalibrated cameras, that is, to calculate the camera parameters as part of the 3D analysis procedure. This requires some knowledge of the scene — typically, the 3D coordinates of a number tracked points must be known — but the amount of required knowledge should be as small as possible.

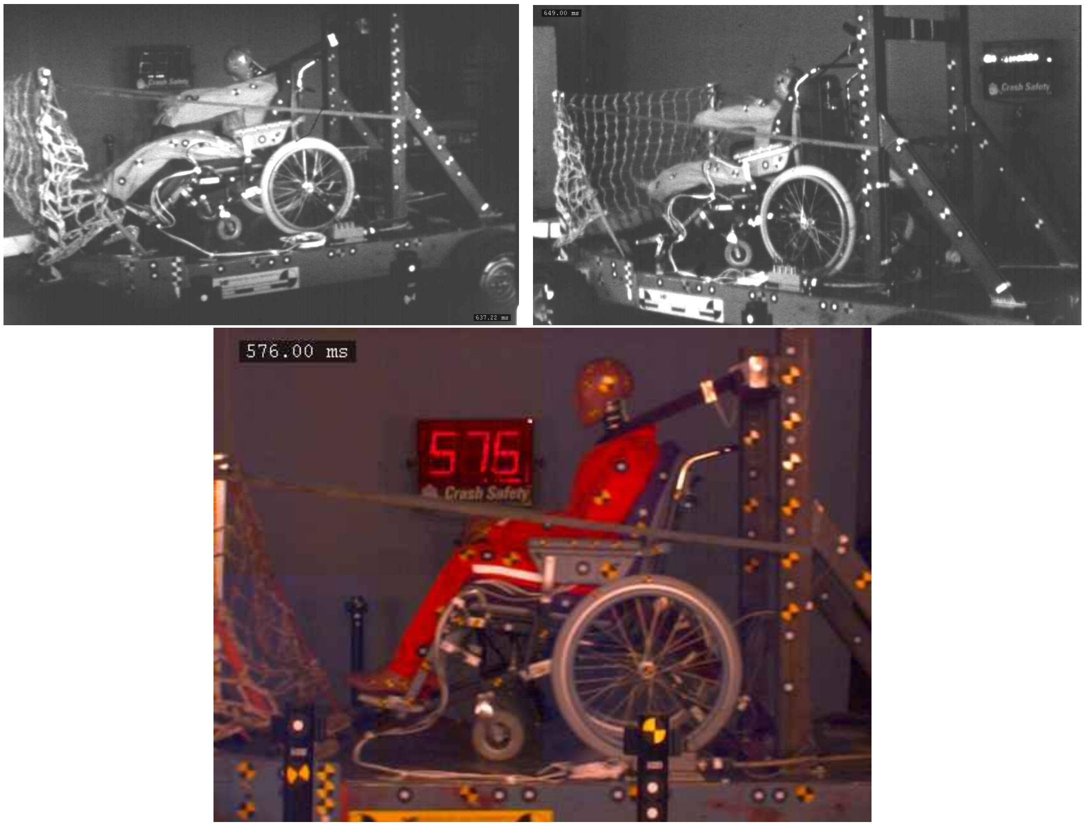


Figure B.1: Images from the left and right (top) and middle (bottom) cameras of the Image Systems “Wheelchair” test data. The two top images were from rotating prism film cameras, and were digitized using a TrackEye^{TM} film scanner. The bottom image was from a high speed digital video camera.

The purpose of the CUMULI WP2 demonstrator is to demonstrate how CUMULI algorithms can be used to perform 3D analysis of crash test images, using uncalibrated cameras.

B.2 Test Data Set

The test data used for the demonstrator consists of three image sequences of a crash test, where a wheel chair was mounted on a wagon that crashed into a stationary obstacle. The test was filmed from three angles using two rotating-prism film cameras and one high-speed video camera. All cameras were running at approximately 500 frames per second. They were not synchronized, but each image is time stamped. Figure B.1 shows some typical images from this data set.

A peculiarity of the rotating-prism cameras is that the entire image moves relative to the film from frame to frame. To compensate for this we track one or more fixed points in the laboratory and translate the image coordinates so that the fixed points get the same coordinates in all frames.

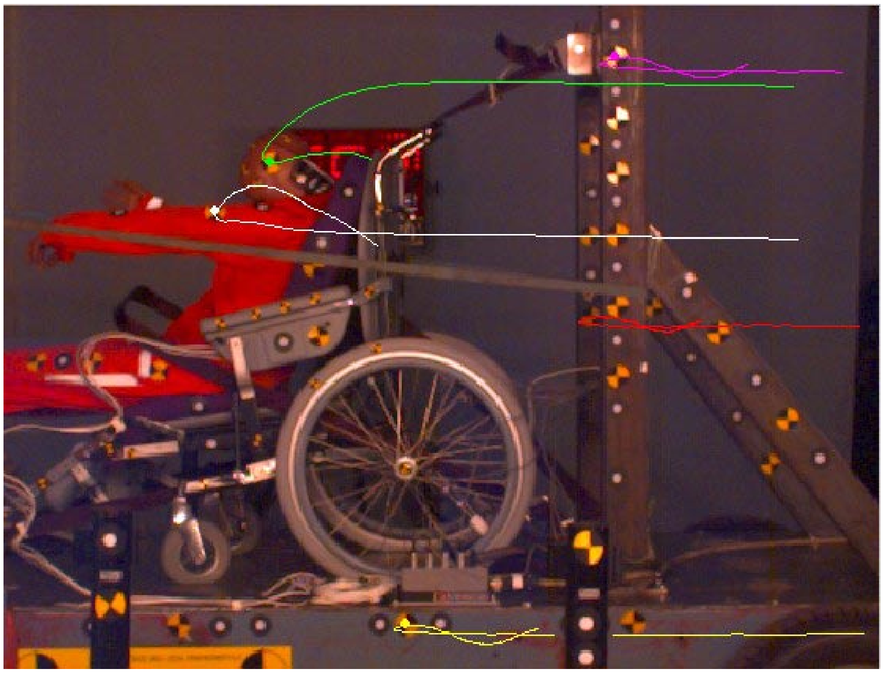


Figure B.2: Tracks of a few points on the wagon and dummy.

B.3 Tracking

The first step in the analysis process is to track the points we are interested in through each image sequence. For quadrant targets (the black and yellow cross targets commonly used in crash testing), the demonstrator uses a tracker developed by Image Systems within CUMULI. This tracker uses pattern recognition techniques to detect the typical symmetry properties of the quad target and find the target's centre of symmetry. Figure B.2 shows a few of the point tracks obtained by the tracker.

Once the tracks have been estimated, they are postprocessed. Tracks from the rotating-prism cameras are compensated for the image motion described above, and all tracks are synchronized to a common set of sample times by interpolating the track coordinates to these times.

B.4 3D Analysis

B.4.1 Bundle Adjustment

The main part of the 3D analysis in the demonstrator consists of a bundle adjustment: a least-squares fit of camera parameters and point coordinates to the tracked image coordinates. One of the algorithms developed at Lund is used. This algorithm computes point coordinates, the cameras' external parameters, focal lengths and principal points. The aspect ratio and skew of all cameras are set to 1 and 0, respectively.

We also assume that the camera parameters are constant, *i.e.* the cameras do not move or change their optical properties during the test. All of these assumptions are satisfied in our test data (and in most crash test setups), provided we compensate for the image motion of the rotating-prism cameras as described above.

Since we have constant camera parameters, we can use the image of the same point at different times as input to the bundle adjustment (in effect creating a “multiple exposure” of the track). This will increase the overdetermination in the equations and lead to more accurate values for the camera parameters.

B.4.2 Initial Values

The bundle adjustment algorithm requires approximate initial values for the camera parameters and point coordinates.

In the demonstrator, the Grenoble 4-point DLT like calibration algorithms are used to compute initial camera parameters. These algorithms require that the 3D coordinates of at least four tracked points be known in one image for each camera (different points can be used for different cameras), and calculate the external camera parameters and the focal length. Since the principal point isn’t known, we set this to the image centre.

In the test data, the 3D coordinates of a number of points have been measured in a coordinate system that moves together with the wagon. If we use these coordinates to calibrate the camera parameters at a certain time t_0 (and make sure that the same t_0 is used for all the cameras) the output from the 3D analysis will be in a coordinate system that coincides with the wagon coordinate system at the time t_0.

Having computed the approximate camera parameters, the demonstrator then computes approximate point positions by intersection, *i.e.* by finding the point in space that gives the smallest deviation from the observed image coordinates when projected back onto the image planes. The computed coordinates are then used as starting values for the bundle adjustment.

B.4.3 Filling in the Gaps

We could in principle perform a bundle adjustment on all the tracked coordinates, over the entire sequences (100 frames or so per camera). To reduce the running time of the bundle adjustment, the demonstrator uses only a subset of the total sequences. The point coordinates for the rest of the frames are then calculated by intersection, using the camera parameters computed by the bundle adjustment.

Appendix C

Demonstrator 3: Image and Video Augmentation

This demonstrator was integrated by Fraunhofer IGD. Elements of it are currently being ported to their commercial VR-system *Virtual Design 2*.

During CUMULI, Fraunhofer IGD developed a new tool for Off-line Augmented Reality (AR). The tool allows the insertion of virtual objects into images and videos with minimal user interaction, and without requiring other knowledge than the images. It also makes it possible to create a 3D model of the scene and to handle occlusion between real and virtual objects.

The first step of the augmentation process consists of the computation of the projection matrix of the real camera. One solution would be to apply classical camera calibration algorithms, which are well-known in Computer Vision and Photogrammetry. The problem is that these algorithms require many 3D scene points to provide accurate results. In practice, only a few reference points may be available. As a result, the calibration process would be fragile and most of the images could not be exploited.

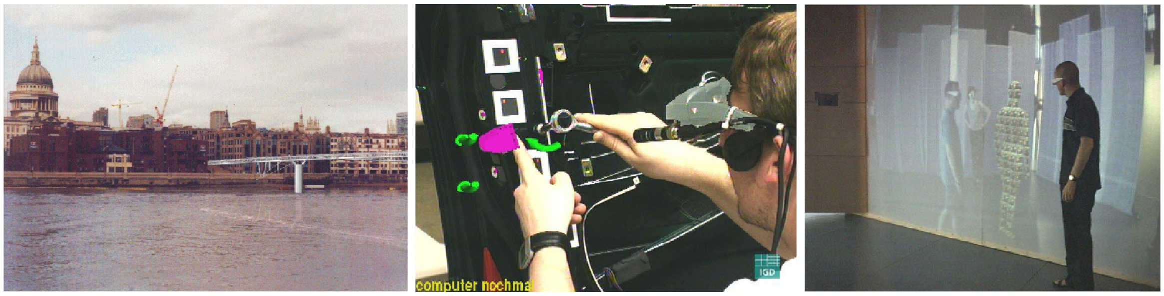
New flexible approaches have been developed in CUMULI to facilitate image calibration. The Fraunhofer IGD tool offers a wide selection of different algorithms that allow most of the common image types and scene configurations to be handled. It applies new academic results in the area of structure from motion, auto-calibration, absolute orientation and calibration with help of a 2D map.

In this description, we briefly present the Augmented Reality technology. Then, we give an overview of the Fraunhofer IGD tool. Afterwards, we present the typical augmentation process for three kinds of situations: augmentation of a single image, of multiple images and of video sequences.

C.1 Augmented Reality

Augmented Reality (AR) is a new research technology, which explores various approaches to augmenting a user's view of the real world with additional information.

By combining live video input with immersive display technology, AR allows users to work with and examine real 3D objects while seeing additional information about the objects or the task at hand. AR can be applied in many sectors, *e.g.*, visualization of architectural CAD-models in their real environments (figure C.1a), repair and maintenance of complex machines and facilities (figure C.1b), and entertainment (figure C.1c).



(a) (b) (c)

Figure C.1: Three AR-applications: (a) architecture; (b) machine maintenance; and (c) entertainment.

C.2 Overview of the Fraunhofer IGD AR tool

The Fraunhofer IGD AR tool consists of four image viewers (figure C.2). In each viewer the user can interactively set 2D image points or lines and enter 3D coordinates, if available.

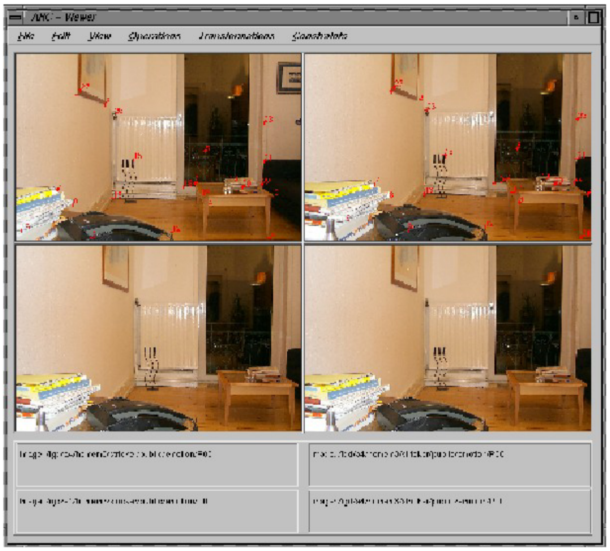


Figure C.2: Tool interface

The user can also scroll through an image sequence and set points for every image. The tool offers common input and output functionalities for images, features and virtual model (Inventor/VRML). Interactions like image zooming, feature selection and deletion, *etc.*, are also supported.

The main CUMULI-specific menus are “Transformation” and “Constraints”.

The “Transformation” menu collects algorithms related to camera and scene geometry, including several methods developed in CUMULI WP 1. These algorithms can be used to determine the relative transformations between different cameras, calibrate them or do 3D reconstruction of the scene.

The “Constraints” menu gives access to methods developed under WP 3. It allows constraints to be set on image features (*e.g.*, that points lie in a plane or on a line) and 3D coordinates to be entered. It also allows the approximate position of the camera to be entered on a map. The map is handled like any other image and is shown in one of the four viewers. An example of image and map is presented in figure C.4.

C.3 Image Augmentation

C.3.1 Calibration

The computation of the camera parameters (*i.e.* focal length, principal point, aspect ratio, orientation and position) is the first step of the image augmentation process.

In practice, it is often convenient to pre-calibrate the camera, *i.e.* to estimate its intrinsic parameters (focal length, principal point, aspect ratio) beforehand. We focused on flexible and easy to use algorithms, such as camera calibration with a simple planar grid, printed on standard paper as shown in figure C.3. For single images, other calibration algorithms have also been implemented. These are based on the Direct Linear Transformation (DLT) and are intended for general 3D point configurations.

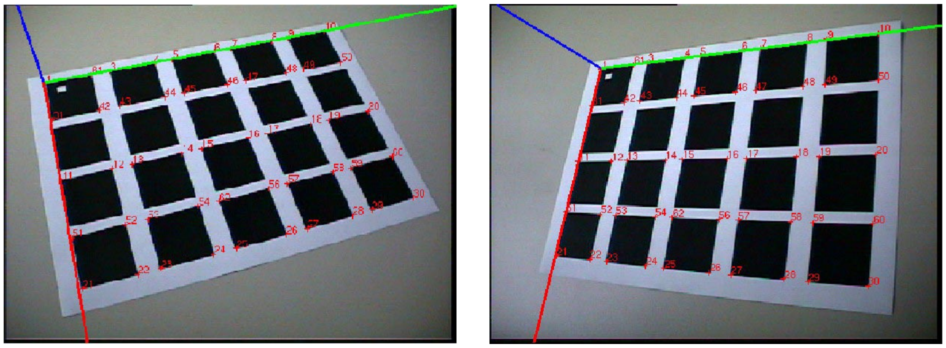


Figure C.3: Flexible camera calibration using a planar grid.

If the intrinsic camera parameters are already known, algorithms to compute absolute camera pose (*i.e.* 3D position and orientation) can be applied. Because fewer parameters have to be recovered, this approach is more robust and provides good results with as few as four 3D points.

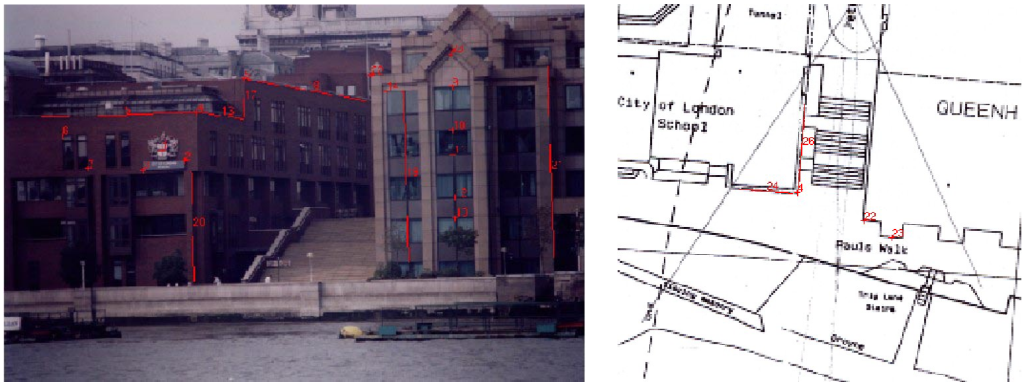


Figure C.4: Correspondence settings between a map and one image.

If no 3D points are available, the user has the possibility of working with a standard 2D map (figure C.4). By setting the approximate position of the camera on the map and giving some supplementary map points in the image, he calibrates the camera. The academic partners have shown that this can give very good calibration results. We expect this approach to be particularly interesting for outdoor scenes, where accurate 3D measurements are very difficult to obtain.

C.3.2 Self-calibration and Relative Motion

If several images of the scene are available, we use self-calibration and structure from motion techniques to augment the images. These techniques use multiple views of the scene to recover camera parameters and the 3D information.

The first group of algorithms computes the image geometry in projective space. Then a self-calibration module based on the Kruppa equations gives a first estimate of the intrinsic camera parameters. Bundle adjustment methods are supplied to refine these estimates. This method assumes that all of the images have the same intrinsic parameters.

Once the intrinsic parameters are available, the relative camera motion can be recovered. These “relative orientation” computations are based on the essential matrix, or in the case of a planar scene on the inter-image homography.

Finally, points of the scene are reconstructed in 3D using triangulation from all the views. Here also, the results are improved by applying a bundle adjustment on the camera parameters and the 3D triangulation.

C.3.3 Occlusion Handling

The next step of image augmentation handles occlusions between real and virtual objects in order to get a consistent rendering of the scene.

This is done by creating a simple model of the scene using triangulation methods. The user defines additional points in several views. These points are grouped into polygons and used to build a coarse 3D model of the scene. The model is rendered transparently on the images. By using the OpenGL depth buffer, occlusions between virtual and real objects can be handled. In figure C.5, we needed to reconstruct only the panels to get the correct occlusion of the left arm of the virtual man.

C.3.4 Image Augmentation

The last step of augmentation is the insertion of the object into the real scene. We defined two main approaches. The first uses a modeling tool to load the reconstructed scene model and the new object and place it at the right position in the scene.

The other possibility is to use the CUMULI-tool directly. The user can move the new object interactively and with help of the different views and the occlusion handling, he places it at the right 3D position. In order to make the interaction easier, it is possible to reduce the degrees of freedom of the virtual object in the scene. The user chooses a reference point of the virtual object in one image. This image point defines a optical ray of the camera. The virtual object is then attached to this ray and can only slide along it. With help of the other views, the object is then moved until it reach the required position in the scene. Then it is scaled and rotated until the desired images are created.

Figure C.5: Augmentation of two images.

C.4 Video Augmentation

For video augmentation, the user first selects some images of the sequence. Then, he proceeds as before, *i.e*. he gives some point correspondences and calibrates the camera. The selected points are then reconstructed in 3D.

Figure C.6: Virtual tower on an outdoor construction site.

After this initialization step, the selected points are automatically tracked through the whole sequence. Using the 3D coordinates of the points, the pose of the camera is updated from one image to the next. Finally, a bundle adjustment is applied in order to refine the results, reject eventual outliers and get a precise calibration of the whole sequence. Some results of the automatic augmentation of a video are shown in figure C.6.

C.5 Conclusions

The CUMULI-tool allows fast and robust augmentation of images and videos. It provides a selection of different algorithms adapted to various different types of camera and scene knowledge. Most of the common situations can be managed successfully. This tool was initially designed as an experimental environment for testing new computer vision algorithms in Augmented Reality. The next version will be oriented more towards the user. Our effort will focus on making the tool acceptable for practical use, and a new module will support the user by guiding him step by step through each stage from calibration to final object insertion.

Appendix D

Scientific Publications from CUMULI

This bibliography attempts to list all of the scientific papers that were supported in full or in part by CUMULI. There may be some omissions and some additional papers that have not yet appeared.

[1] M.-A. Ameller, B. Triggs, and L. Quan. Camera pose revisited — new linear algorithms. Submitted to ECCV'00, 1999.

[2] K. Åström. Using combinations of points, lines and conics to estimate structure and motion. In *Proc. Symposium on Image Analysis, SSAB, Uppsala, Sweden*, pages 61–64, 1998.

[3] K. Åström and A. Heyden. Continuous time matching constraints for image streams. *Int. Journal of Computer Vision*, 28(1):85–96, 1998.

[4] K. Åström and A. Heyden. Flexible calibration: Minimal cases for auto-calibration. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[5] K. Åström and A. Heyden. Stochastic analysis of scale space smoothing. *Advances in Applied Probability*, 30(1), 1999.

[6] K. Åström, A. Heyden, F. Kahl, and M. Oskarsson. Structure and motion from lines under affine projections. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[7] K. Åström and F. Kahl. Motion estimation in image sequences using the deformation of apparent contours. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(2):114–127, 1999.

[8] K. Åström, F. Kahl, A. Heyden, and R. Berthilsson. A statistical approach to structure and motion from image features. In *Statistical Techiques in Pattern Recognition, Sydney, Australia*, pages 929–935, 1998.

[9] K. Åström and M. Oskarsson. Solutions and ambiguities of the structure and motion problem for 1d retinal vision. In *Scandinavian Conf. on Image Analysis, Greenland*, 1999.

[10] R. Berthilsson. Affine correlation. In *Proc. Int. Conf. Pattern Recognition, Brisbane, Australia*, pages 1458–1467, 1998.

[11] R. Berthilsson. A statistical theory of shape. In *Statistical Techiques in Pattern Recognition, Sydney, Australia*, pages 677–682, 1998.

[12] R. Berthilsson. Densities and maximum likelihood estimation of matching constraints. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[13] R. Berthilsson. Finding correspondences of patches by means of affine transformations. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[14] R. Berthilsson, K. Åström, and A. Heyden. Reconstruction of curves in R^3, using factorization and bundle adjustment. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[15] R. Berthilsson and A. Heyden. Recognition of planar point configurations using the density of affine shape. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany*, 1998.

[16] R. Berthilsson, A. Heyden, and G. Sparr. Recursive structure and motion from image sequences using shape and depth spaces. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 444–449. IEEE Computer Society Press, 1997.

[17] R. Berthilsson and K Åström. Reconstruction of 3d-curves from 2d-images using affine shape methods for curves. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 476–481. IEEE Computer Society Press, 1997.

[18] R. Berthilsson, K. Åström, and A. Heyden. Projective reconstruction of 3d-curves from its 2d-images using error models and bundle adjustments. In *Proc. 1st Scandinavian Conf. on Image Analysis, Lappeenranta, Finland*, pages 581–588, 1997.

[19] Rikard Berthilsson. *Extensions and Applications of Affine Shape.* Phd thesis, Centre for Mathematical Sciences, Lund University, December 1999.

[20] D. Bondyfalat, B. Mourrain, and T. Papadopoulo. An application of automatic theorem proving in computer vision. In X.-S. Gao, D. Wang, and L. Yang, editors, *Automated Deduction in Geometry*, number 1669 in LNAI. Springer-Verlag, Berlin Heidelberg, 1999.

[21] Didier Bondyfalat and Sylvain Bougnoux. Imposing Euclidean constraints during self-calibration processes. In R. Koch and L. Van Gool, editors, *Proceedings of SMILE Workshop on Structure from Multiple Images*, Lecture Notes in Computer Science. Springer Verlag, Lecture Notes in Computer Science, 1998.

[22] Sylvain Bougnoux. From projective to Euclidean space under any practical situation, a criticism of self-calibration. In *Proc. 6th Int. Conf. on Computer Vision, Mumbai, India*, 1998.

[23] Thierry Boy de la Tour, Stéphane Fèvre, and Dongming Wang. Clifford term rewriting for geometric reasoning in 3D. In X.-S. Gao, D. Wang, and L. Yang, editors, *Automated Deduction in Geometry — Proceedings of ADG'98*, LNAI 1669, pages 130–155. Springer-Verlag, 1998.

[24] A. Heyden *et al.* Discrete and continuous multi-camera projective reconstruction: A summary of work done within CUMULI. unpublished, 1998.

[25] 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. In *Proc. 5th European Conference on Computer Vision, Freiburg, Germany*, pages 36–52, June 1998.

[26] Olivier Faugeras and Théodore Papadopoulo. Grassmann-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. RR 3225, INRIA, July 1997. A shorter version appeared in Transactions of the Royal Society A.

[27] Olivier Faugeras and Théodore Papadopoulo. A nonlinear method for estimating the projective geometry of three views. RR 3221, INRIA, July 1997.

[28] Olivier Faugeras and Théodore Papadopoulo. Grassmann-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. *Transactions of the Royal society A*, May 1998.

[29] Olivier Faugeras and Théodore Papadopoulo. A nonlinear method for estimating the projective geometry of three views. In *Proc. 6^{th} Int. Conf. Computer Vision*, pages 477–484, Bombay, India, January 1998. IEEE Computer Society Press.

[30] Stéphane Fèvre and Dongming Wang. Combining algebraic computing and term-rewriting for geometry theorem proving. In *Proceedings of the 4th International Conference on Artificial Intelligence and Symbolic Computation*, LNAI **1476**. Springer-Verlag, Berlin Heidelberg, 1998.

[31] Stéphane Fèvre and Dongming Wang. Proving geometric theorems using clifford algebra and rewrite rules. In *Proceedings of the 15th International Conference on Automated Deduction*, LNAI **1421**, pages 17–32. Springer-Verlag, Berlin Heidelberg, 1998.

[32] Stéphane Fèvre and Dongming Wang. Combining clifford algebraic computing and term-rewriting for geometric theorem proving. *Fundamenta Informaticae*, 39(1–2):85–104, 1999.

[33] P Hallberg and A. Vegh. Feature detection, tracking and iterative reconstruction from image sequences. Master's thesis, Dept. of Mathematics, Lund University, October 1997.

[34] A. Heyden. Projective structure and motion from image sequences using subspace methods. In *Proc. 1st Scandinavian Conf. on Image Analysis, Lappeenranta, Finland*, pages 963–968, 1997.

[35] A. Heyden. A common framework for multiple-view tensors. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany*, 1998.

[36] A. Heyden. Tensorial properties of multilinear constraints. *Mathematical Methods in Applied Sciences*, 1999. Submitted.

[37] A. Heyden and K Åström. Minimal conditions on intrinsic parameters for Euclidean reconstruction. In *Proc. 2nd Asian Conf. on Computer Vision, Hong Kong, China*, pages 169–176, 1998.

[38] A. Heyden and R. Berthilsson. Recognition of planar objects using the density of affine shape. *Computer Vision and Image Understanding*, 76(2):135–145, 1999.

[39] A. Heyden, R. Berthilsson, and G. Sparr. Recursive structure and motion from image sequences using shape and depth spaces. *Int. Journal of Computer Vision*, 1998. submitted.

[40] A. Heyden, R. Berthilsson, and G. Sparr. An iterative factorization method for projective structure and motion from image sequences. *Image and Vision Computing*, 17(13):981–991, 1999.

[41] A. Heyden and F. Kahl. Reconstruction from affine cameras using closure constraints. In *Proc. Int. Conf. Pattern Recognition, Brisbane, Australia*, pages 56–65, 1998.

[42] A. Heyden, G. Sparr, and K. Åström. Perception and action using multilinear forms. In *International Workshop on Algebraic Frames for the Perception-Action Cycle, Keil, Germany*, 1997.

[43] A. Heyden and K. Åström. Algebraic properties of multilinear constraints. *Mathematical Methods in the Applied Sciences*, 20:1135–1162, 1997.

[44] A. Heyden and K Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 438–443. IEEE Computer Society Press, 1997.

[45] A. Heyden and K. Åström. Simplifications of multilinear forms for sequences of images. *Image and Vision Computing*, 1997.

[46] F. Kahl. Critical motion sequences. unpublished, 1998.

[47] F. Kahl. Critical motions and ambiguous euclidean reconstructions in auto-calibration. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[48] F. Kahl and K. Åström. Motion estimation in image sequences using the deformation of apparent contours. In *Proc. 6th Int. Conf. on Computer Vision, Mumbai, India*, pages 939–944, 1998.

[49] F. Kahl and A. Heyden. Robust self-calibration and Euclidean reconstruction via affine approximation. In *Proc. Int. Conf. Pattern Recognition, Brisbane, Australia*, pages 47–55, 1998.

[50] F. Kahl and A. Heyden. Structure and motion from points, lines and conics with affine cameras. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany*, 1998.

[51] F. Kahl and A. Heyden. Structure and motion with affine cameras. In *Proc. Symposium on Image Analysis, SSAB, Uppsala, Sweden*, pages 57–60, 1998.

[52] F. Kahl and A. Heyden. Using conic correspondences to estimate the epipolar geometry. In *Proc. 6th Int. Conf. on Computer Vision, Mumbai, India*, pages 761–766, 1998.

[53] F. Kahl and A. Heyden. Affine structure and motion from points, lines and conics. *Int. Journal of Computer Vision*, 33(3):163–180, 1999.

[54] F. Kahl and B. Triggs. Critical motions in euclidean structure from motion. In *Conf. Computer Vision and Pattern Recognition, Fort Collins, USA*, 1999.

[55] F. Kahl, B. Triggs, and K. Åström. Critical motions for auto-calibration when some intrinsic parameters can vary. *Journal of Mathematical Imaging and Vision*, 2000. To appear in October 2000.

[56] Fredrik Kahl. Geometric stratification in computational vision. Licentiate thesis, Centre for Mathematical Sciences, Lund University, April 1999.

[57] G. Klinker, D. Stricker, and D. Reiners. The use of reality models in augmented reality applications. In R. Koch and L. Van Gool, editors, *Proceedings of SMILE Workshop on Structure from Multiple Images*, Lecture Notes in Computer Science. Springer Verlag, Lecture Notes in Computer Science, 1998.

[58] Diane Lingrand. *Analyse Adaptative du Mouvement dans des Séquences Monoculaires non Calibrées*. PhD thesis, Université de Nice - Sophia Antipolis, July 1999.

[59] Diane Lingrand. Perspective projection's approximations: combined model and homographic singularities. Technical Report RR-3682, INRIA Sophia Antipolis, April 1999.

[60] Diane Lingrand. Etude des singularités des homographies dans des séquences monoculaires non calibrées. In *12^e^ Congrès Reconnaissance des Formes et Intelligence Artificielle*, volume 1, February 2000.

[61] S. Mueller. Computer vision based modelling and augmented reality. Presentation of results of REALISE and CUMULI at industrial workshop *From images to knowledge*, SIRA Technology Centre, London, 1–2 October 1997.

[62] M. Oskarsson and K. Åström. Automatic geometric reasoning in structure and motion estimation. In *Scandinavian Conf. on Image Analysis, Greenland*, 1999.

[63] Théo Papadopoulo and Olivier Faugeras. A new characterization of the trifocal tensor. In *Proc. 5^{th}European Conf. Computer Vision*, Fribourg, Germany, June 1998.

[64] L. Quan. Uncalibrated 1d projective camera and 3d affine reconstruction of lines. In *Proc. Conf. Computer Vision and Pattern Recognition*, 1997.

[65] L. Quan. Inherent two-way ambiguity in 2d projective reconstruction from three uncalibrated 1d images. In *Int. Conf. Computer Vision*, pages 344–349, Corfu, Greece, 1999.

[66] L. Quan, A. Heyden, and F. Kahl. Minimal projective reconstruction with missing data. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 210–216. IEEE Computer Society Press, June 1999.

[67] L. Quan and T. Kanade. Affine structure from line correspondences with uncalibrated affine cameras. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 1997.

[68] L. Quan and Z. D. Lan. Linear $\geq N$ point pose determination. In *Int. Conf. Computer Vision*, 1998.

[69] L. Quan and Z.D. Lan. Linear n-point camera pose determination. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 21(8):774–780, August 1999.

[70] L. Quan and M. Lhuillier. Structure from motion from three affine views. Submitted to ECCV'00, 1999.

[71] L. Quan and R. Mohr. Uniqueness of 3d affine reconstruction of lines with affine cameras. In *Int. Conf. on Computer Analysis of Images and Patterns*, pages 231–238, Kiel, Germany, 1997. Springer Verlag.

[72] L. Quan and B. Triggs. A unification of autocalibration methods. In *Asian Conf. Computer Vision*, January 2000.

[73] P. Renault, O. Faugeras, and T. Viéville. Continuous multi-image preprocessing for Euclidean reconstruction. RR 3482, INRIA, September 1998.

[74] G. Sparr. Euclidean and affine structure/motion for uncalibrated cameras from affine shape and subsidiary information. In R. Koch and L. Van Gool, editors, *Proceedings of SMILE Workshop on Structure from Multiple Images*, Lecture Notes in Computer Science. Springer Verlag, Lecture Notes in Computer Science, 1998.

[75] D. Stricker, G. Klinker, and D. Reiners. A fast and robust line-based optical tracker for augmented reality applications. In *Proceedings of first International Workshop on Augmented Reality*, 1998.

[76] D. Stricker and N. Navab. Calibration propagation for image augmentation. In *Proceedings of Second International Workshop on Augmented Reality*, San Francisco, 1999.

[77] K. Åström and A. Heyden. Stochastic analysis of image acquisition and scale-space smoothing. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*. Kluwer Academic Publishers, 1997.

[78] K. Åström, A. Heyden, F. Kahl, R. Berthilsson, and G. Sparr. A computer vision toolbox. In *Proc. Nordic Matlab Conference*, 1997.

[79] P. Sturm. Critical motion sequences and conjugacy of ambiguous Euclidean reconstructions. In *Scandinavian Conf. on Image Analysis*, volume I, pages 439–46, June 1997.

[80] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. Conf. Computer Vision and Pattern Recognition*, Puerto Rico, 1997.

[81] P. Sturm. Self-calibration of a moving zoom-lens camera by pre-calibration. *Image and Vision Computing*, 15(8):583–90, August 1997.

[82] B. Triggs. Autocalibration and the absolute quadric. In *Proc. Conf. Computer Vision and Pattern Recognition*, Puerto Rico, 1997.

[83] B. Triggs. Linear projective reconstruction from matching tensors. *Image and Vision Computing*, 15(8):617–26, August 1997.

[84] B. Triggs. Autocalibration from planar scenes. In *European Conf. Computer Vision*, pages I 89–105, Freiburg, June 1998.

[85] B. Triggs. Optimal estimation of matching constraints. In R. Koch and L. Van Gool, editors, *Workshop on 3D Structure from Multiple Images of Large-scale Environments SMILE'98*, Lecture Notes in Computer Science. Springer Verlag, 1998.

[86] B. Triggs. Camera pose and calibration from 4 or 5 known 3d points. In *Int. Conf. Computer Vision*, pages 278–284, Kerkyra, Greece, September 1999.

[87] B. Triggs. Differential matching constraints. In *Int. Conf. Computer Vision*, pages 370–376, Kerkyra, Greece, September 1999. IEEE press.

[88] B. Triggs. Plane + parallax, tensors and factorization. Submitted to ECCV'00, 1999.

[89] B. Triggs, A. Fitzgibbon, P. McLauchlan, and R. Hartley. Bundle adjustment for structure from motion. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*. Springer-Verlag, 2000.

[90] B. Triggs, R. Szeliski, and A. Zisserman, editors. *Vision Algorithms : Theory and Practice*, Workshop held during ICCV'99 in Kerkyra (Corfu), Greece, September 1999. INRIA. Final proceedings will appear in Springer LNCS in September 2000.

[91] B. Triggs, A. Zisserman, and R. Szeliski, editors. *Vision Algorithms: Theory and Practice*. Springer-Verlag, 2000.

[92] T. Viéville and D. Lingrand. Using specific displacements to analyze motion without calibration. *Int. Journal of Computer Vision*, 1998. Accepted.

[93] Dongming Wang. Clifford algebraic calculus for geometric reasoning with application to computer vision. In R. Caferra, L. Fari nas del Cerro, H. Shi, and D. Wang, editors, *Automated Deduction in Geometry*, Berlin Heidelberg, 1997. Springer-Verlag.

[94] Dongming Wang. Clifford algebraic computing and term-rewriting for geometric theorem proving. In *Proceedings of the 2nd International Theorema Workshop*, 1998. available as RISC-Linz Report no. 98-10, Johannes Kepler University, Austria.

[95] Dongming Wang. Geometric algebra and reasoning. In *Geometric Algebra: A Geometric Approach to Computer Vision, Quantum and Neural Computing, Robotics and Engineering*. Birkhäuser, Boston, 1999. To appear.

Appendix E

Bibliography

This bibliography lists the CUMULI and non-CUMULI papers cited in the body of this report.

[1] K. Åström. Using combinations of points, lines and conics to estimate structure and motion. In *Proc. Symposium on Image Analysis, SSAB, Uppsala, Sweden*, pages 61–64, 1998.

[2] K. Åström and A. Heyden. Continuous time matching constraints for image streams. *Int. Journal of Computer Vision*, 28(1):85–96, 1998.

[3] K. Åström and A. Heyden. Flexible calibration: Minimal cases for auto-calibration. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[4] K. Åström, A. Heyden, F. Kahl, and M. Oskarsson. Structure and motion from lines under affine projections. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[5] K. Åström, F. Kahl, A. Heyden, and R. Berthilsson. A statistical approach to structure and motion from image features. In *Statistical Techiques in Pattern Recognition, Sydney, Australia*, pages 929–935, 1998.

[6] R. Berthilsson. Affine correlation. In *Proc. Int. Conf. Pattern Recognition, Brisbane, Australia*, pages 1458–1467, 1998.

[7] R. Berthilsson. A statistical theory of shape. In *Statistical Techiques in Pattern Recognition, Sydney, Australia*, pages 677–682, 1998.

[8] R. Berthilsson, K. Åström, and A. Heyden. Reconstruction of curves in R^3, using factorization and bundle adjustment. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[9] R. Berthilsson and A. Heyden. Recognition of planar point configurations using the density of affine shape. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany*, 1998.

[10] R. Berthilsson, A. Heyden, and G. Sparr. Recursive structure and motion from image sequences using shape and depth spaces. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 444–449. IEEE Computer Society Press, 1997.

[11] R. Berthilsson and K Åström. Reconstruction of 3d-curves from 2d-images using affine shape methods for curves. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 476–481. IEEE Computer Society Press, 1997.

[12] R. Berthilsson, K. Åström, and A. Heyden. Projective reconstruction of 3d-curves from its 2d-images using error models and bundle adjustments. In *Proc. 1st Scandinavian Conf. on Image Analysis, Lappeenranta, Finland*, pages 581–588, 1997.

[13] D. Bondyfalat, Mourrain. B, and T. Papadopoulo. Automated method for theorem discovering. Workshop ADG'98, China, 1998.

[14] D. Bondyfalat, B. Mourrain, and T. Papadopoulo. An application of automatic theorem proving in computer vision. In *2nd International Workshop on Automated Deduction in Geometry*, 1999.

[15] Didier Bondyfalat. *Interaction Symbolique et Numérique; Application a la Vision Artificielle*. PhD thesis, Université de Nice Sophia-Antipolis, 2000.

[16] Didier Bondyfalat and Sylvain Bougnoux. Imposing Euclidean constraints during self-calibration processes. In R. Koch and L. Van Gool, editors, *Proceedings of SMILE Workshop on Structure from Multiple Images*, Lecture Notes in Computer Science. Springer Verlag, Lecture Notes in Computer Science, 1998.

[17] Thierry Boy de la Tour, Stéphane Fèvre, and Dongming Wang. Clifford term rewriting for geometric reasoning in 3d. Workshop ADG'98, China, 1998.

[18] R. Cipolla, K. Åström, and P. J. Giblin. Motion from the frontier of curved surfaces. In *Proc. 5th Int. Conf. on Computer Vision, MIT, Boston, MA*, pages 269–275. IEEE Computer Society Press, 1995.

[19] P.E. Debevec, C.J. Taylor, and J. Malik. Modeling and rendering architecture from photographs: a hybrid geometry- and image-based approach. In *SIGGRAPH*, pages 11–20, New Orleans, August 1996.

[20] A. Heyden *et al*. Discrete and continuous multi-camera projective reconstruction: A summary of work done within CUMULI. unpublished, 1998.

[21] O. Faugeras and B. Mourrain. About the correspondences of points between n images. In *IEEE Workshop on Representations of Visual Scenes*, pages 37–44, Cambridge, MA, June 1995.

[22] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *Int. Conf. Computer Vision*, pages 951–6, 1995.

[23] 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. In *Proc. 5th European Conference on Computer Vision, Freiburg, Germany*, pages 36–52, June 1998.

[24] Olivier Faugeras and Théodore Papadopoulo. Grassmann-Cayley algebra for modeling systems of cameras and the algebraic equations of the manifold of trifocal tensors. *Transactions of the Royal Society A*, May 1998.

[25] Olivier Faugeras and Théodore Papadopoulo. A nonlinear method for estimating the projective geometry of three views. In *Proc.* 6^{th} *Int. Conf. Computer Vision*, pages 477–484, Bombay, India, January 1998. IEEE Computer Society Press.

[26] Stéphane Fèvre and Dongming Wang. Combining algebraic computing and term-rewriting for geometry theorem proving. In *Proceedings of the 4th International Conference on Artificial Intelligence and Symbolic Computation*, LNAI **1476**. Springer-Verlag, Berlin Heidelberg, 1998.

[27] Stéphane Fèvre and Dongming Wang. Proving geometric theorems using clifford algebra and rewrite rules. In *Proceedings of the 15th International Conference on Automated Deduction*, LNAI **1421**, pages 17–32. Springer-Verlag, Berlin Heidelberg, 1998.

[28] A. Heyden. Projective structure and motion from image sequences using subspace methods. In *Proc. 1st Scandinavian Conf. on Image Analysis, Lappeenranta, Finland*, pages 963–968, 1997.

[29] A. Heyden. A common framework for multiple-view tensors. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany*, 1998.

[30] A. Heyden and K. Åström. A canonical framework for sequences of images. In *IEEE Workshop on Representations of Visual Scenes*, Cambridge, MA, June 1995.

[31] A. Heyden and K Åström. Minimal conditions on intrinsic parameters for Euclidean reconstruction. In *Proc. 2nd Asian Conf. on Computer Vision, Hong Kong, China*, pages 169–176, 1998.

[32] A. Heyden, R. Berthilsson, and G. Sparr. An iterative factorization method for projective structure and motion from image sequences. *Image and Vision Computing*, 17(13):981–991, 1999.

[33] A. Heyden and F. Kahl. Reconstruction from affine cameras using closure constraints. In *Proc. Int. Conf. Pattern Recognition, Brisbane, Australia*, pages 56–65, 1998.

[34] A. Heyden, G. Sparr, and K. Åström. Perception and action using multilinear forms. In *International Workshop on Algebraic Frames for the Perception-Action Cycle, Keil, Germany*, 1997.

[35] A. Heyden and K. Åström. Euclidean reconstruction from constant intrinsic parameters. In *Proc. International Conference on Pattern Recognition, Vienna, Austria*, volume 1, pages 339–343. IEEE Computer Society Press, 1996.

[36] A. Heyden and K. Åström. Algebraic properties of multilinear constraints. *Mathematical Methods in the Applied Sciences*, 20:1135–1162, 1997.

[37] A. Heyden and K Åström. Euclidean reconstruction from image sequences with varying and unknown focal length and principal point. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 438–443. IEEE Computer Society Press, 1997.

[38] F. Kahl. Critical motion sequences. unpublished, 1998.

[39] F. Kahl. Critical motions and ambiguous euclidean reconstructions in auto-calibration. In *Int. Conf. Computer Vision, Kerkyra, Greece*, 1999.

[40] F. Kahl and K. Åström. Motion estimation in image sequences using the deformation of apparent contours. In *Proc. 6th Int. Conf. on Computer Vision, Mumbai, India*, pages 939–944, 1998.

[41] F. Kahl and A. Heyden. Robust self-calibration and Euclidean reconstruction via affine approximation. In *Proc. Int. Conf. Pattern Recognition, Brisbane, Australia*, pages 47–55, 1998.

[42] F. Kahl and A. Heyden. Structure and motion from points, lines and conics with affine cameras. In *Proc. 5th European Conf. on Computer Vision, Freiburg, Germany*, 1998.

[43] F. Kahl and A. Heyden. Using conic correspondences to estimate the epipolar geometry. In *Proc. 6th Int. Conf. on Computer Vision, Mumbai, India*, pages 761–766, 1998.

[44] F. Kahl and B. Triggs. Critical motions in euclidean structure from motion. In *Conf. Computer Vision and Pattern Recognition, Fort Collins, USA*, 1999.

[45] F. Kahl, B. Triggs, and K. Åström. Critical motions for auto-calibration when some intrinsic parameters can vary. *Journal of Mathematical Imaging and Vision*, 1999. To appear in October 2000.

[46] D. Kapur and J.L. Mundy. *Geometric Reasoning*. MIT Press, Cambridge, 1989.

[47] Diane Lingrand. *Analyse Adaptative du Mouvement dans des Séquences Monoculaires non Calibrées*. PhD thesis, Université de Nice - Sophia Antipolis, July 1999.

[48] Diane Lingrand. Perspective projection's approximations: combined model and homographic singularities. Technical Report RR-3682, INRIA Sophia Antipolis, April 1999.

[49] Diane Lingrand. Etude des singularités des homographies dans des séquences monoculaires non calibrées. In *12^e Congrès Reconnaissance des Formes et Intelligence Artificielle*, volume 1, February 2000.

[50] S. Maybank. *Theory of Reconstruction from Image Motion*. Springer Verlag, Berlin, Heidelberg, New York, 1993.

[51] Théo Papadopoulo and Olivier Faugeras. A new characterization of the trifocal tensor. In *Proc. 5thEuropean Conf. Computer Vision*, Fribourg, Germany, June 1998.

[52] J. Philip. A non-iterative algorithm for determining all essential matrices corresponding to five point pairs. *Photogrammetric Record*, 15(88):589–599, October 1996.

[53] L. Quan. Inherent two-way ambiguity in 2d projective reconstruction from three uncalibrated 1d images. In *Int. Conf. Computer Vision*, pages 344–349, Corfu, Greece, 1999.

[54] L. Quan, A. Heyden, and F. Kahl. Minimal projective reconstruction with missing data. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 210–216. IEEE Computer Society Press, June 1999.

[55] L. Quan and M. Lhuillier. Structure from motion from three affine views. Submitted to ECCV'00, 1999.

[56] C. C. Slama. *Manual of Photogrammetry*. American Society of Photogrammetry, Falls Church, VA, 1980.

[57] G. Sparr. Simultaneous reconstruction of scene structure and camera locations from uncalibrated image sequences. In *Proc. International Conference on Pattern Recognition, Vienna, Austria*, volume 1, pages 328–333. IEEE Computer Society Press, 1996.

[58] G. Sparr. Euclidean and affine structure/motion for uncalibrated cameras from affine shape and subsidiary information. In R. Koch and L. Van Gool, editors, *Proceedings of SMILE Workshop on Structure from Multiple Images*, Lecture Notes in Computer Science. Springer Verlag, Lecture Notes in Computer Science, 1998.

[59] K. Åström, R. Cipolla, and P. J. Giblin. Generalised epipolar constraints. In R.C̃ipolla B.B̃uxton, editor, *Proc. 4th European Conf. on Computer Vision, Cambridge, UK*, volume 1065 of *Lecture notes in Computer Science*, pages 97–108. Springer-Verlag, 1996.

[60] K. Åström and A. Heyden. Stochastic modelling and analysis of image acquisition and sub-pixel edge detection. In *Proc. International Conference on Pattern Recognition, Vienna, Austria*, volume 2, pages 86–91. IEEE Computer Society Press, 1996.

[61] K. Åström and A. Heyden. Stochastic analysis of image acquisition and scale-space smoothing. In J. Sporring, M. Nielsen, L. Florack, and P. Johansen, editors, *Gaussian Scale-Space Theory*. Kluwer Academic Publishers, 1997.

[62] K. Åström, A. Heyden, F. Kahl, R. Berthilsson, and G. Sparr. A computer vision toolbox. In *Proc. Nordic Matlab Conference*, 1997.

[63] P. Sturm. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. Conf. Computer Vision and Pattern Recognition*, Puerto Rico, 1997.

[64] B. Triggs. Matching constraints and the joint image. In E. Grimson, editor, *Int. Conf. Computer Vision*, pages 338–43, Cambridge, MA, June 1995.

[65] B. Triggs. Autocalibration and the absolute quadric. In *Proc. Conf. Computer Vision and Pattern Recognition*, Puerto Rico, 1997.

[66] B. Triggs. Autocalibration from planar scenes. In *European Conf. Computer Vision*, pages I 89–105, Freiburg, June 1998.

[67] B. Triggs. Optimal estimation of matching constraints. In R. Koch and L. Van Gool, editors, *Workshop on 3D Structure from Multiple Images of Large-scale Environments SMILE'98*, Lecture Notes in Computer Science. Springer Verlag, 1998.

[68] B. Triggs. Camera pose and calibration from 4 or 5 known 3d points. In *Int. Conf. Computer Vision*, pages 278–284, Kerkyra, Greece, September 1999.

[69] B. Triggs. Differential matching constraints. In *Int. Conf. Computer Vision*, pages 370–6, 1999.

[70] B. Triggs. Plane + parallax, tensors and factorization. Submitted to ECCV'00, 1999.

[71] B. Triggs, A. Fitzgibbon, P. McLauchlan, and R. Hartley. Bundle adjustment for structure from motion. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*. Springer-Verlag, 2000.

[72] B. Triggs, A. Zisserman, and R. Szeliski, editors. *Vision Algorithms: Theory and Practice*. Springer-Verlag, 2000.

[73] Thierry Viéville and Olivier Faugeras. The first order expansion of motion equations in the uncalibrated case. *Computer Vision, Graphics and Image Processing*, 64(1):128–146, July 1996.

[74] Dongming Wang. Clifford algebraic computing and term-rewriting for geometric theorem proving. In *Proceedings of the 2nd International Theorema Workshop*, 1998. available as RISC-Linz Report no. 98-10, Johannes Kepler University, Austria.

[75] W. Wunderlich. Rechnerische rekonstruktion eines ebenen objekts aus zwei photographien. In *Mittsilungen Geodät. Inst. TU Gras*, Folge 40 (festschrift K. Rimmer sum 70. Geburstag), pages 365–77, 1982.